

**UNIVERSIDADE NOVE DE JULHO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA REABILITAÇÃO**

Thamiê Cristina Stella

**PROPRIEDADES DE MEDIDAS DOS INSTRUMENTOS DE AVALIAÇÃO DE
ATIVIDADE FÍSICA EM INDIVÍDUOS PORTADORES DE DOENÇA RENAL
CRÔNICA: UMA REVISÃO SISTEMÁTICA**

São Paulo, SP

2023

THAMIÊ CRISTINA STELLA

**PROPRIEDADES DE MEDIDAS DOS INSTRUMENTOS DE AVALIAÇÃO DE
ATIVIDADE FÍSICA EM INDIVÍDUOS PORTADORES DE DOENÇA RENAL
CRÔNICA: UMA REVISÃO SISTEMÁTICA**

Qualificação apresentada à
Universidade Nove de Julho para
obtenção do título de Mestre em
Ciências da Reabilitação.

Orientadora: Prof.^a Dr.^a Luciana Malosá
Sampaio

**São Paulo, SP
2023**

FICHA CATALOGRÁFICA

Stella, Thamiê Cristina.

Propriedades de medidas dos instrumentos de avaliação de atividade física em indivíduos portadores de doença renal crônica: uma revisão sistemática. / Thamiê Cristina Stella. 2023.

89 f.

Dissertação (Mestrado) - Universidade Nove de Julho - UNINOVE, São Paulo, 2023.

Orientador (a): Prof^a. Dr^a. Luciana Maria Malosá Sampaio.

1. Doença renal crônica. 2. Revisão sistemática. 3. Atividade física. 4. COSMIN, 5. EMPRO.

I. Sampaio, Luciana Maria Malosá. II. Título.

CDU 615.8

São Paulo, 05 de dezembro de 2023.

TERMO DE APROVAÇÃO

Aluno (a): THAMIÊ CRISTINA STELLA

Título da Dissertação: "PROPRIEDADES DE MEDIDAS DOS INSTRUMENTOS DE AVALIAÇÃO DE ATIVIDADE FÍSICA EM INDIVÍDUOS PORTADORES DE DOENÇA RENAL CRÔNICA: UMA REVISÃO SISTEMÁTICA"

Presidente: PROFA. DRA. LUCIANA MARIA MALOSÁ SAMPAIO JORGE



Documento assinado digitalmente
CID ANDRÉ FIDELIS DE PAULA GOMES
Data: 05/12/2023 16:12:03-0300
Verifique em <https://validar.it.gov.br>

Membro: PROF. DR. CID ANDRÉ FIDELIS DE PAULA GOMES



Documento assinado digitalmente
ADRIANA SANCHES GARCIA DE ARAUJO
Data: 05/12/2023 16:30:44-0300
Verifique em <https://validar.it.gov.br>

Membro: PROFA. DRA. ADRIANA SANCHES GARCIA DE ARAÚJO

AGRADECIMENTOS

Agradeço primeiramente à minha família, que são meu principal alicerce e a razão de tudo o que conquistei até aqui e que futuramente conquistarei. Agradeço aos meus pais, Adilson Stella e Maria Aparecida Piffer Stella, por sempre me apoiarem e acreditarem em mim, inclusive em muitos momentos que eu mesma não acreditava; aos meus irmãos e padrinhos, Tatiana Cristina Stella e Adilson Stella Júnior, por serem meus melhores amigos e minha referência desde pequena; à minha cunhada e meu sobrinho, Francinara Rezende Reis Stella e João Victor Reis Stella, pelos momentos de carinho e descontração que sempre me proporcionaram.

Em seguida, agradeço especialmente à minha orientadora, Prof^a Dr^a Luciana Malosá Sampaio, por me aceitar no programa de mestrado, pela paciência que dispendeu comigo desde o início, por todo auxílio e orientação ao longo dessa etapa, e por ser sempre uma fonte de apoio e informação, fazendo com que a experiência no meio acadêmico não fosse nada menos do que prazerosa e receptiva. Agradeço ao Prof^o Dr. Cid André Fidelis de Paula Gomes e à Prof^a Dr^a Adriana Sanches Garcia de Araújo que compuseram a banca para defesa deste trabalho, e cujas contribuições engrandeceram não só esta dissertação como auxiliaram no meu crescimento enquanto pesquisadora.

Agradeço em especial alguns amigos que foram fundamentais nessa jornada. Obrigada Stephanie de Brito, pelo apoio emocional e intelectual que dispendeu ao longo de toda essa jornada, pela amizade inestimável que construímos nesses últimos anos. Obrigada Vanessa Kanashiro, por ser uma presença constante e firme em todos os nossos anos de amizade, por ser uma das minhas maiores apoiadoras tanto no âmbito profissional quanto pessoal.

Agradeço aos colegas e amigos do Hospital do Rim, que tiveram imensa paciência e compreensão comigo durante o período do mestrado. Destaco em especial à Graziella Alves da Silva, que foi quem me estimulou a ingressar no mestrado e me deu suporte nos meus momentos de dúvida.

Agradeço à Universidade Nove de Julho que sempre disponibilizou infraestrutura e assistência ao longo do período.

Agradeço à CAPES pela concessão da bolsa de estudos.

Muito obrigada a todos!

RESUMO

INTRODUÇÃO: Na doença renal crônica (DRC), existem vários estudos que avaliam a atividade física através de medidas relatadas pelo próprio paciente (PROMs), mas com grande heterogeneidade na escolha da ferramenta. **OBJETIVOS:** Identificar, avaliar e sintetizar as propriedades psicométricas das PROMs utilizadas na avaliação de atividade física na população portadora de DRC. **METODOLOGIA:** Foi realizada revisão sistemática nas bases de dados PubMed, SciELO, BIREME, Lilacs e EMBASE. Utilizaram-se os descritores: *Chronic Kidney Disease, Chronic Renal Insufficiency, Chronic Kidney Failure, Physical Activity, Physical Activities, Psychometrics, Patient Outcome Assessment, Outcome Assessment, Health Care, Reproducibility of Results, Reliability of Result, Validity of Result, Reliability and Validity, Test Retest Reliability, Validity, Face, Finding Reproducibility, Validation studies, Cross Cultural Comparison*; com os operadores booleanos AND ou OR. A qualidade dos estudos foi avaliada pelo COSMIN RoB e EMPRO. **RESULTADOS:** Foram encontrados 13638 estudos, com 15 ao final incluídos. Das 11 PROMs encontradas, 2 são específicas para DRC. *Human Activity Profile* (HAP) foi a PROM não-específica com melhor desempenho, e *Chronic Kidney Disease Physical Activity Questionnaire* (CKD-PAQ) foi a única PROM com resultado satisfatório nas duas ferramentas de avaliação. COSMIN RoB e EMPRO apresentaram discordâncias em 5 PROMs. **CONCLUSÃO:** Não há consenso na melhor PROM de avaliar atividade física na DRC. HAP e CKD-PAQ são as PROMs que indicam melhor desempenho. As divergências entre COSMIN RoB e EMPRO demonstram que a escolha da ferramenta de avaliação depende da expertise e objetivo do avaliador. Há a necessidade de mais pesquisas no tema, principalmente nos estágios mais leves da DRC.

Palavras-chave: Doença Renal Crônica, Revisão Sistemática, Atividade Física, COSMIN, EMPRO.

ABSTRACT

INTRODUCTION: In chronic kidney disease (CKD), there are several studies that assess physical activity through patient-reported outcome measures (PROMs), but with large heterogeneity in the tool's choice. **OBJECTIVES:** To identify, evaluate and synthesize the psychometric properties of PROMs used to assess physical activity in the CKD population. **METHODOLOGY:** A systematic review was carried out in PubMed, SciELO, Medline, Lilacs and EMBASE databases. The descriptors used: *Chronic Kidney Disease, Chronic Renal Insufficiency, Chronic Kidney Failure, Physical Activity, Physical Activities, Psychometrics, Patient Outcome Assessment, Outcome Assessment, Health Care, Reproducibility of Results, Reliability of Result, Validity of Result, Reliability and Validity, Test Retest Reliability, Validity, Face, Finding Reproducibility, Validation studies, Cross Cultural Comparison*; with boolean operators AND or OR. The quality of the studies was assessed by COSMIN RoB and EMPRO. **RESULTS:** 13,638 studies were found, with 15 included in the end. Of the 11 PROMs found, 2 are specific to CKD. *Human Activity Profile* (HAP) was the non-specific PROM with the best performance, while *Chronic Kidney Disease Physical Activity Questionnaire* (CKD-PAQ) was the only PROM with a satisfactory result in assessment's tools. COSMIN RoB and EMPRO presented disagreements in 5 PROMs. **CONCLUSION:** There is no consensus on the best PROM to assess physical activity in CKD. HAP and CKD-PAQ are the PROMs that indicate better performance. The divergences between COSMIN RoB and EMPRO demonstrate that the choice of assessment tool depends on the evaluator's expertise and objective. More research on the topic is needed, especially in the milder stages of CKD.

Key-words: Chronic Kidney Disease, Systematic Review, Physical Activity, COSMIN, EMPRO.

LISTA DE FIGURAS

Figura 1. Estratégia de elaboração de pergunta de pesquisa conforme iniciativa COSMIN.....	24
Figura 2. Fluxograma PRISMA do processo de seleção dos artigos incluídos na revisão.....	30

LISTA DE GRÁFICOS

Gráfico 1. Distribuição das publicações incluídas ao longo dos anos.....	31
Gráfico 2. Número de estudos em que cada PROM foi avaliada.....	31

LISTA DE QUADROS

Quadro 1. Estratégia de busca utilizada para revisão sistemática.....	25
--	----

LISTA DE TABELAS

Tabela 1. Características metodológicas dos estudos inclusos.....	33
Tabela 2. Sumarização dos resultados das propriedades psicométricas por PROM.....	39
Tabela 3. Análise da qualidade metodológica dos artigos através do COSMIN RoB.....	43
Tabela 4. Sumarização do nível de evidência das PROMs através do COSMIN RoB.....	44
Tabela 5. Avaliação metodológica dos artigos inclusos pela ferramenta EMPRO.....	45
Tabela 6. Sumarização da avaliação das propriedades psicométricas das PROMs pela ferramenta EMPRO.....	48
Tabela 7. Comparação dos resultados obtidos pelo COSMIN RoB e pela ferramenta EMPRO.....	49
Tabela 1A. Descrição das PROM avaliadas nos estudos incluídos.....	96

LISTA DE ABREVIATURAS E SÍMBOLOS

% – Porcentagem	do <i>Human Activity Profile</i>
= – Igual	HAP-MAS – Escore de atividade máxima do <i>Human Activity Profile</i>
± – Mais ou menos	HAS – Hipertensão Arterial Sistêmica
> – Maior que	HD – Hemodiálise
< – Menor que	IPAQ – <i>International Physical Questionnaire</i>
≥ – Maior ou igual a	kcal – Quilocaloria
≤ – Menor ou igual a	LILACS – Literatura Latino-Americana e do Caribe em Ciências da Saúde
BIREME – Biblioteca Regional de Medicina	LoPAQ – <i>Low Physical Activity Questionnaire</i>
CKD-PAQ – <i>Chronic Kidney Disease Physical Activity Questionnaire</i>	LTA – <i>Minnesota Leisure Time Activities questionnaire</i>
cm – Centímetro	MET – Equivalente Metabólico da Tarefa
COSMIN – <i>CO</i> n <i>SENS</i> us-based <i>ST</i> andards for the selection of health <i>M</i> easurement <i>I</i> nstruments	min – Minuto
COVID-19 – Doença causada pelo coronavírus 2019	ml – Mililitros
DASI – <i>Duke Activity Status Index</i>	NHANES – National Health and Nutrition Examination Survey
DCNT – Doenças Crônicas Não-Transmissíveis	O ₂ – Oxigênio
DLW – Método da Água Duplamente Marcada (<i>Doubly Labelled Water</i>)	OMS – Organização Mundial da Saúde
DM – <i>Diabetes mellitus</i>	PAR – <i>Stanford 7-Day Physical Activity Recall Questionnaire</i>
DMCI – Diferença Mínima Clinicamente Importante	PASE – <i>Physical Activity Scale for the Elderly</i>
DP – Diálise Peritoneal	PRISMA – <i>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</i>
DPOC – Doença Pulmonar Obstrutiva Crônica	PROM – <i>Patient-Reported Outcome Measures</i>
DRC – Doença Renal Crônica	PROSPERO – <i>International Prospective Register of Systematic Reviews</i>
EMBASE – <i>Excerpta Medica dataBASE</i>	RPAQ – <i>Recent Physical Activity Questionnaire</i>
EMPRO – <i>Evaluating Measures of Patient-Reported Outcomes</i>	SciELO – <i>Scientific Electronic Library Online</i>
ESES – <i>Exercise Self-Efficacy Scale</i>	SUS – Sistema Único de Saúde
FWH – <i>Four Week Physical Activity History Questionnaire</i>	SUP – Sistema Ubiquitina-Proteassoma
GPPAQ – <i>General Practice Physical Activity Questionnaire</i>	TGF – Taxa de Filtração Glomerular
HAP – <i>Human Activity Profile</i>	TRS – Terapia Renal Substitutiva
HAP-AAS – Escore de atividade ajustado	

SUMÁRIO

1. INTRODUÇÃO.....	13
1.1 Doença Renal Crônica.....	13
1.2 Desempenho funcional, atividade física e alterações na DRC.....	15
1.3 Avaliação da atividade física.....	16
1.3.1 Instrumentos de autorrelato.....	17
1.4 Avaliação da atividade física na DRC.....	18
1.5 Avaliação psicométrica das PROMs.....	19
1.5.1 Confiabilidade.....	20
1.5.2 Validade.....	20
1.5.3 Responsividade.....	21
1.5.4 Outras características psicométricas.....	21
1.6 Justificativa.....	22
2. OBJETIVOS.....	23
2.1 Objetivo geral.....	23
2.2 Objetivos específicos.....	23
3. MATERIAIS E MÉTODOS.....	24
3.1 Delineamento do estudo.....	24
3.1.1 Elegibilidade.....	25
3.2 Extração de dados.....	26
3.3 Avaliação de qualidade dos estudos.....	27
3.3.1 <i>Checklist</i> COSMIN de Risco de Viés.....	27
3.3.2 Ferramenta EMPRO.....	28
4. RESULTADOS.....	30
5. DISCUSSÃO.....	50
6. CONCLUSÃO.....	58
7. REFERÊNCIAS.....	59
8. ANEXOS.....	67
8.1 Tabela PRISMA.....	67
8.2 Ficha de extração de dados.....	69
8.3 <i>Checklist</i> COSMIN de Risco de Viés.....	70
8.4 Ferramenta EMPRO.....	78
8.5 Tabela com a descrição das PROMs avaliadas nos estudos incluídos.....	96

1. INTRODUÇÃO

1.1 Doença Renal Crônica

A Doença Renal Crônica (DRC) é definida como alterações na estrutura ou função renal por tempo maior que três meses e com impacto à saúde do indivíduo.¹ A função renal é avaliada por meio da taxa de filtração glomerular (TGF), enquanto os principais marcadores de dano parenquimatoso são alterações eletrolíticas, albuminúria, hematúria de origem glomerular e alterações histológicas.²

Os principais fatores de risco para o desenvolvimento de DRC são *diabetes mellitus* (DM) tanto do tipo 1 quanto do tipo 2, hipertensão arterial sistêmica (HAS), idade, obesidade, histórico pessoal de doença cardiovascular, histórico familiar de DRC, tabagismo e uso de agentes nefrotóxicos.²

A DRC é dividida em seis estágios funcionais, delimitados de acordo com a TGF¹:

- Estágio 1 (G1): caracterizada por função renal preservada (TGF \geq 90 ml/min/1,73cm²), porém com a presença de alterações estruturais e/ou eletrolíticas;
- Estágio 2 (G2): presença de alteração leve na função renal (TGF 60-89 ml/min/1,73cm²);
- Estágio 3a (G3a): presença de alteração leve à moderada na função renal (TGF 59-45 ml/min/1,73cm²);
- Estágio 3b (G3b): presença de alteração moderada à grave na função renal (TGF 44-30 ml/min/1,73cm²);
- Estágio 4 (G4): presença de alteração grave na função renal (TGF 29-15 ml/min/1,73cm²);
- Estágio 5 (G5): estágio terminal, onde já se apresenta falência renal (TGF $<$ 15 ml/min/1,73cm²).

O estágio terminal da DRC é caracterizado pela necessidade de terapia renal substitutiva (TRS), termo que abrange os métodos disponíveis para suprir a função renal. A TRS inclui as modalidades de hemodiálise (HD), diálise

peritoneal (DP) e transplante renal.^{1,2} Dentre as TRS, o transplante é a modalidade que garante melhor qualidade de vida para os indivíduos.³

Acredita-se que a prevalência global da doença esteja em torno de 11,1%, acometendo assim mais de 840 milhões de pessoas no mundo.⁴ No Brasil, a prevalência da DRC na população adulta é de 8,9%, com crescimento anual de 3,6% do estágio terminal da doença. De acordo com a Sociedade Brasileira de Nefrologia, em 2017 havia aproximadamente 100 mil portadores de DRC dependentes de diálise e hemodiálise, sendo que 75% realizam tratamento via Sistema Único de Saúde (SUS).^{5,6}

A DRC acarreta alterações metabólicas, cardiovasculares e musculoesqueléticas, tornando-se assim uma doença de caráter multissistêmico. A complicação mais frequente associada à DRC é a anemia, ocasionada pela redução da produção de eritropoietina pelos fibroblastos peritubulares renais e pela deficiência de ferro.⁷ A redução do transporte de oxigênio aos tecidos pela hemoglobina não consegue ser compensada a longo prazo, ocasionando fadiga, palpitações, dispneia e sobrecarga cardiovascular.⁸

A disfunção renal ocasiona desregulação na concentração sanguínea de eletrólitos. As alterações nos níveis de fósforo, cálcio, sódio e potássio impactam o funcionamento orgânico a nível celular.⁹ Paralelamente, o desequilíbrio na homeostase mineral compromete a deposição mineral óssea e afeta diretamente a qualidade do tecido ósseo.¹⁰

Na DRC em estágio terminal há alta incidência de doenças cardiovasculares, tanto por comorbidades previamente existentes, como por alterações promovidas pela própria doença renal. A desregulação da homeostase mineral também ocasiona deposição de cálcio nas artérias.¹⁰ A hipervolemia decorrente da redução da eficiência renal gera sobrecarga às câmaras cardíacas e grandes vasos, impactando a função do coração a longo prazo.¹¹ As alterações da concentração de potássio também interferem diretamente na contração da musculatura cardíaca.⁹

Quanto às alterações musculoesqueléticas, a chamada sarcopenia urêmica é uma complicação comumente encontrada principalmente no estágio

terminal da doença e está associada a maior mortalidade.¹² A uremia presente nos portadores de DRC ocasiona inflamação e estresse oxidativo, alterando a função mitocondrial da musculatura esquelética. A disfunção mitocondrial ativa vias proteolíticas que estimulam a atrofia muscular.¹³ Concomitantemente, o sistema ubiquitina-proteassoma (SUP), a principal via catabólica de proteínas no citosol, é estimulada pela inflamação crônica e acidose metabólica decorrentes da DRC.⁹

1.2 Desempenho funcional, atividade física e alterações na DRC

A funcionalidade é um termo abrangente que inclui aspectos físicos, psicológicos e sociais do ser humano, e traz em sua avaliação os conceitos de capacidade e desempenho funcional. A capacidade funcional traduz a aptidão do indivíduo em executar tarefas de atividades e/ou participação em um ambiente padronizado; enquanto desempenho funcional descreve o que o indivíduo faz em seu contexto real.¹⁴

Dentro do desempenho funcional, há o conceito de atividade física, definida como qualquer movimento corporal produzido pela musculatura esquelética que resulte em gasto energético superior ao repouso.¹⁵ A Organização Mundial da Saúde (OMS) indica que a atividade física regular é um importante fator de controle e proteção de doenças crônicas não transmissíveis (DCNT).¹⁶

No contexto da DRC, as alterações sistêmicas supracitadas geram declínio funcional, redução da mobilidade e fragilidade.¹³ A redução da atividade física nessa população tem relação diretamente proporcional com a redução da TGF,¹³ contudo mesmo nos estágios iniciais da doença é possível observar impacto nos níveis de atividade física.¹⁷

Os benefícios do estímulo à atividade física nos portadores de DRC já são documentados na literatura. Além do aumento de sobrevida,¹⁸ a atividade física regular aumenta a qualidade de vida e impacta positivamente no controle de comorbidades associadas, como HAS, DM e obesidade.¹⁹ Por fim, a performance funcional influencia nas escolhas terapêuticas na DRC. A aptidão

para o transplante renal não depende tão somente da compatibilidade entre receptor e órgão, mas também do estado clínico e funcional do receptor.^{2,5}

1.3 Avaliação da atividade física

A avaliação do nível de atividade física pode ser realizada de diferentes formas. O método da água duplamente marcada (*doubly labelled water* – DLW) é considerado o padrão-ouro para avaliar o gasto energético, porém o alto custo, intrincada aplicabilidade e demanda de tempo dificultam seu uso na pesquisa clínica.^{20,21} Deste modo, outras ferramentas são comumente utilizadas em estudos.

Dentre os dispositivos testados na avaliação da atividade física, o acelerômetro é considerado como o instrumento mais fidedigno no contexto clínico devido sua acurácia, porém ainda possui disponibilidade limitada.^{21,22} Pedômetros, monitores cardíacos e actigrafia são outros exemplos de dispositivos utilizados que avaliam indiretamente o gasto energético do indivíduo.²¹

Outro método de avaliação da atividade física é através de instrumentos de autorrelato, como diários e questionários.^{21,22} Embora os questionários possuam baixo custo e aplicabilidade mais fácil, diversos fatores influenciam sua acurácia, tais como alterações cognitivas, compreensão do indivíduo, diferenças culturais e linguísticas etc.²³

O DLW apresenta sua unidade de medida do gasto energético em quilocalorias (kcal) por dia.²⁰ Já os dispositivos de avaliação indireta utilizam diferentes tipos de desfechos: contagem de passos por dia, variação de temperatura, alteração dos batimentos cardíacos ao longo do período de avaliação e movimentação do indivíduo.²¹ Os questionários e demais instrumentos de autorrelato costumam quantificar a frequência e intensidade da atividade física ao longo do tempo avaliado e discriminá-las pela estratificação da intensidade da atividade física a partir do dispêndio de energia categorizado em equivalentes metabólicos da tarefa (MET).²²⁻²⁴

1.3.1 Instrumentos de autorrelato

Os instrumentos de autorrelato são chamados também de instrumentos de desfechos relatados pelo paciente (*Patient-Reported Outcomes – PROs*), e são definidos como qualquer informação em desfechos de saúde obtida diretamente do paciente sem qualquer modificação feita por profissionais. Não há a necessidade de que os PROs sejam padronizados ou estruturados.²⁵

Por outro lado, as medidas relatadas pelo próprio paciente (*Patient-Reported Outcome Measures – PROMs*) são questionários estruturados e padronizados que avaliam o estado ou condição de saúde do indivíduo. Os dados quantitativos produzidos pelas PROMs permitem a avaliação e comparação de diferentes grupos, sendo preferencialmente escolhidas entre os instrumentos de autorrelato.^{25,26}

As PROMs são categorizadas em: 1) avaliação de qualidade de vida; 2) avaliação de *status* funcional; 3) avaliação de sintomas e seus impactos; 4) avaliação de comportamentos de saúde; e 5) avaliação da experiência do paciente. Os instrumentos que avaliam a atividade física são classificados como PROMs de avaliação de comportamentos de saúde.²⁵

A avaliação da atividade física por esses instrumentos pode ser através de domínios contextuais (lazer, trabalho etc.) ou tipo de atividade performada (caminhada, corrida, subida de escadas etc.).²⁷ Ela então é categorizada sua frequência, duração e intensidade. A sumarização do resultado, então, pode ser apresentada em variáveis dicotômicas, como, por exemplo, em PROMs que categorizam perfis “ativos” e “inativos”, variáveis contínuas ordinais (geralmente pela intensidade) e variáveis contínuas numéricas, tais como METs ou kcal por dia ou semana ou ainda escores numéricos próprios da ferramenta.²⁷

As PROMs de atividade física são importantes instrumentos na prática clínica atual. Em grandes grupos amostrais, as PROMs apresentam boa acurácia e baixo custo, sendo uma alternativa viável para pesquisas epidemiológicas.^{23,26,27} Um exemplo é o seu uso padronizado nas pesquisas *survey* conduzidas pelo *National Center of Health Statistics*, as chamadas *National Health and Nutrition Examination Survey* (NHANES), que proporcionam

avaliações epidemiológicas periódicas de grandes grupos populacionais estadunidenses.²⁸

Sattler e colaboradores defendem o uso das PROMs em contextos clínicos específicos: monitorização e vigilância dos níveis de atividade física em grandes grupos populacionais; triagem e promoção de atividade física na assistência à saúde; identificar determinantes socioambientais que limitam ou facilitam a atividade física e traçar estratégias; avaliação em larga escala de intervenções de saúde pública etc.²³ Ademais, a possibilidade de aplicação de forma remota do instrumento (via eletrônica, por exemplo) facilitou seu uso no contexto atual de pesquisas durante a pandemia de doença do coronavírus 2019 (COVID-19).^{29,30}

1.4 Avaliação da atividade física na DRC

A avaliação dos níveis de atividade física em portadores de DRC é importante por algumas razões. Em princípio, destaca-se a correlação entre atividade física e mortalidade. Assim como em indivíduos sem doença renal, a inatividade física está associada ao aumento de mortalidade nessa população.^{17,18}

A escolha da melhor ferramenta de avaliação em portadores de DCNT é desafiadora, uma vez que esse grupo populacional apresenta valores de referência menores do que a população geral.³¹ Desse modo, existe o risco de PROMs de abordagem mais generalista apresentarem o chamado “efeito chão”: quando os valores da variável de análise pontuam muito próximos (ou abaixo) do nível inferior de corte, o que faz a ferramenta perder a capacidade de discriminar as diferenças entre esses valores.^{27,31}

Dentro do contexto da DRC, observa-se o uso massivo de PROMs gerais, principalmente o componente físico do questionário *Short Form Health Survey 36* (SF-36), que avalia o *status* funcional. Em PROMs específicas para mensuração do nível de atividade física, observa-se na literatura o uso principalmente do *Human Activity Profile* (HAP) e do *Duke Activity Status Index* (DASI), ambas PROMs não específicas para o perfil renal crônico.³² A base de

dados NHANES também apresenta avaliações de subgrupos populacionais, incluindo portadores de DRC, e utiliza um questionário padronizado para mensuração de atividade física.^{18,33}

Na literatura, há duas PROMs elaboradas especificamente para os portadores de DRC: *Low Physical Activity Questionnaire (LoPAQ)*³⁴ e o *Chronic Kidney Disease Physical Activity Questionnaire (CKD-PAQ)*³⁵. No entanto, o LoPAQ é restrito apenas para os indivíduos dialíticos, enquanto o CKD-PAQ foi desenvolvido recentemente e carece de estudos sobre aplicabilidade.

1.5 Avaliação psicométrica das PROMs

A escolha de uma PROM depende de sua qualidade, e esta é avaliada por suas propriedades de medidas, as chamadas propriedades psicométricas.^{25,36}

A psicometria é o campo de estudo que busca analisar as características matemáticas constantes de dados empíricos,³⁶ conceitualmente tratados como construto. O construto é um atributo subjetivo como, por exemplo, a percepção de esforço. O objetivo do construto teórico é associar os conceitos subjetivos com referências numéricas para quantificação e comparação.²⁵

Para critérios de padronização de avaliação de PROMs e suas propriedades psicométricas, foi criada a iniciativa COSMIN (**C**onsensus-based **S**tandards for the selection of health **M**easurement **I**nstruments). A iniciativa foi desenvolvida por uma equipe multidisciplinar e internacional de especialistas com o objetivo de aprimorar a seleção de instrumentos de medidas clínicas relacionadas à saúde, realizar um consenso sobre quais propriedades de medida devem ser avaliadas e desenvolver normas de como elas devem ser avaliadas em termos de delineamento de estudo e análise estatística.³⁷

As variáveis psicométricas a serem avaliadas em instrumentos de medidas são a confiabilidade, a validade e a responsividade.^{25,36,38-43}

1.5.1 Confiabilidade

A confiabilidade diz respeito ao grau ou qualidade de consistência com que os itens do instrumento medem o atributo proposto sem erro de medida e o grau em que o instrumento permite a reprodução e obtenção de resultados consistentes quando aplicado em diferentes ocasiões. Dentro desta propriedade, temos os seguintes aspectos:³⁹⁻⁴¹

- Consistência interna: grau de correlação e coerência que existe entre os itens de um instrumento ou os itens que compõem uma dimensão (em casos de escalas multidimensionais);
- Confiabilidade intraobservador ou Confiabilidade teste-reteste: a capacidade de obter pontuações semelhantes ao aplicar a ferramenta nos mesmos sujeitos, pelos mesmos examinadores, utilizando o mesmo método, porém em momentos distintos. É afetada pelo tempo transcorrido entre as avaliações;
- Confiabilidade interobservador: grau de concordância entre examinadores diferentes em avaliação realizada com os mesmos sujeitos, utilizando o mesmo instrumento e em uma mesma ocasião. Não é avaliada em instrumentos de autoadministração;
- Erro de medida: erros sistemáticos e aleatórios na pontuação do instrumento que não são atribuídos a reais mudanças do construto avaliado.

1.5.2 Validade

A validade refere-se à capacidade do instrumento em medir o construto proposto e para qual foi delineado.⁴⁰⁻⁴³ Esse critério é composto por cinco propriedades, sendo que se pode avaliar a validade do instrumento através de todas ou de apenas algumas propriedades. São elas:^{40,41}

- Validade de face (lógica): grau em que os itens da ferramenta medem de forma lógica o construto que se pretende medir. Não é um conceito estatístico, uma vez que depende do julgamento de pesquisadores experientes sobre a conveniência dos itens;

- Validade de conteúdo: grau em que os itens do instrumento representam adequadamente o construto. Avalia a estrutura da ferramenta e se os itens e domínios são capazes de representar o construto proposto como um todo;
- Validade de critério: grau em que a pontuação obtida em uma escala é válida considerando o padrão de referência da avaliação do construto. É necessário, então, a comparação da ferramenta estudada com uma ferramenta de referência ou equivalente que seja amplamente aceita, confiável e já validada;
- Validade de construto: grau em que o instrumento reflete adequadamente a teoria subjacente do construto que se propõem medir. Demanda a definição prévia do conteúdo do instrumento a ser validado e da elaboração de um marco teórico-conceitual que permita a interpretação dos resultados;
- Validade convergente e divergente: grau da correlação da pontuação obtida em diferentes instrumentos. A validade convergente avalia a similaridade conceitual entre os instrumentos, de modo que é avaliada quando as ferramentas buscam quantificar o mesmo construto. Já a validade divergente indica a não-associação significativa entre as ferramentas, e é avaliada em instrumentos que não avaliam o mesmo construto.

1.5.3 Responsividade

A responsividade é definida como a capacidade do instrumento detectar mudanças ao longo do tempo no construto medido. Pode ser avaliada tanto entre indivíduos quanto como resposta do mesmo indivíduo.^{39,41,42}

1.5.4 Outras características psicométricas

Embora não sejam consideradas propriedades psicométricas, a interpretabilidade e a utilidade são características importantes dos instrumentos.^{41,42}

A interpretabilidade refere-se ao grau em que os valores obtidos na aplicação do instrumento produzem informações clínicas relevantes ao construto mensurado.⁴²

A utilidade, por sua vez, avalia se a ferramenta é de fácil aplicação, pouco complexa e de baixo custo. Esta característica depende de aspectos como tempo necessário para aplicação, formato, clareza das perguntas/enunciados, necessidade de treinamento dos aplicadores, interpretação de resultados, codificação etc.⁴¹

1.6 Justificativa

A avaliação da atividade física por meio de PROMs é rotineira pela sua fácil aplicabilidade e baixo custo.^{25,26,44} Dentro do contexto de portadores de DRC, há vários estudos que avaliam a atividade física por meio de PROM, porém observa-se uma grande heterogeneidade na escolha de ferramenta.³²

A utilização de PROMs mais generalistas pode ocasionar menor fidedignidade dos resultados em populações específicas devido o “efeito chão” em relações a subgrupos com performance funcional reduzida por condições crônicas.³¹ Deste modo, é necessária a avaliação psicométrica dessas ferramentas nas populações portadoras de DCNT.

A iniciativa COSMIN destaca que a realização de revisões sistemáticas de alta qualidade sobre propriedades de medidas de PROMs é importante para a padronização da escolha de clínicos por ferramentas que sejam embasadas em pesquisas e melhor performance técnica.³⁷

Nesse sentido, considerando o uso rotineiro de questionários de avaliação de atividade física dentro do contexto da DRC,^{32,33} porém sem homogeneidade na escolha, a avaliação das propriedades de medidas de questionários de atividade física na população doente renal crônica se faz necessária para a melhor escolha de instrumento de avaliação.

2. OBJETIVOS

2.1 Objetivo geral

- Identificar, avaliar e sintetizar as propriedades psicométricas das PROMs utilizadas na avaliação de atividade física na população portadora de DRC.

2.2 Objetivos específicos

- Identificar as propriedades psicométricas avaliadas em PROMs de avaliação de atividade física no contexto da DRC a partir da busca sistematizada em bases de pesquisa;
- Avaliar as propriedades psicométricas das PROMs identificadas nos estudos disponíveis na literatura através de duas ferramentas específicas;
- Sintetizar os resultados das avaliações das propriedades psicométricas das PROMs pelas duas ferramentas e comparar a concordância entre elas.

3. MATERIAIS E MÉTODOS

3.1 Delineamento do estudo

O presente estudo é uma revisão sistemática realizada a partir das diretrizes da PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*)⁴⁵ e registrada na PROSPERO (*International Prospective Register of Systematic Reviews*) sob o ID CRD42022312143. A tabela PRISMA preenchida está disponível no Anexo 1.

Dois revisores independentes (TCS e LMMS) realizaram a busca e seleção dos estudos de forma paralela. As buscas foram realizadas nas bases de dados PubMed, SciELO, BIREME, LILACS e EMBASE, nos idiomas inglês, português e espanhol. Por tratar-se de revisão sistemática de propriedades de medida de questionários, a pesquisa se baseou na iniciativa COSMIN⁴² para traçar a estratégia de pergunta de pesquisa (Figura 1) e estratégia de busca (Quadro 1).

Construto	Atividade física
População	Portadores de DRC
Tipo de instrumento	Questionários
Propriedades de medida	<ul style="list-style-type: none">• Validade• Confiabilidade• Responsividade

Figura 1. Estratégia de elaboração de pergunta de pesquisa conforme iniciativa COSMIN.

<p><i>Chronic Kidney Disease</i> OR <i>Chronic Renal Insufficiency</i> OR <i>Chronic Renal Disease</i> OR <i>Dialysis</i></p>	<p>AND</p>	<p><i>Physical Activity</i> OR <i>Physical Activities</i></p>	<p>AND</p>	<p><i>Psychometrics</i> OR <i>Patient Outcome Assessment</i> OR <i>Outcome Assessment, Health Care</i> OR <i>Reproducibility of Results</i> OR <i>Reliability of Results</i> OR <i>Validity of Results</i> OR <i>Reliability and Validity</i> OR <i>Test Retest Reliability</i> OR <i>Validity, Face</i> OR <i>Finding Reproducibility</i> OR <i>Validation studies</i> OR <i>Cross Cultural Comparison</i></p>
--	-------------------	--	-------------------	--

Quadro 1. Estratégia de busca utilizada para revisão sistemática.

Os resumos dos artigos encontrados nas buscas foram incluídos e armazenados na base de dados RAYYAN⁴⁶, onde foram triados de acordo com a seguinte ordem:

- Leitura do título do estudo;
- Leitura do resumo do estudo;
- Leitura do estudo na íntegra.

Foi então realizada uma busca indireta nas referências bibliográficas dos estudos incluídos, seguindo a mesma triagem supracitada.

Em caso de discordância entre os avaliadores da seleção de um ou mais artigos, foi realizada a resolução por consenso ou utilizada a avaliação de um terceiro revisor independente (GAS).

3.1.1 Elegibilidade

Foram incluídos na revisão:

- Estudos observacionais de avaliação de propriedade de medida publicados em periódicos revisados por pares;
- Estudos nos idiomas inglês, espanhol e português;
- Estudos cuja amostra é composta por doentes renais crônicos em qualquer estágio da doença;
- Estudos cuja amostra é composta por indivíduos com idade maior ou igual a 18 anos;
- Estudos cuja variável de desfecho principal seja a atividade física;
- Estudos com avaliação de no mínimo uma das seguintes propriedades dos instrumentos: validade, confiabilidade e/ou responsividade.

Foram excluídos da revisão:

- Estudos incompletos ou que não seja possível o acesso na íntegra;
- Resumos de anais de congresso;
- Revisões sistemáticas e da literatura;
- Estudos cuja avaliação da variável de desfecho não seja por meio de PROMs.

3.2 Extração de dados

Os dados de cada artigo selecionado foram extraídos pelos revisores de

forma independente. A fim de padronizar o processo, foi elaborada uma tabela contendo as informações a serem retiradas. Essa tabela foi baseada nas recomendações descritas na diretriz PRISMA⁴⁵ e em elementos da tabela de extração de dados de ferramentas exemplificadas no portal da iniciativa COSMIN⁴², de modo a direcionar a extração para os dados de propriedades psicométricas.

As informações retiradas foram:

- Identificação do artigo: nome dos autores, ano de publicação, país de origem, idioma do artigo e periódico de publicação;
- Detalhes de materiais e métodos: delineamento do estudo, tamanho da amostra, perfil dos participantes e instrumentos avaliados;
- Psicometria: propriedades psicométricas avaliadas, estatística utilizada, intervalo de tempo utilizado (em avaliações de confiabilidade teste-reteste), teste utilizado como referência (em avaliações de validade) e pontos de corte dos testes estatísticos;
- Desfechos primários e secundários.

A tabela de extração utilizada está disponível no Anexo 2.

3.3 Avaliação de qualidade dos estudos

A qualidade dos estudos foi avaliada a partir do *Checklist* COSMIN de Risco de Viés (COSMIN *Risk of Bias* – RoB)⁴⁷⁻⁵⁰ e da ferramenta EMPRO (*Evaluating Measures of Patient-Reported Outcomes*).^{51,52} A escolha pelo uso de ambas as ferramentas foi pautada em estudos metodológicos prévios que verificaram a performance de diferentes ferramentas de avaliação de propriedades psicométricas, inclusive dentro do contexto de instrumentos de avaliação de saúde.^{36,53}

A aplicação dos instrumentos foi realizada de forma independente por cada revisor e os casos de discordância foram resolvidos por meio de consenso ou pela avaliação de um terceiro revisor.

3.3.1 Checklist COSMIN de Risco de Viés

O *checklist* COSMIN de Risco de Viés (COSMIN RoB) foi elaborado a partir de um estudo Delphi da iniciativa, com a finalidade de padronizar e otimizar a escolha de ferramentas de medidas clínicas. Ele possui 10 seções, contendo de 4 a 35 itens cada, e cuja avaliação se baseia nas seguintes taxonomias:^{42,47,48,54}

- Confiabilidade, que contém as propriedades de medida consistência interna, confiabilidade e medida de erro;
- Validade, que contém as propriedades de validade de conteúdo (incluindo validade de face), validade de critério e validade de construto (que inclui validade estrutural, testagem de hipótese e validade transcultural);
- Responsividade, que contém as propriedades de medida de responsividade.

Embora a interpretabilidade tenha sido reconhecida como um importante aspecto de medida de instrumentos, ela não é considerada como uma propriedade de medida.^{42,48,54}

Cada aspecto de propriedade de medida é avaliado em “+” (suficiente), “-” (insuficiente) ou “?” (indeterminado). Ao final, a propriedade de medida é classificada em um sistema que leva em consideração os seguintes fatores: risco de viés, inconsistência, imprecisão e ambiguidade. A partir disso, a propriedade de medida é classificada como qualidade “Alta”, “Moderada”, “Baixa” ou “Muito Baixa”.^{42,48,54}

O COSMIN RoB é amplamente utilizado nas revisões sistemáticas de avaliação de psicométrica de PROMs. A ferramenta completa e seus manuais estão disponíveis para acesso e uso gratuitamente na língua inglesa através do portal <https://www.cosmin.nl/>. A versão utilizada neste estudo está presente no Anexo 3.

3.3.2 Ferramenta EMPRO

A EMPRO foi desenvolvida por Valderas e col. em 2008⁵² com o objetivo de avaliar a qualidade de *PROs*. A ferramenta é composta pelo total de 39 itens distribuídos em 8 atributos, nomeadamente:⁵²

- Modelo conceitual e de medição: conceitos e população que se pretende avaliar;
- Confiabilidade;
- Validade;
- Responsividade;
- Interpretabilidade;
- Sobrecarga: demandas da administração e resposta do instrumento, como tempo despendido e empenho em aplicação;
- Modelos alternativos de administração: autoadministrado ou por avaliador, administrado via telefônica, eletrônica ou pessoalmente etc.;
- Adaptações linguísticas e transculturais: versões traduzidas equivalentes.

O conteúdo de cada item encontra-se sumarizado, e o resultado depende do grau de concordância do estudo avaliado com este conteúdo. Esta concordância é descrita em uma escala Likert de 4 pontos, onde “4” corresponde a “Concorda totalmente” e 1 a “Discorda totalmente”. Há também a opção de “Sem informação” para casos de informação insuficiente, além de 5 itens que permitem a resposta “Não aplicável”.⁵²

A pontuação de cada atributo passa por transformação logarítmica, podendo atingir pontuação de “0,0” (pior pontuação) a “100,0” (melhor pontuação), sendo valores iguais ou maiores a “50,0” considerados aceitáveis para o atributo.⁵¹

Embora ao final a escala permita sumarizar os resultados em indicação de uso ou não da *PRO* avaliada, no presente estudo não utilizamos esta sumarização, uma vez que o objetivo deste estudo foi avaliar a qualidade dos artigos e das propriedades psicométricas das *PROMs* pesquisadas.

A ferramenta EMPRO necessita do pedido de licença através do portal <https://www.bibliopro.org/>, e está disponível nos idiomas inglês e espanhol. Ainda existem poucos estudos publicados utilizando-a e nenhum até o momento avaliando questionários de atividade física. A ferramenta completa e as instruções para pontuação estão apresentadas no Anexo 4.

4. RESULTADOS

A busca nas bases de pesquisa foi realizada entre janeiro de 2022 e julho de 2022, com atualização em maio de 2023 e em novembro de 2023. Foram encontrados no total 13638 artigos, dentre os quais 13 foram selecionados para inclusão na revisão sistemática. Na busca secundária das referências bibliográficas dos artigos incluídos, houve a triagem de 520 artigos, com a inclusão de outros 2 artigos não encontrados na busca inicial, totalizando ao final 15 artigos selecionados para avaliação. Os detalhes da busca, seleção e identificação estão descritos em fluxograma conforme preconizado pela PRISMA (Figura 2).

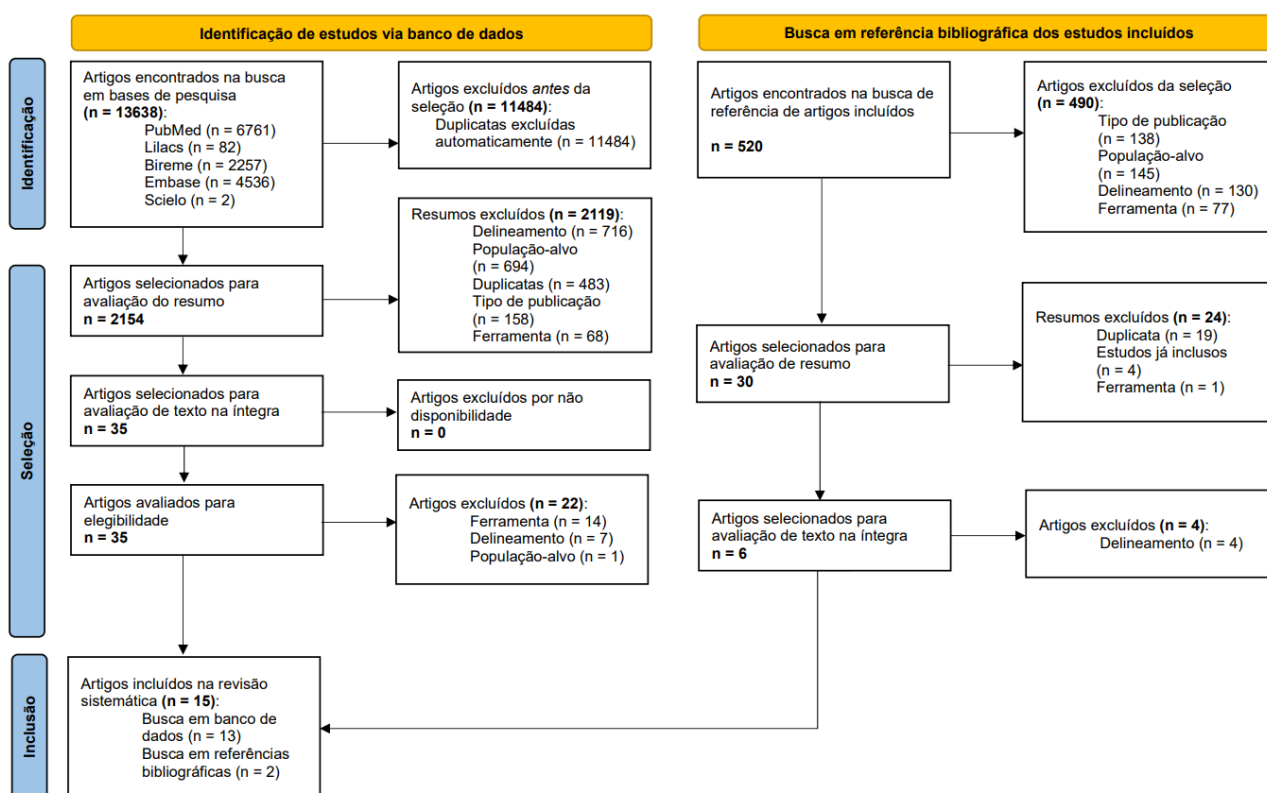


Figura 2. Fluxograma PRISMA do processo de seleção dos artigos incluídos na revisão.

Todos os estudos foram publicados em língua inglesa, no entanto a origem deles é diversa: 4 artigos (26%) são oriundos dos Estados Unidos da América^{34,55–57}, 4 artigos do Reino Unido^{35,58–60}, 2 artigos (13%) da China^{61,62}, 2 do Canadá^{63,64}, 1 artigo (6%) oriundo do Brasil⁶⁵, 1 do Irã⁶⁶ e 1 da Austrália⁶⁷. O ano de publicação variou entre 2001 e 2022, sendo os anos de 2015 e 2019 apresentando maior pico de publicação, com 3 artigos (23%) em cada ano. A distribuição das publicações pelo tempo é detalhada no Gráfico 1.



Gráfico 1. Distribuição das publicações incluídas ao longo dos anos.

Foram encontradas avaliações de propriedades psicométricas de 11 PROMs distintas de avaliação de atividade física, das quais 9 são ferramentas não específicas para doença renal e 2 específicas para portadores de DRC. O Gráfico 2 demonstra o número de estudos em que cada PROM foi avaliada.

Em relação às PROMs específicas para DRC, foi encontrada avaliação psicométrica das anteriormente comentadas CKD-PAQ³⁵ e LoPAQ^{34,55,61}. Já nas PROMs não específicas, foi encontrada avaliação psicométrica do DASI^{60,64}, *Exercise Self-Efficacy Scale (ESES)*⁶⁶, *Four Week Physical History Questionnaire (FWH)*⁵⁷, *General Practice Physical Activity Questionnaire (GPPAQ)*⁵⁹, HAP^{56,57,63,67}, *International Physical Questionnaire (IPAQ)*^{57,62,65}, *Stanford 7-Day Physical Activity Recall (PAR)*^{56,58}, *Physical Activity Scale for the Elderly (PASE)*^{56,57} e da *Recent Physical Questionnaire (RPAQ)*⁵⁸.

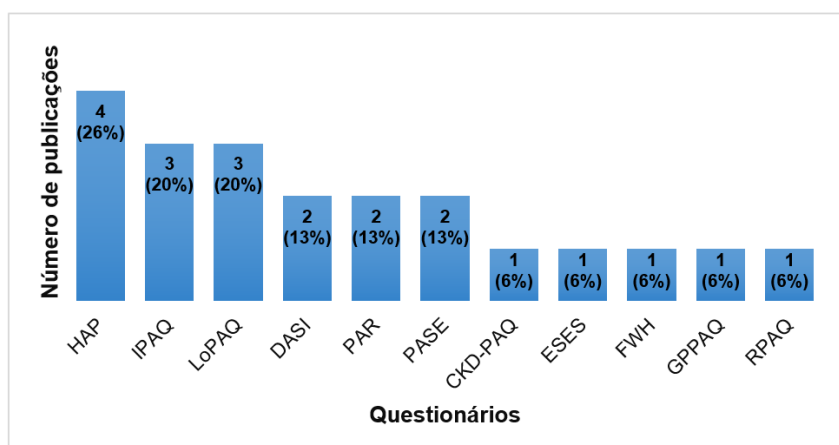


Gráfico 2. Número de estudos em que cada PROM foi avaliada.

A descrição sobre cada PROM encontrada está mais detalhada na Tabela 1A disponível no Anexo 5. O CKD-PAQ, FWH, LoPAQ, PAR e RPAQ sumarizam

o resultado de acordo com a estimativa indireta de gasto energético^{34,35,68-72}, enquanto DASI, ESES, HAP e PASE apresentam escore numérico próprio^{70,73-76} e o GPPAQ e IPAQ categorizam o resultado conforme conceitos próprios de atividade e inatividade.^{77,78}

A descrição dos estudos incluídos na análise está apresentada na Tabela 1. A sumarização dos resultados das propriedades psicométricas por PROM está descrita na Tabela 2. A média de idade dos participantes nos estudos está na faixa de 50 e 60 anos. Todos os estágios da DRC foram englobados nos estudos, porém com maior ênfase aos estágios 3b a 5, caracterizando níveis de maior acometimento clínico no perfil estudado. Os questionários ESES e LoPAQ foram as PROMs que apresentaram avaliação do maior número de propriedades na população portadora DRC.

Das propriedades psicométricas avaliadas, a validade de critério foi a mais estudada, sendo avaliada em 9 (60%) estudos e em 8 (81%) PROMs. Os instrumentos de comparação para a validade de critério foram em sua maioria dispositivos como actigrafia e acelerômetros, com 3 estudos (20%) cada. Apenas 1 (6%) estudo utilizou o método da água duplamente marcada⁵⁸, considerado padrão-ouro, para validação. Na avaliação de validade de construto, ambas as PROMs específicas para a população portadora de DRC avaliaram a propriedade por meio de testagem de hipótese com ferramentas comparativas^{34,35,55}, enquanto o ESES apresentou avaliação de validade estrutural⁶⁶.

Na avaliação de validade de critério, a maior parte das PROMs apresentou correlação com valores abaixo de 0,7, com exceção da comparação entre o questionário HAP e acelerômetro⁵⁶. O questionário GPPAQ apresentou bons valores de especificidade e acurácia em comparação com acelerômetro, porém valor mediano de sensibilidade⁵⁹. Na avaliação de validade de construto em comparação com acelerômetro, o questionário CKD-PAQ apresentou coeficientes de correlação de 0,45 e 0,75³⁵.

Por outro lado, nas avaliações das propriedades de confiabilidade – tanto teste-reteste e consistência interna – apresentaram resultados valores de correlação maiores que 0,7, com exceção do questionário LoPAQ por falta de

informações estatísticas⁶¹.

Tabela 1. Características metodológicas dos estudos inclusos.

Estudo	Ano	Delineamento do estudo	População avaliada	n do estudo	PROM(s) utilizada(s)	Propriedades psicométricas avaliadas	Teste comparativo	Intervalo de reavaliação	Estatística utilizada	Pontos de corte	Desfechos	
Hadjiioannou e col. ⁶⁰	2019	Ensaio clínico pragmático controlado, não-randomizado e unicêntrico	DRC não-dialítico	40 indivíduos (total)	DASI	Confiabilidade - Absoluta - Teste-reteste	Não se aplica	16 dias	Confiabilidade absoluta: EPM e DMCI	Não descritos	Grupo não-dialítico: EPM = 3,95 DMCI = 9,21 CCI = 0,9 (0,75-0,96)	
			TGF 33,0 ± 18,1 ml/min.1,73m ²	12 transplantados								
			DRC dialítico	28 indivíduos (avaliados)								
			TGF 6 ± 4,12 ml/min.1,73m ²	56,7 anos 20 do sexo masculino				Confiabilidade teste-reteste: CCI		Grupo dialítico: EPM = 3,67 DMCI = 8,57 CCI = 0,87 (0,57-0,97)		
			Transplantados (excluídos da avaliação)	Amostra não foi calculada								
Hatef e col. ⁶⁶	2018	Estudo observacional transversal	DRC dialítico (HD)	404 indivíduos	ESES (versão em idioma persa)	Validade - Face - Conteúdo - Construto (estrutural) - Convergente	Não se aplica	Não se aplica	Validade de face: avaliação qualitativa e quantitativa (Escore de Impacto de Cada Item)	Escore de impacto > 1,5 para cada item	Escore de impacto de cada item > 1,5 CVR e CVI não descritos KMO = 0,91	
				56,8 ± 12,04 anos								58% do sexo masculino
				202 selecionados para avaliar as propriedades de estatística exploratória e 202 selecionados para a análise confirmatória								
				A amostra foi calculada								
								Validade de conteúdo: CVR e CVI	CVR > 0,49			
								Validade de Construto: AFE e KMO para avaliar adequação da amostra	KMO > 0,70			
								Validade convergente: AVE	AFE: - x ² > 0,05 - AGFI > 0,8 - RMSEA bom > 0,08 ruim < 0,1			
								Consistência interna: Coeficiente α de Cronbach, Coeficiente Ω de McDonald e Coeficiente θ	AVE > 0,5			
								Coeficiente α = 0,94 Coeficiente θ = 0,96 Coeficiente Ω = 0,91				
								Confiabilidade de construto: Coeficiente α de Cronbach	Coeficientes > 0,7			
										Confiabilidade de construto = 0,894		

Dados numéricos apresentados como Média ± Desvio Padrão ou %

AFE: Análise Fatorial Exploratória; AGFI: Índice de Qualidade de Ajuste (*Adjusted Goodness-of-Fit Index*); AVE: Variância Média Extraída (*Average Variance Extracted*); CCI: Coeficiente de Correlação Intraclasse; CVI: Índice de Validade de Conteúdo (*Content Validity Index*); CVR: Razão de Validade de Conteúdo (*Content Validity Ratio*); DMCI: Diferença Mínima Clinicamente Importante; DRC: Doença Renal Crônica; EPM: Erro Padrão de Medida; HD: Hemodiálise; KMO: teste *Kaiser-Meyer-Olkin*; RMSEA: Raiz Quadrada Média do Erro de Aproximação (*Root Mean Square Error of Approximation*); TGF: Taxa de Filtração Glomerular; x²: Qui-Quadrado.

Tabela 1. Características metodológicas dos estudos incluídos. (Continuação)

Estudo	Ano	Delineamento do estudo	População avaliada	n do estudo	PROM(s) utilizada(s)	Propriedades psicométricas avaliadas	Teste comparativo	Intervalo de reavaliação	Estatística utilizada	Pontos de corte	Desfechos			
Huang e col. ⁶¹	2021	Estudo observacional transversal multicêntrico	DRC dialítico (HD)	85 indivíduos (total)	LoPAQ (versão em mandarim)	Validade - Conteúdo - Critério	Actigrafia via <i>ActiGraph</i> GT3X+ por 7 dias consecutivos	7 dias	Validade de conteúdo: CVR e CVI	CVI > 0,78 CVR > 0,75	CVI = 0,91 CVR = 0,98 <i>Spearman</i> = 0,44* (gasto energético total) 0,47* (tempo de caminhada) 0,22** (tempo sedentário)			
				62,3 ± 11,8 anos 48 do sexo masculino								Confabilidade interna - Teste-reteste	Validade de critério: Correlação de <i>Spearman</i> e teste de <i>Bland-Altman</i>	CCI excelente > 0,75 boa = 0,60 – 0,74 aceitável = 0,40 – 0,59 ruim < 0,40
Johansen e col. ⁵⁶	2001	Estudo observacional transversal e multicêntrico	DRC dialítico (HD)	39 indivíduos	HAP - MAS: escore de atividade máxima - AAS: escore de atividade ajustado	Validade de critério	Acelerômetro <i>TriTrac-R3D</i> utilizado por 7 dias consecutivos	4 meses	Validade de critério: Coeficiente de Correlação de <i>Pearson</i>	Não descritos	HAP-MAS: Correlação acelerômetro = 0,78* Correlação performance = 0,72*			
				52 ± 16 anos 26 do sexo masculino								Confabilidade teste-reteste	Performance física através de avaliação de tempo para atividades	HAP-AAS: Correlação acelerômetro = 0,73* Correlação performance = 0,75* Confabilidade teste-reteste = 0,91
				29 realizaram avaliação teste-reteste								PAR	Confabilidade teste-reteste: não especificado	PASE: Correlação acelerômetro = 0,66* Correlação performance = 0,50*
				80 indivíduos saudáveis para comparativo (literatura prévia)	PASE						PAR: Correlação acelerômetro = 0,59* Correlação performance = 0,36**			
				Amostra não foi calculada										

Dados numéricos apresentados como Média ± Desvio Padrão

CCI: Coeficiente de Correlação Intraclasse; CVI: Índice de Validade de Conteúdo (*Content Validity Index*); CVR: Razão de Validade de Conteúdo (*Content Validity Ratio*); DRC: Doença Renal Crônica; HD: Hemodiálise.

* Correlação significativa ao nível de 0,01; ** Correlação significativa a nível 0,05.

Tabela 1. Características metodológicas dos estudos inclusos. (Continuação)

Estudo	Ano	Delineamento do estudo	População avaliada	n do estudo	PROM(s) utilizada(s)	Propriedades psicométricas avaliadas	Teste comparativo	Intervalo de reavaliação	Estatística utilizada	Pontos de corte	Desfechos
Johansen e col. ³⁴	2015	Estudo observacional transversal e multicêntrico	DRC dialítico (HD)	68 indivíduos 59,0 ± 14 anos 59% do sexo masculino	LoPAQ	Validade - Construto (testagem de hipótese) - Convergente	<i>Short Form (36) Health Survey (SF-36)</i> , componente físico <i>Short Physical Performance Battery (SPPB)</i> Questionário <i>Minnesota Leisure Time Activities (LTA)</i>	Não se aplica	Coefficiente de Correlação de <i>Spearman</i>	Não descritos	<i>Spearman</i> LTA (atividade reportada) = 0,62* <i>Spearman</i> LTA (gasto energético caminhando) = 0,58* <i>Spearman</i> SF-36 = 0,64* <i>Spearman</i> SPPB = 0,51*
Kittiskulnam, Sheshadri e Johansen ⁵⁵	2019	Estudo observacional transversal e multicêntrico	DRC dialítico (HD e DP)	60 indivíduos (total) 58,0 ± 12,7 anos 78,3% do sexo masculino 48 indivíduos em HD e 12 em DP	LoPAQ	Validade - Construto (testagem de hipótese) - Critério	<i>Short Form (36) Health Survey (SF-36)</i> , componente físico <i>Short Physical Performance Battery (SPPB)</i> Pedômetro <i>Accusplit AE 120 TriTrac-R3D</i> utilizado por 7 dias consecutivos	Não se aplica	Coefficiente de Correlação de <i>Pearson</i>	Não descritos	<i>Pearson</i> pedômetro (gasto energético caminhando) = 0,53* <i>Pearson</i> pedômetro (atividade total) = 0,35* <i>Pearson</i> SF-36 = 0,29** <i>Pearson</i> SPPB = 0,26** Sem diferença estatística entre as modalidades de diálise
Lou e He ⁶²	2019	Estudo observacional transversal multicêntrico	DRC dialítico (HD)	320 indivíduos (total) 58,6 anos 200 do sexo masculino 103 indivíduos realizaram avaliação de teste-reteste	IPAQ versão curta (em mandarim)	Validade de critério Confiabilidade teste-reteste	Pedômetro OMRON HJ-328 utilizado por 7 dias consecutivos	7 dias	Validade de critério: Coefficiente de Correlação de <i>Spearman</i>	Correlação de <i>Spearman</i> : grande: 0,8 moderada: 0,5 pequena: 0,2 Teste-reteste: CCI	Atividade vigorosa: <i>Spearman</i> = 0,28* CCI = 0,99* Atividade moderada: <i>Spearman</i> = 0,43* CCI = 0,62* Caminhada: <i>Spearman</i> = 0,407* CCI = 0,90* Total: <i>Spearman</i> = 0,561* CCI = 0,84*

Dados numéricos apresentados como Média ± Desvio Padrão ou %

CCI: Coeficiente de Correlação Intraclasse; DP: Diálise Peritoneal; DRC: Doença Renal Crônica; HD: Hemodiálise.

* Correlação significativa ao nível de 0,01

Tabela 1. Características metodológicas dos estudos incluídos. (Continuação)

Estudo	Ano	Delimitação do estudo	População avaliada	n do estudo	PROM(s) utilizada(s)	Propriedades psicométricas avaliadas	Teste comparativo	Intervalo de reavaliação	Estatística utilizada	Pontos de corte	Desfechos
Overend e col. ⁶³	2010	Estudo observacional de teste-reteste e unicêntrico	DRC dialítico (HD)	25 indivíduos (total) 67,2 anos 14 do sexo masculino Amostra foi calculada	HAP (MAS e AAS)	Confiabilidade - Absoluta - Teste-reteste	Não se aplica	7 dias	Confiabilidade absoluta: EPM e DMCI Confiabilidade teste-reteste: CCI	EPM calculado a partir do desvio padrão e CCI DMCI calculado a partir do EPM CCI > 0,8	HAP-MAS: EPM = 5,5 DMCI = 15,1 CCI = 0,76 (0,53-0,89) HAP-AAS: EPM = 4,1 DMCI = 11,4 CCI = 0,92 (0,83-0,97)
Ravani e col. ⁶⁴	2012	Estudo observacional transversal e unicêntrico	DRC não-dialítico (estágio 3 e 4)	43 indivíduos (total) 58,0 ± 12,7 anos 78,3% do sexo masculino 23 indivíduos (avaliação de confiabilidade) 60 (±14) anos 11 de sexo feminino TGF 35,5 ± 15,0 ml/min.1,73m ² 20 indivíduos (avaliação de validade) 59 (±13) anos 8 do sexo feminino TGF 36,8 ± 10,8 ml/min.1,73m ² Amostra foi calculada	DASI	Confiabilidade teste- reteste Validade de critério	Teste Cardiopulmonar de Exercício (TCPE)	6 meses	Confiabilidade teste-reteste: CCI Validade de critério: CCC	Não descritos	CCC = 0,67 CCI = 0,71 (estágio 3) CCI = 0,81 (estágio 4)

Dados numéricos apresentados como Média ± Desvio Padrão ou %

CCI: Coeficiente de Correlação Intraclasse; CCC: Coeficiente de Correlação de Concordância; DMCI: Diferença Mínima Clinicamente Importante; DRC: Doença Renal Crônica; EPM: Erro Padrão de Medida; HD: Hemodiálise; TGF: Taxa de Filtração Glomerular.

Tabela 1. Características metodológicas dos estudos incluídos. (Continuação)

Estudo	Ano	Delineamento do estudo	População avaliada	n do estudo	PROM(s) utilizada(s)	Propriedades psicométricas avaliadas	Teste comparativo	Intervalo de reavaliação	Estatística utilizada	Pontos de corte	Desfechos
Robinson-Cohen e col. ⁵⁷	2013	Estudo prospectivo de coorte unicêntrico	DRC não-dialítico (TGF 42 ± 15 ml/min.1,73m ²)	46 indivíduos (total) 55 ± 11 anos 54% do sexo masculino	FWH	Validade de critério	Actigrafia via <i>ActiGraph GT3X+</i> por 14 dias consecutivos	Não se aplica	Coeficiente de Correlação de <i>Pearson</i>	Não descritos	PASE = 0,36*
				Amostra não foi calculada	HAP (MAS e AAS)						HAP-MAS = 0,56*
Rosa e col. ⁶⁵	2015	Estudo observacional transversal multicêntrico	DRC dialítico (HD)	40 indivíduos (total) 54 ± 16 anos 21 do sexo feminino	IPAQ versão curta (em idioma português brasileiro)	Validade de critério	Actigrafia via <i>ActiGraph GT3X+</i> por 8 dias consecutivos	Não se aplica	Coeficiente de Correlação de <i>Spearman</i>	Não descritos	FWH = 0,28*
				Amostra não foi calculada							PASE
Sridharan e col. ⁵⁸	2015	Estudo prospectivo transversal	DRC não-dialítico (TGF 44,5 ± 34,0 ml/min.1,73m ²)	40 indivíduos (total) 54,1 ± 17,4 anos 22 do sexo masculino	PAR RPAQ	Validade de critério	Medida de gasto energético: Gasto energético de repouso (REE) a partir de Calorimetria via <i>Sensormedics VMax series 29n metabolic cart.</i>	Não se aplica	Teste de <i>Bland-Altman</i>	Não descritos	Atividade leve = 0,34**
				21 indivíduos no estágio 1-3 19 indivíduos no estágio 4-5 Amostra não foi calculada							Gasto energético total (TEE) a partir do Teste de Água Duplamente Marcada.
											Atividade total = 0,41**
											Correlação TEE com RPAQ = 0,57
											Correlação TEE com PAR = 0,23

Dados numéricos apresentados como Média ± Desvio Padrão ou %

DRC: Doença Renal Crônica; EPM: Erro Padrão de Medida; HD: Hemodiálise; TGF: Taxa de Filtração Glomerular.

* Correlação significante ao nível de 0,01; ** Correlação significante a nível 0,05.

Tabela 1. Características metodológicas dos estudos incluídos. (Continuação)

Estudo	Ano	Delineamento do estudo	População avaliada	n do estudo	PROM(s) utilizada(s)	Propriedades psicométricas avaliadas	Teste comparativo	Intervalo de reavaliação	Estatística utilizada	Pontos de corte	Desfechos
Sridharan e col. ³⁵	2022	Estudo longitudinal de coorte	DRC todos os estágios	Fase de desenvolvimento: 266 indivíduos 40 para avaliação qualitativa 58,3 ± 15,1 anos 57,6% do sexo masculino Avaliação final: 523 indivíduos 60,8 ± 16,1 anos 63,7% do sexo masculino Amostra foi calculada	CKD-PAQ - Valor de MET - Valor de TEE	Validade - Construto (testagem de hipótese) - Convergente	Acelerômetro <i>ActiGraph GT9X+</i> Link por 7 dias consecutivos Questionário RPAQ	Não se aplica	Validade de construto: CCI e teste de <i>Bland-Altman</i> Validade de critério e convergente: CCI e teste de <i>Bland-Altman</i>	Não descritos	Fase de desenvolvimento: CCI (acelerômetro) para TEE = 0,75* CCI (RPAQ) para TEE = 0,91* CCI (acelerômetro) para MET = 0,45* CCI (RPAQ) para MET = 0,90* Fase final: CCI (RPAQ) para TEE = 0,92* CCI (RPAQ) para MET = 0,83*
Wellard, Sally ⁶⁷	2003	Estudo observacional transversal	DRC dialítico (HD)	65 indivíduos 62 ± 13,4 anos 47 do sexo masculino Amostra não foi calculada	HAP (AAS)	Validade convergente	Questionário <i>Sickness Impact Profile (SIP)</i>	Não se aplica	Coefficiente de Correlação de <i>Spearman</i>	Não descritos	SIP (componente físico) = 0.81* SIP (total) = 0.76*
Wilkinson e col. ⁵⁹	2020	Análise secundária do banco de dados de estudos prévios do grupo de pesquisa	DRC todos os estágios	40 indivíduos (11 para avaliação de confiabilidade teste-reteste) 62,5 ± 11,1 anos 21 do sexo feminino Embora seja amostra de conveniência, foi realizado cálculo amostral	GPPAQ GPPAQ-caminhada	Validade de critério Confiabilidade teste-reteste	Acelerômetro triaxial <i>GENEActiv</i> por 7 dias consecutivos	6 semanas	Validade de critério: sensibilidade, especificidade, valor preditivo positivo (VPP) e valor preditivo negativo (VPN) Confiabilidade teste-reteste: Índice Kappa Ponderado kw)	Não descritos	GPPAQ: Sensibilidade = 54,5% Especificidade = 96,6% VPP = 85,7% VPN = 84,9% Acurácia = 85% GPPAQ-Caminhada: Sensibilidade = 50,0% Especificidade = 96,4% VPP = 85,7% VPN = com 81,8% Acurácia = 82,5% kw = 0,74

Dados numéricos apresentados como Média ± Desvio Padrão

CCI: Coeficiente de Correlação Intraclasse; DRC: Doença Renal Crônica; HD: Hemodiálise; MET: Equivalente Metabólico de Tarefa; TEE: Gasto Energético Total; TGF: Taxa de Filtração Glomerular.

* Correlação significativa ao nível de 0,01; ** Correlação significativa a nível 0,05.

Tabela 2. Sumarização dos resultados das propriedades psicométricas por PROM.

PROM	Características da amostra			Validade de construto (estrutural)	Validade de construto (testagem de hipótese)	Validade de conteúdo	Propriedades psicométricas					Erro de medida
	n	Idade (anos)	Estadiamento da doença				Validade de critério	Validade convergente	Validação transcultural	Consistência interna	Confiabilidade	
CKD-PAQ ³⁵	789	58,3 – 60,8	1 à 5	-	n = 266 ³⁵ CCI (Acc) = 0,45 [†] CCI (Acc) = 0,75 [#] CCI (PROM) = 0,90 [†] CCI (PROM) = 0,91 [#]	-	-	n = 523 ³⁵ CCI (PROM) = 0,83 [†] CCI (PROM) = 0,92 [#]	-	-	-	-
DASI ^{60,64}	71	56,7 – 58,0	3b à 5	-	-	-	n = 20 ⁶⁴ CCC (TCPE) = 0,67	-	-	-	n = 28 ⁶⁰ CCI _{TR} = 0,9 ^a CCI _{TR} = 0,87 ^b n = 23 ⁶⁴ CCI _{TR} = 0,71 [*] CCI _{TR} = 0,81 [§]	n = 28 ⁶⁰ EPM = 3,95 ^a EPM = 3,67 ^b DMCI = 9,21 ^a DMCI = 8,57 ^b
ESES ⁶⁶	404	56,8 (±12,04)	5	n = 202 ⁶⁶ x ² = 106,70 AGFI = 0,76 RMSEA = 0,06	-	n = 202 ⁶⁶ CVR e CVI não descritos	-	n = 202 ⁶⁶ AVE = 0,63	Realizada validação no idioma persa: - Tradução - Retradução - Revisão técnica - Avaliação psicométrica ⁶⁶	n = 202 ⁶⁶ Coef α = 0,94 Coef Ω = 0,96 Coef θ = 0,91	n = 202 ⁶⁶ Coef α = 0,89	-
FWH ⁵⁷	46	55 (±11)	3b à 4	-	-	-	n = 46 ⁵⁷ Crr (actigrafia): r _p = 0,28	-	-	-	-	-

Dados numéricos de idade apresentados em Média e Desvio Padrão.

[†]Valores referentes ao equivalente metabólico (MET) pelo CKD-PAQ; [#]Valores referentes ao gasto energético total (TEE) pelo CKD-PAQ; ^aValores referentes ao grupo não-dialítico; ^bValores referentes ao grupo dialítico;

^{*}Valores referentes aos pacientes DRC estágio 3; [§]Valores referentes aos pacientes DRC estágio.

Acc: Acelerômetro; AFE: Análise Fatorial Exploratória; AGFI: Índice de Qualidade de Ajuste (*Adjusted Goodness-of-Fit Index*); AVE: Variância Média Extraída (*Average Variance Extracted*); CCC: Coeficiente de Correlação de Concordância; CCI: Coeficiente de Correlação Intraclasse; CCI_{TR}: Coeficiente de Correlação Intraclasse via Teste-Retest; Coef α : Coeficiente α de *Cronbach*; Coef Ω : Coeficiente Ω de *McDonald*; Coef θ : Coeficiente θ ; Crr: Correlação; CVI: Índice de Validade de Conteúdo (*Content Validity Index*); CVR: Razão de Validade de Conteúdo (*Content Validity Ratio*); DMCI: Diferença Mínima Clinicamente Importante; EPM: Erro Padrão de Medida; PROM: Medidas Relatadas pelo Próprio Paciente (*Patient-Reported Outcome Measures*); n: número da amostra; r_p: Correlação de *Pearson*; RMSEA: Raiz Quadrada Média do Erro de Aproximação (*Root Mean Square Error of Aproximation*); TCPE: Teste Cardiopulmonar de Exercício; x²: Qui-Quadrado.

Tabela 2. Sumarização dos resultados das propriedades psicométricas por PROM. (Continuação)

PROM	Características da amostra			Propriedades psicométricas								
	n	Idade (anos)	Estadiamento da doença	Validade de construto (estrutural)	Validade de construto (testagem de hipótese)	Validade de conteúdo	Validade de critério	Validade convergente	Validação transcultural	Consistência interna	Confiabilidade	Erro de medida
GPPAQ ⁵⁹	40	62,5 (±11)	1 à 5	-	-	-	n = 40 ⁵⁹ S = 54,2% S = 50% ^a E = 96,6% E = 96,4% ^a Acu = 85% Acu = 82,5% ^a	-	-	-	n = 11 ⁵⁹ kw = 0,74	-
HAP _{56,57,63,67}	174	52 – 67,2	3b à 5	-	-	-	n = 39 ⁵⁶ Crr (Acc): r _p = 0,73 [†] r _p = 0,78 [#] Crr (teste clínico): r _p = 0,72 [†] r _p = 0,75 [#] n = 46 ⁵⁷ Crr (actigrafia): r _p = 0,49 [†] r _p = 0,56 [#]	n = 65 ⁶⁷ Crr (PROM): r _s = 0,76	-	-	n = 29 ⁵⁶ CCI _{TR} = 0,91 [†] n = 25 ⁶³ CCI _{TR} = 0,76 [#] CCI _{TR} = 0,92 [†]	n = 25 ⁶³ EPM = 4,1 [†] EPM = 5,5 [#] DMCI = 11,4 [†] DMCI = 15,1 [#]
IPAQ _{57,62,65}	509	54,0 – 58,6	3b à 5	-	-	-	n = 320 ⁶² Crr (pedômetro): r _s = 0,56 n = 40 ⁶⁵ Crr (actigrafia): r _s = 0,41 n = 46 ⁵⁷ Crr (actigrafia): r _p = -0,26 [§]	-	-	-	n = 103 ⁶² CCI _{TR} = 0,84	-

Dados numéricos de idade apresentados em Média e Desvio Padrão.

^aValores referentes ao GPPAQ-Caminhada; [†]Valores referentes ao HAP-AAS; [#]Valores referentes ao HAP-MAS; [§]Valores referentes ao tempo sentado do questionário IPAQ;

Acc: Acelerômetro; Acu: Acurácia; CCI_{TR}: Coeficiente de Correlação Intraclasse via Teste-Reteste; DMCI: Diferença Mínima Clinicamente Importante; E: Especificidade; EPM: Erro Padrão de Medida; kw: Índice Kappa Ponderado; PROM: Medidas Relatadas pelo Próprio Paciente (*Patient-Reported Outcome Measures*); n: número da amostra; r_p: Correlação de *Pearson*; r_s: Correlação de *Spearman*; S: Sensibilidade.

Tabela 2. Sumarização dos resultados das propriedades psicométricas por PROM. (Continuação)

PROM	Características da amostra			Propriedades psicométricas								
	n	Idade (anos)	Estadiamento da doença	Validade de construto (estrutural)	Validade de construto (testagem de hipótese)	Validade de conteúdo	Validade de critério	Validade convergente	Validação transcultural	Consistência interna	Confiabilidade	Erro de medida
LoPAQ 34,55,61	213	58,0 – 62,3	5	-	n = 68 ³⁴ Crr (teste clínico): r _s = 0,51 Crr (PROM): r _s = 0,64 n = 60 ⁵⁵ Crr (teste clínico): r _p = 0,26 Crr (PROM): r _p = 0,29	n = 85 ⁶¹ CVI = 0,91 CVR = 0,98	n = 85 ⁶¹ Crr (actigrafia): r _s = 0,47 n = 60 ⁵⁵ Crr (pedômetro): r _p = 0,35	n = 68 ³⁴ Crr (PROM): r _s = 0,62	Realizada validação no idioma mandarim: - Tradução - Retradução - Revisão técnica - Avaliação psicométrica ⁶¹	Avaliado, porém não descrito ⁶¹	n = 29 ⁶¹ CCI _{TR} = 0,47	-
PAR ^{56,58}	71	52 – 54,1	1 à 5	-	-	-	n = 39 ⁵⁶ Crr (Acc): r _p = 0,59 Crr (teste clínico): r _p = 0,36	-	-	-	-	-
PASE ^{56,57}	77	52 – 55	3b à 5	-	-	-	n = 40 ⁵⁸ Teste <i>Bland-Altman</i> (DLW) = 0,23	-	-	-	-	-
PASE ^{56,57}	77	52 – 55	3b à 5	-	-	-	n = 39 ⁵⁶ Crr (Acc): r _p = 0,66 Crr (teste clínico): r _p = 0,50	-	-	-	-	-
RPAQ ⁵⁸	40	54,1 (±17,4)	1 à 5	-	-	-	n = 46 ⁵⁷ Crr (actigrafia): r _p = 0,36	-	-	-	-	-
RPAQ ⁵⁸	40	54,1 (±17,4)	1 à 5	-	-	-	n = 40 ⁵⁸ Teste <i>Bland-Altman</i> (DLW) = 0,57	-	-	-	-	-

Dados numéricos de idade apresentados em Média e Desvio Padrão.

Acc: Acelerômetro; CCI_{TR}: Coeficiente de Correlação Intraclasse via Teste-Retestes; Crr: Correlação; CVI: Índice de Validade de Conteúdo (*Content Validity Index*); CVR: Razão de Validade de Conteúdo (*Content Validity Ratio*); Crr: Correlação; DLW: Método da Água Duplamente Marcada (*Doubly Labelled Water*); PROM: Medidas Relatadas pelo Próprio Paciente (*Patient-Reported Outcome Measures*); n: número da amostra; r_p: Correlação de *Pearson*.

A avaliação de qualidade metodológica dos artigos pelo COSMIN RoB está apresentada na Tabela 3. Enquanto as avaliações de validade de construto, convergente e critério apresentaram-se “muito boa” ou “adequada”, a avaliação da validade de conteúdo se mostrou “incerta”. A validação transcultural apresentou qualidade “adequada” para o ESES (no idioma persa) e “inadequada” para o LoPAQ (no idioma mandarim). Nas propriedades relacionadas à confiabilidade, o erro de medida apresentou qualidade metodológica “adequada”, enquanto as demais propriedades apresentaram qualidade menor, com algumas sendo classificadas como “inadequada”.

A sumarização do nível de evidência de acordo com o COSMIN RoB (Tabela 4) demonstra que as melhores performances na propriedade de validade foram obtidas no questionário CKD-PAQ para validade de construto e convergente (resultado suficiente, com qualidade de evidência moderada) e no questionário HAP para validade convergente (resultado suficiente, com qualidade de evidência moderada). No escopo da confiabilidade, os questionários DASI e IPAQ apresentaram resultado suficiente (qualidade de evidência moderada), enquanto o ESES também apresentou resultado satisfatório na avaliação da consistência interna (qualidade de evidência alta).

A respeito da avaliação metodológica dos artigos pela ferramenta EMPRO (Tabela 5), apenas dois artigos apresentaram resultado satisfatório na avaliação de validade. Na avaliação da confiabilidade e nos atributos específicos da EMPRO de modelo conceitual e interpretabilidade, os artigos apresentaram em sua maioria pontuação satisfatória. A sumarização do resultado de acordo com a EMPRO está apresentada na Tabela 6. Todas as PROMs estudadas apresentaram pontuação no atributo validade, porém apenas CKD-PAQ e GPPAQ apresentaram valores satisfatórios. Os questionários DASI, ESES, GPPAQ, IPAQ e HAP apresentaram valores de confiabilidade aceitáveis. Todas as PROMs que avaliaram modelo conceitual (CKD-PAQ, ESES e LoPAQ) apresentaram valores acima de 50,0.

Na comparação entre as duas ferramentas de análise psicométrica (Tabela 7), houve concordância em todas as propriedades psicométricas para as PROMs CKD-PAQ, DASI, FWH, PAR, PASE e RPAQ. Dentre estas, apenas o CKD-PAQ apresentou evidência satisfatória pelas duas ferramentas.

Tabela 3. Análise da qualidade metodológica dos artigos através do COSMIN RoB.

Artigo	PROM	Propriedades psicométricas								
		Validade de construto (estrutural)	Validade de construto (testagem de hipótese)	Validade de conteúdo	Validade de critério	Validade convergente	Validação transcultural	Consistência interna	Confiabilidade	Erro de medida
Hadjiioannou e col., 2019 ⁶⁰	DASI	-	-	-	-	-	-	-	Adequado	Adequado
Hatef e col., 2018 ⁶⁶	ESES	Muito bom	-	Incerto	-	Muito bom	Adequado ^a	Muito bom	Incerto	-
Huang e col., 2021 ⁶¹	LoPAQ	-	-	Incerto	Muito bom	-	Inadequado ^b	Inadequado	Adequado	-
Johansen e col., 2001 ⁵⁶	HAP PAR PASE	-	-	-	Muito bom	-	-	-	Inadequado	-
Johansen e col., 2015 ³⁴	LoPAQ	-	Muito bom	-	-	Muito bom	-	-	-	-
Kittiskulnam, Sheshadri e Johansen., 2019 ⁵⁵	LoPAQ	-	Adequado	-	Muito bom	-	-	-	-	-
Lou e He., 2019 ⁶²	IPAQ	-	-	-	Muito bom	-	-	-	Adequado	-
Overend e col., 2010 ⁶³	HAP	-	-	-	-	-	-	-	Adequado	Adequado
Ravani e col., 2012 ⁶⁴	DASI	-	-	-	Muito bom	-	-	-	Adequado	-
Robinson-Cohen e col., 2013 ⁵⁷	FWH HAP IPAQ [†] PASE	-	-	-	Muito bom	-	-	-	-	-
Rosa e col., 2015 ⁶⁵	IPAQ	-	-	-	Muito bom	-	-	-	-	-
Sridharan e col., 2015 ⁵⁸	PAR RPAQ	-	-	-	Muito bom	-	-	-	-	-
Sridharan e col., 2022 ³⁵	CKD-PAQ	-	Adequado	-	-	Adequado	-	-	-	-
Wellard, Sally. 2003 ⁶⁷	HAP [†]	-	-	-	-	Adequado	-	-	-	-
Wilkinson e col., 2020 ⁵⁹	GPPAQ	-	-	-	Muito bom	-	-	-	Adequado	-

[†]análise de domínios específicos da ferramenta; ^aReferente à validação transcultural da PROM em idioma persa; ^bReferente à validação transcultural da PROM em idioma mandarim.

Tabela 4. Sumarização do nível de evidência das PROMs através do COSMIN RoB.

PROM	Propriedades psicométricas								
	Validade de construto (estrutural)	Validade de construto (testagem de hipótese)	Validade de conteúdo	Validade de critério	Validade convergente	Validação transcultural	Consistência interna	Confiabilidade	Erro de medida
CKD-PAQ ³⁵	-	(+) Moderado	-	-	(+) Moderado	-	-	-	-
DASI ^{60,64}	-	-	-	(-) Baixo	-	-	-	(+) Moderado	(+) Baixo
ESES ⁶⁶	(-) Alto	-	(?) Baixo	-	(-) Alto	(?) ^a Moderado	(+) Alto	(+) Baixo	-
FWH ⁵⁷	-	-	-	(-) Baixo	-	-	-	-	-
GPPAQ ⁵⁹	-	-	-	(±) Baixo	-	-	-	(+) Baixo	-
HAP ^{56,57,63,67}	-	-	-	(±) Moderado	(+) Moderado	-	-	(+) Baixo	(+) Baixo
IPAQ ^{57,62,65}	-	-	-	(-) Alto	-	-	-	(+) Moderado	-
LoPAQ ^{34,55,61}	-	(-) Alto	(+) Muito baixo	(-) Alto	(-) Alto	(?) ^b Muito baixo	(?) Muito baixo	(-) Baixo	-
PAR ^{56,58}	-	-	-	(-) Moderado	-	-	-	-	-
PASE ^{56,57}	-	-	-	(-) Moderado	-	-	-	-	-
RPAQ ⁵⁸	-	-	-	(-) Baixo	-	-	-	-	-

(+): Suficiente; (-): Insuficiente; (±): Inconsistente; (?): Indeterminado.

^aAvaliação referente à validação transcultural no idioma persa; ^bAvaliação referente à validação transcultural no idioma mandarim.

Tabela 5. Avaliação metodológica dos artigos inclusos pela ferramenta EMPRO.

Atributos da EMPRO	PROM														
	Hadjiioannou e col., 2019 ⁶⁰	Hatef e col., 2018 ⁶⁶	Huang e col., 2021 ⁶¹	Johansen e col., 2001 ⁵⁶	Johansen e col., 2015 ³⁴	Kittiskulnam, Sheshadri e Johansen, 2019 ⁵⁵	Lou e He, 2019 ⁶²	Overend e col., 2010 ⁶³	Ravani e col., 2012 ⁶⁴	Robinson-Cohen e col., 2013 ⁵⁷	Rosa e col., 2015 ⁶⁵	Sridharan e col., 2015 ⁵⁸	Sridharan e col., 2022 ³⁵	Wellard, Sally, 2003 ⁶⁷	Wilkinson e col., 2020 ⁵⁹
1. MODELO CONCEITUAL E DE MEDIÇÃO															
1.a Conceito de medição claramente descrito	-	++++	-	-	++++	-	-	-	-	-	-	-	++++	-	-
1.b Conceito para obtenção e combinação dos itens descritos	-	++	-	-	++++	-	-	-	-	-	-	-	++	-	-
1.c Racionalidade para dimensão de escalas	-	++++	-	-	++++	-	-	-	-	-	-	-	++	-	-
1.d Envolvimento da população-alvo	-	++++	-	-	++++	-	-	-	-	-	-	-	++++	-	-
1.e Variabilidade da escala descrita e adequada	-	+++	-	-	+	-	-	-	-	-	-	-	++	-	-
1.f Nível de medida descrito	-	++	-	-	+++	-	-	-	-	-	-	-	++	-	-
1.g Procedimento para derivação de pontuação	-	+++	-	-	+	-	-	-	-	-	-	-	++	-	-
Pontuação do atributo	-	71,4^a	-	-	66,7^a	-	-	-	-	-	-	-	52,4^a	-	-
2. CONFIABILIDADE															
2.1. CONFIABILIDADE (Consistência Interna)															
2.1.a Métodos de coleta dos dados descritos	+++	++++	+++	NA	-	-	NA	+++	NA	-	-	-	-	-	NA
2.1.b Coeficiente Alfa de Cronbach adequado	NI	+++	+	NA	-	-	NA	NA	NA	-	-	-	-	-	NA
2.1.c Estimativas de IRT fornecidas	++++	NA	+	NA	-	-	NA	+++	NA	-	-	-	-	-	NA
2.1.d Aplicação em diferentes populações	NA	++++	NA	NA	-	-	NA	NA	NA	-	-	-	-	-	NA
Pontuação do atributo	33,3	58,3^a	8,3	NA	-	-	NA	16,6	NA	-	-	-	-	-	NA
2.2. CONFIABILIDADE (Reprodutibilidade)															
2.2.a Métodos de coleta dos dados descritos	+++	++++	+++	++	-	-	+++	++++	+++	-	-	-	-	-	+++
2.2.b Tempo de intervalo para teste-reteste adequado	+++	++	+++	+	-	-	+++	+++	+++	-	-	-	-	-	++++
2.3.c Coeficientes de reprodutibilidade adequados	++++	NA	+	++++	-	-	+++	+++	+++	-	-	-	-	-	++++
2.4.d Estimativas de IRT fornecidas	++++	NA	NA	+	-	-	+	++	+	-	-	-	-	-	NA
Pontuação do atributo	83,3^a	16,7	25,0	33,3	-	-	50,0^a	66,6^a	50,0^a	-	-	-	-	-	58,3^a

++++: 4 (concorda totalmente); +++: 3; ++: 2; +: 1 (discorda totalmente); NA: Não aplicável; NI: Não informado. ^avalor considerado aceitável

Tabela 5. Avaliação metodológica dos artigos incluídos pela ferramenta EMPRO. (Continuação)

Atributos da EMPRO	PROM														
	Hadjiioannou e col., 2019 ⁶⁰	Hatef e col., 2018 ⁶⁶	Huang e col., 2021 ⁶¹	Johansen e col., 2001 ⁵⁶	Johansen e col., 2015 ³⁴	Kittiskulnam, Sheshadri e Johansen, 2019 ⁵⁵	Lou e He, 2019 ⁶²	Overend e col., 2010 ⁶³	Ravani e col., 2012 ⁶⁴	Robinson-Cohen e col., 2013 ⁵⁷	Rosa e col., 2015 ⁶⁵	Sridharan e col., 2015 ⁵⁸	Sridharan e col., 2022 ³⁵	Wellard, Sally, 2003 ⁶⁷	Wilkinson e col., 2020 ⁵⁹
3. VALIDADE															
3.a Validade de conteúdo adequado	-	++++	++++	NA	NA	NA	NA	-	NA	NA	NA	NA	++	NA	NA
3.b Validade de construto/critério adequado	-	+++	++	+++	+++	+++	+++	-	+++	+++	+++	+++	++++	+++	+++
3.c Composição da amostra descrita	-	+++	+++	+	++	++	+	-	++++	+++	++	+++	+++	+	++++
3.d Hipótese prévia descrita	-	++	NA	NA	++++	++++	NA	-	NA	NA	NA	NA	+++	++++	NA
3.e Racionalidade para validade de critério	-	NA	+++	+++	NA	+++	+++	-	+++	++++	+++	++++	+++	NA	++++
3.f Avaliação em diferentes populações	-	NI	NA	NA	NA	NA	NA	-	NA	NA	NA	NA	NA	NA	++++
Pontuação do atributo	-	38,9	33,3	5,5	16,6	33,3	5,5	-	22,2	22,2	11,1	22,2	50,0^a	11,1	50,0^a
4. RESPONSABILIDADE															
4.a Adequação dos métodos	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4.b Descrição da magnitude estimada de mudança	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4.c Comparação de grupos estáveis e não-estáveis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pontuação do atributo	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5. INTERPRETABILIDADE															
5.a Racional dos critérios externos	-	-	-	+++	++	+++	-	+++	+++	++	-	++	+	-	++++
5.b Descrição das estratégias de interpretação	-	-	-	+++	+++	++++	-	++	+++	+++	-	++++	+++	-	+++
5.c Descrição de como reportar os dados	-	-	-	++	+++	++	-	+++	++	++	-	++++	++++	-	++
Pontuação do atributo	-	-	-	55,5^a	55,5^a	66,6^a	-	55,5^a	55,5^a	44,4	-	77,7^a	55,5^a	-	66,6^a

++++: 4 (concorda totalmente); +++: 3; ++: 2; +: 1 (discorda totalmente); NA: Não aplicável; NI: Não informado.

^avalor considerado aceitável

++++: 4 (concorda totalmente); +++: 3; ++: 2; +: 1 (discorda totalmente); NA: Não aplicável; NI: Não informado.
^avalor considerado aceitável

Tabela 6. Sumarização da avaliação das propriedades psicométricas das PROMs pela ferramenta EMPRO.

PROM	Modelo conceitual e de medição	Atributos da EMPRO								
		Confiabilidade (consistência interna)	Confiabilidade (reprodutibilidade)	Validade	Responsividade	Interpretabilidade	Sobrecarga (resposta)	Sobrecarga (administração)	Modos alternativos de administração	Adaptações linguísticas e transculturais
CKD-PAQ ³⁵	52,4*	-	-	50,0*	-	55,5*	-	-	-	-
DASI ^{60,64}	-	33,3	66,6*	22,2	-	55,5*	-	-	-	-
ESES ⁶⁶	71,4*	58,3*	16,7	38,9	-	-	22,2	22,2	-	100,0*^a
FWH ⁵⁷	-	-	-	22,2	-	44,4	-	-	-	-
GPPAQ ⁵⁹	-	NA	58,3*	50,0*	-	66,6*	-	-	-	-
HAP ^{56,57,63,67}	-	16,6	66,6*	12,9	-	51,8*	-	-	-	-
IPAQ ^{57,62,65}	-	NA	50,0*	12,9	-	44,4	-	-	-	-
LoPAQ ^{34,55,61}	66,7*	8,3	25,0	27,7	-	61,0*	-	-	-	88,9*^b
PAR ^{56,58}	-	NA	33,3	13,9	-	66,6*	-	-	-	-
PASE ^{56,57}	-	NA	33,3	13,9	-	49,9	-	-	-	-
RPAQ ⁵⁸	-	-	-	22,2	-	77,7*	-	-	-	-

NA: Não aplicável

*Valores acima do limite aceitável; ^aValor referente à validação no idioma persa; ^bValor referente à validação no idioma mandarim.

Tabela 7. Comparação dos resultados obtidos pelo COSMIN RoB e pela ferramenta EMPRO.

PROM		Delineamento e conceito	Validade	Propriedades psicométricas			Validação transcultural
				Confiabilidade	Responsividade	Interpretabilidade [#]	
CKD-PAQ ³⁵	COSMIN RoB	✓	✓	-	-	-	-
	EMPRO	✓	✓	-	-	✓	-
DASI ^{60,64}	COSMIN RoB	-	✗	✓	-	-	-
	EMPRO	-	✗	✓	-	✓	-
ESES ⁶⁶	COSMIN RoB	✗	✗	✓	-	-	✗
	EMPRO	✓	✗	✓	-	-	✓
FWH ⁵⁷	COSMIN RoB	-	✗	-	-	-	-
	EMPRO	-	✗	-	-	✗	-
GPPAQ ⁵⁹	COSMIN RoB	-	✗	✗	-	-	-
	EMPRO	-	✓	✓	-	✓	-
HAP ^{56,57,63,67}	COSMIN RoB	-	✓	✗	-	-	-
	EMPRO	-	✗	✓	-	✓	-
IPAQ ^{57,62,65}	COSMIN RoB	-	✗	✓	-	-	-
	EMPRO	-	✗	✗	-	✗	-
LoPAQ ^{34,55,61}	COSMIN RoB	✗	✗	✗	-	-	✗
	EMPRO	✓	✗	✗	-	✓	✓
PAR ^{56,58}	COSMIN RoB	-	✗	-	-	-	-
	EMPRO	-	✗	-	-	✓	-
PASE ^{56,57}	COSMIN RoB	-	✗	-	-	-	-
	EMPRO	-	✗	-	-	✗	-
RPAQ ⁵⁸	COSMIN RoB	-	✗	-	-	-	-
	EMPRO	-	✗	-	-	✓	-

✓ - Resultado satisfatório: sumarização de resultado suficiente e de qualidade alta ou moderada pelo COSMIN RoB, OU sumarização maior ou igual à 50,0 pelo EMPRO

✗ - Resultado insatisfatório: sumarização de resultado insuficiente ou inconsistente OU suficiente, porém com qualidade baixa ou moderada pelo COSMIN RoB, OU sumarização menor à 50,0 pelo EMPRO

5. DISCUSSÃO

O presente estudo é o primeiro a avaliar e sumarizar a evidência disponível da avaliação psicométrica das PROMs para avaliação da atividade física na DRC nos moldes de revisão sistemática. A literatura já dispõe de revisões sistemáticas na DRC no contexto de avaliação psicométrica de testes clínicos de função física e de PROMs de avaliação de qualidade de vida e sintomas, mas não da atividade física em si.⁷⁹⁻⁸¹

Conforme supracitado, a atividade física é um importante marcador de saúde tanto na população geral quanto nos perfis de DCNT, onde a DRC se encaixa. A OMS destaca a importância do estímulo à atividade física como política pública de promoção à saúde.¹⁶ O uso de PROMs na avaliação da atividade física ainda é atual devido sua boa aplicabilidade em contextos epidemiológicos. Sattler e colaboradores destacam a utilidade das PROMs para triagem e promoção de atividade física na assistência à saúde.²³ O estudo epidemiológico NHANES utiliza uma PROM própria para a avaliação tanto da população geral dos EUA quanto de perfis clínicos específicos, auxiliando assim o direcionamento de políticas públicas de saúde.²⁸ Nesse sentido, as PROMs de atividade física no contexto da DRC possuem clara importância clínica por possibilitar a aquisição de informações sobre o perfil de atividade em doentes renais crônicos e auxiliar no direcionamento de condutas de promoção de saúde.

No entanto, o presente trabalho demonstra que a avaliação de PROMs no contexto da DRC ainda é pouco estudada. A evidência disponível é relativamente nova (com publicação a partir de 2001 e um intervalo de 6 anos sem publicações entre 2004 e 2010), com poucos estudos sobre a temática (apenas 15 artigos foram encontrados), mas com grande heterogeneidade na escolha da PROM (11 PROMs diferentes encontradas), na forma de avaliação psicométrica e na qualidade metodológica.

A contradição da grande prevalência global de DRC,⁴ das evidências bem documentadas da redução de funcionalidade no doente renal crônico¹⁷ e da falta de estudos avaliando as PROMs de atividade física nessa população pode ser explicada justamente pela falta de consenso sobre quais seriam melhores ferramentas de avaliação. Koufaki e Mercer em 2009 pontuaram que a avaliação

funcional de indivíduos portadores de DRC não era rotineira, e que provavelmente a falta de diretrizes acerca de quais testes utilizar poderiam contribuir para a subavaliação percebida.³²

Outro ponto de destaque é a distribuição geográfica da evidência encontrada. A maior parte dos estudos foi oriunda da Europa e da América do Norte, a despeito da alta prevalência mundial da DRC.⁴ Tal achado é importante discutir, uma vez que indica que há sub-representação dos resultados em diferentes grupos populacionais e culturais.

A propriedade psicométrica mais avaliada foi a validade, sendo que todas as PROMs tiveram um ou mais aspectos desta propriedade estudados. Por outro lado, a confiabilidade foi avaliada em apenas 6 (54%) PROMs. A validade diz respeito à capacidade da PROM em reproduzir adequadamente o construto a ser avaliado, enquanto a confiabilidade trata da consistência e precisão de sua medição.^{39,40} Portanto, a evidência analisada demonstra o interesse em avaliar a representação real ou não do construto de atividade física no contexto da DRC. No entanto, um instrumento cuja confiabilidade não é satisfatória, ainda que seja a representação válida do construto, pode confundir a percepção de mudanças após intervenção ou ao longo do tempo²⁶. Destarte, a evidência disponível, embora busque embasar o uso das PROMs na avaliação da atividade física no contexto da DRC, carece de maiores informações acerca de sua precisão na avaliação deste construto.

As principais razões que reduziram a qualidade dos estudos encontrados foram o número amostral pequeno, os valores numéricos não satisfatórios dos resultados e a descrição metodológica incompleta. Portanto, a necessidade de estudos com maior número amostral e melhor delineamento metodológico são necessários.

Na avaliação da validade, a maior parte das PROMs apresentou correlação baixa para os valores de corte (menor que 0,7), o que demonstra a pouca habilidade dessas PROMs de conseguirem reproduzir adequadamente o construto no contexto da DRC. O melhor desempenho foi do CKD-PAQ³⁵, uma PROM específica para DRC, que apresentou boa correlação na validade de construto e convergente. Dentre as PROMs não específicas para DRC, o

HAP^{56,57,63,67} foi quem apresentou melhor desempenho nessa propriedade, porém com resultados inconsistentes e contraditórios entre as ferramentas de avaliação. O desempenho na confiabilidade foi melhor nas PROMs avaliadas, com valores de coeficiente majoritariamente acima de 0,8. No entanto, dentre as PROMs específicas, o CKD-PAQ não apresentou avaliação de confiabilidade e o LoPAQ⁶¹ não apresentou valores satisfatórios. Nesta propriedade, o melhor desempenho foi das PROMs não específicas DASI^{60,64} e IPAQ⁶².

Dentre as 9 PROMs não específicas para DRC que foram avaliadas, 2 foram desenvolvidas especificamente para portadores de DCNT (ESES e HAP)^{74,75} e outras 3 (DASI, PASE e PAR),^{69,82,83} apesar do perfil generalista, já foram validadas para perfis crônicos. Essa diferenciação é importante, uma vez que um dos maiores riscos no uso de PROMs em perfis clínicos específicos é o chamado “efeito chão”, ou seja, a não capacidade da PROM de identificar os limiares desses indivíduos pelos valores serem muito baixos.³¹

No presente estudo, a PROM que apareceu em mais estudos foi o HAP, num total de 4 artigos. Tal dado entra em conformidade com a pesquisa em literatura prévia, que demonstrou que o HAP era a PROM específica para atividade física mais utilizada na avaliação de portadores de DRC.³² O fato de o HAP já possuir validação para diferentes idiomas e bem como vasta literatura embasando sua utilização em diferentes DCNT podem justificar sua utilização.^{84,85}

De modo geral, entre as PROMs não específicas, foi observada validade satisfatória para HAP (via COSMIN RoB) e GPPAQ (via EMPRO) e confiabilidade satisfatória para DASI e ESES (ambas com concordância nas ferramentas) e para IPAQ (via COSMIN RoB), HAP e GPPAQ (ambas via EMPRO). Em valores numéricos, o HAP foi a PROM não específica com melhor desempenho em validade na comparação com instrumentos de avaliação objetiva, tais como acelerômetros. Sua menor correlação foi com a actigrafia,⁵⁷ e o baixo número amostral dos estudos encontrados também fez sua qualidade metodológica ficar menor.

A performance do HAP pode ter diferentes razões. A princípio, destaca-se o fato de ser uma PROM elaborada originalmente para uma DCNT (DPOC) que,

tal qual a DRC, gera alterações multissistêmicas no indivíduo com consequente limitação ao exercício.⁷⁵ Ademais, a disposição dos itens do HAP, abrangendo atividades de muito leves até extenuantes, aparenta ser uma boa estratégia a fim de evitar os efeitos teto e chão⁸⁴. Já há evidências de validação do HAP em indivíduos cardiopatas, neurológicos e hospitalizados, traduzindo assim sua boa performance na avaliação de perfis sedentários e inativos.⁸⁴⁻⁸⁶

A respeito das PROMs específicas para doentes renais crônicos, observamos que são duas ferramentas recentes e ainda com pouca literatura: o LoPAQ foi descrito pela primeira vez por Johansen e col. em 2015³⁴ e o CKD-PAQ teve sua primeira publicação por Sridharan e col. em 2022³⁵. Em busca nas bases de dados indexadas, constatou-se que toda a literatura a respeito de ambas as PROMs foram incluídas no presente estudo. Ou seja, no momento dispomos de 3 estudos sobre o LoPAQ e 1 estudo sobre o CKD-PAQ, evidenciando, assim, a necessidade da realização de mais estudos utilizando PROMs específicas para este perfil clínico.

Essas PROMs diferenciam-se entre si no perfil alvo: enquanto o LoPAQ avalia a atividade física exclusivamente em indivíduos dialíticos,³⁴ o CKD-PAQ permite avaliação de todos os estadiamentos da DRC.³⁵ Nesse sentido, a CKD-PAQ propõe ser mais abrangente e, portanto, com um papel importante na avaliação, triagem e prevenção ao declínio funcional de estadiamentos mais leves; o LoPAQ, por outro lado, ao se concentrar no estágio terminal, busca distinguir com maior acurácia os diferentes limiares de atividade física entre dialíticos.

Chama a atenção que o LoPAQ apresentou desempenho insatisfatório em todas as propriedades avaliadas. Os autores dos estudos inclusos que avaliaram o LoPAQ pouco discutem esse resultado, porém reiteram a importância da PROM para a avaliação de um perfil clínico acentuadamente sedentário, no caso o estágio dialítico da DRC. Dentre as razões para as correlações fracas e moderadas do LoPAQ com diferentes instrumentos, os autores desses estudos apontam a similaridade desses resultados com outros trabalhos que avaliaram e validaram PROMs contra métodos objetivos de avaliação de atividade física^{55,61}.

De fato, já foi anteriormente discutido a correlação somente razoável entre

os métodos objetivos de avaliação de atividade física (pedômetros, acelerômetros e actigrafia, por exemplo) e as PROMs. No entanto, uma vez que a premissa da LoPAQ seja justamente a avaliação e categorização da atividade física em perfil clínico e sedentário, a expectativa era que a primeira PROM elaborada para DRC e publicada apresentasse bons níveis de validade e confiabilidade para esta população, o que não foi sustentado pela evidência disponível. Por conseguinte, há ainda a necessidade da validação da LoPAQ com instrumentos mais sensíveis ao perfil mais sedentário da população dialítica, de modo a corroborar seu uso clínico.

Na sumarização dos resultados, apenas o CKD-PAQ apresentou resultados satisfatórios em ambas as ferramentas. Esse é um achado interessante, uma vez que, além de ser PROM específica para DRC, ela também abrange todos os estadiamentos da doença. Deste modo, o CKD-PAQ se apresenta como uma PROM promissora para a avaliação da DRC em seus diferentes níveis.

No entanto, alguns pontos sobre o CKD-PAQ precisam ser destacados. O primeiro ponto é que há apenas 1 estudo disponível sobre a PROM, de modo que a qualidade tanto numérica quanto metodológica dos resultados pode mudar conforme novos estudos sejam realizados. Além disso, foi avaliada apenas a propriedade de validade (de construto e convergente), carecendo então a evidência sobre confiabilidade.³⁵ Por fim, o estudo disponível não apresentou dados adicionais discriminando os diferentes estágios da DRC. Com isso, é necessário questionar sobre se haveria diferenças na performance da PROM para cada estágio da doença.

Ambas as PROMs específicas para DRC, no entanto, carecem de maior validação transcultural: o CKD-PAQ se encontra disponível somente em seu idioma de origem (inglês) e o LoPAQ apresenta-se disponível somente no idioma original (inglês) e com validação transcultural para o mandarim, com qualidade dessa validade contraditória⁶¹.

As ferramentas de avaliação de qualidade metodológica COSMIN RoB e EMPRO apresentaram discordâncias entre si na avaliação de uma ou mais propriedades psicométricas de 5 PROMs (45,5% do total).

Dentre as PROMs não específicas, o GPPAQ não apresentou bons resultados na COSMIN RoB, mas na avaliação da EMPRO obteve resultados satisfatórios nas propriedades psicométricas avaliadas. O HAP apresentou conflito entre as ferramentas para as mesmas propriedades psicométricas: enquanto teve resultado suficiente na validade, mas não na confiabilidade, pelo COSMIN RoB, enquanto na EMPRO apresentou resultado satisfatório na confiabilidade, porém não na validade. Já no ESES houve discordância no delineamento e conceito e na validação transcultural, onde apresentou resultado insatisfatório na COSMIN RoB e satisfatório na EMPRO. Nas PROMs específicas, houve discordância entre as ferramentas com relação ao LoPAQ também em delineamento e conceito e validação transcultural, também com resultados satisfatórios na EMPRO e insatisfatórios na COSMIN RoB.

Essa distinção na sumarização entre as ferramentas pode ter algumas justificativas. Em avaliação da composição dos itens de diferentes ferramentas de avaliação de PROMs, Rosenkoetter e Tate observaram que os itens da COSMIN RoB avaliam em sua maior parte (cerca de 70%) aspectos de delineamento metodológico das propriedades psicométricas, enquanto a EMPRO possui composição mista de itens a respeito de delineamento, relato, resultados e outros aspectos, tais como tempo de administração e sobrecarga³⁶. A COSMIN RoB apresenta racional lógico estruturado e orientações detalhadas e específicas para a determinação do resultado.^{47,48} A EMPRO, por outro lado, apresenta racional lógico mais simples e com pontuação baseada na escala Likert e, portanto, mais suscetível a percepção subjetiva dos avaliadores.⁵² Essas questões podem justificar a diferença observada no resultado de algumas PROMs.

O presente estudo também se diferencia pelo uso de duas ferramentas de avaliação metodológica, sendo que a iniciativa COSMIN é amplamente utilizada e considerada o padrão ouro nos estudos referentes à propriedade psicométrica.^{42,48,87} A EMPRO, por sua vez, é uma ferramenta nova e ainda pouco utilizada, porém com diferenciais em sua avaliação que a tornam uma opção interessante, como avaliação do tempo de administração, sobrecarga na aplicação e modos alternativos de aplicação.^{51,52}

Deste modo, observa-se que cada ferramenta possui características que

impactam na aplicação e interpretação dos resultados e, portanto, a escolha da ferramenta de avaliação deve ser baseada nos objetivos dos pesquisadores bem como sua expertise de avaliação.

Dentro da realidade brasileira, não possuímos até o presente momento nenhuma das PROMs específicas está traduzida e validada para o idioma e cultura local. O HAP, por outro lado, já é traduzido e validado no país,⁸⁶ porém não exclui a necessidade de validação transcultural de PROM específica, considerando a prevalência da DRC no país e seu impacto social e econômico.^{5,6}

Outro ponto importante a se destacar é que a maioria dos estudos focou nos estágios 3b a 5, o que confere um perfil clínico de maior gravidade. Embora já haja evidência na literatura que demonstre que há alterações de funcionalidade mesmo nos estágios iniciais da DRC, a pouca evidência disponível demonstra a concentração de estudos em avaliar a psicométrica em indivíduos mais graves e, portanto, mais inativos. Sridharan e colaboradores comentam que os dispositivos comumente utilizados para avaliação de níveis de atividade física (tais como acelerômetros) possuem baixa acurácia na determinação do comportamento sedentário de indivíduos predominantemente inativos, o que seria uma justificativa para os baixos valores de correlação observados na grande maioria das PROMs.³⁵

Se por um lado este foco no perfil mais grave explique a performance insuficiente de muitas PROMs na validade, por outro não há informações sobre os perfis mais leves, que poderiam se beneficiar de PROMs tanto para rastreamento de redução de atividade física como no auxílio para traçar propostas de promoção ao comportamento físico ativo. Deste modo, o presente estudo também observou que há uma lacuna na pesquisa quanto ao comportamento de estágios iniciais da DRC, bem como na performance das PROMs em diferentes populações, tais como a brasileira.

O presente estudo apresenta algumas limitações. Em primeiro lugar, a pesquisa foi restrita a bases de dados indexadas, não incluindo materiais de “literatura cinza” (*gray literature*). Em segundo lugar, por limitações da evidência encontrada, não foi possível realizar a discriminação da psicométrica das PROMs em cada estágio da DRC, o que poderia influenciar principalmente nos valores

numéricos de validade e confiabilidade.

6. CONCLUSÃO

Não há consenso na literatura existente sobre a melhor PROM para avaliar atividade física na DRC, uma vez que a evidência é pequena e com grande heterogeneidade na escolha das PROMs e na qualidade metodológica. Há indicação de que HAP e CKD-PAQ são as PROMs não específica e específica, respectivamente, com melhor desempenho na DRC.

As ferramentas de avaliação COSMIN RoB e EMPRO divergiram entre si, por conta das diferentes abordagens de cada ferramenta, evidenciando que a escolha pela ferramenta de avaliação deve ser pautada pela expertise e objetivo do avaliador.

Há a necessidade de mais pesquisas no tema, principalmente no que concerne aos estágios mais leves da DRC.

7. REFERÊNCIAS

1. Kidney Disease Improve Global Outcomes (KDIGO). KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney International Supplements* 2013; 3(1). Disponível em: <https://kdigo.org/guidelines/ckd-evaluation-and-management/>
2. Ministério da Saúde. Diretrizes Clínicas para o Cuidado ao Paciente com Doença Renal Crônica (DRC) no Sistema Único de Saúde. Brasília, DF: 2014. Disponível em: https://bvsmis.saude.gov.br/bvs/publicacoes/diretrizes_clinicas_cuidado_paciente_renal.pdf
3. Chadban, S. J. et al. KDIGO Clinical Practice Guideline on the Evaluation and Management of Candidates for Kidney Transplantation. *Transplantation* 2020;104(4SI Supl I): S11-S103.
4. Jager KJ, Kovesdy C, Langham R, Rosenberg M, Jha V, Zoccali C. A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases. *Kidney Int.* 2019 Nov;96(5):1048-1050. doi: 10.1016/j.kint.2019.07.012. Epub 2019 Sep 30. PMID: 31582227.
5. Ministério da Saúde. Doença Renal Crônica no Sistema Único de Saúde. [Apresentação; Brasil, 2017].
6. Marinho AWGB, Penha A da P, Silva MT, Galvão TF. Prevalência de doença renal crônica em adultos no Brasil: revisão sistemática da literatura. *Cad saúde colet* [Internet]. 2017 Jul;25(3):379–88. Disponível em: <https://doi.org/10.1590/1414-462X201700030134>
7. Nangaku M, Eckardt KU. Pathogenesis of renal anemia. *Semin Nephrol.* 2006 Jul;26(4):261-8. doi: 10.1016/j.semnephrol.2006.06.001. PMID: 16949463.
8. Dugdale M. Anemia. *Obstet Gynecol Clin North Am.* 2001 Jun;28(2):363-81. doi: 10.1016/s0889-8545(05)70206-0. PMID: 11430182.
9. Souza VA, Oliveira Dd, Mansur HN, Fernandes NM, Bastos MG. Sarcopenia in chronic kidney disease. *J Bras Nefrol.* 2015 Jan-Mar;37(1):98-105. doi: 10.5935/0101-2800.20150014. PMID: 25923756.
10. Kidney Disease: Improving Global Outcomes (KDIGO) CKD-MBD Update Work Group. KDIGO 2017 Clinical Practice Guideline Update for the Diagnosis, Evaluation, Prevention, and Treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). *Kidney Int Suppl* (2011). 2017 Jul;7(1):1-59. doi: 10.1016/j.kisu.2017.04.001. Epub 2017 Jun 21. Erratum in: *Kidney Int Suppl* (2011). 2017 Dec;7(3):e1. PMID: 30675420; PMCID: PMC6340919.
11. Bolignano D, Rastelli S, Agarwal R, Fliser D, Massy Z, Ortiz A et al. Pulmonary hypertension in CKD. *Am J Kidney Dis.* 2013 Abr;61(4):612-22. doi: 10.1053/j.ajkd.2012.07.029. Epub 2012 Nov 17. Erratum in: *Am J Kidney Dis.* 2015 Mar;65(3):524. PMID: 23164943.
12. Wathanavasin W, Banjongjit A, Avihingsanon Y, Praditpornsilpa K, Tungsanga K, Eiam-Ong S et al. Prevalence of Sarcopenia and Its Impact on Cardiovascular Events and Mortality among Dialysis Patients: A Systematic Review and Meta-Analysis. *Nutrients.* 2022 Set 30;14(19):4077. doi: 10.3390/nu14194077. PMID:

36235729; PMID: PMC9572026.

13. Roshanravan B, Gamboa J, Wilund K. Exercise and CKD: Skeletal Muscle Dysfunction and Practical Application of Exercise to Prevent and Treat Physical Impairments in CKD. *Am J Kidney Dis.* 2017 Jun;69(6):837-852. doi: 10.1053/j.ajkd.2017.01.051. Epub 2017 Apr 18. PMID: 28427790; PMID: PMC5441955.
14. Organização Mundial da Saúde (OMS). CIF – Classificação Internacional de Funcionalidade, Incapacidade e Saúde. São Paulo:EDUSP; 2003.
15. Caspersen CJ, Kriska AM, Dearwater SR. Physical activity epidemiology as applied to elderly populations. *Baillieres Clin Rheumatol.* 1994 Feb;8(1):7-27. doi: 10.1016/s0950-3579(05)80222-5. PMID: 8149451.
16. Organização Mundial da Saúde (OMS). Diretrizes da OMS para Atividade Física e Comportamento Sedentário: num piscar de olhos. 2020. Disponível em: https://ws.santabarbara.sp.gov.br/instar/esportes/downloads/guia_AF_OMS.pdf
17. Clarke AL, Zaccardi F, Gould DW, Hull KL, Smith AC, Burton JO et al. Association of self-reported physical function with survival in patients with chronic kidney disease. *Clin Kidney J.* 2019 Feb;12(1):122-128. doi: 10.1093/ckj/sfy080. Epub 2018 Ago 28. PMID: 30746139; PMID: PMC6366129.
18. Beddhu S, Baird BC, Zitterkoph J, Neilson J, Greene T. Physical activity and mortality in chronic kidney disease (NHANES III). *Clin J Am Soc Nephrol.* 2009 Dez;4(12):1901-6. doi: 10.2215/CJN.01970309. Epub 2009 Out 9. PMID: 19820134; PMID: PMC2798872.
19. Aucella F, Valente GL, Catizone L. The role of physical activity in the CKD setting. *Kidney Blood Press Res.* 2014;39(2-3):97-106. doi: 10.1159/000355783. Epub 2014 Jul 29. PMID: 25117619.
20. Westerterp KR. Doubly labelled water assessment of energy expenditure: principle, practice, and promise. *Eur J Appl Physiol.* 2017 Jul;117(7):1277-1285. doi: 10.1007/s00421-017-3641-x. Epub 2017 Mai 15. PMID: 28508113; PMID: PMC5486561.
21. Sylvia LG, Bernstein EE, Hubbard JL, Keating L, Anderson EJ. Practical guide to measuring physical activity. *J Acad Nutr Diet.* 2014 Feb;114(2):199-208. doi: 10.1016/j.jand.2013.09.018. Epub 2013 Nov 28. PMID: 24290836; PMID: PMC3915355.
22. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. *Prog Cardiovasc Dis.* 2015 Jan-Feb;57(4):387-95. doi: 10.1016/j.pcad.2014.10.005. Epub 2014 Oct 31. PMID: 25446555.
23. Sattler MC, Ainsworth BE, Andersen LB, Foster C, Hagströmer M, Jaunig J et al. Physical activity self-reports: past or future? *Br J Sports Med.* 2021 Ago;55(16):889-890. doi: 10.1136/bjsports-2020-103595. Epub 2021 Feb 3. PMID: 33536193; PMID: PMC8477753.
24. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc.* 2000 Set;32(9 Suppl):S498-504. doi: 10.1097/00005768-200009001-00009. PMID: 10993420.

25. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N et al. Patient-Reported Outcomes in Performance Measurement. Research Triangle Park (NC): RTI Press; 2015 Set. PMID: 28211667.
26. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). Cochrane, 2023. Disponível em: www.training.cochrane.org/handbook.
27. Sternfeld B, Goldman-Rosas L. A systematic approach to selecting an appropriate measure of self-reported physical activity or sedentary behavior. *J Phys Act Health*. 2012 Jan;9 Suppl 1:S19-28. doi: 10.1123/jpah.9.s1.s19. PMID: 22287444.
28. National Center for Health Statistics (NCHS). NHANES – Interviewer Procedures Manual (October 2022). 2022. Disponível em: <https://wwwn.cdc.gov/nchs/data/nhanes/2021-2023/manuals/2022-Interviewer-Procedures-508.pdf>
29. Zhang B, Lei SM, Le S, Gong Q, Cheng S, Wang X. Changes in health behaviors and conditions during COVID-19 pandemic strict campus lockdown among Chinese university students. *Front Psychol*. 2022 Oct 17;13:1022966. doi: 10.3389/fpsyg.2022.1022966. PMID: 36324783; PMCID: PMC9621116.
30. Merellano-Navarro E, Bustamante-Ara N, Russell-Guzmán J, Lagos-Hernández R, Uribe N, Godoy-Cumillaf A. Association between Sleep Quality and Physical Activity in Physical Education Students in Chile in the Pandemic Context: A Cross-Sectional Study. *Healthcare (Basel)*. 2022 Oct 1;10(10):1930. doi: 10.3390/healthcare10101930. PMID: 36292377; PMCID: PMC9601690.
31. Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. *Sports Med*. 2001 Feb;31(2):91-100. doi: 10.2165/00007256-200131020-00002. PMID: 11227981.
32. Koufaki P, Mercer T. Assessment and monitoring of physical function for people with CKD. *Adv Chronic Kidney Dis*. 2009 Nov;16(6):410-9. doi: 10.1053/j.ackd.2009.08.010. PMID: 19801131.
33. Zhang NH, Luo R, Cheng YC, Ge SW, Xu G. Leisure-Time Physical Activity and Mortality in CKD: A 1999-2012 NHANES Analysis. *Am J Nephrol*. 2020;51(11):919-929. doi: 10.1159/000511685. Epub 2020 Nov 18. PMID: 33207339.
34. Johansen KL, Painter P, Delgado C, Doyle J. Characterization of physical activity and sitting time among patients on hemodialysis using a new physical activity instrument. *J Ren Nutr*. 2015 Jan;25(1):25-30. doi: 10.1053/j.jrn.2014.06.012. Epub 2014 Set 8. Erratum in: *J Ren Nutr*. 2015 Jul;25(4):397. PMID: 25213326; PMCID: PMC4282813.
35. Sridharan S, Vilar E, Ramanarayanan S, Davenport A, Farrington K. Energy expenditure estimates in chronic kidney disease using a novel physical activity questionnaire. *Nephrol Dial Transplant*. 2022 Feb 25;37(3):515-521. doi: 10.1093/ndt/gfaa377. PMID: 33416874.
36. Rosenkoetter U, Tate RL. Assessing Features of Psychometric Assessment Instruments: A Comparison of the COSMIN Checklist with Other Critical Appraisal Tools. *Brain Impairment*. Cambridge University Press; 2018;19(1):103–18.

37. Mokkink LB, Prinsen CAC, Bouter LM, Vet HCW de, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther* [Internet]. 2016 Mar;20(2):105–13. Disponível em: <https://doi.org/10.1590/bjpt-rbf.2014.0143>
38. Pasquali L. *Psicometria. Rev esc enferm USP* [Internet]. 2009 Dez;43(spe):992–9. Disponível em: <https://doi.org/10.1590/S0080-62342009000500002>
39. Echevarría-Guanilo ME, Gonçalves N, Romanoski PJ. Propriedades Psicométricas de Instrumentos de Medidas: Bases Conceituais e Métodos de Avaliação - Parte I. *Texto contexto - enferm* [Internet]. 2017;26(4). Disponível em: <https://doi.org/10.1590/0104-07072017001600017>
40. Echevarría-Guanilo ME, Gonçalves N, Romanoski PJ. Psychometric Properties of Measurement Instruments: Conceptual Basis and Evaluation Methods - Part II. *Texto contexto - enferm* [Internet]. 2019;28:e20170311Disponível em: <https://doi.org/10.1590/1980-265X-tce-2017-0311>
41. Luján-Tangarife, JA, Cardona-Arias, JA. Construcción y Validación de Escalas de Medición en Salud: Revisión de Propiedades Psicométricas. *Archivos de Medicina*. 2015;11(3): 1–10. doi: 10.3823/1251
42. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018 Mai;27(5):1147-1157. doi: 10.1007/s11136-018-1798-3. Epub 2018 Feb 12. PMID: 29435801; PMCID: PMC5891568.
43. Souza AC de, Alexandre NMC, Guirardello E de B. Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiol Serv Saúde* [Internet]. 2017 Jul;26(3):649–59. Disponível em: <https://doi.org/10.5123/S1679-49742017000300022>
44. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. *Lancet Glob Health*. 2018 Out;6(10):e1077-e1086. doi: 10.1016/S2214-109X(18)30357-7. Epub 2018 Stp 4. Erratum in: *Lancet Glob Health*. 2019 Jan;7(1):e36. PMID: 30193830.
45. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021 Mar 29;372:n71. doi: 10.1136/bmj.n71. PMID: 33782057; PMCID: PMC8005924.
46. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan – A Web and Mobile App for Systematic Reviews. *Syst Rev* 2016;5:210. <https://doi.org/10.1186/s13643-016-0384-4>
47. Mokkink, LB. COSMIN Risk of Bias Checklist. 2018. Disponível em: www.cosmin.nl.
48. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Risk of Bias Checklist for Systematic Reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018 Mai;27(5):1171-1179. doi: 10.1007/s11136-017-1765-4. Epub 2017 Dez 19. PMID: 29260445; PMCID: PMC5891552.

49. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL et al. COSMIN Risk of Bias Tool to Assess the Quality of Studies on Reliability or Measurement Error of Outcome Measurement Instruments: A Delphi Study. *BMC Med Res Methodol* 2020; 20:293. Disponível em: <https://doi.org/10.1186/s12874-020-01179-5>
50. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability and measurement error of outcome measurement instrument user manual (Version 1.0 dated January 2021). 2021. Disponível em: https://www.cosmin.nl/wp-content/uploads/user-manual-COSMIN-Risk-of-Bias-tool_v4_JAN_final.pdf
51. BiblioPRO [homepage]. EMPRO - Evaluating the Measurement of Patient-Reported Outcomes (English version) [acessado em Dezembro de 2022]. <https://bibliopro.org/buscador/986/evaluating-the-measurement-of-patient-reported-outcomes-english-version>.
52. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M et al; Scientific Committee on "Patient-Reported Outcomes" of the IRYSS Network. Development of EMPRO: A Tool for the Standardized Assessment of Patient-Reported Outcome Measures. *Value Health*. 2008 Jul-Ago;11(4):700-8. doi: 10.1111/j.1524-4733.2007.00309.x. Epub 2008 Jan 8. PMID: 18194398.
53. Lorente S, Viladrich C, Vives J, Losilla JM. Tools to Assess the Measurement Properties of Quality of Life Instruments: A Meta-Review. *BMJ Open*. 2020 Ago 11;10(8):e036038. doi: 10.1136/bmjopen-2019-036038. PMID: 32788186; PMCID: PMC7422655.
54. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN Methodology for Evaluating the Content Validity of Patient-Reported Outcome Measures: A Delphi Study. *Qual Life Res*. 2018 Mai;27(5):1159-1170. doi: 10.1007/s11136-018-1829-0. Epub 2018 Mar 17. PMID: 29550964; PMCID: PMC5891557.
55. Kittiskulnam P, Sheshadri A, Johansen KL. Validation of a New Physical Activity Instrument Against Pedometers Among Dialysis Patients. *J Ren Nutr*. 2019 Nov;29(6):498-503. doi: 10.1053/j.jrn.2019.03.081. Epub 2019 Mai 8. PMID: 31078404; PMCID: PMC6825523.
56. Johansen KL, Painter P, Kent-Braun JA, Ng AV, Carey S, Da Silva M et al. Validation of Questionnaires to Estimate Physical Activity and Functioning in End-Stage Renal Disease. *Kidney Int*. 2001 Mar;59(3):1121-7. doi: 10.1046/j.1523-1755.2001.0590031121.x. PMID: 11231369.
57. Robinson-Cohen C, Littman AJ, Duncan GE, Roshanravan B, Ikizler TA, Himmelfarb J et al. Assessment of Physical Activity in Chronic Kidney Disease. *J Ren Nutr*. 2013 Mar;23(2):123-31. doi: 10.1053/j.jrn.2012.04.008. Epub 2012 Jun 26. PMID: 22739659; PMCID: PMC3496802.
58. Sridharan S, Wong J, Vilar E, Farrington K. Comparison of Energy Estimates in Chronic Kidney Disease Using Doubly-Labelled Water. *J Hum Nutr Diet*. 2016 Fev;29(1):59-66. doi: 10.1111/jhn.12326. Epub 2015 Jul 15. PMID: 26173618.
59. Wilkinson TJ, Palmer J, Gore EF, Smith AC. The Validity of the 'General Practice Physical Activity Questionnaire' Against Accelerometry in Patients with Chronic Kidney Disease. *Physiother Theory Pract*. 2022 Out;38(10):1528-1537. doi:

- 10.1080/09593985.2020.1855684. Epub 2020 Dez 2. PMID: 33263260.
60. Hadjioannou I, Wong K, Lindup H, Mayes J, Castle E, Greenwood S. Test-Retest Reliability for Physical Function Measures in Patients with Chronic Kidney Disease. *J Ren Care*. 2020 Mar;46(1):25-34. doi: 10.1111/jorc.12310. Epub 2019 Dez 16. PMID: 31845535.
 61. Huang R, Zhang H, Yang Y, Fang N, Liu Q, Ma J et al. Validation of the Chinese Version of the Low Physical Activity Questionnaire (LoPAQ) with ActiGraph Accelerometer in Hemodialysis Patients. *BMC Nephrol* 2021;22:(17). Disponível em: <https://doi.org/10.1186/s12882-021-02230-3>
 62. Lou X, He Q. Validity and Reliability of the International Physical Activity Questionnaire in Chinese Hemodialysis Patients: A Multicenter Study in China. *Med Sci Monit*. 2019 Dez 10;25:9402-9408. doi: 10.12659/MSM.920900. PMID: 31819029; PMCID: PMC6916147.
 63. Overend T, Anderson C, Sawant A, Perryman B, Locking-Cusolito H. Relative and Absolute Reliability of Physical Function Measures in People with End-Stage Renal Disease. *Physiother Can*. 2010 Spring;62(2):122-8. doi: 10.3138/physio.62.2.122. Epub 2010 Abr 23. PMID: 21359043; PMCID: PMC2871020.
 64. Ravani P, Kilb B, Bedi H, Groeneveld S, Yilmaz S, Mustata S; Alberta Kidney Disease Network. The Duke Activity Status Index in Patients with Chronic Kidney Disease: A Reliability Study. *Clin J Am Soc Nephrol*. 2012 Abr;7(4):573-80. doi: 10.2215/CJN.07990811. Epub 2012 Feb 16. PMID: 22344510.
 65. Rosa CS, Gracia-Marco L, Barker AR, Freitas IF Jr, Monteiro HL. Assessment of Physical Activity by Accelerometer and IPAQ-Short Version in Patients with Chronic Kidney Disease Undergoing Hemodialysis. *Blood Purif*. 2015;40(3):250-5. doi: 10.1159/000437040. Epub 2015 Set 26. PMID: 26406554.
 66. Hatf M, Sharif Nia H, Boyle C, Shafipour V. The Validity and Reliability of the Exercise Self-Efficacy Scale in a Sample of Hemodialysis Patients. *J Nurs Meas*. 2018 Dez;26(3):566-578. doi: 10.1891/1061-3749.26.3.566. PMID: 30593578.
 67. Wellard S. Validation of Physical Activity Measurement for People on Dialysis Treatment. *EDTNA ERCA J*. 2003 Jul-Set;29(3):140-2. doi: 10.1111/j.1755-6686.2003.tb00295.x. PMID: 14552088.
 68. Golubic R, May AM, Benjaminsen Borch K, Overvad K, Charles MA, Diaz MJ et al. Validity of Electronically Administered Recent Physical Activity Questionnaire (RPAQ) in Ten European Countries. *PLoS One*. 2014 Mar 25;9(3):e92829. doi: 10.1371/journal.pone.0092829. Erratum in: *PLoS One*. 2014;9(11):e114103. PMID: 24667343; PMCID: PMC3965465.
 69. Garfield BE, Canavan JL, Smith CJ, Ingram KA, Fowler RP, Clark AL et al. Stanford Seven-Day Physical Activity Recall questionnaire in COPD. *Eur Respir J*. 2012 Ago;40(2):356-62. doi: 10.1183/09031936.00113611. Epub 2011 Dez 19. PMID: 22183486.
 70. InterAct Consortium; Peters T, Brage S, Westgate K, Franks PW, Gradmark A, Tormo Diaz MJ et al. Validity of a Short Questionnaire to Assess Physical Activity in 10 European Countries. *Eur J Epidemiol*. 2012 Jan;27(1):15-25. doi: 10.1007/s10654-011-9625-y. Epub 2011 Nov 17. PMID: 22089423; PMCID: PMC3292724.

71. Sallis JF, Haskell WL, Wood PD, Fortmann SP, Rogers T, Blair SN et al. Physical Activity Assessment Methodology in the Five-City Project. *Am J Epidemiol*. 1985 Jan;121(1):91-106. doi: 10.1093/oxfordjournals.aje.a113987. PMID: 3964995.
72. Blair SN, Haskell WL, Ho P, Paffenbarger RS Jr, Vranizan KM, Farquhar JW et al. Assessment of Habitual Physical Activity by a Seven-Day Recall in a Community Survey and Controlled Experiments. *Am J Epidemiol*. 1985 Nov;122(5):794-804. doi: 10.1093/oxfordjournals.aje.a114163. PMID: 3876763.
73. Hlatky MA, Boineau RE, Higginbotham MB, Lee KL, Mark DB, Califf RM, et al. A Brief Self-Administered Questionnaire to Determine Functional Capacity (the Duke Activity Status Index). *Am J Cardiol*. 1989 Set 15;64(10):651-4. doi: 10.1016/0002-9149(89)90496-7. PMID: 2782256.
74. Kroll T, Kehn M, Ho PS, Groah S. The SCI Exercise Self-Efficacy Scale (ESES): Development and Psychometric Properties. *Int J Behav Nutr Phys Act*. 2007 Ago 30;4:34. doi: 10.1186/1479-5868-4-34. PMID: 17760999; PMCID: PMC2034591.
75. Fix AJ, Daughton DM. Human activity profile – professional manual. Nebraska: Psychological Assessment Resources; 1988.
76. New England Research Institutes. Physical Activity Scale for the Elderly (PASE) Administration and Scoring Instruction Manual. 1991. Disponível em: <https://meetinstrumentenzorg.nl/wp-content/uploads/instrumenten/PASE-handl.pdf>
77. Gov UK [homepage]. General practice physical activity questionnaire (GPPAQ) [acessado em Março de 2023]. <https://www.gov.uk/government/publications/general-practice-physical-activity-questionnaire-gppaq>
78. National Health Service (NHS). General Practice Physical Activity Questionnaire. Disponível em: https://assets.publishing.service.gov.uk/media/5a7b530140f0b6464693564f/GPPAQ_-_pdf_version.pdf
79. Aiyegbusi OL, Kyte D, Cockwell P, Marshall T, Gheorghe A, Keeley T, et al. Measurement properties of patient-reported outcome measures (PROMs) used in adult patients with chronic kidney disease: A systematic review. *PLoS One*. 2017 Jun 21;12(6):e0179733. doi: 10.1371/journal.pone.0179733.
80. MacRae JM, Harasemiw O, Lightfoot CJ, Thompson S, Wytsma-Fisher K, Koufaki P, et al. Measurement properties of performance-based measures to assess physical function in chronic kidney disease: recommendations from a COSMIN systematic review. *Clin Kidney J*. 2023 Jul 18;16(11):2108-2128. doi: 10.1093/ckj/sfad170.
81. Yangöz ŞT, Turan Kavradim S, Özer Z, Boz İ. Psychometric properties of the Kidney Disease Quality of Life-36 instrument: A systematic review using COSMIN methodology. *Nurs Health Sci*. 2021 Dec;23(4):792-806. doi: 10.1111/nhs.12877.
82. Coutinho-Myrrha MA, Dias RC, Fernandes AA, Araújo CG, Hlatky MA, Pereira DG et al. Duke Activity Status Index for Cardiovascular Diseases: Validation of the Portuguese Translation. *Arq Bras Cardiol*. 2014 Abr;102(4):383-90. doi: 10.5935/abc.20140031. Epub 2014 Fev 17. PMID: 24652056; PMCID: PMC4028943.

83. Granger CL, Parry SM, Denehy L. The Self-Reported Physical Activity Scale for the Elderly (PASE) is a Valid and Clinically Applicable Measure in Lung Cancer. *Support Care Cancer*. 2015 Nov;23(11):3211-8. doi: 10.1007/s00520-015-2707-8. Epub 2015 Mar 26. PMID: 25808104.
84. Davidson M, de Morton N. A Systematic Review of the Human Activity Profile. *Clin Rehabil*. 2007 Feb;21(2):151-62. doi: 10.1177/0269215506069475. PMID: 17264109.
85. Ribeiro-Samora GA, Pereira DA, Vieira OA, de Alencar MC, Rodrigues RS, Carvalho ML, et al. Using the Human Activity Profile to Assess Functional Performance in Heart Failure. *J Cardiopulm Rehabil Prev*. 2016 Mai-Jun;36(3):180-5. doi: 10.1097/HCR.000000000000162. PMID: 26784733.
86. Souza DC, Wegner F, Costa LCM, Chiavegato LD, Lunardi AC. Measurement properties of the Human Activity Profile questionnaire in hospitalized patients. *Braz J Phys Ther*. 2017 May-Jun;21(3):153-158. doi: 10.1016/j.bjpt.2017.03.011. Epub 2017 Apr 9. PMID: 28473282; PMCID: PMC5537466.
87. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) user manual (Version 1.0 dated February 2018). 2018. Disponível em: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf

8. ANEXOS

8.1 Tabela PRISMA

para cada item: 2 = adequadamente descrito, 1 = incompletamente descrito, 0 = não descrito, -1 = descrito, mas inadequado

Título	Título	2/1/ 0/-1	Identifica o artigo como revisão sistemática, meta-análise ou ambos.
Resumo	Resumo Estruturado	2/1/ 0/-1	Oferece um resumo estruturado incluindo, se aplicável, fundamentação, objetivos, fontes de dados, critério de elegibilidade, participantes e intervenções; avaliação do estudo e métodos de síntese; resultados; limitações; conclusões e implicações dos resultados principais; número de registro da revisão sistemática.
	Racional	2/1/ 0/-1	Descrever as razões para a revisão no contexto do que já é conhecido.
Introdução	Objetivos	2/1/ 0/-1	Fornecer uma declaração explícita de questões a ser abordadas com referência aos participantes, intervenções, comparações, resultados e desenho do estudo (dentro da estrutura PICOD).
Métodos	Protocolo e registro	2/1/ 0/-1	Indica se o protocolo da revisão existe, como e onde pode ser acessado (endereço eletrônico, p.ex), e se possível oferece informação do registro incluindo número de registro.
	Critérios de elegibilidade	2/1/ 0/-1	Especifica as características do estudo (ex.: estrutura PICO, período de acompanhamento) e indica características (ex.: anos considerados, idioma, status de publicação) utilizadas como critério para elegibilidade, oferecendo um racional
	Fontes de informação	2/1/ 0/-1	Descreve todas as fontes de informação (ex.: bases de dados com os períodos da(s) busca(s), contato com autores com fim de identificar estudos adicionais) na busca e a sua data limite.
	Busca	2/1/ 0/-1	Apresenta a estratégia completa da busca, para pelo menos uma base de dados, incluindo eventuais limites utilizados, tal que possa ser repetida.
	Seleção de estudos	2/1/ 0/-1	Relata o processo para selecionar os estudos (ex.: rastreamento, elegibilidade, incluídos na revisão sistemática, e, se aplicável, incluídos na meta-análise).
	Processo de coleta de dados	2/1/ 0/-1	Descreve o método da extração de dados a partir de relatos (ex.: formas piloto, independentes, em duplicata) e qualquer proceddo para obter e conformar os dados por parte dos investigadores.
	Ítems de dados	2/1/ 0/-1	Lista e define todas as variáveis para que os dados fossem obtidos (ex.: estratégia PICOD, fontes de financiamento) assim como todos os pressupostos e simplificações feitas.
	Risco de viés nos estudos individuais	2/1/ 0/-1	Descreve os métodos utilizados para avaliar o risco de viés de estudos individuais (incluindo a especificação de que se fora feita em nível de estudo ou em nível de resultado), e como essa informação é para ser usada em qualquer síntese de dados.
	Medidas de sumarização	2/1/ 0/-1	Expõe as principais medidas de sumarização (ex.: razão de risco, diferença nas médias).
	Síntese de Resultados	2/1/ 0/-1	Descreve os métodos para manipulação de dados e combinação de resultados dos estudos, se feita, incluindo as medidas de consistência (como a estatística I ²) para cada meta-análise.

	Risco de viés através dos estudos	2/1/ 0/-1	Especifica qualquer avaliação de risco de viés que pode afetar a evidência cumulativa (ex.: viés de publicação, informação seletiva entre estudos)
	Análises adicionais	2/1/ 0/-1	Descrever os métodos de análises adicionais (ex.: sensibilidade ou a análise de subgrupo, meta-regressão), se feita, indicando que foram pré-especificadas.
Resultados	Seleção de estudos	2/1/ 0/-1	Oferece o número dos estudos extraídos, avaliados por elegibilidade, e incluídos na revisão, explicando os motivos de exclusão em cada estágio, preferencialmente apresentados em um fluxograma.
	Características do estudo	2/1/ 0/-1	Para cada estudo, apresenta características para cada dado extraídos (ex.: tamanho de estudo, estratégia PICOD, período de acompanhamento) e fornece as citações.
	Risco de viés nos estudos	2/1/ 0/-1	Apresentar dados sobre o risco de viés de cada estudo e, se possível, qualquer avaliação em nível de resultado (ver item 12).
	Resultados dos estudos individuais	2/1/ 0/-1	Para todos os resultados considerados (benefícios ou prejuízos), presentes em cada estudo: (a) simples resumo para cada grupo de intervenção e (b) estimativas de efeito e intervalos de confiança, preferencialmente apresentados em um gráfico "forest-plot".
	Síntese dos Resultados	2/1/ 0/-1	Apresentar os resultados de cada meta-análise feita, incluindo intervalos de confiança e as medidas de consistência.
	Risco de viés através dos estudos	2/1/ 0/-1	Apresenta os resultados de qualquer avaliação de risco de viés através dos estudos (ver item 15)
	Análises adicionais	2/1/ 0/-1	Oferece os resultados de análises adicionais, se feitas (ex.: sensibilidade ou análise de subgrupo, meta-regressão [ver item 16]).
Discussão	Sumarização das Evidências	2/1/ 0/-1	Resume as principais conclusões, incluindo a força da evidência para cada resultado principal; considerar sua relevância para grupos-chave (ex.: prestadores de cuidados em saúde, usuários e decisores políticos).
	Limitações	2/1/ 0/-1	Discute as limitações no estudo e o nível dos resultados (ex.: risco de viés), e o nível de revisão (ex.: recuperação incompleta da pesquisa identificada, relatando o viés)
	Conclusões	2/1/ 0/-1	Oferece uma interpretação geral dos resultados no contexto de outras evidências, e implicações para pesquisas posteriores
Financiamento	Financiamento	2/1/ 0/-1	Descreve as fontes de financiamento para a revisão sistemática e outro suporte (como suprimento de dados) e o papel dos financiadores na revisão sistemática.

8.2 Ficha de extração de dados

Informações gerais

Título do artigo:

Autores:

Ano de publicação:

País de origem:

Idioma publicado:

Periódico de publicação:

Delineamento do estudo

Tipo de estudo:

n da amostra:

Perfil da amostra:

Avaliação Psicométrica

Questionário utilizado:

Propriedade(s) psicométrica(s) avaliada(s):

Teste de comparação:

Intervalo para reavaliação:

Estatística utilizada:

Pontos de corte:

Desfechos do estudo

Desfecho principal:

Desfechos secundários:

8.3 Checklist COSMIN de Risco de Viés

COSMIN content validity methodology

Instructions for completing the COSMIN boxes for content validity

- 1 CHECK the COSMIN website if the quality of the PROM development was already rated in another review. In that case, you can skip box 1A and use the quality ratings from the previous review
- 2 We recommend to score all PROMS with two raters, independently, and reach consensus afterwards. You can change "rater 1" and "rater 2" into the names of the raters
- 4 Add extra rows, columns or tables if needed
Tables 1, 2, and 3 will be filled automatically (you may need to add links to the other tabs). They can be included in a
- 5 systematic review

COSMIN box 1. Standards for evaluating the quality of PROM development

Check the COSMIN website to see if the quality of the PROM development was already rated in another review

Ratings: V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable

		PROM		
		Rater 1	Rater 2	Consensus
1a. PROM design				
<i>General design requirements</i>				
1	Is a clear description provided of the construct to be measured?			
2	Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured?			
3	Is a clear description provided of the target population for which the PROM was developed?			
4	Is a clear description provided of the context of use (i.e. discriminative, evaluative purpose, and/or predictive)			
5	Was the PROM development study performed in a sample representing the target population for which the PROM was developed?			
<i>Concept elicitation (relevance and comprehensiveness)</i>				
6	Was an appropriate qualitative data collection method used to identify relevant items for a new PROM?			
7	Were skilled group moderators/ interviewers used?			
8	Were the group meetings or interviews based on an appropriate topic or interview guide?			
9	Were the group meetings or interviews recorded and transcribed verbatim?			
10	Was an appropriate approach used to analyse the data?			
11	Was at least part of the data coded independently?			
12	Was data collection continued until saturation was reached?			
13	For quantitative studies: was the sample size appropriate?			
SUBTOTAL QUALITY CONCEPT ELICITATION STUDY <i>Lowest score of items 6-13</i>				
TOTAL QUALITY OF THE PROM DESIGN <i>Lowest score of items 1-13</i>				

1b. Cognitive interview study or other pilot test

		Rater 1	Rater 2	Consensus
14	Was a cognitive interview study or other pilot test performed? <i>If NO skip items 15-35</i>			
<i>General design requirements</i>				
15	Was the cognitive interview study or other pilot test performed in a sample representing the target population?			
<i>Comprehensibility</i>				
16	Were patients asked about the <u>comprehensibility</u> of the PROM? <i>If NO or not clear, skip items 17-25</i>			
17	Were all items tested in their final form?			
18	Was an appropriate qualitative method used to assess the <u>comprehensibility</u> of the PROM instructions, items, response options, and recall period?			
19	Was each item tested in an appropriate number of patients?			
20	Were skilled interviewers used?			
21	Were the interviews based on an appropriate interview guide?			
22	Were the interviews recorded and transcribed verbatim?			
23	Was an appropriate approach used to analyse the data?			
24	Were at least two researchers involved in the analysis?			
25	Were problems regarding the comprehensibility of the PROM instructions, items, response options, and recall period appropriately addressed by adapting the PROM?			
SUBTOTAL QUALITY OF COMPREHENSIBILITY STUDY <i>Lowest score of items 15-25</i>				
<i>Comprehensiveness</i>				
26	Were patients asked about the <u>comprehensiveness</u> of the PROM? <i>If NO or not clear, skip items 27-35</i>			
27	Was the final set of items tested?			
28	Was an appropriate method used for assessing the <u>comprehensiveness</u> of the PROM?			
29	Was each item tested in an appropriate number of patients?			
30	Were skilled interviewers used?			
31	Were the interviews based on an appropriate interview guide?			
32	Were the interviews recorded and transcribed verbatim?			
33	Was an appropriate approach used to analyse the data?			
34	Were at least two researchers involved in the analysis?			
35	Were problems regarding the <u>comprehensiveness</u> of the PROM appropriately addressed by adapting the PROM?			
SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY <i>Lowest score of items 15, 26-35</i>				
TOTAL QUALITY OF THE PILOT STUDY <i>Lowest score of items 14-35</i>				
TOTAL QUALITY OF THE PROM DEVELOPMENT STUDY <i>Lowest score of items 1-35</i>				

COSMIN box 2. Standards for evaluating the quality of content validity studies of PROMs

Only those parts of the box need to be completed for which information is available

Score: V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable

PROM

				PROM		
				rater 1	rater 2	Consensus
2a. Asking patient about relevance						
1	Was an appropriate method used to ask patients whether each item is <u>relevant</u> for their experience with the condition?					
2	Was each item tested in an appropriate number of patients?					
3	Were skilled group moderators/interviewers used?					
4	Were the group meetings or interviews based on an appropriate topic or interview guide?					
5	Were the group meetings or interviews recorded and transcribed verbatim?					
6	Was an appropriate approach used to analyse the data?					
7	Were at least two researchers involved in the analysis?					
SUBTOTAL QUALITY OF RELEVANCE STUDY <i>Lowest score of items 1-7</i>						
2b. Asking patients about comprehensiveness						
8	Was an appropriate method used for assessing the <u>comprehensiveness</u> of the PROM?					
9	Was each item tested in an appropriate number of patients?					
10	Were skilled group moderators/interviewers used?					
11	Were the group meetings or interviews based on an appropriate topic or interview guide?					
12	Were the group meetings or interviews recorded and transcribed verbatim?					
13	Was an appropriate approach used to analyse the data?					
14	Were at least two researchers involved in the analysis?					
SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY <i>Lowest score of items 8-14</i>						
2c. Asking patients about comprehensibility						
15	Was an appropriate qualitative method used for assessing the <u>comprehensibility</u> of the PROM instructions, items, response options, and recall period?					
16	Was each item tested in an appropriate number of patients?					
17	Were skilled group moderators/interviewers used?					
18	Were the group meetings or interviews based on an appropriate topic or interview guide?					
19	Were the group meetings or interviews recorded and transcribed verbatim?					
20	Was an appropriate approach used to analyse the data?					
21	Were at least two researchers involved in the analysis?					
SUBTOTAL QUALITY OF COMPREHENSIBILITY STUDY <i>Lowest score of items 15-21</i>						
2d. Asking professionals about relevance						
22	Was an appropriate method used to ask professionals whether each item is <u>relevant</u> for the construct of interest?					
23	Were professionals from all relevant disciplines included?					

24	Was each item tested in an appropriate number of professionals?			
25	Was an appropriate approach used to analyse the data?			
26	Were at least two researchers involved in the analysis?			
SUBTOTAL QUALITY OF RELEVANCE STUDY <i>Lowest score of items 22-26</i>				
2e. Asking professionals about comprehensiveness				
		rater 1	rater 2	Consensus
27	Was an appropriate method used for assessing the <u>comprehensiveness</u> of the PROM?			
28	Were professionals from all relevant disciplines included?			
29	Was each item tested in an appropriate number of professionals?			
30	Was an appropriate approach used to analyse the data?			
31	Were at least two researchers involved in the analysis?			
SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY <i>Lowest score of items 27-31</i>				

Criteria for content validity

To fill in ratings use apostrophe (') before the + / - / ± / ? signs

Score: + = sufficient; - = insufficient; ? = indeterminate; ± = inconsistent

PROM (subscale)		PROM development study	PROM development study	PROM development study
		+ / - / ?	+ / - / ?	+ / - / ?
		rater 1	rater 2	consensus
Relevance				
1	Are the included items relevant for the construct of interest? ¹			
2	Are the included items relevant for the target population of interest? ¹			
3	Are the included items relevant for the context of use of interest? ¹			
4	Are the response options appropriate?			
5	Is the recall period appropriate?			
RELEVANCE RATING (+ / - / ± / ?)				
Comprehensiveness				
6	Are all key concepts included?			
COMPREHENSIVENESS RATING (+ / - / ± / ?)				
Comprehensibility				
7	Are the PROM instructions understood by the population of interest as intended?			
8	Are the PROM items and response options understood by the population of interest as intended?			
9	Are the PROM items appropriately worded?			
10	Do the response options match the question?			
COMPREHENSIBILITY RATING (+ / - / ± / ?)				
CONTENT VALIDITY RATING (+ / - / ± / ?)				

¹ These criteria refer to the construct, population, and context of use of interest in the systematic review.

² Add more columns if more content validity studies are available

³ If ratings are inconsistent between studies, consider using separate tables for subgroups of studies with consistent results.

COSMIN Risk of Bias checklist

Only those parts of the boxes need to be completed for which information is available

Score: V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable

PROM

Article reference:

		PROM		
3. Structural validity		rater 1	rater 2	Consensus
unidimensionality or structural validity?				
1	For CTT: Was exploratory or confirmatory factor analysis performed?			
2	For IRT/Rasch: does the chosen model fit to the research question?			
3	Was the sample size included in the analysis adequate?			
4	Were there any other important flaws?			
TOTAL Lowest score of items 1-4				
4. Internal consistency		rater 1	rater 2	Consensus
1	Was an internal consistency statistic calculated for each unidimensional (sub)scale separately?			
2	For continuous scores: Was Cronbach's alpha or omega calculated?			
3	For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?			
4	For IRT-based scores: Was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?			
5	Were there any other important flaws?			
TOTAL Lowest score of items 1-5				
5. Cross-cultural validity/measurement invariance		rater 1	rater 2	Consensus
1	Were the samples similar for relevant characteristics except for the group variable?			
2	Was an adequate approach used to analyse the data?			
3	Was the sample size included in the analysis adequate?			
4	Were there any other important flaws?			
TOTAL Lowest score of items 1-4				
6. Reliability		rater 1	rater 2	Consensus
1	Were patients stable in the interim period on the construct to be measured?			
2	Was the time interval appropriate?			
3	Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions			
4	For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?			
5	For dichotomous/nominal/ordinal scores: Was kappa calculated?			
6	For ordinal scores: Was a weighted kappa calculated?			

7	For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic			
8	Were there any other important flaws?			
TOTAL <i>Lowest score of items 1-8</i>				

7. Measurement error		rater 1	rater 2	Consensus
1	Were patients stable in the interim period on the construct to be measured?			
2	Was the time interval appropriate?			
3	Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions			
4	For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?			
5	For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?			
6	Were there any other important flaws?			
TOTAL <i>Lowest score of items 1-6</i>				

8. Criterion validity		rater 1	rater 2	Consensus
1	For continuous scores: Were correlations, or the area under the receiver operating curve calculated?			
2	For dichotomous scores: Were sensitivity and specificity determined?			
3	Were there any other important flaws?			
TOTAL <i>Lowest score of items 1-3</i>				

9. Hypotheses testing for construct validity				
9a. Comparison with other outcome measurement instruments (convergent validity)		rater 1	rater 2	Consensus
1	Is it clear what the comparator instrument(s) measure(s)?			
2	Were the measurement properties of the comparator instrument(s) adequate?			
3	Was the statistical method appropriate for the hypotheses to be tested?			
4	Were there any other important flaws?			
TOTAL <i>Lowest score of items 1-4</i>				
9b. Comparison between subgroups (discriminative or known-groups validity)		rater 1	rater 2	Consensus
5	Was an adequate description provided of important characteristics of the subgroups?			
6	Was the statistical method appropriate for the hypotheses to be tested?			
7	Were there any other important flaws?			
TOTAL <i>Lowest score of items 5-7</i>				

10. Responsiveness				
10a. Criterion approach (i.e. comparison to a gold standard)		rater 1	rater 2	Consensus
1	For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?			
2	For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?			
3	Were there any other important flaws?			
TOTAL <i>Lowest score of items 1-3</i>				

Quality of the evidence for measurement properties of the PROMS

	OVERALL RATING	QUALITY OF EVIDENCE	OVERALL RATING	QUALITY OF EVIDENCE	OVERALL RATING	QUALITY OF EVIDENCE
	+ / - / ?	High, moderate, low, very low	+ / - / ?	High, moderate, low, very low	+ / - / ?	High, moderate, low, very low
Content validity						
<i>Relevance</i>						
<i>Comprehensiveness</i>						
<i>Comprehensibility</i>						
Structural validity						
Internal consistency						
Cross-cultural validity						
Measurement invariance						
Reliability						
Measurement error						
Criterion validity						
Construct validity						
Responsiveness						

8.4 Ferramenta EMPRO

I. CONCEPTUAL AND MEASUREMENT MODEL

The rationale for the concept and populations a measure is intended to assess and the relationship between these concepts must be clearly presented.

○ **Mark only one box**

1. The concept to be measured is clearly stated.

Aspects to be considered:

- *The broad concept the instrument is trying to measure (e.g., functional status, well-being, health-related quality of life, satisfaction, etc.) is clearly stated.*
- *If the instrument is designed to assess multiple domains, a listing of all the domains or dimensions is provided.*

Strongly Agree Strongly Disagree No information available

Reference: **Comments:**

2. The conceptual and empirical basis for obtaining the items for the instrument and for combining them into one or more dimensions is clearly stated and appropriate.

Aspects to be considered:

- *The conceptual basis and methods for generating items (e.g., by consulting existing literature, using expert group judgments, other questionnaires, etc.)*
- *The rationale for combining different items into one or more dimensions (e.g., by means of factor analysis, expert panel judgments) is clearly stated.*
- *The rationale for including an item in one dimension and not another is clearly stated.*

Strongly Agree Strongly Disagree No information available

Reference: **Comments:**

3. The dimensionality and distinctiveness of the scales is specifically described and well-supported.

Aspects to be considered:

- *The rationale for using single or multiple domains is clearly stated.*
- *The construct to be measured in each domain is well defined.*
- *If there is an index, the reasons for having it are justified.*

Strongly Agree Strongly Disagree No information available

Reference: **Comments:**

4. The involvement of the target population in obtaining the final content of the instrument is clearly described, the methods are appropriate and the results are satisfactory.

Aspects to be considered:

- *The target population is involved in the process of generating the item pool (e.g., through the use of focus groups).*
- *The appropriateness of the instrument's content for the target population has been evaluated (e.g., by means of pretesting in target populations, assessing clinical impact).*

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

5. Evidence of scale variability in the population is specifically described and appropriate to its intended use.

Aspects to be considered:

- Distribution, central tendency and dispersion of scores.
- Ceiling and floor effects.
- Pattern of missing data.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

6. The intended level of measurement is clearly defined and supportive evidence is provided.

Aspects to be considered:

- The intended level of measurement (e.g., ordinal, interval, or ratio scales) is clearly stated.
- Empirical evidence that justifies this level of measurement is provided.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

7. The rationale and procedures for deriving scale scores (global or by domains) from raw scores is clearly described.

Aspects to be considered:

- The procedures and transformations for deriving domain scores from raw scores are clearly described.
- Evidence is provided to justify whether or not certain items or domains should be weighted.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

II. RELIABILITY

II.1. Internal consistency.

Internal consistency refers to the precision of a scale based on the homogeneity of the scale's items at one point in time. The index used in classical psychometrics is Cronbach's coefficient alpha. Kuder-Richardson's formula 20 (KR-20) is appropriate for instruments employing dichotomous response choices. The standard error of measurement (SEM) is computed by multiplying the standard deviation of the score by the square root of 1-reliability. In the Item Response Theory (IRT) approach, measurement precision is evaluated at different points along the scale. Marginal reliability, a single value, can be estimated as an analog to the classical reliability coefficient.

8. The methods employed to collect internal consistency data are clearly described and adequate.*Aspects to be considered:*

- *Methods of sample accrual and sample size.*
- *Characteristics of the samples (socio-demographics, clinical characteristics if drawn from a patient population, etc.).*
- *The testing conditions (where and how the instrument was administered)*
- *Descriptive statistics for the results obtained with the instrument (means, standard deviation, floor and ceiling effects).*

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

9. Cronbach's coefficient alpha and/or KR-20 values are acceptable.

A rating of 4 should be given for total scores >0.9 and for sub-scale scores >0.7. If most of the coefficients for total and sub-scale scores are below 0.7, the rating cannot be higher than a 1 total score subscale score

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

10. Reliability estimates employing the IRT approach are clearly reported and acceptable.*Aspects to be considered:*

- *The plot showing the SEM over the scale range.*
- *The marginal reliability of the IRT application for each of the scores.*

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

11. Internal consistency data for each population of interest are clearly described.

If there is reason to believe that internal consistency estimates may differ for the various populations in which an instrument is to be used (e.g., different disease populations, cultural groups, age groups), data for each major population of interest must be presented. If no information is presented because there is no reason to believe that there will be such differences between populations, a rating of Not Applicable should be given.

Strongly Agree Strongly Disagree Not Applicable
No information available

Reference:	Comments:
-------------------	------------------

II.2. Reproducibility

Reproducibility refers to the stability of an instrument over time (test-retest among respondents who are assumed not to have changed) and inter-rater agreement at one point in time (for interviewer-administered instruments). In classical psychometrics the index typically used is the intraclass correlation coefficient (ICC) for interval-level data variables, and Kappa (unweighted and weighted for nominal and ordinal scales, respectively). In Item Response Theory (IRT) applications, stability is evaluated at different points along the scale.

12. The methods employed to collect reproducibility data are clearly described and appropriate-

Aspects to be considered:

- Methods of sample accrual and sample size.
- Characteristics of the sample (socio-demographics, clinical characteristics if drawn from a patient population, etc.).
- The testing conditions (where and how the instrument was administered).
- Descriptive statistics for the results obtained with the instrument.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

13. Well-argued rationale is provided to support the design of the test-retest comparison and the interval between first and subsequent administrations.

Aspects to be considered:

- The time interval between administrations.
- The method employed to determine the stability of the sample (self-report about perceived change in health or other general or specific measures).
- Independence between interviewers.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

14. Coefficients for test-retest reliability and/or inter-rater reliability are specifically described and adequate for all scores.

A rating of 4 should be given when the intraclass correlation coefficient (ICC) is >0.9 for total scores and >0.7 for subscale scores. If most of the scores of either kind are below 0.7, a rating of 1 should be given. Value reference estimates do not exist for the Kappa coefficient. When estimators other than ICC or Kappa are supplied (e.g., test-retest means, correlations, limits of agreement, or ROC curves) a maximum rating of 2 can be given.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

15. Item parameter estimates, using IRT applications, are adequately described and appropriate.

Aspects to be considered:

- Comparability of the item parameter estimates.
- Measurement precision over repeated administrations.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

III. VALIDITY

The validity of an instrument is defined as the degree to which the instrument measures what it purports to measure. Content-related validity refers to the appropriateness of the items and the domains, and is evaluated by means of panels of experts and patients. Construct-related validity refers to the hypotheses concerning relations with other measures and patterns of scores for known groups. Criterion-related validity refers to the extent to which scores of the instrument are related to a criterion measure or gold standard.

16. Sufficient evidence is presented regarding content-related validity of the instrument for its intended use.

Aspects to be considered:

- *Assessment of content-related validity in terms of clarity, comprehensiveness, relevance and redundancy of items and scales.*
- *Assessment of content-related validity by a panel of experts and/or target population.*
- *If the measure has been evaluated by only one type of panel, a maximum rating of 2 can be given.*

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

17. The methods employed to assess construct and criterion-related validity are clearly described and appropriate.

Aspects to be considered:

- *Methods of sample accrual and sample size.*
- *Characteristics of the sample (socio-demographics, clinical characteristics if drawn from a patient population, etc.).*
- *The testing conditions (where and how the instrument was administered).*
- *Descriptive statistics for the results obtained with the instrument (means, standard deviation, ceiling and floor effects).*

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

18. The composition of the sample used to examine construct and criterion validity is described in sufficient detail and appropriate.

The factors that might be expected to influence validity, such as gender, age, ethnicity and language, have been examined in sufficient detail.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

19. The hypotheses regarding construct validity are specifically described and the results are consistent with them.

Aspects to be considered:

- Hypotheses concerning the logical relations that should exist with other measures and/or patterns of scores for known groups. If no explicit hypotheses exist, a maximum rating of 2 can be given.
- Consistency between the hypotheses and the results with regard to the correlation with other measures.
- Consistency between the hypotheses and the results with regard to the pattern of scores for groups known to differ on relevant variables.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

20. A clear rationale and support for the choice of criteria measures or gold standards for criterion validity is provided.

When testing shorter versions of measures against longer ones and for testing screening instruments, information should be provided regarding the results obtained in comparisons with criterion measures (sensitivity, specificity and ROC curves). In other cases, it is acceptable for criterion validity not to be tested when no widely-accepted criterion measure exists, which is likely to be the case in most health-related quality of life questionnaires.

Strongly Agree Strongly Disagree Not Applicable No information available

Reference:	Comments:
-------------------	------------------

21. The testing of the validity of the instrument for each population of interest is clearly described.

When reasons exist to believe that validity will differ for the various groups or populations in which the instrument is to be used, (e.g., different disease groups, or cultural groups) validity data should be provided for each of them. If there is no evidence to suggest that differences in validity may exist for different populations, choose the rating of Not Applicable for this item.

Strongly Agree Strongly Disagree Not Applicable No information available

Reference:	Comments:
-------------------	------------------

IV. RESPONSIVENESS

Responsiveness refers to an instrument's ability to detect change over time

22. The methods employed to assess responsiveness are clearly described and appropriate.**Aspects to be considered:**

- Methods of sample accrual and sample size.
- Characteristics of the sample (socio-demographics, clinical characteristics if drawn from a patient population, etc.)
- The testing conditions (where and how the instrument was administered), the time interval and the type of intervention (if there is one) between administrations.
- Descriptive statistics for the results obtained with the instrument.

4 3 2 1

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

23. The estimated magnitude of change is clearly described and the results are acceptable.

Aspects to be considered:

- The estimated magnitude of change is provided for instrument scores in longitudinal studies.
- Adjusted or standardized measures are provided for the distance between before and after scores, e.g., effect sizes, standardized response means, standard error of measurement.
- The formulae used to calculate the responsiveness statistics are specified.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

24. The magnitude of change in a group that is expected to change has been compared with that of a group that is expected to remain stable in longitudinal studies.

Aspects to be considered:

- Longitudinal data are provided for the groups that have changed.
- Longitudinal data are provided for the groups that have remained stable.
- The difference in the magnitude of change between both groups has been evaluated.

A maximum rating of 3 can be given when the groups that are compared come from different studies.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

V. INTERPRETABILITY

The degree to which one can assign easily understood meaning to an instrument's quantitative scores. The most widely-used methods are:

- a. "Reference scores": derived from a representative sample of the general population or other populations;
- b. "Benchmarks": which make it possible to compare differences or changes in scores with other characteristics (e.g., clinical condition, presence or absence of a disease, minimally important changes, life events, therapy treatments, etc.);
- c. "External criteria" which make it possible to estimate the ability of scores to predict known relevant events, such as death or need for institutional care.

25. The rationale for the selection and evaluation of the external criteria are specifically described and well-supported.

Aspects to be considered:

- The rationale for selection of external criteria, "benchmarks" and/or comparison populations.
- The methods of sample accrual and sample size.
- Characteristics of the sample.
- The testing conditions.

Strongly Agree Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

26. The strategies to facilitate interpretation are clearly described and appropriate.

Aspects to be considered:

- Reference norms based on representative sample of the general population or other population groups are provided.
- "Benchmarks" are cited to facilitate the interpretation of scores with regard to other personal characteristics (e.g., clinical status, presence or absence of disease, etc).
- "Benchmarks" are cited that make it possible to interpret longitudinal changes (e.g., definition of minimally important changes, differences related to a specific intervention or life event).
- Information concerning the ability of the score to predict known relevant events, such as death or need for institutional care, is provided.

4
3
2
1

Strongly Agree
 Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

27. The way in which data from the instrument should be reported and displayed is clearly described.

Aspects to be considered:

- Theoretically expected rank and observed rank.
- Characteristics of the reference population used to calculate standardized scores.

4
3
2
1

Strongly Agree
 Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

VI. BURDEN

The resources needed to administer the instrument (time, effort, and other demands) should be assessed, in general terms. No standards have been defined for these characteristics.

VI.1. Respondent burden

28. The skills and time needed to complete the instrument are clearly described and acceptable.

Aspects to be considered:

- The average and range of time the respondent needs in order to complete the instrument (on a self-administered basis, in a face-to-face interview, by telephone or with a computer-assisted format/application).
- The reading and comprehension level needed.
- The need to consult health care records with the interviewee (medications used, diagnostic tests and others).
- The adequacy of the design of the instrument for the population it is intended for (font size, response options, breaks, etc.)

4
3
2
1

Strongly Agree
 Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

29. The assessment of the acceptability of the instrument and evidence that the instrument places no undue physical or emotional strain on the respondent are clearly described and acceptable.

Aspects to be considered:

- *Refusal rates (for the questionnaire as a whole and for the individual items) and the reasons for it.*
- *Rate of missing data (for the questionnaire as a whole and for the individual items) and the reasons for it.*
- *Mode of administration: (mail, self-administered, face-to-face interview, by telephone or computer-assisted formats/applications)*

4
3
2
1

Strongly Agree
 Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

30. Indications as to when or under what circumstances the instrument is not suitable for respondents.

Aspects to be considered:

- *Minimum cognitive level required for administration.*
- *Reading level in child population.*
- *Appropriate age-range for administration.*
- *Appropriate domain for administration (home, hospital and institution).*

4
3
2
1

Strongly Agree
 Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

VI.2. Administrative burden and scoring of the instrument

31. The resources required for administration of the instrument are specified.

The need for the following should be assessed: manual for the interviewer, support material (cards with response options, a calendar, photographs, etc.), need to consult the health records of the interviewee (list of medications, diagnostic tests and others), computers, telephones, specific computer software needed to administer or score the instrument.

4
3
2
1

Strongly Agree
 Strongly Disagree No information available

Reference:	Comments:
-------------------	------------------

32. The time required of a trained interviewer to administer the instrument is clearly indicated and it is acceptable.

*The average and range of time required of a trained interviewer to administer the instrument should be assessed.
In case of self-administered instruments, choose the rating of Not Applicable for this item.*

4
3
2
1
Strongly Disagree
Not Applicable

Strongly Agree
No information available

Reference:	Comments:
-------------------	------------------

33. The amount of training and level of education or professional expertise and experience needed to administer the instrument are clearly defined and acceptable.

Aspects to be considered:

- Amount of training required.
- Level of education or specialization required.

In case of self-administered instruments, choose the rating of Not Applicable for this item.

4
3
2
1
Strongly Disagree
Not Applicable

Strongly Agree
No information available

Reference:	Comments:
-------------------	------------------

34. Adequate information is provided for scoring the instrument and the burden associated with it is acceptable.

Aspects to be considered:

- The availability of scoring instructions.
- The need for computer software to calculate the scores.
- The need for external criteria to calculate the scores.

4
3
2
1
Strongly Disagree
No information available

Strongly Agree

Reference:	Comments:
-------------------	------------------

VII. ALTERNATIVE MODES OF ADMINISTRATION

Alternative modes of administration are those that differ from the ones the instrument was originally designed for, such as self-report, interviewer-administered, computer-assisted, performance-based measures or other modes of administration. Alternative modes may include self-administered versions of instruments originally intended to be completed directly by patients, when they are completed instead by proxy respondents such as parents, spouses, providers, etc.).

NOTE: the instrument you are assessing may not have any alternative modes of administration. If this is the case, mark the following option:

Not Applicable

35. The metric characteristics and use of each alternative mode of administration are specifically described and are adequate.

For each alternative mode of administration, the information available regarding the reliability, validity, responsiveness, interpretability and burden, should be assessed.

Strongly Agree Strongly Disagree Not Applicable
 No information available

Reference:	Comments:
-------------------	------------------

36. Information is provided concerning comparability of alternative modes of administration and the original and the results are acceptable.

Aspects to be considered:

- Comparison of the metric properties of alternative forms with the original.
- Appraisal of the equivalence of scores from alternative modes compared to each other.

Strongly Agree Strongly Disagree Not Applicable
 No information available

Reference:	Comments:
-------------------	------------------

VIII. CULTURAL AND LANGUAGE ADAPTATIONS

IMPORTANT:
This section will be assessed only for adapted instruments.

The cross-cultural adaptation of an instrument involves the assessment of conceptual and linguistic equivalence and the evaluation of measurement properties.

37. The methods applied to achieve and assess linguistic equivalence to the original are adequately described and appropriate.

Aspects to be considered:

- At least two independent forward translations from the source language have been carried out.
- At least one, preferably more, backward translations to the source language have been carried out.
- The translated versions have been reviewed by expert panels and patients (from among the population which the instrument is intended for).
- Field tests have been carried out to provide evidence of acceptability and intelligibility of the instrument.

Strongly Agree Strongly Disagree Not Applicable
 No information available

Reference:	Comments:
-------------------	------------------

38. The methods applied to achieve and assess conceptual equivalence of the items are adequately described and appropriate.

Aspects to be considered:

- Content validity has been assessed (equivalence in relevance and meaning of the same concepts being measured) in each cultural or linguistic group to which the instrument is to be applied, by means of expert panels and patients (or the population it is intended for).
- Possible differential item functioning (some items may function differently in one language than in the original and in the adapted version of the instrument), have been evaluated e.g., with item response theory, confirmatory or exploratory factor analysis, or other methods.

4
3
2
1

Strongly Agree

 Strongly Disagree
 No information available

Reference:	Comments:
-------------------	------------------

39. The significant differences between the original and the adapted versions are clearly identified and resolved satisfactorily.

Aspects to be considered:

- The most significant differences between the original and adapted versions have been identified.
- These differences are explained.
- Inconsistencies have been satisfactorily reconciled.

4
3
2
1

Strongly Agree

 Strongly Disagree
 No information available

Reference:	Comments:
-------------------	------------------

OVERALL ASSESSMENT

Degree of recommendation for questionnaires:

Considering the information available, would you recommend this instrument for measuring outcomes?

- 1 Strongly recommended
- 2 Recommended with provisos or alterations
- 3 Would not recommend
- 4 Not enough information available to determine the degree of recommendation

Comments:

Recommendations for the use of the questionnaire:

This instrument is **APPROPRIATE** for *(mark all applicable)*

- 1 Longitudinal use (monitoring over time)
- 2 Transversal use (discrimination between patients and groups)

Comments

--

APPENDIX:

This appendix aims to assess a more extensive format of recommendations than the one that is currently included in this instrument. For this reason, we would appreciate it if you could specify the recommended use for the questionnaire in terms of populations and applications endorsed by the review material.

1. Type of population:

- | | |
|---|--|
| <input type="checkbox"/> 1 General child population | <input type="checkbox"/> 3 General adolescent population |
| <input type="checkbox"/> 2 General adult population | <input type="checkbox"/> 4 General elderly population |

Patients (clinical domain)

- | | |
|---|--|
| <input type="checkbox"/> 1 Cardiovascular | <input type="checkbox"/> 11 Ophtalmology |
| <input type="checkbox"/> 2 Dermatology | <input type="checkbox"/> 12 Osteoarticular |

- 3 Digestive
- 4 Endocrinology
- 5 Infectious diseases
- 6 Geriatrics
- 7 Gynecology
- 8 Intensive care medicine
- 9 Nephrology
- 10 Neurology

- 13 Otorhynolaryngology
- 14 Pediatrics
- 15 Respiratory diseases
- 16 Mental health
- 17 Oral health
- 18 Sexuality
- 19 Urology
- 20 Other population types
(please specify in the box marked OTHER)

OTHER

Comments:

2. Use of the questionnaire:

- 1 Group comparison
- 2 Individual monitoring

3. Type of study:

- 1 Descriptive transversal study (questionnaire)
- 2 Cohort study
- 3 Population monitoring

4. Type of assessment:

- 1 Effectiveness studies
- 2 Cost-effectiveness studies (economic evaluation)
- 3 Cost-efficiency studies
- 4 Other types of studies (please specify in the box marked OTHER)

OTHER

Comments:

Appraisal completion date:

		/			/		
<i>day</i>			<i>month</i>			<i>year</i>	

The EMPRO (Evaluating the Measurement of Patient-Reported Outcomes) tool.

The search for outcomes which incorporate the patient's perspective has led to the production of an increasing number of "patient-reported outcome" (PRO) measures. This situation derives in two questions: 1) how to identify which questionnaires are available for a specific use; and 2) how to choose the most appropriate measure from among those available. Some initiatives have addressed the first question by compiling catalogues of questionnaires in different formats (including books and electronic databases). To solve the second question there have been some publications with defined criteria to identify the strengths and weaknesses of the PRO measures. One of the most exhaustive proposals was published by the Scientific Advisory Committee of the Medical Outcomes Trust (MOT), a non-profit organization from the United States which aims to provide quality control guidelines for the development of measures incorporating the patient's perspective in the field of health service research. Based initially on sound theoretical grounds, the MOT initiative was recently expanded to include an explicit guide (QualLife Res 2002;11:193–205).

The updated version of the MOT proposal was selected for providing the most comprehensive and explicit set of recommendations available. A panel of four experts was nominated to define the quality of a PRO instrument considering the "degree of confidence that all possible bias has been minimized and that the information about the process which led to its development and evaluation is clear and accessible". The content of the new tool was

generated by converting each of the specific criteria in the MOT proposal into individual items. The expert panel aimed to respect the original wording and to maximize the standardization of the assessment by incorporating quantitative criteria based on the literature.

The EMPRO tool was designed to measure the quality of patient reported outcome instruments. It is composed by eight attributes (39 items), assessing how well the development process of the instrument has been carried out ("conceptual and measurement model"), how well it performs in terms of metric properties ("reliability", "validity", "responsiveness", and "interpretability"), and how it is administrated ("burden", "alternative modes of administration", and "cross-cultural and linguistic adaptations").

EMPRO attributes and items are accompanied by a short description to facilitate understanding of the intended meaning and of its application in the evaluation process. Agreement with each item can be answered on a 4-point Likert scale, from 4 (strongly agree) to 1 (strongly disagree). Experts can choose the "no information" box, in case of insufficient information. Some items allow replying with "not applicable". Experts are asked to provide detailed comments to justify their rating on each item. These comments are helpful in the interpretation of the EMPRO scores and also in the consensus rounds.

1) APPLICATIONS

The most common use of the EMPRO tool is to evaluate the performance of a PRO instrument according to the scientific evidence available. EMPRO evaluations could be useful to select the most suitable PRO measure for a new study or a new program. Additionally, EMPRO could also be useful to evaluate the information on a PRO reported just in an article, or to evaluate original instruments or adapted versions, and could be even used as a checklist when preparing the protocol to develop a new instrument or a manuscript describing metric properties assessment.

In case of applying the EMPRO tool for selecting an instrument according to the psychometric properties reported in the available scientific publications, it is recommended that at least two experts conduct the evaluation of each instrument. We encourage to perform two rounds in the EMPRO evaluation. In the first round, every expert independently evaluates the assigned instrument by reviewing the scientific publications available and applying the EMPRO tool. In the second round, each expert is provided with the rating results of the other expert. In case of discrepancies, they are invited to reach a consensus. A third reviewer could be required to solve discrepancies.

2) SCORING EMPRO

The attribute-specific scores are obtained by calculating the mean response of the applicable items when at least 50% of the items are rated. Items for which the "no information" response option has been selected are assigned a score of 1 (lowest possible score). Afterwards, the scores are linearly transformed to a range of 0 (worst possible score) to 100 (best possible score). Separate sub-scores for the "reliability" attribute can be calculated as this attribute is formed by two components, internal consistency and reproducibility. The highest sub-score is then chosen for the "reliability" score. In addition to the attribute-specific scores, an overall score is obtained by calculating the mean of the five metric-related attributes: "conceptual and measurement model", "reliability", "validity", "responsiveness" and "interpretability". The overall score is only calculated when at least three of these five attributes have a score. EMPRO scores are considered reasonably acceptable if they reach at least 50 points (half the maximum score).

Please find attached the SPSS scoring algorithm for the EMPRO tool.

SCORING EMPRO

Step by step instructions for scoring the EMPRO tool.

Please note that all the algorithms would need the statistical software SPSS installed on your PC.

Step 1: Open your SPSS data file containing the EMPRO item responses.

Step 2: Copy the algorithm and paste it in SPSS syntax file.

Step 3: If the data file is in Excel or Access or in STATA, you need to convert your data file into SPSS. It can be done from SPSS ->File-> Menu.

Step 4: Open the SPSS Syntax file containing the Algorithm from SPSS File Menu.

Step 5: Run the Algorithm using the Run Icon in SPSS Syntax file or run all or selection from the Run Menu.

Scoring instructions

*The items should be labelled finalp1, finalp2, etc.

*Scores range from 1 to 4. The item responses should be entered as 1, 2, 3 or 4 according to the presentation order in the questionnaire.

*If more than the half of item responses in each domain are not in the score range (1-4), will not be a dimension score.

*5 is considered a missing value and 6 is considered not applicable.

*Missing values are recoded as 1 (the lowest score).

*Not applicable are handled by imputing the mean value within each domain using items before imputing missing values .

*To calculate the global scores, it is considered that missing values of the domains are 0 (less score) and include the scales concept, validity, reliability, responsiveness and interpretation.

NUMERIC finalp1_rec finalp2_rec finalp3_rec finalp4_rec finalp5_rec finalp6_rec finalp7_rec finalp8_rec


```

finalp9_rec finalp10_rec finalp11_rec
finalp12_rec finalp13_rec finalp14_rec finalp15_rec finalp16_rec finalp17_rec finalp18_rec finalp19_rec
finalp20_rec finalp21_rec finalp22_rec finalp23_rec finalp24_rec finalp25_rec finalp26_rec
finalp27_rec finalp28_rec finalp29_rec finalp30_rec finalp31_rec finalp32_rec finalp33_rec finalp34_rec
finalp35_rec finalp36_rec finalp37_rec finalp38_rec finalp39_rec (F8.2).
EXECUTE.

```

```

COMPUTE ninf_conc=sum(finalP1=5,finalP2=5,finalP3=5,finalP4=5,finalP5=5,finalP6=5,finalP7=5).
ninf_conc=sum(finalP1=5,finalP2=5,finalP3=5,finalP4=5,finalP5=5,finalP6=5,finalP7=5).
COMPUTE ninf_internalcons=sum(finalp8=5, finalp9=5, finalp10=5, finalp11=5).
COMPUTE ninf_reproduc=sum(finalp12=5, finalp13=5, finalp14=5, finalp15=5).
COMPUTE ninf_valid=sum(finalp16=5, finalp17=5, finalp18=5, finalp19=5, finalp20=5, finalp21=5).
COMPUTE ninf_respons=sum(finalp22=5, finalp23=5, finalp24=5).
COMPUTE ninf_interpr=sum(finalp25=5, finalp26=5, finalp27=5).
COMPUTE ninf_burden1=sum(finalp28=5, finalp29=5, finalp30=5).
COMPUTE ninf_burden2=sum(finalp31=5, finalp32=5, finalp33=5, finalp34=5).
COMPUTE ninf_modes=sum(finalp35=5, finalp36=5).
COMPUTE ninf_adapt=sum(finalp37=5,finalp38=5,finalp39=5).
EXECUTE.

```

```

COMPUTE NV_conc=7-ninf_conc-na_conc.
COMPUTE NV_internalcons=4-ninf_internalcons-NA_internalcons.
COMPUTE NV_reproduc=4-ninf_reproduc-NA_reproduc.
COMPUTE NV_valid=6-ninf_valid-NA_valid.
COMPUTE NV_respons=3-ninf_respons-NA_respons.
COMPUTE NV_interpr=3-ninf_interpr-na_interpr.
COMPUTE NV_burden1=3-ninf_burden1-na_burden1.
COMPUTE NV_burden2=4-ninf_burden2-na_burden2.
COMPUTE NV_modes=2-ninf_modes-NA_modes.
COMPUTE NV_adap=3-ninf_adapt-na_adapt.
EXECUTE.

```

```

RECODE
finalp1 finalp2 finalp3 finalp4 finalp5 finalp6 finalp7 finalp37 finalp38 finalp39 finalp8 finalp9 finalp10 finalp11
finalp12
finalp13 finalp14 finalp15 finalp16 finalp17 finalp18 finalp19 finalp20 finalp21 finalp22 finalp23 finalp24 finalp25
finalp26 finalp27 finalp28 finalp29 finalp30 finalp31 finalp32 finalp33 finalp34 finalp35 finalp36
(1=1) (2=2) (3=3) (4=4) (5=SYSMIS) (6=SYSMIS) INTO arg1 arg2 arg3 arg4 arg5 arg6 arg7 arg37 arg38
arg39 arg8 arg9 arg10
arg11 arg12 arg13 arg14 arg15 arg16 arg17 arg18 arg19 arg20 arg21 arg22 arg23 arg24 arg25 arg26 arg27
arg28 arg29
arg30 arg31 arg32 arg33 arg34 arg35 arg36.
EXECUTE .

```

```

COMPUTE mean_C=MEAN(arg1,arg2,arg3,arg4,arg5,arg6,arg7).
COMPUTE mean_IC = MEAN(arg8,arg9,arg10,arg11) .
COMPUTE mean_RP = MEAN(arg12,arg13,arg14,arg15).
COMPUTE mean_V=MEAN(arg16,arg17,arg18,arg19,arg20,arg21).
COMPUTE mean_Resp=MEAN(arg22,arg23,arg24).
COMPUTE mean_I=MEAN(arg25,arg26,arg27).
COMPUTE mean_B1=MEAN(arg28,arg29,arg30).
COMPUTE mean_B2=MEAN(arg31,arg32,arg33,arg34).
COMPUTE mean_M=MEAN(arg35,arg36).
COMPUTE mean_A=MEAN(arg37,arg38,arg39).
EXECUTE .

```

DELETE VARIABLES arg1 TO arg36.

RECODE

```
finalp1 finalp2 finalp3 finalp4 finalp5 finalp6 finalp7 finalp37 finalp38 finalp39 finalp8 finalp9 finalp10 finalp11
finalp12 finalp13 finalp14 finalp15 finalp16 finalp17 finalp18 finalp19 finalp20
finalp21 finalp22 finalp23 finalp24 finalp25 finalp26 finalp27 finalp28 finalp29 finalp30 finalp31 finalp32 finalp33
finalp34 finalp35 finalp36
(1=1) (2=2) (3=3) (4=4) (5=1) INTO finalp1_rec finalp2_rec finalp3_rec finalp4_rec finalp5_rec finalp6_rec
finalp7_rec finalp37_rec finalp38_rec finalp39_rec finalp8_rec
finalp9_rec finalp10_rec finalp11_rec finalp12_rec finalp13_rec finalp14_rec finalp15_rec finalp16_rec
finalp17_rec finalp18_rec finalp19_rec finalp20_rec finalp21_rec finalp22_rec
finalp23_rec finalp24_rec finalp25_rec finalp26_rec finalp27_rec finalp28_rec finalp29_rec finalp30_rec
finalp31_rec finalp32_rec finalp33_rec finalp34_rec finalp35_rec finalp36_rec .
EXECUTE .
```

```
IF (finalp1=6) finalp1_rec = mean_C .
  IF (finalp2=6) finalp2_rec = mean_C .
IF (finalp3=6) finalp3_rec = mean_C .
  IF (finalp4=6) finalp4_rec = mean_C .
IF (finalp5=6) finalp5_rec = mean_C .
  IF (finalp6=6) finalp6_rec = mean_C .
IF (finalp7=6) finalp7_rec = mean_C .
  IF (finalp8=6) finalp8_rec = mean_A .
IF (finalp9=6) finalp9_rec = mean_A .
  IF (finalp10=6) finalp10_rec = mean_A .
IF (finalp11=6) finalp11_rec = mean_IC .
  IF (finalp12=6) finalp12_rec = mean_IC .
IF (finalp13=6) finalp13_rec = mean_IC .
  IF (finalp14=6) finalp14_rec = mean_IC .
IF (finalp15=6) finalp15_rec = mean_RP .
  IF (finalp16=6) finalp16_rec = mean_RP .
IF (finalp17=6) finalp17_rec = mean_RP .
  IF (finalp18=6) finalp18_rec = mean_RP .
IF (finalp19=6) finalp19_rec = mean_V .
  IF (finalp20=6) finalp20_rec = mean_V .
IF (finalp21=6) finalp21_rec = mean_V .
  IF (finalp22=6) finalp22_rec = mean_V .
IF (finalp23=6) finalp23_rec = mean_V .
  IF (finalp24=6) finalp24_rec = mean_V .
IF (finalp25=6) finalp25_rec = mean_Resp .
  IF (finalp26=6) finalp26_rec = mean_Resp .
IF (finalp27=6) finalp27_rec = mean_Resp .
  IF (finalp28=6) finalp28_rec = mean_I .
IF (finalp29=6) finalp29_rec = mean_I .
  IF (finalp30=6) finalp30_rec = mean_I .
IF (finalp31=6) finalp31_rec = mean_B1 .
  IF (finalp32=6) finalp32_rec = mean_B1 .
IF (finalp33=6) finalp33_rec = mean_B1 .
  IF (finalp34=6) finalp34_rec = mean_B2 .
IF (finalp35=6) finalp35_rec = mean_B2 .
  IF (finalp36=6) finalp36_rec = mean_B2 .
IF (finalp37=6) finalp37_rec = mean_B2 .
  IF (finalp38=6) finalp38_rec = mean_M .
IF (finalp39=6) finalp39_rec = mean_M .
EXECUTE .
```

DO IF NV_CONC>=4.

COMPUTE

Concept=(MEAN(finalp1_rec,finalp2_rec,finalp3_rec,finalp4_rec,finalp5_rec,finalp6_rec,finalp7_rec)-1)*100/3.

ELSE.

COMPUTE Concept=9999.

END IF.

EXECUTE.

```

DO IF NV_internalcons>=2 .
    COMPUTE InternalConsistency=(MEAN(finalp8_rec,finalp9_rec,finalp10_rec,finalp11_rec)-1)*100/3.
ELSE.
    COMPUTE InternalConsistency=9999.
END IF.
EXECUTE.

DO IF NV_reproduc>=2 .
    COMPUTE Reproducibility=(MEAN(finalp12_rec,finalp13_rec,finalp14_rec,finalp15_rec)-1)*100/3.
ELSE.
    COMPUTE Reproducibility=9999.
END IF.
EXECUTE.

MISSING VALUES InternalConsistency Reproducibility ( 9999 ).

COMPUTE Reliability=MAX(InternalConsistency,Reproducibility).
EXECUTE.

DO IF NV_valid>=3 .
    COMPUTE
Validity=(MEAN(finalp16_rec,finalp17_rec,finalp18_rec,finalp19_rec,finalp20_rec,finalp21_rec)-1)*100/3.
ELSE.
    COMPUTE Validity=9999.
END IF.
EXECUTE.

DO IF NV_respons>=2 .
    COMPUTE Responsiveness=(MEAN(finalp22_rec,finalp23_rec,finalp24_rec)-1)*100/3.
ELSE.
    COMPUTE Responsiveness=9999.
END IF.
EXECUTE.

DO IF NV_interpr>=2 .
    COMPUTE Interpretation=(MEAN(finalp25_rec,finalp26_rec,finalp27_rec)-1)*100/3.
ELSE.
    COMPUTE Interpretation=9999.
END IF.
EXECUTE.

DO IF NV_burden1>=2 .
    COMPUTE Burden1=(MEAN(finalp28_rec,finalp29_rec,finalp30_rec)-1)*100/3.
ELSE.
    COMPUTE Burden1=9999.
END IF.
EXECUTE.

DO IF NV_burden2>=2 .
    COMPUTE Burden2=(MEAN(finalp31_rec,finalp32_rec,finalp33_rec,finalp34_rec)-1)*100/3.
ELSE.
    COMPUTE Burden2=9999.
END IF.
EXECUTE.

DO IF NV_modes>=2 .
    COMPUTE AdministrationMode=(MEAN(finalp35_rec,finalp36_rec)-1)*100/3.
ELSE.
    COMPUTE AdministrationMode=9999.
END IF.
EXECUTE.

```

```

DO IF NV_ADAP>=2 .
    COMPUTE Adaptations=(MEAN(finalp37_rec,finalp38_rec,finalp39_rec)-1)*100/3.
ELSE.
    COMPUTE Adaptations=9999.
END IF.
EXECUTE.

```

```

MISSING VALUES Concept Adaptations InternalConsistency Reproducibility Reliability Validity Responsiveness
Interpretation Burden1 Burden2
AdministrationMode( 9999 ).

```

```

RECODE
    Concept Reliability Validity Responsiveness Interpretation
    (MISSING=0) (ELSE=Copy) INTO concept_0 reliability_0 validity_0 responsiveness_0 Interpretation_0.
EXECUTE .

```

```

COMPUTE Overall_EMPRO=MEAN(concept_0,reliability_0,validity_0,responsiveness_0,interpretation_0).
EXECUTE.

```

```

COMPUTE n_WITHMISS_5=sum(MISSING(Concept), MISSING(Reliability), MISSING(Validity),
MISSING(Responsiveness), MISSING(Interpretation)).
EXECUTE.

```

```

DELETE VARIABLES finalp1_rec finalp2_rec finalp3_rec finalp4_rec finalp5_rec finalp6_rec finalp7_rec
finalp8_rec finalp9_rec finalp10_rec
finalp11_rec finalp12_rec finalp13_rec finalp14_rec finalp15_rec finalp16_rec finalp17_rec finalp18_rec
finalp19_rec
finalp20_rec finalp21_rec finalp22_rec finalp23_rec finalp24_rec finalp25_rec finalp26_rec finalp27_rec
finalp28_rec
finalp29_rec finalp30_rec finalp31_rec finalp32_rec finalp33_rec finalp34_rec finalp35_rec finalp36_rec
finalp37_rec
finalp38_rec finalp39_rec concept_0 reliability_0 validity_0 responsiveness_0 Interpretation_0.

```

8.5 Tabela de descrição das PROMs avaliadas nos estudos incluídos

Tabela 1A. Descrição das PROMs avaliadas nos estudos incluídos.

PROM	Descrição	Modo de administração	Composição	Pontuação
<i>Chronic Kidney Disease Physical Activity Questionnaire</i> (CKD-PAQ)	Criado a partir do RPAQ direcionada especificamente para DRC tanto dialítico quanto não-dialítico	Autoadministrado	27 itens organizados em 3 seções: “Atividades domésticas” (A) / “Atividades laborais” (B) / “Atividades recreativas” (C) ^{45,46} e 5 itens adicionais para indivíduos que realizam HD a respeito dos dias de diálise Contabiliza a intensidade (vezes/semana) e duração (em horas e minutos) de cada atividade nas últimas 4 semanas	É realizada a estimativa de gasto metabólico a partir do valor de MET correspondente para cada tipo de atividade.
<i>Duke Activity Status Index</i> (DASI)	Criado para avaliar a população geral, já validado para doenças cardiovasculares e Doença Pulmonar Obstrutiva Crônica (DPOC)	Por avaliador Autoadministrado	Possui no total 12 questões a respeito de atividades de diferentes intensidades que o indivíduo pode realizar Cada questão possui um valor de MET correspondente	É realizada a somatória do peso de cada questão Pontuação mínima: 0 Pontuação máxima: 58,2
<i>Exercise Self-Efficacy Scale</i> (ESES)	Desenvolvido para avaliar população em geral a respeito da percepção de eficácia e confiança em realizar atividade física. Carece de validação em perfis clínicos	Autoadministrado	Possui 10 questões a respeito da percepção de eficácia e confiança em realizar atividade física em diferentes intensidades Cada questão possui uma escala Likert de 4 pontos: “Nunca” (1) / “Raramente” (2) / “Às vezes” (3) / “Sempre” (4)	É realizada a somatória das questões Pontuação mínima: 10 Pontuação máxima: 40
<i>Four Week Physical Activity History Questionnaire</i> (FWH)	Modificação do <i>Minnesota Leisure Time Activities Questionnaire</i> , com a diferença que aborda os hábitos de atividade física do indivíduo considerando as últimas 4 semanas	Autoadministrado	Possui 63 questões agrupadas em 9 categorias de diferentes tipos de atividades (doméstica, recreativa e ocupacional) Para cada atividade é indicado o número médio de vezes que foi realizada nas últimas 4 semanas e o tempo médio (em minutos) gasto na mesma	A partir das respostas é realizada a estima de gasto energético: Gasto energético = $\sum (I \times M \times F \times T)$ <ul style="list-style-type: none"> • \sum = somatória de cada tipo de atividade (doméstica, recreativa e ocupacional) • I = intensidade de cada atividade em equivalente metabólico (MET) • M = número de vezes por semana em que a atividade foi realizada • F = número médio de vezes em que foi realizada nas últimas 4 semanas • T = duração média da atividade em cada ocasião

Tabela 1A. Descrição das PROMs avaliadas nos estudos incluídos. (Continuação)

PROM	Descrição	Modo de administração	Composição	Pontuação
<i>General Practice Physical Activity Questionnaire</i> (GPPAQ)	Elaborado como um instrumento de avaliação rápida da atividade física na atenção primária para a população em geral. Validado para avaliação de indivíduos de 16 a 74 anos Existe a versão caminhada do questionário (GPPAQ-Caminhada) que considera caminhada como atividade física	Autoadministrado	Composto por 3 questões: <ul style="list-style-type: none"> Atividade física no trabalho Atividade física na última semana Velocidade de caminhada usual (para a versão do questionário GPPAQ-Caminhada) 	A partir da sumarização das respostas, o resultado é categorizado conforme os 4 níveis: <ul style="list-style-type: none"> Inativo: trabalho sedentário e nenhuma atividade física/ciclismo Moderadamente inativo: trabalho sedentário e < 1 hora por semana de atividade física e/ou ciclismo, ou trabalho em pé e sem atividade física/ciclismo Moderadamente ativo: trabalho sedentário e 1-2,9 horas por semana de atividade física e/ou ciclismo, ou trabalho em pé e < 1 hora por semana de atividade física e/ou ciclismo, ou trabalho físico e nenhuma atividade física/ciclismo Ativo: trabalho sedentário e ≥ 3 horas por semana de atividade física e/ou ciclismo, ou trabalho em pé e 1-2,9 horas por semana de atividade física e/ou ciclismo, ou trabalho físico e < 1 hora por semana de atividade física e/ou ciclismo, ou trabalho manual pesado Na versão GPPAQ-Caminhada, considera-se o ponto de corte de ≥ 3 horas por semana de caminhada em velocidade rápida como categorização de "Ativo"
<i>Human Activity Profile</i> (HAP)	Criado para avaliar atividade física na DPOC, já foi utilizado na literatura em diferentes patologias, com boas propriedades psicométricas	Por avaliador	Dois domínios: Atividade (94 questões) Dispneia (8 questões) No domínio atividade, as questões estão organizadas de forma ascendente ao nível de energia requerido para performá-la, considerando o equivalente metabólico (MET) Cada questão possui 3 opções de resposta: "Ainda realiza" (3) / "Parou de realizar" (2) / "Nunca realizou" (1)	Domínio Atividade: Existem 2 formas de pontuação: <ul style="list-style-type: none"> Escore de atividade máxima (MAS): valor numérico que identifica a atividade com maior consumo de O₂ que o indivíduo performa Escore de atividade ajustado (AAS): diferença do MAS e o número de atividades abaixo da pontuação máxima que o indivíduo parou de realizar Pontuação mínima: 94 Pontuação máxima: 282
<i>International Physical Activity Questionnaire</i> (IPAQ)	Elaborado para avaliar o nível de atividade física em diferentes populações com diferentes contextos socioeconômicos	Autoadministrado	Avalia o tempo despendido (em minutos) em uma semana habitual nas atividades sentado, caminhada, moderada e vigorosa. Versão longa: 27 questões Versão curta: 8 questões	Classificação em quatro níveis de atividade física: <ul style="list-style-type: none"> Muito ativo: pelo menos 5 dias/semana de atividade física vigorosa por no mínimo 30 minutos/dia ou pelo menos 3 dias/semana de atividade física vigorosa com no mínimo 20 minutos/dia e atividade moderada e/ou leve por pelo menos 5 dias/ com no mínimo 30 minutos/dia Ativo: atividade vigorosa pelo menos 3 dias/semana, com no mínimo de 20 minutos/dia; ou atividade moderada ou leve por pelo menos 5 dias/semana com no mínimo 30 minutos/dia; ou que realiza atividade física combinada (leve, moderada e vigorosa) por mais de 5 dias/semana, com no mínimo 150 minutos/semana Irregularmente ativo: realiza atividade física, porém insuficiente para ser classificado como ativo Sedentário: não realiza nenhuma atividade física por pelo menos 10 minutos contínuos

Tabela 1A. Descrição das PROMs avaliadas nos estudos incluídos. (Continuação)

PROM	Descrição	Modo de administração	Composição	Pontuação
<i>Low Physical Activity Questionnaire</i> (LoPAQ)	Desenvolvido especificamente para indivíduos dialíticos, com o objetivo de avaliar atividades físicas de (com enfoque na baixa intensidade) na última semana	Por avaliador Autoadministrado	23 itens organizados em 11 questões acerca de hábitos de atividade (em diferentes intensidades), tempo sentado e dormindo na última semana Concentra-se em atividades domésticas e recreativas Utiliza-se o tempo (em minutos) dispendido em atividade, sentado e dormindo; bem como o gasto em quilocalorias (kcal) em cada modalidade de atividade	Contabiliza o tempo em cada tipo de intensidade de atividade, que possui MET correspondente, e assim calculado o gasto energético em kcal
<i>Stanford 7-Day Physical Activity Recall Questionnaire</i> (PAR)	Criado para mensuração de atividade física na população em geral, já validado em DPOC	Por avaliador	9 questões sobre tempo dispendido (em horas) em diferentes tipos de intensidade de atividades (tanto laborais quanto recreativas) Cada questão possui uma escala Likert de 4 pontos: "Nunca" (1) / "Raramente" (2) / "Às vezes" (3) / "Sempre" (4)	Contabiliza o tempo em cada tipo de intensidade de atividade, que possui MET correspondente, e assim calculado o gasto energético em kcal
<i>Physical Activity Scale for the Elderly</i> (PASE)	Criado para avaliar indivíduos acima de 65 anos, já foi validado em diferentes perfis clínicos	Por avaliador Autoadministrado	12 questões sobre frequência e duração de atividade recreativa, doméstica e laborativa nos últimos 7 dias Cada questão possui 4 opções de resposta: "Nunca" (0) / "Raramente" (1) / "Algumas vezes" (2) / "Frequentemente" (3)	Cada opção de resposta (com exceção de "Nunca") possui 4 categorizações a partir do tempo dispendido (horas/dia), e cada categoria possui um peso diferente. A pontuação final é dada pela somatória do peso de cada questão Pontuação mínima: 0 Pontuação máxima: 400
<i>Recent Physical Activity Questionnaire</i> (RPAQ)	Baseado no <i>Norfolk Physical Activity Questionnaire</i> (EPAQ2), tem como objetivo avaliar a população geral quanto aos hábitos de atividade física nas últimas 4 semanas	Autoadministrado	57 itens organizados em 3 seções: "Atividades domésticas" (A) / "Atividades laborais" (B) / "Atividades recreativas" (C) Contabiliza a intensidade (vezes/semana) e duração (em horas e minutos) de cada atividade nas últimas 4 semanas	É realizada a estimativa de gasto metabólico a partir do valor de MET correspondente para cada tipo de atividade.