

NAYARA FERREIRA FREITAS VERMELHO

**Propriedades de medida dos instrumentos de avaliação do
equilíbrio e risco de quedas na doença de Parkinson: uma
revisão sistemática**

**Belo Horizonte
2014**

NAYARA FERREIRA FREITAS VERMELHO

Propriedades de medida dos instrumentos de avaliação do equilíbrio e risco de quedas na doença de Parkinson: uma revisão sistemática

Dissertação apresentada ao Programa de Pós-graduação em Ciências da Reabilitação da Escola de Educação Física, Fisioterapia e Terapia Ocupacional da Universidade Federal de Minas Gerais como requisito parcial á obtenção do título de Mestre em Ciências da Reabilitação.

Área de concentração: Desempenho Funcional Humano

Orientadora: Prof^a. Fátima Rodrigues de Paula, Ph.D

Coorientadora: Prof^a. Aline Scianni, Ph.D

Belo Horizonte
2014

V473p Vermelho, Nayara Ferreira Freitas
2014 Propriedades de medida dos instrumentos de avaliação do equilíbrio e risco de quedas na Doença de Parkinson: uma revisão sistemática. [manuscrito] / Nayara Ferreira Freitas Vermelho – 2014.
93f., enc.: il.

Orientadora: Fátima Valéria Rodrigues de Paula
Co-orientadora: Aline Alvim Scianni

Mestrado (dissertação) – Universidade Federal de Minas Gerais, Escola de Educação Física, Fisioterapia e Terapia Ocupacional.

Bibliografia: f. 60-63

1. Doença de Parkinson - Teses. 2. Acidentes por quedas - Teses. 3. Capacidade motora – Teses. I. Paula, Fátima Valéria Rodrigues de. II. Scianni, Aline Alvim. III. Universidade Federal de Minas Gerais. Escola de Educação Física, Fisioterapia e Terapia Ocupacional. IV. Título.

CDU: 796.015

COLEGIADO DE PÓS-GRADUAÇÃO EM CIÊNCIAS EM REABILITAÇÃO
 DEPARTAMENTOS DE FISIOTERAPIA E DE TERAPIA OCUPACIONAL
 SITE: www.eeffto.ufmg.br/mreab E-MAIL: mreab@eeffto.ufmg.br FONE/FAX: (31) 3409-4781/7395

ATA DE NÚMERO 213 (DUZENTOS E TREZE) DA SESSÃO DE ARGUIÇÃO E DEFESA DE DISSERTAÇÃO APRESENTADA PELA CANDIDATA **NAYARA FERREIRA FREITAS VERMELHO** DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA REABILITAÇÃO.....

Aos 21 (vinte um) dias do mês de maio do ano de dois mil e quatorze, realizou-se na Escola de Educação Física, Fisioterapia e Terapia Ocupacional, a sessão pública para apresentação e defesa da dissertação **“INSTRUMENTOS DE AVALIAÇÃO DE EQUILÍBRIO E RISCO DE QUEDAS NA DOENÇA DE PARKINSON: Uma revisão sistemática das propriedades de medida”**A banca examinadora foi constituída pelos seguintes Professores Doutores: Fatima Valéria Rodrigues de Paula, Paula Luciana Scalzo, Rosângela Corrêa Dias sob a presidência da primeira. Os trabalhos iniciaram-se às 09 horas com apresentação oral da candidata, seguida de arguição dos membros da Comissão Examinadora. **Após avaliação, os examinadores consideraram a candidata aprovada e apta a receber o título de Mestre, após a entrega da versão definitiva da dissertação.** Nada mais havendo a tratar, eu, Marilane Soares, secretária do Colegiado de Pós-Graduação em Ciências da Reabilitação dos Departamentos de Fisioterapia e de Terapia Ocupacional, da Escola de Educação Física, Fisioterapia e Terapia Ocupacional, lavrei à presente Ata, que depois de lida e aprovada será assinada por mim e pelos membros da Comissão Examinadora. Belo Horizonte, 21 de maio de 2014.

Professora Dra. Fátima Valéria Rodrigues de Paula

Fátima R. de Paula

Professora Dra. Paula Luciana Scalzo

Paula Luciana Scalzo

Professora Dra. Rosângela Corrêa Dias

Rosângela Corrêa Dias

Marilane Soares

Marilane Soares

Secretária do Colegiado de Pós-Graduação em Ciências da Reabilitação

COLEGIADO DE PÓS-GRADUAÇÃO EM CIÊNCIAS EM REABILITAÇÃO
DEPARTAMENTOS DE FISIOTERAPIA E DE TERAPIA OCUPACIONAL
SITE: www.eeffto.ufmg.br/mreab E-MAIL: mreab@eeffto.ufmg.br
FONE/FAX: (31) 3409-4781

PARECER

Considerando que a dissertação de mestrado de **NAYARA FERREIRA FREITAS VERMELHO** intitulada "**INSTRUMENTOS DE AVALIAÇÃO DE EQUILÍBRIO E RISCO DE QUEDAS NA DOENÇA DE PARKINSON: Uma revisão sistemática das propriedades de medida**", defendida junto ao Programa de Pós-Graduação em Ciências da Reabilitação, nível mestrado, cumpriu sua função didática, atendendo a todos os critérios científicos, a Comissão Examinadora **APROVOU** a defesa de dissertação, conferindo-lhe as seguintes indicações:

Nome dos Professores/Banca	Aprovação	Assinatura
Fátima Valéria Rodrigues de Paula	X	<i>Fátima Valéria Rodrigues de Paula</i>
Paula Luciana Scalzo	X	<i>Paula Luciana Scalzo</i>
Rosângela Corrêa Dias	X	<i>Rosângela Corrêa Dias</i>

Belo Horizonte, 21 de maio de 2014.

Leani Souza Maximo Pereira
Colegiado de Pós-Graduação em Ciências da Reabilitação/EEFFTO/UFMG

Prof. LEANI SOUZA MAXIMO PEREIRA
Coordenadora do Colegiado
Pós-Graduação em Ciências da Reabilitação
Inscrição UFMG: 06081X Inscrição SIAPE: 0319760

Dedico este trabalho ao meu esposo amado, Samuel, que tem sido meu fiel companheiro. Sua paciência, cumplicidade e suporte fizeram toda a diferença.

Ao meu querido filho Isaías, que me ensinou que todas as barreiras podem ser vencidas quando o amor é verdadeiro.

À minha amada sogra, Marta, o seu suporte foi vital para a realização deste projeto.

AGRADECIMENTOS

Em primeiro lugar agradeço ao meu Deus, o Eterno de Israel, por me dar fôlego de vida, inteligência e força para sonhar e realizar este projeto.

Aos meus pais pelo incentivo e dedicação em momentos em que tanto precisei. Mãe, fico feliz em poder realizar um sonho seu. Pai, seu olhar de admiração e satisfação com este trabalho me abençoou ricamente.

Ao meu sogro e minha sogra que são pais brilhantes e avós fascinantes.

Aos meus irmãos pela alegria compartilhada deste projeto.

Agradeço aos meus cunhados e cunhadas que se transformaram em irmãos amados e cuidaram de mim quando mais precisei.

À minha amada congregação Har Tzion na qual encontrei refúgio e conforto nas horas difíceis.

Em especial á minha querida orientadora Fátima Valéria Rodrigues de Paula, que ousou acreditar em mim, que me abraçou como mãe em um momento importantíssimo: a gestação e a maternidade.

À Aline Alvim Scianni. Como foi bom e agradável te conhecer, como foi importante a sua contribuição neste projeto.

À Lidiane Lima. Obrigada pelo suporte acadêmico e, principalmente, pessoal.

Às professoras Rosângela Corrêa Dias e Paula Luciana Scalzo que de uma maneira tão sensata e carinhosa nos ajudaram a refletir sobre este projeto.

A todos os professores que contribuíram para o meu desenvolvimento acadêmico durante este curso.

Ao meu filho Isaías. Você foi gerado juntamente com este projeto, mas nenhum projeto acadêmico poderia me tornar tão feliz quanto a sua existência.

RESUMO

Contextualização: A instabilidade postural tem sido apontada como um dos fatores de risco para a ocorrência de quedas em indivíduos com doença de Parkinson (DP), sendo mais freqüente em indivíduos com maior tempo de evolução da doença. O impacto da instabilidade postural e quedas está no fato de que suas conseqüências podem ser devastadoras para o indivíduo com DP tais como dores, limitações de atividades, medo da queda, depressão e fraturas. Evidências têm mostrado que o indivíduo com DP pode se beneficiar da fisioterapia, como por exemplo, através de treino de marcha e treino de tarefas específicas de controle postural. A avaliação do equilíbrio é um componente chave da avaliação fisioterápica em pacientes com DP, pois permite ao profissional identificar e quantificar os déficits de equilíbrio e risco de quedas bem como monitorar a progressão dos mesmos e, desta forma, realizar uma adequada tomada de decisão clínica direcionada à intervenção. O instrumento de medida é parte do processo de avaliação, sendo fundamental sua utilização, pois possibilita a obtenção de dados objetivos e quantitativos acerca do que se pretende medir possibilitando comparações entre diferentes populações. A qualidade de um instrumento de medida está diretamente relacionada às suas propriedades de medida tais como confiabilidade, validade e responsividade. Objetivos: Identificar quais instrumentos clínicos de avaliação de equilíbrio e risco de quedas são utilizados na DP, caracterizar as propriedades de medida de tais instrumentos e determinar a qualidade metodológica dos estudos envolvidos. Materiais e método: O presente estudo é uma revisão sistemática de estudos metodológicos sobre as propriedades de medida dos instrumentos de avaliação clínica do equilíbrio e risco de quedas. As bases de dados MEDLINE, LILACS, CINAHL e SPORTdiscus foram pesquisadas até março de 2013. Foram incluídos estudos metodológicos que desenvolveram os instrumentos de avaliação clínica do equilíbrio e risco de quedas na DP bem como avaliaram suas propriedades de medida. A avaliação da qualidade metodológica foi realizada por dois examinadores independentes por meio do *Checklist* COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments). As estimativas foram sintetizadas para cada instrumento e as propriedades de medida

foram classificadas quanto aos níveis de evidência forte, moderada, limitada, conflitante e desconhecida. Resultados: A estratégia de busca inicial identificou 150 estudos, sendo 20 incluídos na revisão sistemática. Quatorze instrumentos foram encontrados, sendo que moderada evidência foi apresentada para uma adequada consistência interna da Berg Balance Scale (BBS), adequada confiabilidade teste-reteste do Survey of Activities and Fear of Falling in the Elderly (SAFFE – NL) e Swedish Falls Efficacy Scale (FES-S) e adequada validade de critério do Balance Evaluation Systems Test (BESTEST) e Functional Gait Assessment (FGA). Conclusão: A qualidade metodológica dos estudos e os níveis de evidência encontrados destacaram cinco instrumentos potencialmente mais adequados, na atualidade, para avaliar o equilíbrio ou o risco de quedas na DP. São eles: BBS, BESTEST, FGA na categoria de instrumentos de desempenho, e SAFFE-NL e FES-S na categoria de questionários de auto-relato.

Palavras-chave: Doença de Parkinson, avaliação clínica, equilíbrio, quedas, propriedades de medida, COSMIN.

ABSTRACT

Background: Postural instability has been implicated as a risk factor for the occurrence of falls in people with Parkinson's disease (PD), being more frequent in the late stages of the disease, leading to impaired balance and frequent falls. The consequences of the postural instability can have devastating effects such as pain, activity limitations, fear of falling, depression and fractures. Evidence has shown that individuals with PD can benefit from physical therapy interventions, such as gait training and specific tasks of postural control training. The measure of balance is a key component of physical therapy evaluation in patients with PD, because it allows the professional to identify and quantify the deficits of balance and risk of falls and monitor the progression of the same and thus conduct a appropriate clinical decision making. Appropriate measurement instruments are essential to obtain objective and quantitative data about what it intends to measure, allowing comparisons between different populations. The quality of a measurement instrument is directly related to its measurement properties such as reliability, validity and responsiveness.

Objectives: To identify the most utilized clinical assessment instruments of balance and risk of falls for PD, to characterize the measurement properties of the instruments, and to determine the methodological quality of the studies analysed.

Materials and methods: This study is a systematic review of methodological studies of measurement properties of clinical assessment instruments of balance and risk of falls. The databases MEDLINE, LILACS, CINAHL and SPORTDiscus were searched until March 2013. Methodological studies which developed the clinical assessment instruments of balance and risk of falls in PD as well as those which rated their measurement properties were included. The assessment of methodological quality was performed by two independent raters using the Checklist COSMIN (Consensus - based Standards for the Selection of Health Measurement Instruments). The estimates were summarized for each instrument and the measurement properties were rated according to the levels of evidence strong, moderate, limited, conflicting and unknown

Results: The initial search strategy identified 150 studies, of which 20 were included in the systematic review. Fourteen instruments were found. Moderate evidence was presented for an adequate internal consistency of the Berg Balance Scale (BBS), adequate test-retest reliability of Survey of Activities and Fear of Falling in the Elderly (SAFFE – NL) and Swedish Falls Efficacy Scale (FES-S) and adequate

criterion validity of the Balance Evaluation Systems Test (BESTEST) and Functional Gait Assessment (FGA). Conclusion: The methodological quality of studies and level of evidence analysis highlighted five potentially most appropriate instruments at present to measure the balance or the risk of falls in PD. They are: BBS, BESTEST, FGA in the category of performance instruments and SAFFE and FES-S in the category of self-report questionnaires.

Keywords: Parkinson's disease, clinical assessment, balance, falls, measurement properties, COSMIN.

PREFÁCIO

Este trabalho foi elaborado no formato opcional para dissertação de mestrado. Sendo assim, na primeira parte encontram-se a introdução ampliada e metodologia, e na segunda parte, o artigo já nas normas do periódico para o qual será enviado, contendo toda a discussão e conclusão deste projeto.

Por se tratar de estudo de revisão sistemática, após o artigo, estão contidas todas as tabelas com os dados que foram coletados e analisados. E após as considerações finais estão disponíveis todos os anexos e apêndices que serviram de ferramenta para a execução desta revisão.

Dado à diversidade de instrumentos encontrados bem como das propriedades de medida analisadas, o grande desafio deste trabalho foi conseguir realizar uma síntese objetiva de todos os dados encontrados de maneira que pudesse ser traduzida em significado clínico que fosse útil tanto para os profissionais da clínica quanto para pesquisadores de distúrbios de equilíbrio na DP.

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Doença de Parkinson.	14
1.2 Controle postural e instabilidade na Doença de Parkinson	16
1.3 Avaliação do equilíbrio	19
1.4 Propriedades de medida	21
1.5 Avaliação da qualidade metodológica	22
1.6 Justificativa	23
2 OBJETIVOS	25
3 MATERIAIS E MÉTODO	26
3.1 Desenho do estudo	26
3.2 Estratégia de busca	26
3.3 Critérios de elegibilidade	27
3.4 Seleção dos estudos	27
3.5 Extração dos dados	28
3.6 Avaliação da qualidade metodológica	28
3.7 Síntese dos resultados	30
3.8 Apresentação dos resultados	32
4 ARTIGO	33
5 CONSIDERAÇÕES FINAIS	58
6 REFERÊNCIAS	60
7 ANEXOS	64
Anexo 1: <i>Checklist</i> COSMIN	64
Anexo 2: Avaliação dos níveis de evidência	87
Anexo 3: Escala de avaliação da utilidade clínica	88
8 APÊNDICES	89
Apêndice 1: Estratégia de busca	89
Apêndice 2: Protocolo para inclusão dos estudos	91
Apêndice 3: Formulário de extração dos dados	92

1. INTRODUÇÃO

1.1 Doença de Parkinson

A doença de Parkinson (DP) é uma doença neurodegenerativa progressiva caracterizada pela perda das células pigmentadas da substância negra parte compacta com presença de corpos de Lewy nas células remanescentes e diminuição nos níveis de dopamina nos núcleos da base com conseqüente diminuição de seu metabolismo (Marsden, 1994; Bartels and Leenders, 2009).

Classificada como a segunda desordem neurodegenerativa, após a doença de Alzheimer, a DP é prevalente em 0,3% da população de países industrializados (Dexter et al., 2013). O avanço da idade eleva esta taxa para um por cento em indivíduos com idade acima de 60 anos e quatro por cento com idade acima de 80 anos (Dexter et al., 2013). Nos Estados Unidos a incidência da DP é de 10,8/100.000 pessoas por ano, na Europa é de 4,9/1000 pessoas por ano com idade acima de 65 anos. Em nível mundial, é esperado um aumento de cem por cento no número de pessoas com DP, principalmente em nações populosas como é o caso da China e Índia (Kleinman and Frank, 2013). No Brasil, apenas um estudo investigou a prevalência da DP, sendo de 3,3/100 casos da doença (Barbosa et al., 2006).

As causas da DP são fatores ambientais e genéticos. Sendo , que a interação destes fatores tem sido relacionada a mecanismos de neurodegeneração tais como neuroinflamação, estresse oxidativo, disfunção mitocondrial, alteração da proteólise, e mecanismos de excitotoxicidade (Bartels and Leenders, 2009; Dexter et al., 2013). Os principais fatores de risco são o envelhecimento, pertencer ao sexo masculino, sofrer trauma na cabeça e exposição a toxinas tais como herbicidas e pesticidas (Bartels and Leenders, 2009; Dexter et al., 2013;).

A deficiência de Dopamina altera a função dos núcleos da base e tem sido associada a uma excessiva inibição talâmica com conseqüente supressão do sistema motor cortical o que leva aos distúrbios motores típicos da DP (Bartels and Leenders, 2009; Languette and Lozano, 1998; Obeso et al., 2000).

Embora não existam marcadores biológicos para a DP, os critérios diagnósticos utilizados apresentam boa confiabilidade (Languette and Lozano,

1998; Teixeira e Cardoso, 2004), são eles a presença de dois dos três sinais cardinais: bradicinesia, rigidez e tremor; Além disso, o estabelecimento dos sinais deve ser assimétrico, e não deve existir déficits cognitivos, autonômicos e instabilidade precoce. Associado a isto, uma boa resposta à medicação dopaminérgica é um fator de reconhecimento da DP (Marsden, 1994; Dexter et al., 2013; Bartels and Leenders, 2009; Languette and Lozano, 1998).

A DP é caracterizada pela presença de sinais e sintomas motores e não motores. Dentre os sintomas motores destacam-se a bradicinesia ou redução da velocidade do movimento; a hipocinesia, caracterizada pela diminuição da amplitude de movimento e, a acinesia ou dificuldade para iniciar o movimento, além de outros como rigidez, tremor e instabilidade postural. Tais sintomas têm estabelecimento assimétrico evoluindo, com a progressão da doença, para alterações bilaterais (Morris, 2000; Dexter et al., 2013; Languette and Lozano, 1998; Teixeira e Cardoso, 2004; Bartels and Leenders, 2009).

Os principais sinais e sintomas não motores da DP são a depressão, distúrbios do sono, anormalidades sensoriais, disfunção autonômica, declínio cognitivo, hiposmia, hipotensão ortostática e constipação. A importância de se conhecer os sintomas não motores está no fato de que alguns deles podem preceder, em alguns anos, o estabelecimento dos sintomas motores, auxiliando no diagnóstico precoce da doença (Dexter et al., 2013; Languette and Lozano, 1998; Teixeira e Cardoso, 2004).

O tratamento da DP é focado nos sinais e sintomas da doença uma vez que ainda não existe cura para a mesma (Dexter et al., 2013; Kulisevsky et al., 2013). O objetivo do tratamento não mudou muito nos últimos 30 anos, o qual consiste basicamente na reposição dos níveis de dopamina utilizando drogas a base de Levodopa (Marsden, 1994; Dexter et al., 2013; Kulisevsky et al., 2013). Além do mais, algumas complicações advêm do uso prolongado de tal medicação, tais como flutuações motoras, discinesias, dificuldades na fala, problemas cognitivos e psiquiátricos bem como instabilidade na marcha e aumento do risco de quedas (Marsden, 1994).

1.2 Controle Postural e Instabilidade na Doença de Parkinson

Dados de uma revisão sistemática sobre determinantes da qualidade de vida na DP demonstraram que as desordens da marcha e a instabilidade postural estão entre os principais preditores da qualidade de vida, ficando atrás da depressão e das complicações decorrentes da terapia medicamentosa (Soh et al., 2011). A instabilidade postural tem sido apontada como um dos fatores de risco para a ocorrência de quedas em indivíduos com DP, sendo mais freqüente em pacientes com maior tempo de evolução da doença (Kim et al., 2013; Bloem et al., 2001; Wood et al., 2002; Krygowska-Wais, 2005).

O impacto da instabilidade postural e quedas está no fato de que suas conseqüências podem ser devastadoras para o indivíduo com DP tais como dores, limitações de atividades, medo da queda, depressão e fraturas, sendo que fraturas de quadril compreendem 46% de todas as fraturas nesta população gerando um impacto importante na mobilidade (Williams et al., 2006; Kim et al., 2013; Krygowska-Wais, 2005). Dados de um estudo sobre o impacto das quedas na DP mostrou que os pacientes caidores apresentaram pior avaliação no PDQ39, principalmente nos domínios de mobilidade e atividades de vida diária (Krygowska-Wais, 2005). Ashburn et al (2001) analisaram a freqüência de quedas em 63 indivíduos com DP e demonstraram que 68% destes indivíduos tiveram pelo menos uma queda em 12 meses de acompanhamento.

A Fisioterapia é um importante tratamento coadjuvante da DP, tendo como principais objetivos maximizar a capacidade funcional e minimizar complicações secundárias por meio da reabilitação do movimento, educação e suporte aos pacientes (Kwakkel et al., 2007). Evidências têm mostrado que o indivíduo com DP pode se beneficiar da fisioterapia, como por exemplo, através de treino de marcha e treino de tarefas específicas de controle postural (Morris, 2000; de Goede, 2001; Kwakkel et al., 2007). Com o objetivo de desenvolver intervenções mais eficazes, os pesquisadores têm procurado entender os mecanismos normais de controle postural, bem como as alterações presentes na DP (Horak et al., 1992; Winter, 1995; Bohnen and Cham, 2006; Horak et al., 2006; Vaugoyeau et al., 2011).

O controle postural envolve dois processos sensoriomotores distintos: a orientação postural e o equilíbrio. A orientação postural se refere ao posicionamento dos segmentos corporais em relação a eles mesmos ou ao ambiente. Dependendo da tarefa, os segmentos corporais podem ser alinhados em relação à linha da gravidade, à percepção visual da verticalidade ou à superfície de suporte (Horak, 2013; Horak et al., 1992; Winter, 1995; Bohnen and Cham, 2006; Horak et al., 2006; Vaugoyeau et al., 2011). As informações vestibular, visual e proprioceptiva são integradas para determinar a posição do corpo no espaço. A orientação postural é importante para auxiliar no equilíbrio, uma vez que alterações antecipadas da orientação corporal podem minimizar o efeito de uma possível ameaça ao equilíbrio (Horak, 2013; Horak et al., 1992; Horak et al., 2006;).

O equilíbrio pode ser considerado um termo genérico para descrever a dinâmica da postura corporal em prevenir queda (Winter, 1995; Pollock et al., 2000). Ele envolve a resistência ativa a forças externas que atuam sobre o corpo, sendo que a gravidade é a principal delas. Para manter o equilíbrio é necessário o controle da posição e do movimento do centro de massa corporal bem como a rotação corporal sobre este centro (Horak, 2013; Winter, 1995; Horak et al., 1992; Horak et al., 2005; Mackinnon and Winter, 1993)

A postura de pé requer a manutenção do centro de massa dentro da base de suporte e para isto, dois requisitos devem ser alcançados: suporte antigravitacional, ou seja, manutenção do centro de massa corporal na altura adequada e controle do equilíbrio, a fim de manter a trajetória do centro de massa no plano horizontal (Horak, 2013; Winter, 1995; Mackinnon and Winter, 1993). O suporte antigravitacional ou tônus postural representa a ativação tônica dos músculos que geram força contra o solo para manter os membros estendidos e o centro de massa na altura adequada. Ele requer a contração ativa de músculos posturais tais como do tornozelo, tronco e extensores da coluna.

O controle da trajetória do centro de massa corporal se dá tanto por respostas posturais automáticas às perturbações inesperadas quanto por ajustes posturais antecipatórios a movimentos voluntários, que são específicos às condições biomecânicas do corpo. Ao contrario da postura de pé, durante a

marcha ou corrida o centro de massa corporal está em constante estado de queda se deslocando anteriormente e lateralmente em direção à perna que esta na fase de balanço. Em termos biomecânicos a queda é prevenida pelo movimento da base de suporte anterior e lateral ao centro de massa (Horak, 2013; Winter, 1995;).

A instabilidade postural é determinada pela rapidez de deslocamento do centro de massa corporal bem como sua proximidade dos limites da base de suporte (Horak, 2013; Mackinnon and Winter, 1993). Ela está presente quando há deficiência nos mecanismos do equilíbrio que comprometem a habilidade de manutenção da postura (Kim et al.,2012).

As desordens do equilíbrio na DP, possivelmente refletem uma falha dos núcleos da base em planejar os movimentos seqüenciais e realizar os ajustes posturais antecipatórios adequados. Tais ajustes servem para minimizar o deslocamento do centro de massa durante os movimentos e reflete a habilidade do individuo em estabelecer um padrão postural adequado (Morris and lansek, 1996). Indivíduos com DP não apenas apresentam dificuldade nos ajustes posturais, como também, têm pobre capacidade de adaptação postural em decorrência de alterações súbitas do ambiente (Kim, 2012; Horak et al., 1992; Horak et al., 2006; Morris and lansek, 1996)

Além disso, rigidez axial e pobre controle de tronco em resposta a perturbações do corpo na direção posterior parecem contribuir para o aumento dos déficits do equilíbrio nessa população (Benatru et al., 2008; Diniz et al., 2011). Horak et al. (2005) observaram que indivíduos com DP apresentam um menor limite de estabilidade, quando comparados a jovens e idosos saudáveis. Alguns estudos tem apontado a alteração na função proprioceptiva como um fator que contribui para déficits de equilíbrio na DP. Tal alteração é caracterizada por uma incapacidade do individuo em ativar respostas posturais adequadas dado um estímulo proprioceptivo (Vaugoyeau et al., 2011; Horak et al. (2005). Mensurar o equilíbrio é um componente chave da avaliação fisioterápica em pacientes com DP, pois permite ao profissional identificar e quantificar os déficits de equilíbrio e risco de quedas bem como monitorar a progressão dos mesmos e, desta forma, realizar uma adequada tomada de decisão clinica (Kwakkel et al.,2007).

1.3 Avaliação do Equilíbrio

A avaliação é parte vital do processo de reabilitação. Ela pode ser realizada tanto para estabelecer o estado funcional do paciente e fornecer informações acerca das deficiências que contribuem para a incapacidade funcional, quanto para monitorar respostas a intervenções e redirecionar tratamentos (Bernhardt and Hill, 2005). Além disso, a avaliação envolve uma adequada seleção de instrumentos de medida, uma efetiva condução do processo e correta interpretação dos resultados (Bernhardt and Hill, 2005).

Como parte do processo de avaliação, o instrumento de medida é fundamental, pois possibilita a obtenção de dados objetivos e quantitativos acerca do que se pretende medir, permitindo assim, comparações entre diferentes populações (Gadotti et al., 2006) .

O controle postural pode ser avaliado por meio de análises posturográficas estática e dinâmica. A posturografia estática avalia apenas um componente do equilíbrio e consiste no registro do deslocamento do centro de pressão, definido como a localização do ponto de força de reação do solo e representa a média de todas as pressões sobre a superfície da área em contato com o solo (Winter, 1995). A posturografia dinâmica analisa respostas posturais tanto reativas quanto antecipatórias através de perturbações inesperadas da superfície de suporte (Benatru et al., 2008).

A posturografia tem sido utilizada em pesquisas de controle postural por apresentar medidas objetivas e quantitativas do equilíbrio. Entretanto, sua utilização na prática clínica do fisioterapeuta é limitada devido à complexidade, necessidade de treinamento específico e alto custo do equipamento (Mancini and Horak, 2010). Somado a isto, sabe-se que a instabilidade postural em pacientes com DP se traduz em alterações nas atividades funcionais como transferências, desordens da marcha e dependência nas atividades de vida diária (Benatru et al., 2008). Tais atividades são mais complexas e tem suas informações reduzidas quando avaliadas apenas em ambientes laboratoriais.

1.3.1 Avaliação clínica do equilíbrio

A avaliação clínica do equilíbrio deve levar em conta a observação de que o controle postural varia dependendo do contexto, especificidade da tarefa e intenções do indivíduo. Tal avaliação deve investigar como as estratégias do equilíbrio mudam com mudanças nas condições de suporte e condições sensoriais, mudanças na expectativa e experiência do indivíduo e com distintas demandas da tarefa (Benatru et al., 2008; Horak, 1997; Mancini and Horak, 2010).

A avaliação clínica do equilíbrio busca também identificar possíveis alterações com o intuito de prever risco de quedas, determinar se uma intervenção é necessária ou tem sido efetiva bem como determinar as causas de base dos problemas do equilíbrio para abordá-las efetivamente (Horak, 1997).

Os instrumentos de avaliação clínica surgem como uma possibilidade para a medida do equilíbrio no dia a dia dos clínicos, uma vez que, tais instrumentos tendem a ser mais fáceis de utilizar, rápidos na sua administração e de baixo custo (Mancini and Horak., 2010). Além disso, dada à sua facilidade e praticidade, é esperado que os instrumentos clínicos permitam um monitoramento adequado das desordens do equilíbrio, minimizando o caráter subjetivo das avaliações ao mesmo tempo que permitem coleta de dados mais objetiva sobre a real situação do indivíduo, uma vez que, podem ser realizadas medidas com maior frequência, seja para acompanhar a evolução das desordens e conseqüente aumento no risco de quedas seja para investigar a eficácia de intervenções (Mancini and Horak., 2010; Berg, 1989).

Visto que os resultados obtidos por meio destes instrumentos direcionam a tomada de decisão clínica, é necessário que tais instrumentos sejam válidos, confiáveis e responsivos, ou do contrário, existe o risco de se obter resultados imprecisos ou inadequados que podem levar a conclusões e condutas equivocadas.

1.4 Propriedades de Medida

A qualidade de um instrumento de medida está diretamente relacionada às suas propriedades de medida tais como confiabilidade, validade e responsividade (Mokkink et al., 2010). Confiabilidade é definida como a magnitude em que um instrumento é livre de erro. A confiabilidade de um instrumento pode ser medida nos seguintes aspectos: erro de medida, consistência interna e confiabilidades teste-reteste, intra e entre - examinadores (Mokkink et al., 2010). O erro de medida se refere aos erros aleatórios e sistemáticos do escore de um indivíduo que não condizem com as verdadeiras mudanças no constructo que está sendo medido (Mokkink et al., 2010; Terwee et al., 2011). A consistência interna considera o grau de relação entre os itens de um instrumento. A confiabilidade teste-reteste se refere à magnitude em que um escore de um paciente não muda ao longo do tempo sob administração de medidas repetidas, e a confiabilidade intra-examinador diz da magnitude em que um escore de um indivíduo não muda quando medido pelo mesmo examinador em diferentes momentos (Mokkink et al., 2010; Terwee et al., 2011). Finalmente, a confiabilidade entre - examinadores se refere a magnitude em que o escore de um indivíduo não muda quando medido por diferentes examinadores sob as mesmas condições (Mokkink et al., 2010; Scholtes et al., 2011; Terwee et al., 2011) .

A validade é definida como a magnitude em que um instrumento mede o constructo que ele se propõe a medir. Tal propriedade pode ser medida nos seguintes aspectos: validade de conteúdo, de face, de constructo, estrutural, trans-cultural e validade de critério (Mokkink et al., 2010). Validade de conteúdo está presente quando o constructo a ser medido está adequadamente representado no instrumento (Mokkink et al., 2010). Validade de face diz do quanto os itens de um instrumento conseguem representar o construto a ser medido. A validade de constructo está presente quando os escores de um instrumento são consistentes com hipóteses, relacionadas a escores de outros instrumentos, por exemplo, baseado na premissa de que o instrumento validamente mede o constructo em questão. A validade estrutural diz da magnitude em que os escores de um instrumento conseguem refletir a

dimensionalidade do construto a ser medido. A validade transcultural se refere ao grau em que os itens de um instrumento traduzido ou adaptado culturalmente representam os itens da versão original do instrumento. E por fim, a validade de critério se refere à magnitude em que os escores de um instrumento são uma adequada reflexão de um padrão ouro (Mokkink et al., 2010; Terwee et al., 2011).

A responsividade é definida como a habilidade de um instrumento detectar mudanças ao longo do tempo no construto a ser medido. Vale ressaltar que a interpretabilidade, embora não seja uma propriedade de medida, é uma característica importante de um instrumento de medida. O grau de interpretabilidade de um instrumento de medida pode ser definido como a extensão em que se pode atribuir significado qualitativo a escores quantitativos, ou a mudanças nos mesmos, (Mokkink et al., 2010; Mokkink et al., 2010; Scholtes et al., 2011; Terwee et al., 2011).

1.5 Avaliação da Qualidade Metodológica

Para desenvolver e avaliar as propriedades de medida de um instrumento é preciso realizar estudos observacionais de caráter exploratório, comumente denominados metodológicos. É importante ressaltar que estes estudos são específicos para a população que está sendo avaliada (Portney and Watkins, 2009). Para a realização de tais estudos, determinadas exigências precisam ser atendidas para que o mesmo possa ser considerado de boa qualidade. Assim, tamanho de amostra, análise estatística, intervalo entre as medidas e comparação com um padrão ouro são aspectos relevantes e que devem estar adequadamente desenhados (Portney and Watkins, 2009).

Uma forma de se avaliar, sistematicamente, a qualidade dos estudos é através da realização de revisões sistemáticas. Revisão sistemática envolve um rigoroso processo de busca, avaliação e síntese da informação existente sobre um determinado tema. Tais revisões são mais comumente focadas na efetividade de intervenções, mas também podem ser direcionadas para a acurácia de instrumentos diagnósticos ou identificação de fatores prognósticos (Terwee et al. 2011; Portney and Watkins. 2009). Sua importância está em

fornecer aos profissionais as evidências necessárias para guiá-los na tomada de decisão clínica. O uso de ferramentas específicas para avaliação da qualidade metodológica é um componente essencial em uma revisão sistemática, pois tais ferramentas direcionam o processo de validação dos resultados encontrados nos estudos (Portney and Watkins, 2009). No caso dos estudos sobre propriedades de medidas, se a qualidade metodológica é adequada, então os resultados podem ser utilizados para avaliar se um instrumento de medida é de boa qualidade ou não (Terwee et al. 2011; Portney and Watkins. 2009).

Algumas iniciativas têm buscado melhorar a informação sobre a qualidade dos estudos que avaliam as propriedades de medida dos instrumentos utilizados em saúde. Para isso, foi desenvolvido um *checklist* contendo diretrizes para a avaliação da qualidade metodológica dos estudos. Publicado em 2010, o *checklist* denominado *Consensus-based Standards for the Selection of Health Measurement Instruments* (COSMIN) foi resultado do estudo com metodologia Delphi desenvolvido por um grupo de pesquisadores com a colaboração de especialistas de áreas relevantes tais como epidemiologia, estatística, psicologia e medicina clínica (Mokkink et al., 2010). O *checklist* COSMIN tem sido uma ferramenta utilizada em revisões sistemáticas com o intuito de avaliar a qualidade dos estudos das propriedades de medida dos instrumentos de avaliação em saúde.

1.6 Justificativa

Até o momento, vários estudos avaliaram diferentes propriedades de medida de instrumentos de avaliação do equilíbrio e risco de quedas na DP. Porém, não existe um consenso na literatura sobre qual ou quais destes são adequados para esta população.

Dado o impacto da instabilidade postural na qualidade de vida de indivíduos com DP, faz-se necessário um adequado manejo com intervenções mais eficazes para estes indivíduos. Isto inclui avaliações acuradas e precisas que irão direcionar este processo.

A importância das propriedades de medida está no fato de que elas caracterizam um instrumento apontando suas potencialidades e debilidades em medir a variável em questão, neste caso, o equilíbrio e o risco de quedas. Além disso, instrumentos de medida são comumente utilizados para medir desfechos em estudos experimentais envolvendo o equilíbrio na DP, o que pode implicar em resultados enviesados caso haja uma utilização inadequada de tais ferramentas.

Dessa forma, faz-se necessário identificar e caracterizar os instrumentos de medida do equilíbrio e risco de quedas disponíveis na literatura envolvendo a DP, a fim de conhecer as propriedades de medidas investigadas, a qualidade das análises realizadas, assim como, direcionar novos estudos que favoreçam a obtenção de mais conhecimento sobre as reais características e possíveis disfunções dos indivíduos com DP.

2. OBJETIVOS

2.1 Geral

Realizar uma revisão sistemática sobre as propriedades de medida dos instrumentos de avaliação clínica do equilíbrio e risco de quedas na DP.

2.2 Específicos

- Identificar os instrumentos clínicos de avaliação do equilíbrio e risco de quedas utilizados na DP;
- Identificar as propriedades de medida avaliadas nos instrumentos encontrados;
- Investigar a qualidade metodológica dos estudos que avaliaram as propriedades de medidas;
- Fornecer a pesquisadores e clínicos evidências sobre a qualidade e utilidade clínica dos instrumentos de avaliação do equilíbrio e risco de quedas na DP.

3. MATERIAIS E MÉTODO

3.1 Desenho do estudo

O presente estudo é de uma revisão sistemática que foi devidamente registrada no International Prospective Register of Systematic Reviews - PROSPERO: nº CRD42013004754.

3.2 Estratégia de Busca

Inicialmente, foi realizada uma busca na literatura para identificar os estudos que investigaram as propriedades de medida dos instrumentos utilizados na avaliação do equilíbrio e risco de quedas em indivíduos com DP. Tal busca se deu até Março de 2013 nas seguintes bases de dados: MEDLINE, LILACS, CINAHL e SPORTdiscus. A estratégia de busca foi otimizada para cada base de dados com a combinação de termos como *Clinical Assessment*, *Clinical Test*, *Clinical Evaluation*, *Balance*, *Postural Balance*, *Postural Instability*, *Falls*, *Parkinson Disease*, *Outcome Assessment*, *Psychometrics*, *Reliability*, *Validity*, *Reproductibility* (apêndice 1). Além disso, uma busca manual foi realizada nas referências dos artigos e em bibliotecas virtuais de teses e dissertações.

Uma vez realizada a estratégia de busca, dois revisores independentes e treinados previamente (L.L e N.V) avaliaram os títulos e resumos dos estudos e aplicaram os critérios de elegibilidade para a seleção dos mesmos. Outros dois revisores (F.R e A.S.) foram requisitados no caso de discordância ou existência de dúvida. Quando necessário, os autores dos estudos foram contatados por e-mail para fornecerem o texto completo ou outras informações necessárias.

3.3 Critérios de Elegibilidade

Para um estudo ser incluído seria preciso atender aos seguintes critérios: ter avaliado indivíduos com DP idiopática, independente do sexo ou nível de incapacidade; ter apontado, claramente, que a medida clínica avaliada era o equilíbrio e/ou risco de quedas em indivíduos com DP; ter investigado apenas instrumentos (teste, escala ou questionário) de avaliação clínica do equilíbrio e/ou risco de quedas na população de indivíduos com DP; serem estudos metodológicos de desenvolvimento de instrumentos e/ou avaliação de no mínimo uma das seguintes propriedades: validade, confiabilidade, responsividade; serem estudos publicados nos seguintes idiomas: Português, Inglês, Francês e Espanhol.

Foram excluídos os estudos que investigaram as propriedades de medida de instrumentos laboratoriais de avaliação do equilíbrio tais como posturografias e ensaios clínicos onde a análise da propriedade de medida foi realizada de maneira complementar ao estudo em questão. Se algum instrumento, conhecido por avaliar o equilíbrio, foi utilizado pelos autores de um dado estudo para avaliar outro tipo de medida clínica como mobilidade ou atividade de vida diária, tal estudo foi excluído

3.4 Seleção dos estudos

Os mesmos revisores independentes (L.L e N.V.) aplicaram os critérios de elegibilidade para a seleção dos estudos. Quando necessário, outros dois revisores (F.R. e A.S.) foram requisitados no caso de discordância ou dúvida. Após a implementação da estratégia de busca, os artigos foram avaliados por meio da leitura dos títulos e resumos. Nessa fase, todas as duplicatas foram retiradas. Após análise dos títulos e resumos, os artigos potencialmente relevantes foram adquiridos na íntegra e lidos até a seção métodos.

Os revisores foram cegados quanto aos autores dos artigos e jornais de publicação utilizando uma tarja preta nos artigos impressos e um protocolo para inclusão foi preenchido para cada artigo com base nos critérios de elegibilidade estabelecidos acima (apêndice 2) . Os estudos que preencheram

os critérios de elegibilidade foram separados para leitura na íntegra e extração dos dados.

3.5 Extração dos dados

Após a leitura dos artigos, os dados relevantes foram extraídos por dois revisores independentes (L.L e N.V). São eles: objetivos do estudo, características da população (idade, gravidade e tempo de evolução da doença), medida clínica avaliada, instrumentos de medida avaliados, examinadores (número, treinamento prévio e experiência), operacionalização das medidas (local, intervalos entre as medidas, gravação de vídeo para análise) e tipo de análise estatística realizada (modelo teórico, pacote estatístico e testes realizados). Tais informações encontram-se no apêndice 3 e foram utilizadas para análise da qualidade metodológica dos estudos, que será descrita a seguir.

Os resultados encontrados nos estudos também foram extraídos e sumarizados com o objetivo de classificação das estimativas das propriedades de medida para cada instrumento.

3.6 Avaliação da qualidade metodológica

A avaliação da qualidade metodológica de cada estudo também foi realizada por dois revisores independentes (L.L e N.V.) através do *Checklist* COSMIN (anexo 1). As divergências foram resolvidas por consenso ou quando necessário com auxílio de outros dois revisores (F.R. e A.S.).

3.6.1 Apresentação do *Checklist*

O *Checklist* COSMIN foi desenvolvido a partir do estudo internacional com metodologia Delphi cujos objetivos foram alcançar um consenso sobre quais propriedades de medida deveriam ou devem ser avaliadas, desenvolver normas de como elas devem ser avaliadas em termos de desenho de estudo e análise estatística e definir instrumentos onde a medida de algum aspecto da

saúde é avaliada pelo próprio paciente (*Health-Related Patient-Reported Outcomes- HR-PROs*).

O foco do desenvolvimento do COSMIN foi a avaliação da qualidade metodológica de estudos de instrumentos HR-PROs. Porém, o *Checklist* pode ser utilizado em outras áreas da saúde para avaliar outros tipos de instrumentos de medida uma vez que as propriedades de medidas avaliadas são comuns e relevantes para instrumentos de saúde de uma forma geral (Mokkink et al., 2010).

3.6.2 Taxonomia e terminologias do *Checklist*

Baseado no consenso alcançado no estudo Delphi, o grupo de pesquisadores que fazem parte do comitê do COSMIN desenvolveu uma taxonomia na qual as propriedades de medida foram distinguidas em três domínios de qualidade: confiabilidade, validade e responsividade. Cada domínio contém uma ou mais propriedades ou aspectos das propriedades de medida. Os conceitos apresentados e utilizados nesse estudo seguem a definição apresentada pelo COSMIN.

3.6.3 Caracterização

O *checklist* é constituído de 12 seções, contendo de 4 a 18 itens cada. Dez seções são utilizadas para avaliar se o estudo atende às normas de boa qualidade metodológica. Dessas seções, nove contém normas para cada uma das propriedades de medida incluídas: consistência interna (seção A), confiabilidade (seção B), medida erro (seção C), validade de conteúdo (incluindo validade de face) (seção D), validade de constructo e estrutural (seção E), testagem de hipóteses (seção F), validade transcultural (seção G), validade de critério (seção H) e responsividade (seção I)). Uma seção contém normas para estudos sobre interpretabilidade (seção J). Além disso, duas seções incluídas no checklist contêm informações gerais do estudo: a seção generabilidade, que contém informações sobre a capacidade de generalização

dos resultados do estudo e a seção que avalia especificamente estudos onde foram utilizados modelos estatísticos baseados na Teoria de Resposta ao Item.

3.6.4 Operacionalização

Para completar o *checklist* quatro passos foram seguidos: identificação de qual a propriedade de medida foi avaliada no estudo; identificação da utilização de teoria de resposta ao item para completar a seção correspondente caso existisse; preenchimento de cada uma das seções correspondentes às propriedades de medida identificadas em cada estudo analisado; utilização da escala de quatro pontos do *Checklist* COSMIN para cada propriedade de medida avaliada. Tal escala classifica cada propriedade de medida como excelente, boa, fraca ou pobre.

3.6.5 Adequações ao COSMIN

Os artigos foram analisados com base na Taxonomia do *Checklist* COSMIN, sendo assim, algumas propriedades de medida encontradas nos estudos foram adequadas à nomenclatura do COSMIN por consenso dos revisores. Por exemplo: alguns autores analisaram validade preditiva, validade convergente e validade concorrente. Como tais termos não existem nas definições do COSMIN, eles foram analisados através da seção H (validade de critério) por se aproximarem mais de sua definição.

3.7 Síntese dos dados

3.7.1 Categorização dos instrumentos

As estimativas das propriedades de medida encontradas foram sintetizadas para cada instrumento os quais foram categorizados em instrumentos de desempenho, testes de componente único e questionários de auto-relato.

3.7.2 Sumarização das propriedades de medida de acordo com a taxonomia do *checklist* COSMIN

Os artigos foram analisados com base na taxonomia do *checklist* COSMIN e a qualidade metodológica de cada estudo foi reportada para cada propriedade de medida analisada.

3.7.3 Avaliação do nível de evidência das propriedades de medida

Para cada estudo, as estimativas das propriedades de medida investigadas foram classificadas como adequada (+), não adequada (-) ou conflitante (\pm) baseado em critérios pré-definidos (Terwee et al., 2007). Visto não haver um critério específico para avaliar as estimativas encontradas para a validade de critério, os autores desta revisão decidiram classificar como adequada a utilização da área sob a curva (*area under the curve* - AUC) com valor igual ou maior que 0,70 e classificar como inadequada a não utilização de tal curva ou sua obtenção com valor menor que 0,70.

A qualidade geral de cada propriedade de medida para cada instrumento foi classificada quanto aos níveis de evidência forte, moderada, limitada, conflitante e desconhecida de acordo com os critérios de Elbers et al (2012). Para isto, uma síntese dos dados qualitativos foi realizada considerando-se a classificação das estimativas para cada propriedade de medida, a consistência dos resultados entre os estudos, o número e a qualidade metodológica dos estudos. Os critérios para a classificação das estimativas encontradas bem como para os níveis de evidência para cada propriedade de medida estão descritos no anexo 2.

3.7.4 Avaliação da utilidade clínica dos instrumentos

Foi realizada uma avaliação da utilidade clínica dos instrumentos encontrados de acordo com os critérios de Tyson and Connell (2009): tempo de administração, análise e interpretação do instrumento; custo; necessidade de equipamento especializado e treinamento específico para utilização; portabilidade da medida (anexo 3). A pontuação máxima para esta avaliação é 10, sendo que, os instrumentos recomendados para utilização na prática clínica

devem apresentar pontuações iguais ou maiores que nove (Tyson and Connell, 2009).

3.8 Apresentação dos resultados

As estimativas das propriedades de medida encontradas foram sintetizadas para cada instrumento e apresentadas em uma tabela descritiva (tabela 1). Com o objetivo de discutir a qualidade metodológica dos estudos analisados, os dados gerais sobre a qualidade metodológica de cada propriedade de medida para cada estudo foram apresentados em formato de tabela (tabela 2); Os níveis de evidência para as propriedades de medida de cada instrumento foram apresentados na tabela 3. A pontuação que cada instrumento de medida obteve na avaliação da utilidade clínica foi apresentada na tabela 4.

4. ARTIGO

Measurement properties of instruments of balance and risk of falls in Parkinson's Disease: a systematic review

Propriedades de medida de instrumentos de avaliação do equilíbrio e risco de quedas na Doença de Parkinson: uma revisão sistemática

ABSTRACT | Objective: To systematically review the measurement properties of clinical assessment instruments of balance and risk of falls for Parkinson's disease (PD). **Method:** The databases MEDLINE, LILACS, CINAHL and SPORTdiscus were searched until March 2013. Methodological studies which developed the clinical assessment instruments of balance and risk of falls in PD as well as those which rated their measurement properties were included. The assessment of methodological quality was performed by two independent raters using the Checklist COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments). The estimates were summarized for each instrument and the measurement properties were rated according to the levels of evidence strong, moderate, limited, conflicting and unknown. **Results:** The initial search strategy identified 150 studies, of which 20 were included in this review. Fourteen instruments were found, and moderate evidence was presented for an adequate internal consistency for the BBS, adequate reliability for SAFFE –NL and FES-S and adequate criterion validity for the BESTEST and FGA. **Conclusion:** The methodological quality of the studies and the level of evidence analysis highlighted five potentially most appropriate instruments at present to measure the balance or the risk of falls in PD. They are: BBS, BESTEST, FGA in the category of performance instruments and SAFFE-NL and FES-S in the category of self-report questionnaires.

Keywords: Clinical assessment, balance, falls, measurement properties, COSMIN.

RESUMO | Objetivo: Revisar sistematicamente os estudos de propriedades de medida dos instrumentos de avaliação clínica do equilíbrio e risco de quedas na Doença de Parkinson (DP). **Método:** As bases de dados MEDLINE, LILACS, CINAHL e SPORTdiscus foram pesquisadas até março de 2013. Foram incluídos estudos metodológicos que desenvolveram os instrumentos de avaliação clínica do equilíbrio e risco de quedas na DP bem como aqueles que avaliaram suas propriedades de medida. A avaliação da qualidade metodológica foi realizada por dois examinadores independentes por meio do Checklist COSMIN (*Consensus-based Standards for the Selection of Health Measurement Instruments*). As estimativas foram sintetizadas para cada instrumento e as propriedades de medida foram classificadas quanto aos níveis de evidência forte, moderada, limitada, conflitante e desconhecida. **Resultados:** A estratégia de busca inicial identificou 150 estudos, sendo 20 artigos incluídos na presente revisão. Quatorze instrumentos foram encontrados, sendo que moderada evidência foi apresentada para uma adequada consistência Interna do BBS, adequada confiabilidade teste –reteste do SAFFE- NL e FES-S e adequada validade de critério do BESTEST e FGA. **Conclusão:** A qualidade metodológica dos estudos e os níveis de evidência encontrados destacaram cinco instrumentos potencialmente mais adequados, na atualidade, para avaliar o equilíbrio ou o risco de quedas na DP. São eles: BBS, BESTEST, FGA na categoria de instrumentos de desempenho e, SAFFE-NL e FES-S na categoria de questionários de auto-relato.

Palavras-chave: Avaliação clínica, equilíbrio, quedas, propriedades de medida, *COSMIN*.

Introdução

Classificada como a segunda desordem neurodegenerativa, após a doença de Alzheimer, a Doença de Parkinson (DP) é prevalente em 0,3% da população de países industrializados¹. O avanço da idade eleva esta taxa para um por cento em indivíduos com idade acima de 60 anos e quatro por cento com idade acima de 80 anos¹. Em nível mundial, é esperado um aumento de 100% no número de casos, principalmente em nações populosas como é o caso da China e Índia².

Desordens da marcha e instabilidade postural estão entre os principais preditores de qualidade de vida na DP, ficando atrás somente da depressão e das complicações decorrentes de terapias medicamentosas³. A instabilidade postural tem sido apontada como um dos fatores de risco para a ocorrência de quedas em indivíduos com DP, sendo mais freqüente em casos com maior tempo de evolução da doença⁴. Dados de uma meta-análise demonstraram que 46% dos indivíduos com DP apresentaram quedas dentro de um período de três meses⁵.

A medida do equilíbrio na DP é um componente chave da avaliação fisioterápica, pois permite ao profissional identificar e quantificar os déficits de equilíbrio e risco de quedas bem como monitorar a progressão dos mesmos e desta forma, realizar uma adequada tomada de decisão clínica direcionada á intervenção⁶⁻⁸. Desse modo, é necessário que os instrumentos de avaliação clínica sejam válidos, confiáveis e responsivos, ou do contrário, existe o risco de se obter resultados imprecisos ou inadequados que podem levar a conclusões e condutas equivocadas^{9,10}. Além disso, é importante na prática fisioterápica que tais instrumentos sejam de fácil utilização, rápidos na sua administração e de baixo custo⁹.

É necessário que os estudos das propriedades de medida de um instrumento atendam a determinadas exigências para serem considerados de boa qualidade. Assim, tamanho de amostra, análise estatística, intervalo entre as medidas e comparação com um padrão ouro são aspectos relevantes e que devem estar adequadamente desenhados¹¹⁻¹³.

O *checklist* denominado *Consensus-based Standards for the Selection of Health Measurement Instruments* (COSMIN), tem sido uma ferramenta utilizada em revisões sistemáticas com o intuito de avaliar a qualidade de estudos que investigam as propriedades de medida de instrumentos de avaliação em saúde¹².

Até o presente momento, vários estudos avaliaram diferentes propriedades de medida de instrumentos de avaliação do equilíbrio e risco de quedas na DP. Porém, não existe um consenso na literatura sobre qual ou quais destes instrumentos são adequados para esta população.

Sendo assim, o objetivo do presente estudo foi fazer uma revisão sistemática para identificar quais instrumentos clínicos de avaliação de equilíbrio e risco de quedas são utilizados na DP, caracterizar as propriedades de medida de tais instrumentos e determinar a qualidade metodológica dos estudos envolvidos.

Método

Estratégia de busca

Esta revisão foi registrada no PROSPERO nº CRD42013004754. Realizou-se uma busca até março de 2013 nas bases de dados MEDLINE, LILACS, CINAHL e SPORTdiscus. A estratégia de busca foi otimizada para cada base de dados com a combinação de termos como *Clinical Assessment, Clinical Test, Clinical Evaluation, Balance, Postural Balance, Postural Instability, Falls, Parkinson Disease, Outcome Assessment, Psychometrics, Reliability, Validity, Reproducibility*. Além disso, uma busca manual foi realizada nas referências dos artigos e bibliotecas virtuais de teses e dissertações.

Seleção dos estudos

Dois revisores (L.L e N.V) aplicaram independentemente os critérios de elegibilidade para a seleção dos estudos. Outros dois revisores (F.R. e A.S.) foram requisitados no caso de discordância ou dúvida. Os critérios de inclusão dos estudos foram: indivíduos com DP idiopática, independente do sexo ou nível de incapacidade; instrumentos clínicos de avaliação do equilíbrio e risco de quedas (testes, escalas ou questionários); estudos metodológicos que desenvolveram os instrumentos e/ou avaliaram suas propriedades de medida reportando a avaliação de, no mínimo, uma das seguintes propriedades: validade, confiabilidade, responsividade; e estudos nos idiomas português, inglês, francês e espanhol.

Foram excluídos estudos que investigaram as propriedades de medida de instrumentos laboratoriais de avaliação do equilíbrio como posturografias e ensaios clínicos com análise de propriedades de medida. Após a análise dos títulos e

resumos, os artigos potencialmente relevantes foram adquiridos na íntegra e lidos até a seção métodos. Os revisores foram mascarados em relação aos autores dos artigos e jornais de publicação. Um protocolo de inclusão foi preenchido para cada estudo com base nos critérios de elegibilidade estabelecidos acima.

Extração dos dados

Os dados relevantes foram extraídos dos estudos incluídos por dois revisores independentemente. Os dados extraídos foram: objetivos do estudo, características da população, medida clínica avaliada, instrumentos de medida avaliados, examinadores, operacionalização das medidas e tipo de análise estatística. Os resultados encontrados também foram extraídos com objetivo de sumarização das estimativas das propriedades de medida.

Avaliação da qualidade metodológica

A avaliação da qualidade metodológica de cada estudo também foi realizada por dois examinadores independentes (L.L e N.V.) por meio do *Checklist* COSMIN. As divergências foram resolvidas por consenso ou, quando necessário, com auxílio de outros dois revisores. O checklist COSMIN é constituído de 12 seções, contendo de 4 a 18 itens cada. Dez seções são utilizadas para avaliar se o estudo atende às normas de boa qualidade metodológica, as outras duas se referem a generalização dos resultados e Teoria de Resposta ao Item¹¹. Os itens de cada seção foram classificados em: excelente, bom, fraco e pobre¹⁴. A qualidade de cada propriedade de medida foi classificada levando em consideração a pior pontuação por item.

Síntese dos dados

Categorização dos instrumentos

As estimativas das propriedades de medida encontradas foram sintetizadas para cada instrumento os quais foram categorizados em instrumentos de desempenho, testes de componente único e questionários de auto-relato.

Sumarização das propriedades de medida de acordo com a taxonomia do *checklist* COSMIN

Os estudos foram analisados com base na taxonomia do *Checklist* COSMIN, sendo assim, algumas propriedades de medida encontradas nos estudos foram adequadas à nomenclatura do COSMIN por consenso dos avaliadores. A qualidade metodológica dos estudos foi reportada para cada propriedade de medida analisada.

Avaliação do nível de evidência das propriedades de medida

Para cada estudo, as estimativas das propriedades de medida investigadas foram classificadas como adequada (+), não adequada (-) ou conflitante (\pm) baseado nos critério de Terwee et al. (2007)¹⁵. Visto não haver um critério específico para avaliar as estimativas encontradas para a validade de critério, os autores desta revisão decidiram classificar como adequada a utilização da área sob a curva (*area under the curve* - AUC) com valor igual ou maior que 0,70 e classificar como inadequada a não utilização de tal curva ou sua obtenção de valor menor que 0,70.

A qualidade geral de cada propriedade de medida para cada instrumento foi classificada quanto aos níveis de evidência forte, moderada, limitada, conflitante e desconhecida de acordo com os critérios de Elbers et al. (2012)¹⁶. Para isto, uma síntese dos dados qualitativos foi realizada considerando-se a classificação das estimativas, a consistência dos resultados entre os estudos, o número e a qualidade metodológica dos estudos.

Avaliação da utilidade clínica dos instrumentos

Foi realizada uma avaliação da utilidade clínica dos instrumentos encontrados de acordo com os critérios de Tyson and Connell (2009)¹⁷: tempo de administração, análise e interpretação do instrumento; custo; necessidade de equipamento especializado e treinamento para utilização; portatibilidade da medida. A pontuação máxima para esta avaliação é 10, sendo que, os instrumentos recomendados para utilização na prática clínica devem apresentar escores iguais ou maiores que nove¹⁷.

Resultados

Seleção dos estudos

A estratégia de busca identificou inicialmente 150 estudos, sendo 20 incluídos nessa revisão sistemática, como pode ser visto na figura 1.

Identificação dos instrumentos

Entre os estudos analisados, encontrou-se os seguintes instrumentos: de desempenho: Berg Balance Scale (BBS), Timed up and Go (TUG), Balance Evaluation Systems Test (BESTEST), MINI-BESTEST, Dynamic Gait Index (DGI), Functional Gait Assessment (FGA), Tinetti Mobility Test (TMT) ou Performance Oriented Mobility Assessment (POMA); testes de componente único: Functional Reach Test (FRT), PULL TEST, ROMBERG TEST; e questionários de auto-relato:

Survey of Activities and Fear of Falling in the Elderly (SAFFE – NL) , Swedish Falls Efficacy Scale (FES-Swedish), Activity Balance Confidence Scale (ABC) e Rapid Assessment of Postural Instability in Parkinson's disease Questionnaire (RAPID QUESTIONNAIRE).

As características dos estudos e as estimativas das propriedades de medida dos instrumentos encontrados estão descritas na Tabela 1. Os instrumentos que apresentaram o maior número de propriedades de medida investigadas (quatro) foram: BBS, BESTEST, Mini-BESTEST, FRT, FES-Swedish e SAFFE– NL. TUG, DGI, e a ABC tiveram três propriedades investigadas e, os demais instrumentos, tiveram menos de três propriedades analisadas.

Qualidade metodológica dos estudos

Os resultados da análise da qualidade metodológica por meio do *checklist* COSMIN para cada estudo estão descritos na tabela 2. As principais propriedades de medida avaliadas foram consistência interna, confiabilidade, erro de media e validade de critério. Apenas três estudos apresentaram propriedades de medida classificadas como boa: a consistência interna para a BBS¹⁹, a confiabilidade para a FES-S e o SAFF-NL³⁶ e o erro de media para o TUG e DGI²⁷. Estes estudos apresentaram bom tamanho de amostra (>50), não apresentaram falhas graves no delineamento e apresentaram adequada análise estatística.

Os demais estudos que avaliaram as propriedades citadas acima foram considerados de fraca a pobre qualidade metodológica, sendo que, tamanho de amostra e falha no delineamento constituíram os principais problemas. Para a confiabilidade, além das falhas citadas acima, número de medidas realizadas e erros de análise estatística foram problemas adicionais. A unidimensionalidade constituiu um problema adicional na análise da consistência interna. Na análise da validade de critério, as principais falhas foram a não utilização de um padrão ouro como comparador e erros de análise estatística.

Caracterização das propriedades de medida dos instrumentos quanto ao nível de evidência

A tabela 3 apresenta os níveis de evidência das propriedades de medida para os instrumentos avaliados.

Evidência forte

Nenhum estudo apresentou forte evidência para as propriedades de medida avaliadas.

Evidência moderada

A BBS apresentou moderada evidência para adequada consistência interna. Moderada evidência também foi encontrada para a adequada confiabilidade teste – reteste do SAFFE- NL e FES-S. Quanto á avaliação do erro de medida este nível de evidência foi demonstrado para TUG e DGI. O BESTEST e o FGA apresentaram moderada evidência para adequada validade de critério.

Evidência limitada

O PULL Test apresentou limitada evidência para adequada confiabilidade entre - examinadores. O MINI-BESTEST apresentou este nível de evidência para uma adequada validade de critério. O TUG, o TMT (POMA) e o PULL TEST apresentaram este nível de evidência para inadequada validade de critério. Limitada evidência foi encontrada para a validade transcultural do MINI-BESTEST.

Evidência conflitante

Evidência conflitante foi demonstrada para a consistência interna do ABC e para a confiabilidade teste-reteste do TUG.

Evidência desconhecida

Evidência desconhecida foi demonstrada para a adequada consistência interna avaliada nos instrumentos FES-S e SAFFE-NL. A BBS, o BESTEST, o MINI-BESTEST, o TUG, o FGA, o DGI, o FRT e o ROMBERG TEST apresentaram evidência desconhecida para adequada confiabilidade teste-reteste. Para a adequada confiabilidade entre - examinadores este nível de evidência foi encontrado nos instrumentos BBS, BESTEST, MINI-BESTEST, FGA, TMT (POMA). Em relação à avaliação do erro de medida, desconhecida evidência foi demonstrada para a BBS, o FRT, o ROMBERG TEST e o FES-S. O RAPID apresentou evidência desconhecida para a validade de critério. Este nível de evidência foi encontrado para a validade transcultural do BESTEST.

Utilidade Clínica dos instrumentos de medida

A maioria dos instrumentos obteve pontuação máxima (10/10) ou submáxima (9/10) na avaliação da utilidade clínica, exceto o BESTEST (5/10) e o Mini-BESTEST (6/10) como pode ser visto na tabela 4.

Discussão

No conhecimento dos autores, essa revisão é a primeira que avalia sistematicamente e sumariza a evidência das propriedades de medida dos instrumentos de avaliação clínica do equilíbrio e risco de quedas na DP.

Quatorze instrumentos tiveram suas propriedades de medida investigadas na população de indivíduos com DP, sendo sete de desempenho^{18-25; 27-230}, três testes de componente único^{20-22;29;31-34} e quatro questionários de auto-relato^{20;28;35-37}. Estes dados sugerem que há uma preferência por parte de pesquisadores em considerar o desempenho de pacientes com DP em atividades funcionais na hora de avaliar déficits do equilíbrio e risco de quedas.

Qualidade metodológica dos estudos e nível de evidência das propriedades de medida investigadas

As propriedades de medida mais frequentemente investigadas foram a consistência interna, a confiabilidade, o erro de medida e a validade de critério (ver tabela 1). A consistência interna foi investigada em cinco estudos^{19-20; 28; 35-36}, sendo que apenas um envolvendo a BBS foi considerado de boa qualidade metodológica¹⁹. Esse instrumento também obteve nível de evidência moderado para uma consistência interna adequada (ver tabelas 1, 2 e 3).

Em geral, a unidimensionalidade foi um item que empobreceu a qualidade metodológica dos estudos, uma vez que a maioria deles não avaliou essa propriedade ou não reportou referência de estudos prévios. A unidimensionalidade é um pré-requisito para analisar a consistência interna em instrumentos baseados em modelos refletivos, ou seja, aqueles em que todos os itens são uma manifestação do mesmo construto, pois ela mostrará a qualidade da definição do constructo e da relação entre os itens da escala^{11, 14}. Portanto, esta propriedade de medida só faz sentido para os instrumentos onde exista grupo de itens cujo objetivo é medir o constructo equilíbrio. Neste caso estão excluídos os testes de componente único e, entre os instrumentos de desempenho, o TUG.

Treze estudos analisaram a confiabilidade e apenas um envolvendo a FES-S e o SAFFE-NL foi considerado de boa qualidade³⁶. Moderada evidência para uma adequada confiabilidade teste-reteste também foi encontrada para esses instrumentos (ver tabelas 1, 2 e 3). Os demais estudos foram considerados de pobre

qualidade metodológica, e 80% deles, classificados como pobre no tratamento estatístico, sendo a utilização inadequada dos índices de correlação (coeficiente de Pearson ou Spearman) a principal limitação encontrada. Muitos instrumentos de medida usam dados ordinais que são adequadamente avaliados através do coeficiente de Kappa por considerar as chances de concordância entre as medidas³⁸. A confiabilidade é fundamental em qualquer instrumento de medida clínica, pois garante que a variabilidade da medida seja devido à verdadeira variabilidade do indivíduo e não ao erro¹¹⁻¹⁴. Sendo assim, a confiabilidade se torna importante para propósitos discriminativos, o que é fundamental para avaliar e distinguir quem tem ou não um risco de quedas.

Sobre o erro de medida, quatro estudos avaliaram esta propriedade, sendo que apenas um estudo teve classificação boa analisando o TUG e DGI²⁷. Embora moderada evidência tenham sido apresentada para o erro de medida destes instrumentos, somente estes dados sem uma forte evidência de adequada confiabilidade podem não ser relevantes do ponto de vista clínico. Em relação às outras propriedades de medida, o TUG apresentou evidência conflitante para confiabilidade e o DGI evidência desconhecida, o que aponta para a necessidade de novos estudos destas propriedades, bem como da validade de critério, uma vez que o TUG apresentou limitada evidência com $AUC < 0,70$ e para o DGI esta análise não foi feita.

Onze estudos avaliaram a validade de critério^{18, 19; 21-24; 29,30; 33,34; 37} sendo que a maioria recebeu classificação fraca na qualidade metodológica (ver Tabela 2). Quanto ao nível de evidência, BESTEST e FGA obtiveram moderada evidência para adequada validade de critério demonstrando potencial capacidade dos mesmos em prever risco de quedas na DP (ver tabelas 1, 2 e 3).

A análise da validade de critério demonstrou que 40% dos estudos foram pobres por não utilizarem um comparador considerado padrão ouro na avaliação do equilíbrio e risco de quedas^{18-19; 24,30-31; 34,37}. Isto pode ser explicado pelo fato de que ainda não há consenso sobre qual instrumento seja padrão ouro para avaliar equilíbrio e risco de quedas na DP. Muitos estudos utilizaram o relato de quedas como padrão ouro^{21, 24,30}, porém sabe-se que a obtenção de dados prospectivos em

estudos longitudinais para observação do risco de quedas na DP são mais adequados do ponto de vista metodológico¹¹⁻¹⁴.

É possível destacar ainda outras limitações que afetaram a qualidade metodológica dos estudos incluídos nessa revisão, tais como tamanho de amostra e falha no delineamento. Uma amostra pequena aponta para dificuldades no recrutamento dos indivíduos com desordens neurológicas para pesquisa em reabilitação³⁹, principalmente quando não há intervenções previstas no projeto, mas apenas a realização de medidas. Além disso, de acordo com o COSMIN¹⁴, um item é considerado pobre caso apresente uma falha grave no delineamento ou método do estudo. Dentre as principais falhas encontradas destacam-se a aplicação de diferentes itens do instrumento em um mesmo indivíduo por diferentes examinadores, levando a viés nos resultados encontrados e a utilização de dados da versão longa de um instrumento para análise de propriedades de medida da versão curta do mesmo instrumento, desconsiderando as mudanças no contexto ao se aplicar uma e outra versão. Outras falhas foram não especificar critérios de inclusão e exclusão, não descrever a operacionalização das medidas bem como a não caracterização dos examinadores (número, experiência e treinamento).

Principais instrumentos e recomendações

Embora nenhum dos instrumentos encontrados tenha apresentado características robustas com forte evidência para suas propriedades de medida, a qualidade metodológica dos estudos e os níveis de evidência encontrados, destacaram cinco instrumentos potencialmente mais adequados, na atualidade, para avaliar o equilíbrio ou o risco de quedas na DP. São eles: BBS, BESTEST, FGA na categoria de instrumentos de desempenho e, SAFFE e FES na categoria de questionários de auto-relato.

BBS

A adequada consistência interna da BBS revela que ela é consistente em avaliar o constructo equilíbrio, ou seja, há uma boa correlação entre os seus itens. Entretanto, ainda existem lacunas referentes a outras propriedades de medida que precisam ser preenchidas, como a confiabilidade e a validade de critério. Embora a BBS tenha adequada confiabilidade (ver tabela 1) o nível de evidência dessa

propriedade foi desconhecido (ver tabela 3) devido á pobre qualidade dos estudos encontrados (ver tabela 2). Apesar de ter moderada evidência para a validade de critério (ver tabela 3), esta foi considerada inadequada pelo fato de nenhum estudo ter utilizado a AUC (ver tabela 1). Tal método é considerado a melhor estimativa para propósitos discriminativos, ou seja, capacidade de predizer quem tem ou não um risco de quedas. Sendo assim, novos estudos com boa qualidade metodológica precisam ser realizados a fim de definir evidências sobre sua confiabilidade e validade não apenas em avaliar o equilíbrio, mas para predizer risco de quedas na DP.

BESTEST

O BESTEST foi válido em predizer risco de quedas na DP com $AUC > 0,80$ (ver tabela 1), mas se tratando de um instrumento com várias dimensões, cabe uma análise de sua consistência interna. Uma vez que este instrumento foi desenvolvido não apenas para avaliar os distúrbios do equilíbrio, mas também identificar qual sistema do equilíbrio apresenta alterações⁴⁰, seria importante estudar a consistência interna do BESTEST a fim de investigar a unidimensionalidade de cada subescala para verificar se todas medem o mesmo constructo, e a relação de todas as dimensões com o constructo equilíbrio. A confiabilidade do BESTEST na DP também carece de novos estudos, pois embora adequada, apenas um estudo avaliou a confiabilidade entre - examinadores e dois estudos avaliaram a confiabilidade teste-reteste.

FGA

O FGA é uma modificação do DGI, cujo objetivo foi melhorar a sua confiabilidade, uma vez que este instrumento produzia efeito teto na população de indivíduos com desordens vestibulares⁴¹. Os estudos também indicaram que o FGA pode ser um bom preditor de quedas na DP ($AUC > 0,80$ - tabela 1). Entretanto, a confiabilidade deste instrumento em indivíduos com DP ainda é desconhecida. Também não se tem informação sobre a presença de efeito teto em tais indivíduos, o que pode ser avaliado analisando sua validade de conteúdo.

SAFFE–NL e FES-S

O SAFFE–NL e FES-S foram questionários de auto-relato confiáveis para serem utilizados em indivíduos com DP. Entretanto, mais estudos são necessários para investigar sua consistência interna e validade de conteúdo, a fim de estabelecer se, de fato, o constructo medo da quedas, está sendo avaliado da forma como estas ferramentas propõem. O medo da queda é apontado como um bom preditor de futuras quedas em indivíduos com DP e instrumentos que medem este constructo podem ser associados a ferramentas de avaliação do equilíbrio nesta população³⁵. Na realidade, pesquisadores e clínicos concordam que apenas um instrumento de medida pode não ser suficiente para capturar as variações nos déficits do equilíbrio e risco de quedas apresentadas por indivíduos com DP^{5,42-44}. Sendo assim, avaliar a validade de critério se torna fundamental para a boa qualidade dos questionários de auto-relato citados.

Outras considerações

A análise da utilidade clínica não diferenciou os instrumentos uma vez que a maioria obteve alta pontuação, indicando facilidade e praticidade na sua utilização. Apenas o BESTEST e o MINI-BESTEST obtiveram uma pontuação mais baixa nos itens tempo de utilização (BESTEST) e necessidade de equipamento e treino específico (BESTEST e MINI-BESTEST).

O Mini- BESTEST e o TMT foram os únicos instrumentos de desempenho que não apresentaram moderada evidência para nenhuma de suas propriedades. O Mini- BESTEST é a versão curta do BESTEST, desenvolvida na tentativa de torná-lo mais útil no quesito tempo de administração²³, porém tal instrumento ainda necessita de equipamento e treino específicos o que dificulta sua utilização na prática clínica. Além disso, o MINI-BESTEST apresentou limitada evidência para a validade de critério (ver tabela 1), cuja análise foi feita por apenas um estudo de fraca qualidade (ver tabela 2). Com base nas evidências, a confiabilidade do Mini-BESTEST ainda é desconhecida^{23, 25}.

Todos os testes de componente único apresentaram evidência limitada ou desconhecida para as propriedades de medida investigadas, inclusive o PULL TEST, frequentemente utilizado na avaliação do equilíbrio de indivíduos com DP³⁴, (ver tabela 2). Entre os questionários de auto-relato, a ABC foi a escala mais

investigada, porém apresentou evidência conflitante para sua consistência interna e desconhecida para confiabilidade (ver tabelas 1 e 3).

Nenhum dos instrumentos encontrados teve a validade de conteúdo investigada em indivíduos com DP. Esta é uma propriedade importante porque analisa a relevância dos itens de um instrumento para a população na qual ele está sendo aplicado. Em relação ao equilíbrio, seria relevante investigar se os itens dos instrumentos utilizados atendem às especificidades e natureza das desordens do equilíbrio encontradas na DP.

Vale ressaltar que o checklist COSMIN avalia a qualidade dos estudos e não dos instrumentos de medida, portanto, a carência de evidências para os instrumentos de avaliação do equilíbrio e risco de quedas na DP, de uma maneira geral, não é resultado da ausência de estudos, mas das falhas metodológicas encontradas. Na classificação das evidências, basta haver um estudo de excelente qualidade para uma forte evidência existir. Dada a freqüente utilização e a boa utilidade clínica dos instrumentos encontrados, esta revisão reforça a necessidade de mais estudos de qualidade investigando as propriedades de medida de tais instrumentos. Nesse sentido, o checklist COSMIN pode ser utilizado como guia no desenho de estudos de qualidade superior¹².

Embora o período de busca tenha sido amplo, a restrição de idiomas pode ser considerada uma limitação da presente revisão.

Conclusões

Esta revisão identificou quatorze instrumentos para avaliação do equilíbrio e risco de quedas na DP, são eles: a BBS, o TUG, o BESTEST, o MINI-BESTEST, o DGI, o FGA, o TMT (POMA), o FRT, o PULL Test, o ROMBERG Test, o SAFFE – NL, a FES-S, a ABC e o RAPID. As principais propriedades de medida investigadas pelos estudos foram consistência interna, confiabilidade, erro de medida e validade de critério.

A qualidade metodológica dos estudos e os níveis de evidência encontrados destacaram cinco instrumentos potencialmente mais adequados, na atualidade, para avaliar o equilíbrio ou o risco de quedas na DP. São eles: BBS, BESTEST, FGA na categoria de instrumentos de desempenho e, SAFFE-NL e FES-S na categoria de questionários de auto-relato. Dos instrumentos citados acima, com exceção do BESTEST, todos apresentaram uma boa utilidade clínica o que reforça a necessidade de serem melhor explorados quanto as suas propriedades de medida em indivíduos com DP.

O *checklist* COSMIN é uma ferramenta que visa a excelência na qualidade metodológica de estudos de propriedades de medida. Sua utilização identificou que a maioria dos estudos apresentaram limitações importantes que influenciaram a evidência sobre os instrumentos utilizados para avaliar o equilíbrio e risco de quedas na DP. Esta revisão reforça a necessidade de novos estudos de qualidade e sugere a utilização do *checklist* COSMIN como um guia para pesquisas futuras.

Referências

1. Dexter D. T. and Peter J.. Parkinson disease: from pathology to molecular disease mechanisms. **Free Radical Biology and Medicine**, v. 62, p.132–144, 2013.
2. Kleinman, M., & Frank, S. Epidemiology and Clinical Diagnosis of Parkinson Disease. **PET Clinics**, v.8, n.4, p.447-458, 2013.
3. Soh SE, Morris ME, McGinley JL. Determinants of health-related quality of life in Parkinson's disease: a systematic review. **Parkinsonism Relat Disord**, v. 17, n.1, p.1-9, 2011.
4. KIM, Samuel D. et al. Postural Instability in Patients with Parkinson's Disease. **CNS drugs**, v. 27, n. 2, p. 97-112, 2013.
5. PICKERING, Ruth M., et al. A meta-analysis of six prospective studies of falling in Parkinson's disease. **Movement Disorders**, v.22, n.13, p.1892-1900, 2007.
6. Kwakkel G., de Goede C.J.T., van Wegen E.E.H. Impact of physical therapy for Parkinson's disease: A critical review of the literature. **Parkinsonism and Related Disorders**, v.13, n. Sup3, p. S478–S487, 2007.
7. Morris M. Movement disorders in people with Parkinson's disease: a model for physical therapy. **Phys Ther**, v.80, n.6, p.578-597, 2000.
8. DE GOEDE, Cees JT et al. The effects of physical therapy in Parkinson's disease: a research synthesis. **Archives of physical medicine and rehabilitation**, v. 82, n. 4, p. 509-515, 2001.
9. Mancini M, Horak F. B. The relevance of clinical balance assessment tools to differentiate balance deficit. **EUR J PHYS REABIL MED**, v.46, n.2, p.239-248, 2010.
10. HORAK, Fay B. Clinical assessment of balance disorders. **Gait & Posture**, v. 6, n. 1, p. 76-84, 1997.
11. MOKKINK, Lidwine B. et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. **Quality of Life Research**, v. 19, n. 4, p. 539-549, 2010.
12. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. **Journal of Clinical Epidemiology**, v. 63, N. 7, P. 737-745, 2010.
13. WATKINS, Mary P.; PORTNEY, L. G. **Foundations of clinical research: applications to practice**. Pearson/Prentice Hall, 2008.

14. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. **Quality of Life Research**, v. 19, n. 4, p. 539-549, 2010.
15. TERWEE, Caroline B. et al. Quality criteria were proposed for measurement properties of health status questionnaires. **Journal of clinical epidemiology**, v. 60, n. 1, p. 34-42, 2007.
16. Elbers, R. G., Rietberg, M. B., van Wegen, E. E., Verhoef, J., Kramer, S. F., Terwee, C. B., & Kwakkel, G.. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. **Quality of Life Research**, v. 21, n. 6, p.925-944, 2012.
17. TYSON, Sarah; CONNELL, Louise. The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review. **Clinical Rehabilitation**, v.23, n.11, p.1018-1033, 2009.
18. Qutubuddin AA, Pegg PO, Cifu DX, Brown R, McNamee S, Carne W. Validating the Berg Balance Scale for patients with Parkinson's disease: a key to rehabilitation evaluation. **Arch Phys Med Rehabil**, v.86, n.4, p. 789-92, 2005.
19. SCALZO, Paula L., et al. Validation of the Brazilian version of the Berg balance scale for patients with Parkinson's disease. **Arquivos de neuro-psiquiatria**, v. 67, n. 3B, p. 831-835, 2009.
20. Steffen T, Seney M. Test- Retest Reliability and Minimal Detectable change on Balance and Ambulation Test, The 36-item Short Form Healthy Survey, and Unified Parkinson Disease Rating Scale in People with Parkinson Disease. **Physical Therapy**, v. 88, n.6, p. 733-746, 2008.
21. DIBBLE, Leland E., et al. Diagnosis of fall risk in Parkinson disease: an analysis of individual and collective clinical balance test interpretation. **Physical therapy**, v.88, v.3, p.323-332, 2008.
22. DIBBLE, Leland E.; LANGE, Mark. Predicting falls in individuals with Parkinson disease: a reconsideration of clinical balance measures. **Journal of Neurologic Physical Therapy**, v. 30, n.2, p.60-67, 2006.
23. Leddy AL, Crowner BE, Earhart GM. Utility of the Mini-BESTest, BESTest, and BESTest sections for balance assessments in individuals with Parkinson disease. **J Neurol Phys Ther**. V.35, p.90-97, 2011.
24. LEDDY, Abigail L.; CROWNER, Beth E.; EARHART, Gammon M. Functional gait assessment and balance evaluation system test: reliability, validity, sensitivity, and

specificity for identifying individuals with Parkinson disease who fall. **Physical Therapy**, v.91, n.1, p.102-113, 2011.

25. Maia, A.C.Tradução e adaptação para o português-Brasil do Balance Evaluation Systems Test e do MINI-BESTest e análise de suas propriedades psicométricas em idosos e indivíduos com Doença de Parkinson.2012. 94 f. Dissertação (Mestrado em Ciências da Reabilitação) - **Escola de Educação Física, Fisioterapia e Terapia Ocupacional da Universidade Federal de Minas Gerais**. 2012.

26. BERGSTRÖM, Malin; LENHOLM, Emma; FRANZÉN, Erika. Translation and validation of the Swedish version of the mini-BESTest in subjects with Parkinson's disease or stroke: A pilot study. **Physiotherapy Theory and Practice**, v.28, n.7, p. 509-514, 2012.

27. HUANG, Sheau-Ling, et al. Minimal detectable change of the Timed "Up & Go" Test and the Dynamic Gait Index in people with Parkinson disease. **Physical Therapy**, v. 91, n.1, p.114-121, 2011.

28. DAL BELLO-HAAS, Vanina, et al. Psychometric Properties of Activity, Self-Efficacy, and Quality-of-Life Measures in Individuals with Parkinson Disease. **Physiotherapy Canada**, v. 63, n.1, p. 47-57, 2011

29. Foreman, K. B., Addison, O., Kim, H. S., & Dibble, L. E.). Testing balance and fall risk in persons with Parkinson disease, an argument for ecologically valid testing. **Parkinsonism & related disorders**, v.17, n.3, p.166-171, 2011.

30. Kegelmeyer, D. A., Kloos, A. D., Thomas, K. M., & Kostyk, S. K. . Reliability and validity of the Tinetti Mobility Test for individuals with Parkinson disease. **Physical Therapy**, v.87, n.10, p.1369-1378, 2007.

31. Schenkman, M., Cutson, T. M., Kuchibhatla, M., Chandler, J., & Pieper, C.). Reliability of impairment and physical performance measures for persons with Parkinson's disease. **Physical therapy**, v.77, n.1, p. 19-27, 1997.

32. SMITHSON, Fiona; MORRIS, Meg E.; IANSEK, Robert. Performance on clinical tests of balance in Parkinson's disease. **Physical therapy**, v. 78,n. 6, p. 577-592, 1998.

33. BEHRMAN, Andrea L., et al. Is the functional reach test useful for identifying falls risk among individuals with Parkinson's disease?. **Archives of physical medicine and rehabilitation**, v.83, n.4, p. 538-542, 2002 .

34. Visser M, Marinus J, Bloem BR, Kisjes H, van den Berg BM, van Hilten JJ. Clinical tests for the evaluation of postural instability in patients with Parkinson's disease. **Arch Phys Med Rehabil**,v. 84, n.11, p. 1669-74, 2003.

35. LOHNES, Corey A.; EARHART, Gammon M. External validation of abbreviated versions of the activities-specific balance confidence scale in Parkinson's disease. **Movement Disorders**, v. 25, n.4, p. 485-489, 2010.
36. NILSSON, Maria H.; DRAKE, Anna-Maria; HAGELL, Peter. Assessment of fall-related self-efficacy and activity avoidance in people with Parkinson's disease. **BMC geriatrics**, v.10, n.1, p.78, 2010.
37. CHONG, R. K. Y., et al. Diagnostic value of the rapid assessment of postural instability in parkinson's disease (RAPID) questionnaire. **International journal of clinical practice**, v.66, n.7, p. 718-721, 2012.
38. TOOTH, Leigh R.; OTTENBACHER, Kenneth J. The κ statistic in rehabilitation research: An examination. **Archives of physical medicine and rehabilitation**, v. 85, n. 8, p. 1371-1376, 2004.
39. LIMA, Lidiane O.; RODRIGUES-DE-PAULA, Fátima. Recruitment rate, feasibility and safety of power training in individuals with Parkinson's disease: a proof-of-concept study. **Brazilian Journal of Physical Therapy**, v. 17, n. 1, p. 49-56, 2013.
40. HORAK, Fay B.; WRISLEY, Diane M.; FRANK, James. The balance evaluation systems test (BESTest) to differentiate balance deficits. **Physical therapy**, v. 89, n. 5, p. 484-498, 2009.
41. WRISLEY, Diane M.; KUMAR, Neeraj A. Functional gait assessment: concurrent, discriminative, and predictive validity in community-dwelling older adults. **Physical therapy**, v. 90, n. 5, p. 761-773, 2010.
42. DIBBLE, Leland E.; LANGE, Mark. Predicting falls in individuals with Parkinson disease: a reconsideration of clinical balance measures. **Journal of Neurologic Physical Therapy**, v. 30, n. 2, p. 60-67, 2006.
43. LIM, L. I. I. K. et al. Measuring gait and gait-related activities in Parkinson's patients own home environment: a reliability, responsiveness and feasibility study. **Parkinsonism & related disorders**, v. 11, n. 1, p. 19-24, 2005.
44. JACOBS, J. V. et al. Multiple balance tests improve the assessment of postural stability in subjects with Parkinson's disease. **Journal of Neurology, Neurosurgery & Psychiatry**, v. 77, n. 3, p. 322-326, 2006.

Figura 1. Diagrama de fluxo para a seleção dos estudos

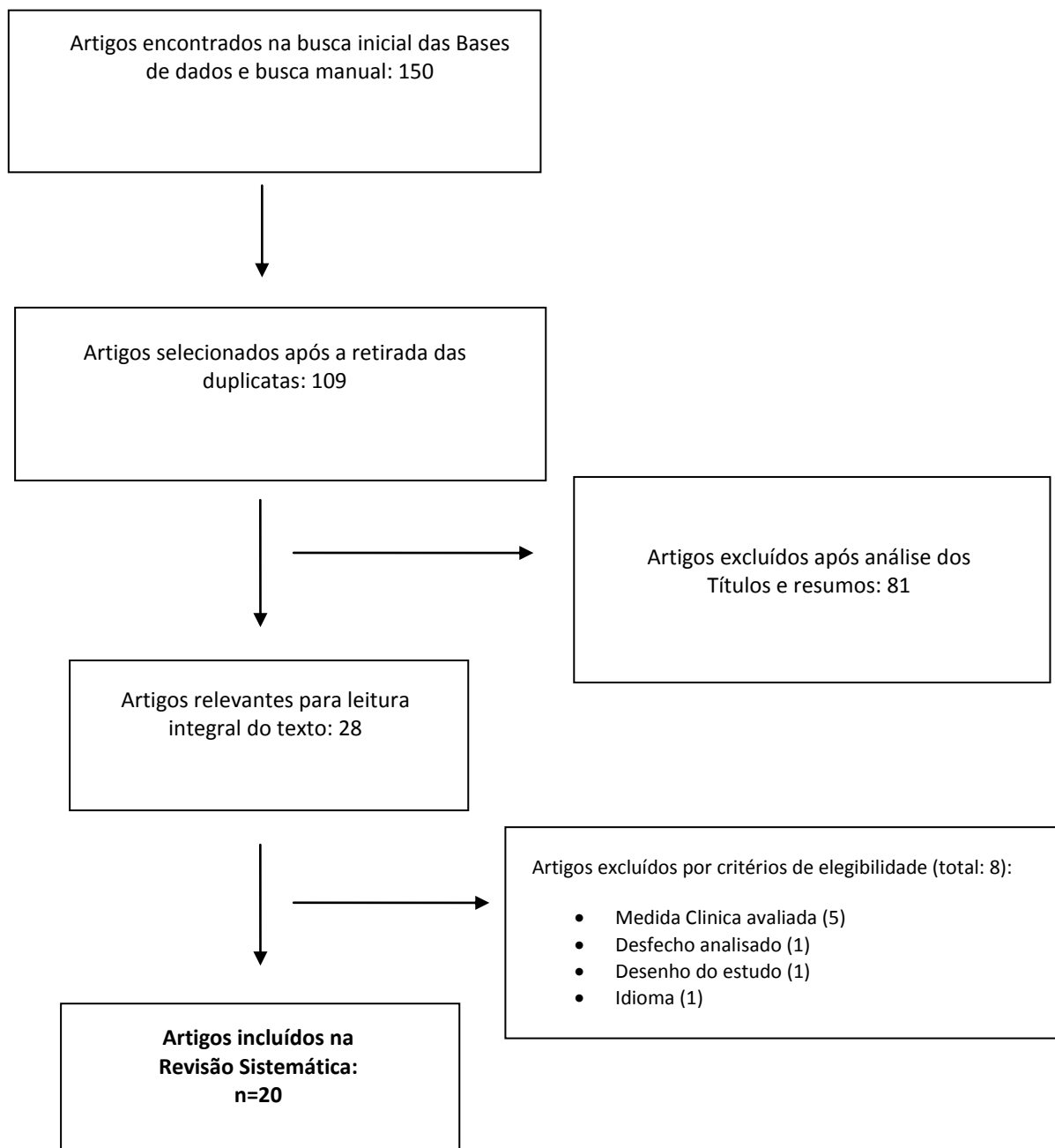


Tabela 1. Características dos estudos e propriedades de medida dos Instrumentos de Avaliação do Equilíbrio e Risco de Quedas na DP.

Instrumento	Característica dos indivíduos			Propriedades de Medida			
	Idade (anos)	Duração da doença (anos)	Gravidade	Consistência Interna	Confiabilidade	Erro de Medida	Validade de Critério
BBS ¹⁸⁻²²	62.0(7.9) -73.91(6.45) ^a	7.43(5.62) -14.0(6) ^a	HY:1-4 ^f	n=37-53 ^C 0.86≤α≤0.92	n=37-53 ^C EE.-CCI=0.84 ²¹ TR.- CCI=0.94 ²²	n=37 MMD=5 ²²	Análise da correlação ^{18, 19} n=37-70 ^C BBS -UPDRS III=-0.37≤r _s ≤-0.58 BBS - HY=-0.45≤ r _s ≤-0.50 BBS- S&E=0.49≤ r _s ≤0.55 Validade preditiva n=37-70 ^C S =0.64-0.79 / E=0.74-0.83 ^d LR=-0.29-0.43 ^{d,e} / LR+=3.07-3.83 ^{d,e}
BESTEST ²³⁻²⁵	66.5(10.3) - 68.2(9.3) ^a	8.5(4.8) -9.9(6.4) ^a	HY:1-4 ^f	—————	n=10-24 ^C EE.-CCI=0.96 ²⁵ TR.-0.88≤CCI≤0.98	—————	Análise de Correlação (r_s): Bestest - HY=r _s 0.73 Bestest -UPDRS III r _s =0.75 Bestest - ABC=r _s 0.75 Bestest - BBS=r _s 0.87 Bestest- FGA=r _s 0.88 Validade preditiva ^{23,24} n=80 AUC=0.84-0.85 ^d Ponto de Corte ≤69% S=0.84 / E=0.76 LR -=0.21 /LR= 3.49 ^e Ponto de Corte ≤84% S: 1.0 / E: 0.39 LR -=0.00 /LR+=1.64 ^e
Mini-BESTEST ^{23,25,26}	60.3[46-85] ^b -68.2(9.3) ^a	8.5(4.8)-9.9(6.4) ^a	HY:1-4 ^f	—————	n=9-23 ^C EE.- CCI=0.91 ²⁵ TR.-0.92≤CCI≤0.95	—————	Análise de correlação ²⁵ Mini-Bestest - BBS: r _s =0.94 Mini-Bestest - TUG: r _s =0.81 Validade preditiva ²³ n=80 AUC=0.86^e Ponto de Corte ≤63%: S=0.88/ E=0.78 LR -=0.15 /LR+: 4.03 ^e Ponto de Corte ≤72%: S=0.96/ E=0.47 LR -=0.08 ^e / LR+=1.82 ^e

Tabela 1. ...Continuação

Instrumento	Característica dos indivíduos			Propriedades de Medida			
	Idade (anos)	Duração da doença (anos)	Gravidade	Consistência Interna	Confiabilidade	Erro de Medida	Validade de Critério
TUG ^{22,27-29}	64.9(8)-69.94(11.28) ^a	4.5(4.3)-7.43(5.62) ^a	HY:1-4 ^f	—————	n=24-72 ^C TR.- 0.69≤CCI≤0.80	n=24-72 ^C MMD=3.5-4.85 ^a EPM=1.75 ²³	Validade Preditiva ^{20,24} n= 36-45 ^C AUC= 0.68 Ponto de Corte= 7.95 ²² S=0.93 / E=0.30 LR -=0.27/ LR +=1.31 ^e
FGA ^{24,29}	Caidores: 68.8(7.8)-70.95(11.41) ^a Não caidores:66.64(10.05)-67.90(10) ^a	Caidores:8.18(4.58)-11.4(5.5) Não caidores:4.64(3.25)-7.15(3.81)	HY:1-4 ^f	—————	n=24 ²⁴ TR.-CCI=0.91 n=15 ²⁴ EE – CCI=0.93	—————	Análise da correlação ²⁴ FGA-HY=r _s 0.67 FGA-UPDRS= r _s 0.66 FGA-BBS= r _s 0.78 FGA-ABC= r _s 30 Validade Preditiva n= 36-80 ^C AUC: 0.80-.81 ^d Ponto de Corte ≤15/30 ²² S= 0.72/ E=0.78 LR -=0.36/ LR+=3.24 ^e Ponto de Corte ≤27/30 S=1.0/ E= 0.19 LR -=0.00/ LR+=1.23 ^e
DGI ^{21,22,27}	67.5(11.6) -73.91(6.45) ^a	5.43(4.25)-9.07(6.13) ^a	HY:1-4 ^f	—————	n=72 TR.-CCI=0.84 ²⁷	n=72 MMD= 2.9 ²⁷	Validade Preditiva ^{21,22} n= 45-70 ^C Ponto de corte=19-22 ^d S=0.64-0.89/E=0.48-0.85 ^d LR -=0.22-0.42 / LR +=1.86-4.26 ^d
TMT ³⁰	65(10.9) – 68.8(11.04)	9.4(7.3)	HY:1-5 ^f	—————	n= 30 EE. – ICC=0.87	—————	Validade Preditiva n= 149 S=0.55-0.85/E=0.53-0.76 ^{d,e} LR -=0.29-0.62/ LR +=1.80-2.45 ^{d,e}
FRT ^{20-22;31-33}	64.3(9.3)-74.5(5.7) ^a	5.43(4.25)-12.9(3.5) ^a	HY:1-4 ^f	—————	n=14-70 ^C TR.- 0.73≤CCI≤0.93	n=37 MMD = 7 ²⁰	Validade Preditiva ^{21,22} n= 45-70 ^C S=0.77-0.86 /E== 0.52- 0.55 ^d LR -=0.30-0.42/ LR += 1.71-1.79 ^{d,e}

Tabela 1. ...Continuação

Instrumento	Característica dos indivíduos			Propriedades de Medida			
	Idade (anos)	Duração da doença (anos)	Gravidade	Consistência Interna	Confiabilidade	Erro de Medida	Validade de Critério
ROMBERG ²⁰	71(12)	14(6)	HY:1-4 ^f	_____	n=37 TR.-0.84≤CCI≤0.86	n=37 MMD=10-19	_____
PULL Test ^{29,34}	62.7(8)-70.95(11.41) ^a	4.64(3.25)-8.18(4.58) ^a	HY:1-4 ^f	_____	n=57 E.E ³⁴ : Pastor - K_w: 0.98 Nutt- K_w: 0.93 SPES- K_w: 0.87 Bloem- K_w: 0.85 UPDRS- K_w: 0.63	_____	Validade Preditiva n=57 S / E ³⁴ Nutt 0.63/0.88 Bloem 0.65/0.85 UPDRS 0.66/0.82 SPES 0.55/0.92 Pastor 0.70/0.69 AUC=0.69 ²⁹
ABC ^{20,28,35}	64.9(8) -71.0(12) ^a	4.5(4.3)-14.0(6) ^a	HY:1-4 ^f	n=24-89 ^C 0.93≤α≤0.96	n=24-89 ^C TR.-0.79≤ CCI ≤0.94	n=22-37 ^C MMD=[11.2-13] ^d EPM=4.01 ²⁸	_____
SAFFE- NL ³⁶	64.5(7.2)	15.9(7.3)	UPDRS III: 15.0 ^b	n= 79 α:0.95	n=79 TR.-CCI=0.92	n=79 EPM: 2.4	_____
FES -S ³⁶	64.5(7.2)	15.9(7.3)	UPDRS III: 15.0 ^b	n= 79 α:0.97	n=79 TR.-CCI=0.87	n=79 EPM=12.3	_____
RAPID ³⁷	68(8) – 70(8) ^d	-----	HY:1-4 ^f	_____	_____	_____	Validade Preditiva AUC= 0.50 – 0.66 ^{d,e} S=0.71/ E=0.44 ^e

HY: Hoehn and Yahr; **UPDRS:** Unified Parkinson's Disease Rating Scale; **EE:** Entre-Examinadores; **TR:** Teste- Reteste; **CCI:** Coeficiente de Correlação Intraclasse; **MMD:** Mínima mudança detectável; **EPM:** Erro padrão de medida; **r_s:** Spearman ; **K_w:** kappa ponderado; **S:** Sensibilidade; **E:** Especificidade; **AUC** (Area Under the Roc Curve): Area sob a Curva ROC (Receiver Operator Characteristic); **LR** (Likelihood Ratio) +/- : Razão de Verossimilhança positivo/negativo.

^a Variação da média (DP) entre os estudos.

^b Expresso em mediana [amplitude].

^c Variação do numero de sujeitos entre os estudos.

^d Variação dos valores encontrados nos estudos.

^e Valores dentro do intervalo de confiança de 95 %.

^f Amplitude.

Tabela 2. Análise da qualidade metodológica dos estudos através do *checklist* COSMIN.

Referência	Instrumentos avaliados	Propriedades de medida investigadas					
		Consistência Interna	Confiabilidade	Erro de Medida	Teste de Hipótese	Validade Transcultural	Validade de Critério
Behrman AL et al 2002³³	FRT						Pobre
Bergstrom M et al 2012²⁶	Mini-BESTEST				Pobre	Fraca	
Chong RK et al 2011³⁷	RAPID						Pobre
Dal Bello-Haas V et al 2010²⁸	ABC /TUG	Pobre	Pobre	Pobre			
Dibble LE et al 2008²¹	BBS / DGI / FRT / TUG						Fraca
Dibble LE and Lang M 2006²²	FRT/ BBS /DGI /TUG						Fraca
Foreman K.B. et al 2011²⁹	FGA / TUG						Fraca
Huang S-L et al 2011²⁷	TUG / DGI		Pobre	Boa			
Kegelmeyer DA et al 2007²⁹	TMT (POMA)		Pobre				Fraca
Leddy AL et al 2011²⁴	FGA / BESTEST		Pobre				Fraca
Leddy AL et al 2011²³	BESTEST / MINI-BESTEST		Pobre				Fraca
Lohnes and Earhart 2010³⁵	ABC	Pobre	Pobre				
Maia AC 2012²⁵	BESTEST/ Mini-BESTEST		Pobre			Pobre	
Nilsson et al 2010³⁶	FES-S / SAFFE-NL	Pobre	Boa	Pobre	Pobre		
Qutubuddin AA et al 2005¹⁸	BBS						Fraca
Scalzo et al 2009¹⁹	BBS (Versão Brasileira)	Boa	Pobre				Fraca
Schenkman et al 1997³¹	FRT		Pobre				
Smithson et al 1998³²	FRT		Pobre				
Steffen and Seney 2008²⁰	ABC / BBS/ FRT / ROMBERG	Pobre	Pobre	Pobre			
Visser M et al 2003³⁴	PULL TEST		Fraca				Pobre

Tabela 3. Caracterização das propriedades de medida quanto ao nível de evidência.

Instrumento	Propriedades de medida investigadas					
	Consistência Interna	Confiabilidade	Erro de Medida	Teste de Hipótese	Validade Transcultural	Validade de Critério
Instrumentos de desempenho						
BBS	+	+				-
	Moderada	Desconhecida	Desconhecida			Moderada
BESTEST		+				+
		Desconhecida			Desconhecida	Moderada
MINI- BESTEST		+				+
		Desconhecida			Limitada	Limitada
TUG		±				-
		Conflitante	Moderada			Limitada
DGI		+				-
		Desconhecida	Moderada			Moderada
FGA		+				+
		Desconhecida				Moderada
TMT(POMA)		+				-
		Desconhecida				Limitada
Testes de componente único						
FRT		+				-
		Desconhecida				Moderada
PULL TEST		+				-
		Limitada				Limitada
ROMBERG TEST		+				
		Desconhecida	Desconhecida			
Questionários de auto-relato						
ABC	±	+				
	Conflitante	Desconhecida	Desconhecida			
SAFFE- NL	+	+				
	Desconhecida	Moderada	Desconhecida			
FES –S	-	+				
	Desconhecida	Moderada	Desconhecida			
RAPID						-
						Desconhecida

(+) adequada; (-) inadequada; (±) conflitante

Tabela 4 - Avaliação da utilidade clínica dos instrumentos de avaliação do equilíbrio e risco de quedas na doença de Parkinson.

Instrumento de Medida	Tempo de Administração	Custo	Portatibilidade	Equipamento e treino específicos	Total (máximo = 10)
<i>Instrumentos de desempenho</i>					
BBS	2	3	2	2	9
BESTEST	1	2	1	1	5
MINI-BESTEST	2	2	1	1	6
TUG	3	3	2	2	10
DGI	3	3	2	2	10
FGA	3	3	2	2	10
TMT(POMA)	2	3	2	2	9
<i>Testes de Componente Único</i>					
FRT	3	3	2	2	10
PULL Test	3	3	2	2	10
ROMBERG	3	3	2	2	10
<i>Questionários de Auto-relato</i>					
ABC Scale	2	3	2	2	9
SAFFE- NL	3	3	2	2	10
FES –S	3	3	2	2	10
RAPID	3	3	2	2	10

5. CONSIDERAÇÕES FINAIS

A instabilidade postural tem sido apontada como um dos fatores de risco para a ocorrência de quedas em indivíduos com DP. A presença de dores, limitações de atividades, medo da queda, depressão e fraturas são algumas das conseqüências que podem ocorrer por causa das quedas. Por tanto, a abordagem do equilíbrio é fundamental na prática fisioterápica durante o manejo de pacientes com DP, visando identificar e quantificar os déficits de equilíbrio e risco de quedas bem como monitorar a progressão dos mesmos e, desta forma, realizar uma adequada tomada de decisão clínica direcionada à intervenção.

Os instrumentos de avaliação clínica surgem como uma alternativa para a medida do equilíbrio, uma vez que, tais instrumentos tendem a ser mais fáceis de utilizar, rápidos na sua administração e de baixo custo. Para uma adequada utilização de tais ferramentas é necessário que as mesmas sejam válidas, confiáveis e responsivas, ou do contrário, existe o risco de se obter resultados imprecisos ou inadequados que podem levar a conclusões e condutas equivocadas.

A qualidade de um instrumento de medida está diretamente relacionada às suas propriedades de medida tais como confiabilidade, validade e responsividade, as quais caracterizam um instrumento apontando suas potencialidades e debilidades em medir a variável em questão, neste caso, o equilíbrio. Neste trabalho pode-se observar que dentre os quatorze instrumentos encontrados para avaliar o equilíbrio e risco de quedas na DP, cinco se destacaram como potencialmente mais adequados. São eles: BBS, BESTEST, FGA na categoria de instrumentos de desempenho e, SAFFE-NL e FES-S na categoria de questionários de auto-relato. Dentro da perspectiva de funcionalidade da Classificação Internacional de Funcionalidade, Incapacidade e Saúde (CIF), os instrumentos BBS, BESTEST e FGA podem ser considerados medidas de capacidade, uma vez que são realizados em ambientes controlados pelo terapeuta. O BESTEST possui itens de dimensão corporal, ou seja, itens específicos de estrutura e função associados a itens da dimensão individual que são itens específicos de atividades de cotidiano; a BBS e o FGA possuem itens específicos da dimensão individual. Em contrapartida, SAFFE-NL e FES-S podem ser considerados medidas de desempenho da dimensão social, uma vez que os itens destes instrumentos são preenchidos pelo próprio indivíduo, através de sua

percepção de como ele lida com o medo da queda. Sendo assim, uma interação de todos os instrumentos citados acima no momento de avaliar um indivíduo com distúrbios de equilíbrio e risco de quedas podem possibilitar uma visão mais holística da real situação do indivíduo podendo abranger todos os aspectos do modelo de Funcionalidade proposto pela CIF.

Dada a freqüente utilização destes instrumentos em indivíduos com DP e a boa utilidade clínica da maioria, Esta revisão reforça a necessidade de novos estudos de qualidade e sugere a utilização do *checklist* COSMIN como um guia para pesquisas futuras.

6. REFERÊNCIAS

ASHBURN, Ann *et al.* A community-dwelling sample of people with Parkinson's disease: characteristics of fallers and non-fallers. **Age and Ageing**, v. 30, n. 1, p. 47-52, 2001.

BARTELS Anna L, LEEDERS Klaus L. Parkinson's disease: The syndrome, the pathogenesis and pathophysiology. **Cortex**, v. 45, n.8, p. 915-92, 2009.

BARBOSA, M.T. Parkinsonism and Parkinson's disease in the elderly: A community-based survey in Brazil (the Bambuí study). **Movement Disorders**, v. 21, n. 6, p. 800-808, 2006.

BENATRU, I.; VAUGOYEAU M.; AZULAY J.-P. Postural disorders in Parkinson's disease. **Neurophysiologie Clinique/Clinical Neurophysiology**, v. 38, n. 6, p. 459-465, 2008.

BERG, K. Balance and its measure in the elderly: a review. **Physiotherapy Canada**, v. 41, n. 5, p. 240-246, 1989.

BERNHARDT, J.; HILL, K. We only treat what it occurs to us to assess: the importance of knowledge-based assessment In: REFSHAUGE, K.; ADA, L.; ELLIS, E. (1st Ed) **Science-Based Rehabilitation.Theories into Practice**. Butterworth-Heinemann, 2005. P. 15-48.

BLOEM, B.R. *et al.* Prospective assessment of falls in Parkinson's disease. **Journal of neurology**, v. 248, n. 11, p. 950-958, 2001.

BOHNEN, N.I.; CHAM, Rakié. Postural control, gait, and dopamine functions in parkinsonian movement disorders. **Clinics in geriatric medicine**, v. 22, n. 4, 2006.

DE GOEDE, C.J.T. *et al.* The effects of physical therapy in Parkinson's disease: a research synthesis. **Archives of physical medicine and rehabilitation**, v. 82, n. 4, p. 509-515, 2001.

DEXTER, D.T.; JENNER, P. Parkinson disease: from pathology to molecular disease mechanisms. **Free radical biology and medicine**, v. 62, p. 132-144, 2013.

DINIZ, C. *et al.* Redução do limite de estabilidade direção-específica em indivíduos leve a moderadamente afetados pela doença de Parkinson. **Fisioterapia e pesquisa**, v. 18, n. 4, p. 306-310, dez. 2011.

ELBERS R.G *et al.* Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. **Quality of life research**, v. 21, n. 6, p.925-944, 2012.

GADOTTI, I.C.; VIEIRA, E.R.; MAGEE, D.J.; Importance and clarification of measurement properties in rehabilitation. **Revista brasileira de fisioterapia**, v.10, n.2, p.137-146, 2006.

HORAK, F.B.; NUTT, J.G.; NASHNER, L.M. Postural inflexibility in parkinsonian subjects. **Journal of the Neurological Sciences**, v. 111, n.1, p. 46-58, 1992.

HORAK, F.B. Clinical assessment of balance disorders. **Gait & Posture**, v. 6, n. 1, p. 76-84, 1997.

HORAK, F.B.; DIMITROVA, D.; NUTT, J.G. Direction-specific postural instability in subjects with Parkinson's disease. **Experimental neurology**, v. 193, n. 2, p. 504-521, 2005.

HORAK, F.B. Posture. In: KANDEL, ER, SCHWARTZ, JH, JESSEL, TM. (5th Ed.) Principles of neural science. New York: **McGraw-Hill**, p. 935-959, 2013.

KIM, S.D. et al. Postural Instability in Patients with Parkinson's Disease. **CNS drugs**, v. 27, n. 2, p. 97-112, 2013.

KLEINMAN M.; FRANK, S. Epidemiology and Clinical Diagnosis of Parkinson Disease. **PET Clinics**, v.8, n.4, p.447-458, 2013.

KRYGOWSKA-WAIS, Anna. Falls in Parkinson's disease. Causes and impact on patients' quality of life. **Functional neurology**, v. 20, n. 4, p. 163-168, 2005.

KULISEVSKY J, LUQUIN M, ARBELO J, BURGUERA J, CARRILLO F, CASTRO A, et al . Advanced Parkinson's disease: Clinical characteristics and treatment. Part I. **Neurologia**, v.28,n.8,p.503-521, 2013.

KWAKKEL G.; de GOEDE C.J.T.; VAN WEGEN, E.E.H. Impact of physical therapy for Parkinson's disease: A critical review of the literature. **Parkinsonism and related disorders**, v.13, n. Sup3, p. S478–S487, 2007.

LANG, A.E.; LOZANO, A.M. Parkinson's disease. **New England Journal of Medicine**, v. 339, n. 15, p. 1044-1053, 1998.

Marsden, C.D. Parkinson's disease. **Journal of Neurology, Neurosurgery & Psychiatry**, v. 57, n.6, p.672-81, 1994.

MACKINNON, C.D.; WINTER, D.A. Control of whole body balance in the frontal plane during human walking. **Journal of biomechanics**, v. 26, n. 6, p. 633-644, 1993.

MANCINI, M.; HORAK, F.B. The relevance of clinical balance assessment tools to differentiate balance deficits. **European journal of physical and rehabilitation medicine**, v. 46, n. 2, p. 239, 2010.

MOKKINK, L.B. et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. **Quality of Life Research**, v. 19, n. 4, p. 539-549, 2010.

MOKKINK, LB. *et al.* International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. **Journal of Clinical Epidemiology**, v. 63, N. 7, P. 737-745, 2010.

MORRIS, M.E. Movement disorders in people with Parkinson disease: a model for physical therapy. **Physical Therapy**, v. 80, n. 6, p. 578-597, 2000.

MORRIS, M.E.; IANSEK, R. Characteristics of motor disturbance in Parkinson's disease and strategies for movement rehabilitation. **Human Movement Science**, v. 15, n. 5, p. 649-669, 1996.

OBESO, J.A. *et al.* Pathophysiology of the basal ganglia in Parkinson's disease. **Trends in neurosciences**, v. 23, p. S8-S19, 2000.

PICKERING, R.M. *et al.* A meta-analysis of six prospective studies of falling in Parkinson's disease. **Movement Disorders**, v.22, n.13, p.1892-1900, 2007.

POLLOCK, A.S. *et al.* What is balance? **Clinical rehabilitation**, v. 14, n. 4, p. 402-406, 2000.

PORTNEY, L.G.; WATKINS, M.P. Foundations Of Clinical Research: Applications To Practice.(3rd Ed.) .New Jersey: **Pearson Prentice hall**. p 892, 2009.

SCHOLTES, V.A.; TERWEE, C.B.; POOLMAN, R.W.; What makes a measurement instrument valid and reliable? **Injury-International Journal of Care of Injured**, v. 42, n.3, p. 236–240, 2011.

SOH, S.E.; MORRIS, M.E.; MCGINLEY, J.L.; Determinants of health-related quality of life in Parkinson's disease: a systematic review. **Parkinsonism and related Disorders**, v. 17, n.1, p.1-9, 2011.

TEIXEIRA, A.L.; CARDOSO, F . Tratamento inicial da doença de Parkinson. **Revista de Neurociências**, v. 12, p. 146-151, 2004.

TERWEE, C.B. *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. **Journal of clinical epidemiology**, v. 60, n. 1, p. 34-42, 2007.

TERWEE, C.B. *et al.* Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. **Quality of Life Research**, v.21, n.4,p. 651-657, 2011.

TYSON, S.; CONNELL, L. The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review. **Clinical Rehabilitation**, v. 23, n. 11, p. 1018-1033, 2009.

WILLIAMS D R, WATT H C, LEES A J. Predictors of falls and fractures in bradykinetic rigid syndromes: a retrospective study. **Journal of Neurology, Neurosurgery & Psychiatry**, v.77, n.4, p.468–473, 2006.

WOOD, B. H. *et al.* Incidence and prediction of falls in Parkinson's disease: a prospective multidisciplinary study. **Journal of Neurology, Neurosurgery & Psychiatry**, v. 72, n. 6, p. 721-725, 2002.

VAUGOYEAU, M. *et al.* Impaired vertical postural control and proprioceptive integration deficits in Parkinson's disease. **Neuroscience**, v.146, n.2, p. 852-863, 2007.

VAUGOYEAU, M.; HAKAM, H.; AZULAY, J-P. Proprioceptive impairment and postural orientation control in Parkinson's disease. **Human movement science**, v. 30, n. 2, p. 405-414, 2011.

7. ANEXOS

Anexo 1: Checklist COSMIN



COSMIN checklist with 4-point scale

Contact

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

Instructions

This version of the COSMIN checklist is recommended for use in systematic reviews of measurement properties. With this version it is possible to calculate overall methodological quality scores per study on a measurement property. A methodological quality score per box is obtained by taking the lowest rating of any item in a box ('worse score counts'). For example, if for a reliability study one item in the box 'Reliability' is scored poor, the methodological quality of that reliability study is rated as poor. The Interpretability box and the Generalizability box are mainly used as data extraction forms. We recommend to use the Interpretability box to extract all information on the interpretability issues described in this box (e.g. norm scores, floor-ceiling effects, minimal important change) of the instruments under study from the included articles. Similar, we recommend to use the Generalizability box to extract data on the characteristics of the study population and sampling procedure. Therefore no scoring system was developed for these boxes.

This scoring system is described in this paper:

Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research* 2011, July 6 [epub ahead of print].

Step 1. Evaluated measurement properties in the article

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

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

Box General requirements for studies that applied Item Response Theory (IRT) models

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

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

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

Box A. Internal consistency

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

7	Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	Internal consistency statistic calculated for each subscale separately			Internal consistency statistic NOT calculated for each subscale separately
8	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor Methodological flaws in the design or execution of the Study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>					
9	for Classical Test Theory (CTT), continuous scores: Was Cronbach's alpha calculated?	Cronbach's alpha calculated		Only item-total Correlations Calculated	No Cronbach's alpha and no item-total correlations calculated
10	for CTT, dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total Correlations Calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated
11	for IRT: Was a goodness of fit statistic at a global level calculated? E.g. χ^2 , reliability coefficient of estimated latent trait value (index of (subject or item) separation)	Goodness of fit statistic at a global level calculated			Goodness of fit statistic at a global level NOT calculated

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

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)

Design requirements

	excellent	good	fair	poor
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥ 100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Were at least two measurements available?	At least two measurements			Only one measurement
5 Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6 Was the time interval stated?	Time interval stated		Time interval NOT stated	
7 Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate

9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>					
11	for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
12	for dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
13	for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
14	for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

Box C. Measurement error: absolute measures

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

9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>					
11	for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	SEM, SDC, or LoA calculated	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population

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

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

Box E. Structural validity

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

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

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

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

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

Box G. Cross-cultural validity

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

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

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

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

Box H. Criterion validity

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

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

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

6	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	Anything that occurred during the interim period (e.g. treatment) adequately described	Assumable what occurred during the interim period	Unclear or NOT described what occurred during the interim period
7	Was a proportion of the patients changed (i.e. improvement or deterioration)?	Part of the patients were changed (evidence provided)	NO evidence provided, but assumable that part of the patients were changed	Unclear if part of the patients were changed Patients were NOT changed

Design requirements for hypotheses testing

For constructs for which a gold standard was not available:

8	Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
9	Was the expected <i>direction</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated	
10	Were the expected absolute or relative <i>magnitude</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated	

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

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

Interpretability

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

Box Interpretability

Percentage of missing items

Description of how missing items were handled

Distribution of the (total) scores

Percentage of the respondents who had the lowest possible (total) score

Percentage of the respondents who had the highest possible (total) score

Scores and change scores (i.e. means and SD) for relevant (sub) groups, e.g. for normative groups, subgroups of patients, or the general population

Minimal Important Change (MIC) or Minimal Important Difference (MID)

Generalizability

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

Box Generalisability

Median or mean age (with standard deviation or range)

Distribution of sex

Important disease characteristics (e.g. severity, status, duration) and description of treatment

Setting(s) in which the study was conducted (e.g. general population, primary care or hospital/rehabilitation care)

Countries in which the study was conducted

Language in which the HR-PRO instrument was evaluated

Method used to select patients (e.g. convenience, consecutive, or random)

Percentage of missing responses (response rate)

Anexo 2: Avaliação dos níveis de evidência das propriedades de medida

Critérios para qualidade geral das propriedades de medida

Classificação das estimativas			
Propriedade de medida	Adequada (+)	Inadequada (-)	Conflitante (±)
Consistência interna	Alpha de Cronbach entre 0.70 e 0.95	Alpha de Cronbach <0.70 ou >0.95	Resultados conflitantes
Confiabilidade	ICC ou Kappa \geq 0.70	ICC ou Kappa <0.70	
Validade de critério	AUC \geq 0.70	AUC<0.70 ou AUC não calculada	

Modificado de Terwee et al., 2007, pg.39.

Níveis de evidência das propriedades de medida

Nível	Classificação das estimativas	Critério
Forte	Adequada (+) ou Inadequada (-)	Resultados consistentes em múltiplos estudos de boa qualidade metodológica ou em um estudo de excelente qualidade metodológica.
Moderada	Adequada (+) ou Inadequada (-)	Resultados consistentes em múltiplos estudos de fraca qualidade metodológica ou em um estudo de boa qualidade metodológica.
Limitada	Adequada (+) ou Inadequada (-)	Um estudo de fraca qualidade metodológica.
Desconhecida	Adequada (+) ou Inadequada (-)	Somente estudos de pobre qualidade metodológica.
Conflitante	Conflitante (±)	Resultados conflitantes.

Modificado de Elbers et al., 2012, pg.928.

Anexo 3: Escala de avaliação da utilidade clínica

Avaliação da utilidade clínica

<u>Questão avaliada</u>	<u>Pontuação</u>
Tempo para administrar, analisar e interpretar a medida.	Menos de 10 minutos. Pontue 3 10-30 minutos. Pontue 2 30-60 minutos. Pontue 1 >1 hora. Pontue 0
Custo.	< R\$376,00. Pontue 3 Entre R\$376,00 e R\$1881,00 . Pontue 2 Entre R\$1881,00 a R\$3762,00. Pontue 1 > R\$3762,00 pontue ou valor desconhecido. Pontue 0
A medida precisa de equipamento e treino específico para ser utilizada?	Não. Pontue 2 Necessita de equipamento mas este é simples e fácil de ser utilizado e não necessita de treino específico. Pontue 1 Sim ou desconhecido. Pontue 0
A medida é portátil? Ela pode ser levada pelo paciente?	Sim, facilmente (pode ser transportada em uma bolsa). Pontue 2 Sim, pode ser levada em uma pasta ou carrinho. Pontue 1 Não ou muito difícil. Pontue 0

Modificado de: Tyson and Connell, 2009, pg. 1020.

8. APÊNDICES

Apêndice 1: Estratégia de busca

1. explode PARKINSON'S DISEASE/ all subheadings {No Related Terms}
2. (PARKINSON* in TI) or (PARKINSON* in AB) {No Related Terms}
3. 1 or 2
4. psychometric properties.mp
5. psychometrics/
6. psychometric.mp
7. psychometric\$.mp
8. reproducibility of Results.sh
9. Reproducibility of Results.mp
10. reproducibilit\$.mp
11. measurement propert\$.mp
12. reliability.mp
13. reliability coefficients.mp
14. reliab\$.mp
15. repeatability.mp
16. repeatab\$.mp
17. validity.mp
18. valid\$.mp
19. validation.mp
20. validation studies.sh
21. validation Studies as Topic.sh
22. repeated measurements.mp
23. (intrarater adj5 relia\$).mp
24. (Interrater adj5 relia\$).mp
25. (int\$ reliability).mp
26. responsiveness.mp
27. appropriateness.mp
28. methodological stud\$.mp
29. (methodological adj5 stud\$).mp
30. Cronbach\$.mp
31. methodological study.mp
32. methodological research.mp
33. consistency.mp
34. test-retest.mp
35. intraclass correlation.mp
36. kappa statistics.mp
37. clinical utility.mp

38. clinical usefulness.mp
39. sensitivity to change.mp
40. or/4-39
41. Balance
42. Postural
43. Postural Balance
44. Equilibrium
45. Postural Equilibrium
46. Accidental falls
47. or/41-46
48. 3 AND 40 AND 47
CINAHL

(PARKINSON'S DISEASE OR PARKINSON) AND (Cronbach OR consistency OR test-retest OR intraclass correlation OR kappa statistics OR clinical utility OR clinical usefulness OR sensitivity to change OR psychometric properties OR reproducibility OR reliability OR validity OR responsiveness OR methodological) AND (Balance OR postural OR equilibrium)

Apêndice 2: Protocolo para inclusão dos estudos

CRITÉRIOS PARA INCLUSÃO DOS ESTUDOS NA REVISÃO SISTEMÁTICA

Artigo (numero): _____

Revisor: _____

1. Participante

- Diagnóstico de Doença de Parkinson

Sim

Não

2. Medida clínica avaliada: Equilíbrio ou Risco de quedas

Sim

Não

3. Desfecho de análise: Propriedade de medida (pelo menos uma)

Sim

Não

4. Tipos de instrumentos – Avaliação clínica (exceto posturografias)

Escala

Teste

Questionário

Posturografias

5. Desenho de estudo Metodológico (exceto ensaios clínicos com análise de propriedades de medida):

Sim

Não

INCLUIR ARTIGO:

Sim

Não

Dúvida

Apêndice 3: Formulário de extração dos dados

FORMULÁRIO DE EXTRAÇÃO DOS DADOS

1- INFORMAÇÕES GERAIS

NOME DO

REVISOR: _____

DATA DA EXTRAÇÃO: ____/____/____

DETALHES DA BIBLIOGRAFIA

TÍTULO DO ESTUDO:

AUTOR:

REVISTA:

DETALHES DA PUBLICAÇÃO

NOTAS

2- CARACTERÍSTICAS DO ESTUDO

OBJETIVOS	
GERAL	ESPECÍFICOS

CRITÉRIOS DE ELIGIBILIDADE	
Inclusão	Exclusão

CARACTERÍSTICAS DA POPULAÇÃO

Severidade da doença:

Idade:

Sexo:

COLETA DE DADOS

Local de coleta:

Dados Prospectivos (houve planejamento) ou retrospectivos:

MENSURAÇÃO

Medida clínica avaliada:

Instrumento avaliado e sua descrição técnica (como medir, dados contínuos, ordinais ou nominais, pontuação, ponto de corte etc.):

Examinadores (numero de examinadores, treinamento previo e experiência)

Utilizou uma referencia padrão?Qual?

ANALISE ESTATISTICA

Modelo teorico: IRT OU CCT

Pacote estatístico utilizado:

Testes realizados:

RESULTADOS

Participantes

Início e fim do recrutamento:

Características clinicas e demográficas da população recrutada:

Diagrama de fluxo (pacientes recrutados, incluídos e avaliados)

Teste

Intervalo entre as medidas

Houve intervenção entre as medidas?

Reportou distribuição da severidade da doença:

Resultados encontrados