

TESTEX: UMA FERRAMENTA PARA AVALIAÇÃO DA QUALIDADE METODOLÓGICA E QUALIDADE DE RELATO EM ESTUDOS DE TREINAMENTO COM EXERCÍCIO FÍSICO

TESTEX: A TOOL FOR ASSESSING METHODOLOGICAL QUALITY AND REPORTING QUALITY IN PHYSICAL EXERCISE TRAINING STUDIES

TESTEX: UNA HERRAMIENTA PARA EVALUAR LA CALIDAD METODOLÓGICA Y LA CALIDAD DE LOS INFORMES EN LOS ESTUDIOS DE ENTRENAMIENTO CON EJERCICIO FÍSICO

Camilo Luis Monteiro Lourenço ¹
Ciro Oliveira Queiroz ²

Manuscrito recebido em: 24 de fevereiro de 2023.

Aprovado em: 03 de dezembro de 2023.

Publicado em: 19 de fevereiro de 2024.

Resumo

A avaliação da qualidade dos estudos indica credibilidade científica e ajuda a fazer uma classificação preliminar dos resultados, garantindo que os estudos abordem as questões e a particularidade das evidências da melhor forma. Diversas ferramentas foram propostas na literatura para avaliar ensaios clínicos. Entretanto, nem todos os critérios são aplicáveis para estudos de treinamento com exercício físico. Nesse sentido, o objetivo deste estudo é apresentar uma versão em português da escala *Tool for the assessment of Study quality and reporting in EXercise* (TESTEX), uma escala específica na área de ciências do exercício. Neste estudo, são apresentados o processo de criação e os critérios comuns da qualidade metodológica e de relato que compõem a escala. A TESTEX foi projetada por especialistas da área exclusivamente para avaliação da qualidade e do relatório de estudos de treinamento com exercício. Essa é uma escala de 15 pontos, que se dividem em dois blocos: 5 pontos para qualidade metodológica do estudo e 10 pontos para qualidade do relatório. Alguns desses critérios são: análise por intenção de tratar, monitoramento das atividades no(s) grupo(s) controle, intensidade relativa do exercício mantida constante, características do volume do exercício e gasto energético etc. A apresentação da escala TESTEX, na versão em português, deverá servir tanto para pesquisadores, na realização de ensaios clínicos e revisões sistemáticas, quanto para profissionais na área de ciências do exercício, como um suporte nas avaliações dos estudos em sua prática clínica.

¹ Doutor em Educação Física pela Universidade Federal de Santa Catarina. Docente no Programa de Pós-Graduação em Educação Física da Universidade Estadual do Sudoeste da Bahia.

ORCID: <https://orcid.org/0000-0003-0231-1702> Contato: camilo_lourenco@outlook.com

² Doutor em Medicina e Saúde Humana pela Escola Bahiana de Medicina e Saúde Pública. Docente no Programa de Pós-Graduação em Educação Física da Universidade Estadual do Sudoeste da Bahia.

ORCID: <https://orcid.org/0000-0002-3596-6062> Contato: ciro.queiroz@uesb.edu.br

Palavras-chave: Exercício físico; Revisão sistemática; Lista de checagem; Prática clínica baseada em evidência.

Abstract

Assessment of the quality of studies indicates scientific credibility and helps to make a preliminary classification of results, ensuring that studies address the issues and the particularity of the evidence in the best way. Several tools have been proposed in the literature to evaluate clinical trials. However, not all criteria are applicable to physical exercise training studies. In this sense, the objective of this study is to present a Portuguese version of the Tool for the assessment of Study quality and reporting in EXercise (TESTEX) scale, a specific scale in the area of exercise science. In this study, the creation process and common criteria for methodological and reporting quality that make up the scale are presented. TESTEX was designed by experts in the field exclusively for evaluating the quality and reporting of exercise training studies. This is a 15-point scale, which is divided into two blocks: 5 points for the methodological quality of the study and 10 points for the quality of the report. Some of these criteria are: analysis by intention to treat, monitoring of activities in the control group(s), relative intensity of exercise kept constant, characteristics of exercise volume and energy expenditure, etc. The presentation of the TESTEX scale, in the Portuguese version, should serve both researchers, when carrying out clinical trials and systematic reviews, and professionals in the area of exercise science, as support in evaluating studies in their clinical practice.

Keywords: Physical exercise; Systematic review; Checklist; Evidence-based clinical practice.

Resumen

La evaluación de calidad de los estudios garantiza la credibilidad científica y facilita una clasificación preliminar de los resultados, garantizando un abordaje óptimo de las preguntas y particularidades de las pruebas. Existen diversas herramientas para evaluar ensayos clínicos, pero no todos los criterios se aplican a los estudios de entrenamiento físico. En este sentido, el objetivo de este estudio es presentar una versión portuguesa de la escala *Tool for the assessment of Study quality and reporting in EXercise* (TESTEX), escala específica en el área de las ciencias del ejercicio. En este estudio se presenta el proceso de creación y los criterios comunes de calidad metodológica y de presentación de informes que componen la escala. TESTEX fue diseñado por expertos en el campo exclusivamente para evaluar la calidad y la presentación de informes de estudios de entrenamiento físico. Se trata de una escala de 15 puntos, que se divide en dos bloques: 5 puntos por la calidad metodológica del estudio y 10 puntos por la calidad del informe. Algunos de estos criterios son: análisis por intención de tratar, seguimiento de las actividades en el grupo(s) de control, intensidad relativa del ejercicio mantenida constante, características del volumen de ejercicio y gasto energético, etc. La presentación de la escala TESTEX, en la versión portuguesa, debe servir tanto a investigadores, en la realización de ensayos clínicos y revisiones sistemáticas, como a profesionales del área de las ciencias del ejercicio, como apoyo en la evaluación de estudios en su práctica clínica.

Palabras clave: Ejercicio físico; Revisión sistemática; Lista de Verificación; Práctica clínica basada en la evidencia.

Introdução

Os questionamentos sobre a qualidade das pesquisas na área da saúde começaram a ser publicados nos anos 80 (Baar; Tannock, 1989; Sonis; Joines, 1994) e, com isso, o volume de críticas foi aumentado com o passar dos anos. Uma estratégia para identificação de relatórios de pesquisas de baixa qualidade e incompletos é a criação e validação de escalas de avaliação que informam aos autores, revisores e leitores sobre quais elementos devem ser incluídos numa publicação científica. Atualmente, a maior parte dos periódicos científicos tornaram obrigatório o uso de escalas de verificação mais apropriadas para os desenhos de estudos (Chan; Heinemann; Roberts, 2014; Dijkers; Millis, 2020).

A ferramenta Consolidated Standards of Reporting Trials (CONSORT) foi uma das primeiras escalas a serem construídas (BEGG, 1996). Essa ferramenta foi desenvolvida para verificação de itens em ensaios clínicos randomizados e sua versão inicial foi baseada em estudos com drogas versus placebo. Mais recentemente, uma versão para ensaios não farmacológicos também foi desenvolvida e atualizada (Boutron et al., 2017). Apesar de a ferramenta CONSORT para ensaios não farmacológicos ser uma tentativa de ampliação do escopo, tal versão ainda parece ser insuficiente para orientação de estudos em determinadas áreas do conhecimento, a exemplo do treinamento com exercício físico.

Há algum tempo a Colaboração Cochrane vem incentivando a utilização da escala JADAD para ensaios clínicos e mais recentemente a escala de risco de viés (ROB) foi desenvolvida com o objetivo de fornecer uma avaliação completa e robusta do nível de viés dos estudos dentro de revisões sistemáticas (Jadad et al., 1996; Whiting et al., 2016). Essas ferramentas, quando aplicadas na área de treinamento com exercício físico, podem não ter especificidade e alguns critérios são redundantes (Smart et al., 2015).

A escala da *Physiotherapy Evidence Database* (PEDro) é uma escala metodológica de avaliação, comumente usada em estudos de exercício físico, e que foi desenvolvida para a área da fisioterapia (Maher et al., 2003). Entretanto, adotar o uso da escala PEDro na

avaliação de estudos de treinamento com exercícios pode incorrer em alguns equívocos, pois alguns critérios dos seus critérios também não apresentam boa aderência para a área de treinamento com exercício. Por exemplo, os critérios de cegamento têm baixíssima aderência à área do treinamento com exercícios, pois tanto o cegamento dos participantes quanto dos investigadores tem baixa viabilidade – o que faz com que as intervenções com exercício tenham menor pontuação na avaliação por essa escala. Critérios como eficácia e risco de uma intervenção com exercício também não são consideradas na escala PEDro, e esses são itens cruciais para estudos testando uma intervenção.

Em estudo recente, Prado, Delevatti e Monteiro (2022) apresentaram um guia básico para a criação e implementação de revisões sistemáticas sobre exercício físico. No estudo, os autores indicaram o uso da *Tool for the assessment of Study quality and reporting in EXercise* (TESTEX) para avaliação da qualidade dos estudos incluídos em revisões sistemáticas na área do exercício, por se tratar de uma ferramenta desenvolvida especificamente para essa área do conhecimento. Apesar da relevância da indicação, Prado, Delevatti e Monteiro (2022) apresentam a ferramenta de maneira superficial, tornando-se necessário uma apresentação mais aprofundada da ferramenta TESTEX. Por exemplo, na TESTEX são incluídos diversos outros critérios que não são abordados nas escalas de ensaios clínicos apresentados nos parágrafos anteriores do presente estudo, alguns exemplos são: reportar retiradas/reações de eventos adversos, comparecimento às sessões, adesão ao programa de exercício e características dos programas de exercícios (Smart et al., 2015).

A escala TESTEX pode ajudar a determinar os caminhos de um atendimento, cobertura de serviço, método de treinamento e assim por diante. Além disso, os tomadores de decisões devem utilizar estudos com padrões consistentes e imparciais em todos os tópicos de um artigo científico, seja ele um ensaio clínico ou uma revisão sistemática. As falhas ou limitações sistemáticas no desenho ou condução do estudo têm grande potencial de influenciar nos resultados. Diante disso, nosso objetivo é apresentar detalhadamente uma escala específica na área de ciências do exercício para avaliar a qualidade e relatórios de ensaios clínicos de treinamento com exercício físico.

Métodos

Trata-se de um ensaio teórico, no qual é apresentada uma versão, em língua portuguesa, da escala TESTEX. Este estudo foi desenvolvido a partir do contato de seus autores com a leitura crítica da escala e a observação de que a referida escala é pouco utilizada no relato dos estudos de exercício físico, sobretudo aqueles relatados em língua portuguesa (Prado; Delevatti; Monteiro, 2022). Desse modo, foi realizada a leitura do artigo original de validação da escala TESTEX (Smart et al., 2015) e formulado, em separado, documentos contendo apontamentos sobre as principais etapas apresentadas no artigo original, a saber: os métodos (composição do grupo de experts, procedimentos de validação e confiabilidade e análises estatísticas usadas para avaliação da confiabilidade) e os resultados (critérios da escala, tabelas e pontuações). A partir de cada documento individual, foram discutidas e compiladas as principais informações em um documento único, o qual é apresentado como resultado do presente artigo.

Os resultados deste estudo são apresentados sob as seguintes seções: a) descrição do processo de criação e de validação da escala TESTEX, b) descrição dos critérios da TESTEX (critérios comuns de qualidade metodológica do estudo e critérios comuns do relatório do estudo). Cabe salientar que a apresentação deste ensaio é uma tradução livre da escala, não tendo sido adotado procedimento de validação (ex.: tradução, retrotradução, adaptação e validação) para a língua portuguesa. Desse modo, destaca-se que a apresentação deste manuscrito não isenta os leitores a realizarem a leitura detalhada e atenta do estudo de validação original da TESTEX (Smart et al., 2015). Apesar da tradução livre, houve esforço para manutenção ao máximo da originalidade e fidedignidade às informações da escala original. Nesse sentido, foram realizadas apenas duas alterações: 1) a tabela de pontuação da escala TESTEX, do presente estudo, mostra apenas os itens pontuáveis, ou seja 12 itens e 15 pontos, ao contrário da tabela da escala original, a qual mostra dois itens, não pontuáveis, referentes ao cegamento de participantes e terapeutas (profissionais) – informações sobre essa modificação são apresentadas no rodapé da tabela; 2) foram atualizados alguns dos exemplos adotados pelos autores da escala original, quando da pontuação dos itens.

Criação e Validação da Escala TESTEX

A escala TESTEX foi criada por um grupo colaborativo, formado por membros com experiência prévia em conduzir metanálises e avaliar a qualidade de estudos. O grupo de especialistas se organizou para desenvolver uma ferramenta específica para avaliar a qualidade de estudos de treinamento com exercícios. Inicialmente, a ferramenta TESTEX foi baseada na escala PEDro e teve como objetivo melhorar a precisão das informações e a sua tradução para a prática clínica. Os membros do grupo identificaram as dificuldades encontradas ao usarem a escala PEDro e propuseram novos critérios para incluir na nova ferramenta. Após uma série de reuniões para avaliarem os critérios, um consenso foi alcançado e a versão final da nova escala foi elaborada, após várias edições do documento. Por fim, um estudo de confiabilidade foi realizado para validar a nova ferramenta. No estudo original da escala TESTEX são apresentados os processos de validação, envolvendo a seleção e avaliação da qualidade de estudos de treinamento com exercícios físicos e apresentação dos resultados da confiabilidade da escala (Smart et al., 2015). Em síntese, relata-se que a escala apresentou resultados de concordância moderada e quase perfeita entre avaliadores. Adicionalmente, as categorias “grupos similares na linha de base” (Critério 4) e “desistências abaixo de 15%” tiveram valores de concordância moderada. Oito dos 15 critérios da escala apresentaram nível de concordância quase perfeita e outros cinco critérios foram classificadas como tendo nível de concordância substancial. Por fim, recomenda-se que os leitores consultem o artigo original para mais detalhes dos resultados da avaliação de concordância da escala TESTEX (Smart et al., 2015).

Em síntese, o resultado desse processo é uma escala de 15 pontos, sendo cinco pontos para avaliação da qualidade do estudo e 10 pontos para avaliação da qualidade de relato. A escala TESTEX é apresentada na Tabela 1 a seguir.

Tabela 1 – escala TESTEX*

Critério	Pontuação	Pontuação**
Qualidade do estudo		
1. Critérios de elegibilidade especificados	1 ponto – se os critérios de elegibilidade forem claramente declarados e cumpridos	<input type="checkbox"/> 1 ponto
2. Randomização especificada	1 ponto - se os métodos forem descritos e forem realmente aleatórios, por exemplo lançamento de moeda, sequência de números gerados aleatoriamente	<input type="checkbox"/> 1 ponto
3. Ocultação da alocação	1 ponto - se a alocação do grupo foi ocultada de pacientes elegíveis para inclusão no estudo (por exemplo, o consentimento deve ser dado antes da randomização)	<input type="checkbox"/> 1 ponto
4. Similaridade dos grupos na linha de base	1 ponto - se os dados da linha de base forem separados por alocação de grupo, apresentados e nenhuma diferença for aparente	<input type="checkbox"/> 1 ponto
5. Cegamento do avaliador (para ao menos um desfecho-chave)	1 ponto - se for declarado inequivocamente que o avaliador de pelo menos 1 medida de resultado primário estava cego para a alocação do grupo	<input type="checkbox"/> 1 ponto
Relatório do estudo		
6. Medidas de desfecho em 85% dos participantes	Sem pontuação – se o abandono for > 15%	
	1 ponto – se a aderência for > 85%	<input type="checkbox"/> 1 ponto
	1 ponto - se os eventos adversos foram relatados	<input type="checkbox"/> 1 ponto
	1 ponto – se a frequência (comparecimento) foi relatada	<input type="checkbox"/> 1 ponto
	Total possível = 3 pontos	
7. Análise por intenção de tratar	1 ponto - se a análise de intenção de tratar foi realizada nos desfechos de interesse	<input type="checkbox"/> 1 ponto
8. Comparações estatísticas entre grupos	1 ponto - se as comparações estatísticas entre grupos foram relatadas para a medida de desfecho primário de interesse	<input type="checkbox"/> 1 ponto
	1 ponto - se as comparações estatísticas entre grupos foram relatadas para ao menos uma medida de desfecho secundário	<input type="checkbox"/> 1 ponto
	Total possível = 2 pontos	

Critério	Pontuação	Pontuação**
Qualidade do estudo		
9. Medidas pontuais e medidas de variabilidade para todas as medidas de desfechos relatadas	1 ponto - se todos os desfechos foram relatados com estimativas pontuais	<input type="checkbox"/> 1 ponto
10. Monitoramento de atividades nos grupos de controle	1 ponto – se os pacientes de controle foram solicitados a relatarem seus níveis de atividade física e os dados foram apresentados	<input type="checkbox"/> 1 ponto
11. Intensidade relativa do exercício mantida constante	1 Ponto – se a carga do exercício for ajustada para manter a intensidade relativa constante	<input type="checkbox"/> 1 ponto
12. Características do volume de exercício e gasto energético	1 ponto – se o volume do exercício e o gasto energético puderem ser calculados	<input type="checkbox"/> 1 ponto
Total de pontos possíveis = 15 pontos	Total de pontos alcançado (soma dos itens marcados com 1 ponto) =	Escolher um item.

Adaptado de Smart et al. (2015).

* Esta ferramenta foi adaptada para mostrar apenas os itens para os quais há atribuição de pontuação na escala original.

** Deve-se marcar a caixa de seleção apenas se o critério for satisfeito, caso contrário, deve-se deixá-la em branco.

Critérios TESTEX

A seguir, são apresentados detalhes descritivos de cada seção e critérios da escala TESTEX, com alguns exemplos e justificativas. Na primeira seção, são apresentados os critérios de avaliação da qualidade metodológica do estudo: a) especificação dos critérios de elegibilidade; b) processo de randomização; c) ocultação da alocação; d) similaridade dos grupos na linha de base; e) cegamento de todos os participantes e terapeutas (profissionais) (critérios não pontuáveis); f) cegamento do avaliador. Na segunda seção, são apresentados os critérios de avaliação do relatório do estudo: a) aderência ao exercício/eventos adversos/frequência nos exercícios; b) análises por intenção de tratar; c) análises estatísticas; d) apresentação dos resultados dos desfechos; e) contaminação de grupos; f) ajuste da intensidade do exercício; g) informações de volume/gasto energético do exercício.

- Critérios Comuns de Qualidade Metodológica do Estudo

Critérios de elegibilidade especificados – Critério 1 da TESTEX

No critério 1 da escala TESTEX, os autores abordam sobre a especificação dos critérios de elegibilidade do estudo. Eles argumentam que os estudos de treinamento com exercícios físicos precisam estabelecer critérios específicos para determinar a existência de uma condição. No estudo original da escala TESTEX, os autores usam a hipertensão arterial como exemplo para especificar um critério de elegibilidade. Para esse exemplo, cabe frisar que o critério atual descreve que a hipertensão arterial é caracterizada por elevação persistente da pressão arterial sistólica maior ou superior a 140 mmHg e/ou diastólica superior a 90 mmHg (BARROSO et al., 2021). No entanto, é comum que alguns participantes não atendam a esses critérios, uma vez que os valores de média e desvio padrão para um dos grupos de estudo estejam fora do intervalo especificado; desse modo, nem todos os participantes atenderiam ao critério de elegibilidade. Esses critérios de elegibilidade afetam a validade externa dos estudos, mas não sua validade interna ou

estatística. A escala PEDro sugere que esses critérios devem ser especificados, mas não os utiliza para calcular a pontuação, uma vez que a PEDro foi projetada para evitar a avaliação da validade externa. A precisão dos critérios de elegibilidade na ciência do exercício é maior, segundo os autores da TESTEX, pois depende menos do julgamento clínico (ex.: um exame físico) e são baseados em valores numéricos de testes diagnósticos objetivos. Além disso, os autores da escala TESTEX argumentam que, no contexto prático, cientistas do exercício físico não têm papel de realizarem ao diagnóstico, o que nem sempre é o caso dos fisioterapeutas. Por esses motivos, os estudos que relatam e atendem aos critérios de elegibilidade incluídos são avaliados positivamente. Por esses motivos, atribui-se **1 ponto** para estudos que relatam e atendem aos critérios de elegibilidade incluídos.

Randomização especificada – Critério 2 da TESTEX

O segundo critério da qualidade metodológica dos estudos, na escala TESTEX, refere-se à especificação detalhada do processo de randomização (alocação aleatória) dos participantes aos grupos do estudo. Em conformidade com o segundo critério da PEDro, os autores da escala TESTEX argumentam que consideram insuficiente apenas a menção de que um estudo utilizou uma alocação aleatória sem a especificação do método preciso de randomização (tais como números aleatórios gerados por computador, lançamento de moeda e lançamento de dados). Acrescenta-se, aqui, a possibilidade de uso de programas (alguns gratuitos) e ferramentas *online* para randomização (BLAND, 2020; MORICE, 2012). Procedimentos de alocação de *quasi-randomização*¹ não são considerados adequados para atender a este critério. A alocação aleatória garante que, dentro dos limites impostos pelo acaso, os grupos de tratamento e controle sejam comparáveis. Isso é de particular importância em estudos de treinamento com exercício físico, pois, em geral, o tamanho amostral nesse campo de investigação é reduzido (muitas vezes não superior a 50 participantes) e a taxa de abandono dos estudos gira em torno de 15%, argumentam os

¹ Acredita-se ser importante elucidar que a *quasi-randomização* se trata da alocação de participantes de ensaios clínicos a um grupo de intervenção ou controle usando métodos que não são verdadeiramente aleatórios, mas ainda destinados a produzir grupos semelhantes. Por exemplo, alocação com base na data de nascimento, no número do prontuário ou na ordem em que os participantes foram recrutados (EUPATI, 2023).

autores da escala TESTEX. Ao se analisarem dados recentes da literatura, verifica-se que o percentual de abandono em estudos de exercício físico varia de 12% a 33% (LOURENÇO, 2022). Por esses motivos, atribui-se **1 ponto** para estudos que relatam e especificam o método de randomização utilizado.

Ocultação da alocação – Critério 3 da TESTEX

O terceiro critério da qualidade metodológica dos estudos, na escala TESTEX, refere-se à ocultação da alocação. Esse critério é complementar ao critério de elegibilidade e ao critério de randomização, uma vez que “a ocultação de alocação adequada garante a implementação estrita de uma sequência de alocação aleatória sem conhecimento prévio das atribuições de tratamento”, adicionando-se o fato de que a “ocultação de alocação refere-se à técnica usada para implementar a sequência, não para gerá-la” (SCHULZ; GRIMES, 2002); ou seja, considera-se que um estudo forneceu a ocultação da alocação se investigadores e pacientes (i.e. participantes) não tinham conhecimento prévio para qual grupo do estudo ele foi alocado, até que o momento do seu consentimento. Nesse sentido, atribui-se **1 ponto** para estudos que relatam e especificam a ocultação da alocação.

Similaridade dos grupos na linha de base – Critério 4 da TESTEX

No critério 4 da escala TESTEX, os autores argumentam que os estudos de intervenções de treinamento com exercício físico devem apresentar, no mínimo, uma medida da gravidade da condição em questão e uma medida de desfecho-chave (diferente) na linha de base. Segundo os autores, é importante que o avaliador, usando a escala TESTEX, tenha a convicção de que não se esperam diferenças significativas nos resultados dos grupos, com base somente nas diferenças nas variáveis de prognóstico na linha de base do estudo. Discrepâncias na linha de base entre os grupos podem indicar a existência de procedimentos de randomização inadequados. Nesse sentido, atribui-se **1 ponto**, se os dados da linha de base são apresentados por grupo de alocação e não há diferenças significativas entre os grupos de estudo nos desfechos principais de interesse e na medida da gravidade da condição a ser tratada.

Cegamento de todos os participantes e todos os terapeutas – critérios não pontuados

Ambos os itens não são pontuados na escala TESTEX, diferentemente da escala PEDro. O cegamento é considerado como uma medida para garantir que os participantes sejam incapazes de determinar se receberam ou não o tratamento. Observa-se que em alguns estudos de fisioterapia, é possível fornecer intervenções "simuladas" (i.e., *sham*) que podem ser percebidas pelos participantes como semelhantes às intervenções reais. No entanto, em estudos de treinamento com exercícios, é difícil alcançar o cegamento dos participantes, com poucas exceções; como em estudos que comparam exercícios físicos com “pedalar com carga de trabalho zerada” ou “estimulação elétrica funcional”. Intervenções nas quais os participantes podem ser submetidos a estímulos que estão abaixo do limiar para fornecer adaptação fisiológica. Entretanto, os autores da TESTEX argumentam que os participantes geralmente têm ciência para qual dos grupos ele foi randomizado, tomando por base as informações fornecidas pelos investigadores no momento de recolher do seu consentimento; ou seja, é quase impossível garantir o verdadeiro cegamento dos participantes em estudos de treinamento com exercícios físicos.

Seguindo a mesma lógica apresentada para o item acima, os autores da escala TESTEX argumentam que cegar o terapeuta que administra intervenções de treinamento com exercícios físicos é uma prática inviável. Eles expõem que, nessa área de investigação, é importante que o terapeuta esteja ciente dos possíveis efeitos do tratamento nos participantes, sobretudo naqueles com doenças crônicas. Além disso, segundo os autores, o papel do terapeuta também estaria ligado a proporcionar motivação para que os participantes continuem se exercitando, o que vincula o profissional ao tratamento. Por essas razões, os autores afirmam que este critério também é inadequado como critério de pontuação na escala TESTEX – diferentemente da escala PEDro.

Cegamento do avaliador (para ao menos um desfecho-chave) - Critério 5 da TESTEX

Neste item 5 da escala TESTEX, os autores abordam o cegamento do avaliador do(s) desfecho(s) do estudo. Diferentemente dos dois critérios anteriores, cegamento dos participantes e cegamento dos terapeutas, é apropriado que o(s) avaliador(es) do(s) desfecho(s) – pessoas envolvidas as avaliações/medições de dados do(s) desfecho(s) – sejam cegados para a alocação da intervenção do participante. Nesse sentido, argumenta-se que a alocação cega do(s) avaliador(es) garante que o aparente efeito (ou ausência de efeito) do tratamento não seja devido aos vieses dos avaliadores que possam interferir nas medidas de desfecho de interesse do estudo. Neste tópico da escala TESTEX, há dois critérios possíveis para a atribuição de **1 ponto**: a) quando, de forma clara e inequívoca, os autores do estudo declaram que o avaliador da medida de desfecho primário foi cegado para a alocação ou b) quando os métodos de medição do estudo são completamente automatizados (ex.: análise de sangue), removendo o potencial viés humano.

- Critérios Comuns do Relatório do Estudo

Medidas de desfecho avaliadas em 85% dos participantes – Critério 6 da TESTEX

Neste critério do relatório do estudo, os autores da escala TESTEX explicam a possibilidade de haver a atribuição de até **3 pontos**, o que é mostrado a seguir. Neste critério da escala TESTEX, os autores argumentam que o volume e a duração de treinamento com exercícios físicos para promover adaptações é variável de acordo com as medidas de desfechos, prescrições de exercícios e características dos participantes. De fato, reafirma-se a declaração dos autores ao ser apontado que a literatura científica confirma que a devida prescrição e a correta manipulação dos componentes do princípio FITT (do inglês, *Frequency, Intensity, Time e Type*) deve assegurar tais adaptações (ACSM, 2010). Os autores argumentam que apesar disso, é amplamente aceito que mudanças significativas em determinadas medidas (ex.: aptidão cardiorrespiratória, função cardíaca, lipídios, controle glicêmico) não são imediatas e que, pelo menos, um mês é necessário para detectar as mudanças. Nesse sentido, considerando as durações de intervenções típicas do treinamento com exercícios físicos, a proporção de participantes que completam

o estudo, em geral, é inferior a 85% e nem todos os que completam o estudo participam de todas as sessões de exercícios (a escala TESTEX aborda o abandono/desistência no critério 7). Diante disso, torna-se relevante a distinção entre aderência ao exercício e frequência ao exercício (do inglês, *attendance*). Na escala TESTEX, aderência ao exercício é definida como o número de desistências e conclusões nos grupos de intervenção e controle do estudo. Por sua vez, a frequência ao exercício é definida como a porcentagem de sessões-alvo concluídas por cada indivíduo que completou o estudo. Como já mencionado no Critério 2 da escala TESTEX, é comum que mais de 15% dos indivíduos desistam de um estudo de treinamento físico durante o período de estudo estabelecido (LOURENÇO, 2022). Além disso, a frequência ao exercício é inferior a 85% em alguns indivíduos que não desistem do estudo. Assim, é desejável que o número de desistências em cada grupo de alocação, bem como a média e o desvio padrão da porcentagem de comparecimento (frequência) às sessões de exercícios sejam relatados para os grupos de intervenção. Os autores da escala TESTEX fazem uma ressalva a esse critério, a qual se dá quando os investigadores do estudo estabelecem uma “meta” a ser cumprida como, por exemplo, uma quantidade de quilocalorias por semana de gasto energético no exercício sendo alcançada com sucesso – neste caso a pontuação para a frequência ao exercício deveria ser computada, independentemente do número de sessões atendidas.

Sobre as pontuações neste critério: **1 ponto** é fornecido se os estudos relatam a aderência ao treinamento com exercício físico em no mínimo 85% - o ponto não é fornecido se a aderência é inferior a 85%; **1 ponto** é fornecido se os estudos relatam os eventos adversos (relato de morte, hospitalizações etc.). A atribuição deste ponto é justificada, uma vez que a avaliação da aderência à terapia de exercícios físicos é geralmente realizada considerando o equilíbrio entre os benefícios esperados e o risco de ocorrência de eventos adversos; **1 ponto** é atribuído se os autores do estudo relatam a frequência ao exercício (i.e., porcentagem de sessões-alvo concluídas por cada indivíduo que completou o estudo) for relatada para os grupos de exercício físico.

Análise de intenção de tratar – Critério 7 da TESTEX

Neste critério da escala TESTEX, é avaliado se todos os pacientes para os quais as medidas de desfecho estavam disponíveis foram submetidos ao tratamento ou condição de controle conforme alocado, ou, de outro modo, se os dados de pelo menos um desfecho-chave de interesse foram analisados por meio da análise intenção de tratar (ITT). Propõe-se que, sempre que possível, a análise de ITT seja realizada. Os autores da escala TESTEX informam que a ITT pode ser realizada utilizando-se o último valor obtido para cada medida de desfecho como o valor após a intervenção ou o valor da linha de base como o valor do pós-intervenção, no caso em que o participante tenha desistido. Conforme indicam os autores, a inclusão deste critério visa determinar se certos dados demográficos do participante, estado clínico, medicamentos, entre outros, predisõem os participantes a desistirem. Por fim, o propósito é auxiliar futuras pesquisas e serviços clínicos a melhorarem a aderência ao exercício físico, identificando fatores predisponentes que levam à desistência do estudo. Nesse aspecto, reafirmam-se as colocações dos autores da escala TESTEX, uma vez que a literatura científica indica que certas características dos participantes têm capacidade preditiva para desistência em estudos de treinamento com exercício físico (SCHMIDT et al., 2000; VANCAMPFORT et al., 2021). A atribuição de **1 ponto** é concedida se uma análise de ITT for realizada. Mais uma vez, os autores da escala TESTEX informam que a análise de ITT é conduzida substituindo a última medição (que pode ser a medição de linha de base) para aqueles que não concluíram o estudo e esses dados são incluídos nas análises dos que concluíram.

Comparações estatísticas entre grupos relatadas – Critério 8 da TESTEX

Neste critério, avalia-se a presença do relato das comparações entre os grupos do estudo para a medida de desfecho primário e ao menos uma medida de desfecho secundário. Em caso de positivo, havendo relato de ambas as medidas, o item é pontuado. Se não forem relatados todos os resultados, será considerado um relato seletivo.

Os autores argumentam que o principal interesse é determinar se os grupos intervenção apresentam melhorias, no entanto, a comparação com um grupo de controle fornece duas informações importantes. Em primeiro lugar, indica se houve ocorrência de

regressão à média, ou seja, os participantes melhoram sem intervenção, talvez devido a melhorias na medicação, se esse for o caso, ou outras razões não relacionadas à intervenção. Em segundo lugar, pode ser que alguns pacientes que deveriam ter recebido a intervenção não o fizeram e alguns que não deveriam ter recebido, tenham participado. De fato, legitima-se as afirmações dos autores, uma vez que o efeito de o participante “estar sendo observado” (i.e. efeito Hawthorne²) é algo com o qual idealizadores de estudos de treinamento com exercícios físicos devem considerar ao planejarem seus estudos (TAI; ILIFFE, 2000). Por fim, é esclarecido que embora essas questões sejam avaliadas em outros critérios, a comparação entre os grupos alerta para a possibilidade de regressão à média ou alocação inadequada de tratamento, e se a diferença entre os grupos é maior do que pode ser atribuída ao acaso. Neste critério, há possibilidade de atribuição de até **2 pontos**: a) **1 ponto** é atribuído para o relato do resultado principal de interesse e b) outro **1 ponto** é atribuído para o relato de pelo menos uma medida de desfecho secundário.

Medidas pontuais e medidas de variabilidade para todas as medidas de desfechos relatadas – Critério 9 da TESTEX

As avaliações precisas (geralmente valores p) do impacto do tratamento fornecem apenas informações restritas sobre o desempenho dos grupos de tratamento e controle. Uma abordagem mais ampla seria também apresentar medidas da variabilidade. Nesse sentido, os autores da escala TESTEX recomendam que isso seja ampliado para todas as medidas dos desfechos descritos, a fim de evitar a apresentação de relato seletivo dos desfechos. Atribui-se **1 ponto** para o critério atendido.

Monitoramento de atividades nos grupos de controle – Critério 10 da TESTEX

² De maneira simples, o efeito Hawthorne é o fenômeno em que os indivíduos em um experimento ou em um estudo mostram melhorias em seus desempenhos ou comportamentos simplesmente porque estão sendo observados ou porque sabem que estão participando de um estudo (MCCAMBRIDGE; WITTON; ELBOURNE, 2014).

O décimo critério da escala TESTEX aborda o viés de contaminação. Acredita-se ser pertinente uma breve explicação sobre esse tipo de viés, o qual ocorre quando os membros do(s) grupo(s) controle inadvertidamente recebem o tratamento ou são expostos à intervenção, potencialmente minimizando a diferença nos resultados entre os dois grupos (MAGILL et al., 2019). Os autores da escala TESTEX utilizam o estudo HF-ACTION, um estudo de treinamento com exercícios físicos em casos de insuficiência cardíaca, esclarecendo que a diferença observada entre os grupos após a intervenção foi menor do que o esperado devido ao fato de que aproximadamente 30% dos pacientes designados para o grupo controle (sedentário) realizaram algum tipo de treinamento com exercício físico (O'CONNOR et al., 2009). De fato, evidência mais recente atesta a presença potencial da contaminação em estudo com atividade física, inclusive na presença de grupo controle ativo (EHLERS et al., 2016).

Diante disso, os autores da escala TESTEX recomendam que futuros estudos robustos quantifiquem esse fenômeno, estabelecendo medidas que visem a avaliar os níveis de atividade dos participantes designados ao grupo controle (incluindo grupo(s) controle ativo(s), *inclusão dos autores*³), com o objetivo de evitar contaminação do grupo de controle e influência sobre o resultado. As medidas para mitigar isso podem incluir a utilização de um diário de atividade, ou ainda dispositivos avançados, como acelerômetros ou monitores de frequência cardíaca, para avaliar o potencial de contaminação do(s) grupo(s) controle e monitorar o comportamento de atividade física dos participantes. Nesse sentido, concede-se **1 ponto** a qualquer relato que inclua informações sobre o monitoramento da atividade dos participantes do grupo controle sedentário (e/ou grupo(s) controle ativo(s)).

Intensidade relativa do exercício mantida constante – Critério 11 da TESTEX

³ Diante de evidência mais recente, sugerindo que potencial contaminação em estudos de treinamento com exercício físico/atividade física em grupo(s) controle ativo(s), acredita-se que seja pertinente a inclusão dessa informação no referido critério.

No critério 11 da escala TESTEX, os autores abordam sobre as adequações/ajustes na intensidade relativa do exercício físico, à medida que o programa de treinamento avança no tempo e os participantes têm ganhos adaptativos. Nesse sentido, os autores afirmam que a evolução dos participantes quanto ao treinamento com exercícios físicos é caracterizada por uma adaptação da sua capacidade física de trabalho. Quando a carga de trabalho é mantida inalterada durante o tempo (semanas, meses) do programa de treinamento, a intensidade relativa do exercício tende a diminuir ao longo do tempo em decorrência da melhoria na capacidade física dos participantes. Por essa razão, os autores apontam que avaliações periódicas da capacidade de trabalho para o exercício, utilizando métodos como taxa de esforço percebido, frequência cardíaca de reserva, percentual cardíaco máximo ou pico de VO_2 , são consideradas necessárias a fim de regular a carga de exercício. Qualquer desvio na intensidade do exercício pode levar a uma estabilização ou até mesmo a uma atenuação das adaptações (em 3-4 semanas, segundo os autores), tornando difícil a avaliação das conclusões sobre a magnitude da adaptação em relação à intensidade prescrita. Diante disso, **1 ponto** é atribuído para o ajuste periódico visando a manutenção da intensidade relativa do exercício constante.

Características do volume de exercício e gasto energético – Critério 12 da TESTEX

O critério 12 da escala TESTEX refere-se ao adequado relato de características fundamentais para os programas de treinamento com exercícios físicos: tipo, duração da sessão e do programa, frequência das sessões, intensidade do treinamento e modalidade. De fato, já foi informado no Critério 6 da TESTEX que os componentes do princípio FITT são cruciais para a correta prescrição do exercício físico, a fim de assegurar as adaptações do organismo (ACSM, 2010). Os autores da escala TESTEX esclarecem que apesar de haver ampla variação dos marcadores clínicos⁴ aos diferentes tipos e doses de exercício, é reconhecida e frequente a existência de uma relação dose-resposta entre o exercício e a melhora clínica. Assim, os relatórios devem ser adequados, com clareza, para possibilitar

⁴ De acordo com o dicionário da *American Psychological Association* (2023), um marcador clínico é um sinal observável indicativo de desordem.

ao leitor quantificar, ao invés de apenas estimar, o volume de exercício em termos de energia gasta durante o programa. Para alcançar a clareza adequada, as características do exercício (tipo, duração da sessão, frequência das sessões, intensidade e modalidade) devem ser claramente descritas de modo que cada prescrição de exercício possa ser avaliada quanto ao alcance dos limites terapêuticos. Desse modo, atribui-se **1 ponto** se todas as características do exercício são relatadas adequadamente, novamente: intensidade, frequência, modo, duração da sessão e duração da intervenção, para que o volume do exercício e o gasto energético possam ser determinados.

Discussão

Este estudo teve por objetivo apresentar uma versão, em língua portuguesa, da escala TESTEX. Essa escala tem por finalidade a avaliação da qualidade metodológica e de relato dos estudos na área de ciências do exercício. A TESTEX foi construída a partir de outras ferramentas disponíveis na literatura para servir como um suporte para uso por pesquisadores, profissionais e outros especialistas em exercício físico para avaliarem a qualidade metodológica de ensaios clínicos randomizados que contenham treinamento com exercícios físicos (SMART et al., 2015).

A TESTEX é composta por 12 itens pontuáveis, podendo chegar ao somatório total de 15 pontos. Os 12 itens são divididos em critérios comuns de qualidade metodológica do estudo (5 itens: 5 pontos) e critérios comuns do relatório do estudo (10 itens: 10 pontos). Ao contrário de outras escalas comumente adotadas para avaliação da qualidade metodológica de estudos de exercício físico, a exemplo da PEDro (MAHER et al., 2003) e JADAD (JADAD et al., 1996), a TESTEX apresenta sensibilidade às particularidades dessa área de pesquisa. Por exemplo, a escala PEDro é composta por 11 itens e pode chegar à pontuação máxima de 10 pontos. Além disso, a TESTEX considera critérios de elegibilidade especificados como critério de pontuação, o que não acontece na PEDro. Cabe salientar que critérios de elegibilidade são passíveis de afetar a validade externa de um estudo (potencial de generalização dos resultados) e, por isso, são pontuadas na TESTEX.

Em contraste com outras ferramentas, alguns itens importantes para a área da ciência do exercício físico são considerados e pontuados na TESTEX. Por exemplo, a escala TESTEX permite uma análise ampliada da avaliação dos desfechos, abrangendo informações sobre a desistência dos participantes dos estudos, a participação nas sessões, análises segundo o princípio de ITT, relato sobre a contaminação de grupos, ajustes periódicos da carga de exercício para manter a intensidade constante, registros de eventos adversos e imprecisões de medição, além da descrição das características do exercício que possibilitam o cálculo do volume do exercício e o gasto energético.

Revisões sistemáticas recentes têm adotado a escala TESTEX na avaliação da qualidade metodológica. Na revisão sistemática de Edwards et al. (2022), os estudos de treinamento com exercícios físicos para pessoas com esclerose múltipla alcançaram mediana de 10 pontos, pontuação superior àquela identificada em outra revisão de exercício físico para pessoas com a mesma condição clínica (DENNETT et al., 2020). A revisão sistemática conduzida por Valenzuela et al. (2023) mostrou que estudos de exercício físico para idosos em acolhimento residencial obtiveram mediana de nove pontos na TESTEX. Por fim, a revisão sistemática conduzida por Perrier-Melo et al. (2020) mostrou que em estudos comparando efeitos agudos do exercício intervalado versus contínuo a mediana de pontuação na escala TESTEX foi 10 pontos. Em alguns desses estudos, os critérios de menor aderência à pontuação na TESTEX incluíram: a contaminação no grupo controle, ausência de detalhes das características das intervenções (i.e., detalhes de volume e intensidade) ou ajuste gradual para manutenção constante da carga de treinamento, ausência de análise segundo princípio ITT, falta de explicação sobre a ocultação da alocação, cegamento de avaliador e ausência de relato de eventos adversos (DENNETT et al., 2020; EDWARDS et al., 2022; PERRIER-MELO et al., 2020; VALENZUELA et al., 2023). Isso é um indicativo de que o uso da escala TESTEX aponta para que estudos na área da ciência do exercício físico falhem em critérios cruciais da qualidade metodológica. Nesse sentido, pesquisadores da área necessitam dedicar maior atenção para relato mais transparente e detalhado, mesmo quando na ausência de determinadas situações indicadas na escala (ex.: ausência de eventos adversos).

Acredita-se que a apresentação da escala TESTEX, na sua versão em língua portuguesa, poderá facilitar a discussão sobre a qualidade metodológica dos estudos dentro da ciência do exercício físico. A facilidade de acesso e leitura desta versão fomentará a discussão sobre qualidade metodológica em programas de pós-graduação, em nível de mestrado e doutorado, por exemplo. Além disso, espera-se que a versão apresentada no presente estudo possa potencializar o uso da TESTEX, desde a concepção, implementação e avaliação dos estudos na ciência do exercício.

Conclusão

A escala TESTEX é uma ferramenta confiável e específica para as ciências do exercício. Essa escala tem potencial de auxiliar na compreensão dos passos de um estudo de intervenção com exercício físico, fornecendo critérios bem delineados e aplicáveis. A TESTEX aborda deficiências comuns nesse tipo de estudo e que não são incluídos em outras escalas como critérios de avaliação. É possível afirmar que a escala TESTEX serve tanto para pesquisadores, na condução de ensaios clínicos e revisões sistemáticas, quanto para tomadores de decisões e profissionais, como uma ferramenta de suporte nas avaliações dos estudos em sua prática cotidiana.

Referências

AMERICAN COLLEGE OF SPORTS MEDICINE. Chapter 7 - General Principles of Exercise Prescription. Em: THOMPSON, W. R.; GORDON, N. F.; PESCATELLO, L. S. (Eds.). **ACSM's guidelines for exercise testing and prescription**. 8. ed. Philadelphia: American College of Sports Medicine, 2010. p. 380.

AMERICAN PSYCHOLOGICAL ASSOCIATION. **APA Dictionary of Psychology - clinical marker**. Disponível em: <<https://dictionary.apa.org/>>. Acesso em: 30 jan. 2023.

BAAR, J.; TANNOCK, I. Analyzing the same data in two ways: a demonstration model to illustrate the reporting and misreporting of clinical trials. **Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology**, v.7, n.7, p.969–978, 1989.

BARROSO, W. K. S. et al. Diretrizes Brasileiras de Hipertensão Arterial – 2020. **Arquivos Brasileiros de Cardiologia**, v.116, n.3, p.516-658, 2021.

BEGG, C. Improving the Quality of Reporting of Randomized Controlled Trials: The CONSORT Statement. **JAMA**, v.276, n.8, p.637, 1996.

BLAND, M. **Directory of randomisation (randomization) software and services**. Disponível em: <<https://www-users.york.ac.uk/~mb55/guide/randsery.htm>>. Acesso em: 27 jan. 2023.

BOUTRON, I. et al. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. **Annals of Internal Medicine**, v.167, n.1, p.40, 2017.

CHAN, L.; HEINEMANN, A. W.; ROBERTS, J. Elevating the quality of disability and rehabilitation research: mandatory use of the reporting guidelines. **The American Journal of Occupational Therapy: Official Publication of the American Occupational Therapy Association**, v.68, n.2, p.127-129, 2014.

DENNETT, R. et al. Adherence and drop-out in randomized controlled trials of exercise interventions in people with multiple sclerosis: A systematic review and meta-analyses. **Multiple Sclerosis and Related Disorders**, v.43, 2020.

DIJKERS, M. P.; MILLIS, S. R. The Template for Intervention Description and Replication as a Measure of Intervention Reporting Quality: Rasch Analysis. **Archives of Rehabilitation Research and Clinical Translation**, v.2, n.3, p.100055, 2020.

EDWARDS, T. et al. Exercise training improves participation in persons with multiple sclerosis: A systematic review and meta-analysis. **Journal of Sport and Health Science**, v.11, n.3, p.393-402, 2022.

EHLERS, D. K. et al. Contamination by an Active Control Condition in a Randomized Exercise Trial. **PLOS ONE**, v.11, n.10, p.e0164246, 2016.

EUROPEAN PATIENTS' ACADEMY ON THERAPEUTIC INNOVATION. **Quasi-randomised trial**. Disponível em: <<https://toolbox.eupati.eu/glossary/quasi-randomised-trial/>>. Acesso em: 27 jan. 2023.

JADAD, A. R. et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? **Controlled Clinical Trials**, v.17, n.1, p. 1-12, 1996.

LOURENÇO, C. L. M. **Treinamento baseado em exergames sobre parâmetros hemodinâmicos em adultos: análise de efeitos e dropouts**. Tese (Doutorado em Educação Física) – Universidade Federal de Santa Catarina, Florianópolis, 2022.

MAGILL, N. et al. A scoping review of the problems and solutions associated with contamination in trials of complex interventions in mental health. **BMC Medical Research Methodology**, v.19, n.1, p.4, 2019.

MAHER, C. G. et al. Reliability of the PEDro Scale for Rating Quality of Randomized Controlled Trials. **Physical Therapy**, v.83, n.8, p.713-721, 2003.

MCCAMBRIDGE, J.; WITTON, J.; ELBOURNE, D. R. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. **Journal of Clinical Epidemiology**, v.67, n.3, p.267-277, 2014.

MORICE, V. RandoWeb, an online randomization tool for clinical trials. **Computer Methods and Programs in Biomedicine**, v.107, n.2, p.308-314, 2012.

O'CONNOR, C. M. et al. Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. **JAMA**, v.301, n.14, p.1439-1450, 2009.

PERRIER-MELO, R. J. et al. Efeito Agudo do Exercício Intervalado versus Contínuo sobre a Pressão Arterial: Revisão Sistemática e Metanálise. **Arquivos Brasileiros de Cardiologia**, v.115, n.1, p.5-14, 2020.

PRADO, A. K. G.; DELEVATTI, R. S.; MONTEIRO, E. P. Revisões sistemáticas para o campo de pesquisa em exercício físico e saúde: um framework de definições, aplicações e guia básico para elaboração. **Corpoconsciência**, v.26, n.3, p.86-112, 2022.

SCHMIDT, J. A. et al. Attrition in an Exercise Intervention: A Comparison of Early and Later Dropouts. **Journal of the American Geriatrics Society**, v.48, n.8, p.952-960, 1 ago. 2000.

SCHULZ, K. F.; GRIMES, D. A. Allocation concealment in randomised trials: defending against deciphering. **Lancet (London, England)**, v.359, n.9306, p.614-618, 2002.

SMART, N. A. et al. Validation of a new tool for the assessment of study quality and reporting in exercise training studies: TESTEX. **International Journal of Evidence-Based Healthcare**, v.13, n.1, p.9-18, 2015.

SONIS, J.; JOINES, J. The quality of clinical trials published in The Journal of Family Practice, 1974-1991. **The Journal of Family Practice**, v.39, n.3, p.225-235, 1994.

TAI, S. S.; ILIFFE, S. Considerations for the design and analysis of experimental studies in physical activity and exercise promotion: advantages of the randomised controlled trial. **British Journal of Sports Medicine**, v.34, n.3, p.220-224, 2000.

VALENZUELA, P. L. et al. Effects of physical exercise on physical function in older adults in residential care: a systematic review and network meta-analysis of randomised controlled trials. **The Lancet Healthy Longevity**, v.4, n.6, p.e247-e256, 2023.

VANCAMPFORT, D. et al. Dropout from exercise randomized controlled trials among people with anxiety and stress-related disorders: A meta-analysis and meta-regression. **Journal of Affective Disorders**, v.282, p.996-1004, 2021.

WHITING, P. et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. **Journal of Clinical Epidemiology**, v.69, p.225-234, 2016.

WHITING, P. et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. **Journal of Clinical Epidemiology**, v.69, p.225-34, 2016.