

SEBASTIÃO HENRIQUE ASSIS DA SILVA

**ANALISANDO A ROBUSTEZ DAS MEDIDAS DE FORÇA MUSCULAR E
DESEMPENHO FÍSICO EM MULHERES COM MAIS DE 40 ANOS: UM ESTUDO
DE CONFIABILIDADE TESTE-RETESTE E ANÁLISE DO EFEITO DO
TREINAMENTO DE FORÇA COM CARGA AUTOSSELECIONADA SOBRE O
DESEMPENHO FÍSICO, FORÇA E MASSA MUSCULAR.**

UBERABA

2024

UNIVERSIDADE FEDERAL DO TRIÂNGULO MINEIRO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM EDUCAÇÃO FÍSICA

Sebastião Henrique Assis Da Silva

Dissertação apresentada ao Programa de Pós-graduação em Educação Física, área de concentração Educação Física, Esporte e Saúde (Linha de Pesquisa: Aspectos Psicobiológicos do Exercício Físico Relacionados à Saúde e ao Desempenho), da Universidade Federal do Triângulo Mineiro, como requisito parcial para obtenção do título de mestre.

Orientador: Dr. Markus Vinicius Campos Souza

UBERABA

2024

**Catálogo na fonte: Biblioteca da Universidade Federal do
Triângulo Mineiro**

S583a	Silva, Sebastião Henrique Assis da Analisando a robustez das medidas de força muscular e desempenho físico em mulheres com mais de 40 anos: um estudo de confiabilidade teste-reteste e análise do efeito do treinamento de força com carga autoselecionada sobre o desempenho físico, força e massa muscular/ Sebastião Henrique Assis da Silva. -- 2024. 84 f. : il., fig., graf., tab. Dissertação (Mestrado em Educação Física) -- Universidade Federal do Triângulo Mineiro, Uberaba, MG, 2024 Orientador: Prof. Dr. Markus Vinicius Campos Souza 1. Saúde da mulher. 2. Exercício físico. 3. Força muscular. 4. Teste de esforço. 5. Reprodutibilidade dos testes I. Souza, Markus Vinicius Campos. II. Universidade Federal do Triângulo Mineiro. III. Título. CDU 613.99
-------	---

Sebastião Henrique Assis Da Silva

ANALISANDO A ROBUSTEZ DAS MEDIDAS DE FORÇA MUSCULAR E
DESEMPENHO FÍSICO EM MULHERES COM MAIS DE 40 ANOS: UM ESTUDO
DE CONFIABILIDADE TESTE-RETESTE E ANÁLISE DO EFEITO DO
TREINAMENTO DE FORÇA COM CARGA AUTOSSELECIONADA SOBRE O
DESEMPENHO FÍSICO, FORÇA E MASSA MUSCULAR.

Dissertação apresentada ao Programa de Pós-graduação em Educação Física, área de concentração Educação Física, Esporte e Saúde (Linha de Pesquisa: Aspectos Psicobiológicos do Exercício Físico Relacionados à Saúde e ao Desempenho), da Universidade Federal do Triângulo Mineiro, como requisito parcial para obtenção do título de mestre.

Aprovada em 29 / fevereiro / 2024

Banca Examinadora:

Dr. Markus Vinícius Campos Souza
Universidade Federal do Triângulo Mineiro

Dr. Cláudio de Oliveira Assumpção
Universidade Federal do Triângulo Mineiro

Dr. Clayton Augusto Libardi
Universidade Federal de São Carlos

RESUMO

O declínio da massa muscular e do desempenho físico em mulheres, especialmente após os 40 anos e durante a transição menopausal, pode levar a casos graves de sarcopenia. Portanto, existe a necessidade de avaliações precisas para identificar indivíduos em risco e propor intervenções como o treinamento de força como uma alternativa eficaz. O desafio da aderência a programas de treinamento, tem ênfase na intolerância da intensidade do exercício. A dissertação tem por objetivos de pesquisa, a investigação da confiabilidade dos testes de força e desempenho físico, bem como a avaliação do efeito de protocolos de carga autosselecionada. Destaca-se a importância de identificar respondedoras além das medidas mínimas detectáveis para obter dados confiáveis sobre intervenções na saúde de mulheres acima de 40 anos. O primeiro estudo tratou-se de um estudo transversal com total de 113 mulheres, com idade média de 58 ± 11 anos (82% pós-menopausa). Foram realizados testes de uma repetição máxima (1RM) para sete exercícios de força: leg press a 45° , supino, extensão de pernas, pec deck, pull-down lateral com pegada larga, flexão de pernas e remada sentada com cabo, com reteste após 48 horas. O Timed-Up-and-Go, caminhada de seis minutos e teste de 30 segundos para levantar e sentar foram realizados no mesmo dia e retestados após 72 horas. Os testes e retestes foram administrados pelos mesmos avaliadores. A confiabilidade relativa (consistência da posição do participante entre teste e reteste) foi avaliada usando o coeficiente de correlação intraclass para consistência e concordância, e a confiabilidade absoluta (precisão da pontuação) foi avaliada usando a MDM com base no erro padrão de predição. Os testes de 1RM e de desempenho apresentaram excelente confiabilidade. Esses testes mostraram valores significativos de MDM, especialmente para o teste de 1RM realizado na máquina. Embora os testes de FM e DF demonstrem excelente confiabilidade relativa em mulheres com mais de 40 anos, eles apresentam baixa confiabilidade absoluta, especialmente o teste de 1RM realizado na máquina e o teste de 30 segundos para levantar e sentar. Mudanças nestes testes menores que a MDM pode não indicar alterações reais em mulheres de meia-idade e mais velhas. No segundo estudo foi realizado um ensaio clínico randomizado conduzido com 90 mulheres acima de 40 anos submetidas a dois protocolos de TF (TF) por 12 semanas. No início, foram aplicados questionários sobre função cognitiva (MOCA), atividade física (IPAC), qualidade de vida (SF36) e aspectos sociodemográficos. Antes e após

a intervenção, foram realizados testes de DF (TUG, LS, 6MWT), avaliação da força muscular (1RM) e composição corporal. Após a avaliação inicial, as participantes foram randomizadas em dois grupos: o grupo tradicional (GT), que treinou com intensidade imposta de 50-80% de 1RM, e o grupo de carga autosselecionada (GA), com autonomia para escolher a intensidade. O protocolo com intensidade autosselecionada demonstrou ser eficaz para ganho de força, massa muscular e melhora no DF. Contudo, não foi possível estabelecer equivalência como o protocolo tradicionalmente recomendado pelo ACSM. A intensidade apresentou-se como variável importante para o ganho de força muscular, adicionalmente o volume foi relevante para respostas na massa muscular. Além disso, as respostas além da AMD mostraram-se influenciadas tanto pela intensidade quanto pelo volume.

ABSTRACT

The decline in muscle mass and physical performance in women, especially after the age of 40 and during the menopausal transition, can lead to severe cases of sarcopenia. Therefore, there is a need for accurate assessments to identify individuals at risk and propose interventions such as strength training as an effective alternative. The challenge of adherence to training programs emphasizes the intolerance of exercise intensity. The dissertation's research objectives include investigating the reliability of strength and physical performance tests, as well as evaluating the effect of self-selected load protocols. Emphasis is placed on the importance of identifying responders beyond the minimum detectable measures to obtain reliable data on interventions in the health of women over 40. The first study was a cross-sectional study with a total of 113 women, with an average age of 58 ± 11 years (82% postmenopausal). One-repetition maximum (1RM) tests were conducted for seven strength exercises: 45° leg press, bench press, leg extension, pec deck, wide-grip lateral pull-down, leg curl, and seated cable row, with a retest after 48 hours. The Timed-Up-and-Go, six-minute walk, and 30-second sit-to-stand tests were performed on the same day and retested after 72 hours. Test-retest reliability was assessed using intra-class correlation coefficients for consistency and agreement, and absolute reliability was assessed using the minimum detectable change based on the standard error of measurement. 1RM and performance tests demonstrated excellent reliability, although significant MDM values were observed, especially for the machine-based 1RM test. Despite excellent relative reliability, FM and DF tests showed low absolute reliability, particularly the machine-based 1RM and the 30-second sit-to-stand test. Changes in these tests smaller than the MDM may not indicate real alterations in middle-aged and older women. In the second study, a randomized clinical trial was conducted with 90 women over 40 undergoing two strength training protocols (ST) for 12 weeks. Initially, questionnaires on cognitive function (MOCA), physical activity (IPAC), quality of life (SF36), and sociodemographic aspects were administered. Before and after the intervention, DF tests (TUG, LS, 6MWT), muscle strength assessment (1RM), and body composition evaluation were conducted. After the initial assessment, participants were randomized into two groups: the traditional group (GT), which trained with an imposed intensity of 50-80% of 1RM, and the self-selected load group (GA), with autonomy to choose intensity. The self-selected intensity protocol

proved effective for strength gain, muscle mass, and DF improvement. However, equivalence with the traditionally recommended ACSM protocol could not be established. Intensity was found to be an important variable for muscle strength gain, with volume being relevant for muscle mass responses. Moreover, responses beyond the MDC were influenced by both intensity and volume.

SUMÁRIO

1	INTRODUÇÃO	10
2	ARTIGOS PRODUZIDOS.....	13
2.1	ARTIGO 1	13
2.2	ARTIGO 2.....	33
3	CONSIDERAÇÕES FINAIS	64
	REFERÊNCIAS.....	65
	ANEXO 01.....	69

1 INTRODUÇÃO

Com o passar dos anos inevitavelmente ocorre a redução da massa muscular consequentemente levando ao declínio da força e do desempenho físico (DF). Na mulher especialmente a partir dos 40 anos, fase onde frequentemente ocorre a transição menopausal, a redução dessas variáveis chega à -0,06kg de massa muscular ao ano, bem como redução mais acelerada da força e DF, cerca de 50% mais evidente quando comparados com homens na mesma idade, (SAMSON et al., 2000^a; GREENDALE et al., 2019). Além disso, declínio dessas variáveis aumentam o risco de sarcopenia, levam a maior suscetibilidade a quedas o que pode contribuir para maiores taxas de mortalidade, nesta população.

Portanto, faz-se necessário uma avaliação precisa da força e do DF funcional, a fim de identificar indivíduos em uma fase inicial de risco, bem como a elaboração de estratégias para mitigar desfechos adversos. Além disso, medições confiáveis são capazes de identificar até mesmo alterações mínimas na força muscular e DF, permitindo diagnósticos mais precoces e melhorias na avaliação, desempenhando um papel crucial no diagnóstico de determinadas condições de risco, além de ser fundamental para a eficaz monitorização e avaliação do efeito das intervenções na prática clínica (ASSIS SILVA et al., 2024).

Embora os testes de força e DF sejam populares devido à sua simplicidade, relação custo-eficácia, capacidade de refletir o desempenho nas atividades da vida diária e apresentarem confiabilidade na avaliação, ainda existe uma escassez de investigação abrangente sobre a confiabilidade destes testes em mulheres com 40 anos e acima. Ademais, a falta de consideração dos estudos sobre as alterações sistemáticas ou alterações mínimas detectáveis são lacunas existentes na literatura. Sendo essa medida importante para avaliação do efeito de intervenções na prática clínica e em ensaios clínicos elaborados com treinamento de força (TF).

Já é evidente que o TF é uma alternativa eficaz para atenuar o declínio do DF, da força e massa muscular em mulheres na pós menopausa (ORSATTI et al., 2008; NASCIMENTO DE OLIVEIRA-JÚNIOR et al., 2022), melhoras na força (CARNEIRO et al., 2020) e no DF (VIKBERG et al., 2019). Os testes de força e de DF são utilizados para verificar o efeito desse tipo de intervenção, considerando que são ferramentas acessíveis e eficazes para monitorar os impactos de ensaios clínicos com TF. Portanto, faz-se necessário analisar a confiabilidade e medidas de erros a fim de obter

resultados fidedignos sobre efeitos que extrapolam os erros de medidas advindos da ferramenta, do avaliador e do avaliado, dado esse que ainda não é contemplado nas pesquisas, entretanto é inovador e importante para detectar nas pesquisas respondedores além das medidas de erro sistemático de mudança mínimas detectáveis.

Sendo assim, detectar respondedores além de medidas de mudança mínima detectável faz com que as pesquisas sejam mais fidedignas, levando a dados relevantes sobre intervenções alternativas para melhorias no DF, na força e massa muscular. Visto que apesar dos benefícios substanciais do TF tradicional sobre o DF, massa e força muscular, a aderência a programas de TF na população de mulheres acima de 40 anos ainda é desafiadora, o que pode ser explicado por diversos fatores como a fragilidade física, problemas de saúde e negativas respostas sensitivas, especialmente relacionadas à percepção de desconforto resultante da intensidade do exercício (ARIKAWA; O'DOUGHERTY; SCHMITZ, 2011). Sendo assim, é necessário estudos com protocolos com intensidade alternativa afim de melhorar a aderência e promover benefícios a saúde.

Nesse sentido, a intensidade, caracterizada pela quantidade de resistência externa usada para se exercitar, é uma das variáveis do TF mais manipuladas em ensaios clínicos para promover maiores benefícios à saúde, como o aumento da força muscular e do DF. Para alcançar uma prescrição segura e eficaz, o American College of Sports Medicine (ACSM) propõe diretrizes que preconizam uma intensidade entre 60% e 80% da repetição máxima (1RM) (AMERICAN COLLEGE OF SPORTS MEDICINE, 2009). A intensidade prescrita pelo percentual de 1RM é efetiva e segura para ganhos de força e massa muscular em mulheres na pós-menopausa (ORSATTI et al., 2008). Recentemente, verificou-se que o TF com baixa intensidade também apresentou resultados positivos no DF, na força e massa muscular (CARNEIRO et al., 2021), ampliando a janela de intensidade para prescrição do TF nessa população, aumentando a aderência por consequência de melhores respostas perceptivas.

Quanto a respostas perceptivas e a tolerância à intensidade, já se sabe que exercitar em intensidades menores apresente maiores scores em escalas perceptivas de prazer, proporcionando o aumento da tolerância ao exercício e, consequentemente, maior aderência (EKKEKAKIS; LIND; VAZOU, 2010). Outra alternativa estabelecida a fim de melhorar respostas perceptivas e aumentar a aderência ao TF é a autosseleção da intensidade, visto que realizar atividades em

intensidade autoselecionada mostrou ser capaz de melhorar a aderência ao treinamento pela individualidade, maior percepção de prazer e menor percepção de esforço (PARFITT; HUGHES, 2009).

O processo de autonomia da escolha da intensidade, promove maior tolerância ao esforço e durante o exercício, pode ser percebida como agradável e, em última análise, pode aumentar a motivação intrínseca para a prática do exercício (ALVES et al., 2022). No TF, uma abordagem de protocolo com carga autoselecionada foi considerada mais prazerosa, associada ao menor esforço percebido em indivíduos inativos ou não treinados (ELSANGEDY et al., 2021). Em contrapartida, estudos envolvendo indivíduos treinados não encontraram diferenças entre protocolos para respostas afetivas ou esforço percebido (HAILE et al., 2019). Além disso, alguns estudos demonstraram que idosos inativos selecionam em média carga a 42% do 1RM, o que está em concordância com a carga recomendada para idosos iniciantes no TF (ELSANGEDY et al., 2013); (GLASS; AHMAD; GABLER, 2020).

Já é claro então que o protocolo com intensidade autoselecionada já demonstrou efeito na composição corporal, no metabolismo e no DF (ALVES et al., 2022). Contudo, os efeitos deste protocolo não foram comparados com os efeitos resultantes de um protocolo tradicional com intensidade recomendada pelo ACSM. Também não é claro se o TF de carga imposta ou carga autoselecionada promove efeitos além das medidas mínimas detectáveis, sendo esses aspectos um resultado inovador para área. Sendo assim, um protocolo com intensidade autoselecionada, no qual o indivíduo escolhe sua própria intensidade do exercício, foi proposto como alternativa a fim de aumentar a autonomia percebida, o prazer e, possivelmente, a aderência contínua ao exercício, promovendo efeitos equivalentes ao protocolo recomendado pelo ACSM para maiores benefícios no DF, força e massa muscular.

Portanto, essa dissertação tem os seguintes objetivos: 1) investigar a confiabilidade teste-reteste e alteração mínima detectável dos testes de força muscular e DF em mulheres acima de 40 anos. 2) 2.1) verificar o efeito do protocolo com carga autoselecionada sobre o DF, massa e força muscular. 2.2) verificar a relevância de variáveis como intensidade e volume sobre as respostas no DF, força e massa muscular. 2.3) verificar as respondedoras além da medida de alteração mínima detectável sobre a força muscular e DF.

2 ARTIGOS PRODUZIDOS

2.1 ARTIGO 1

ANALISANDO A ROBUSTEZ DAS MEDIDAS DE FORÇA MUSCULAR E DESEMPENHO FÍSICO EM MULHERES COM MAIS DE 40 ANOS: UM ESTUDO DE CONFIABILIDADE TESTE-RETESTE

RESUMO

Este estudo investigou a confiabilidade teste-reteste e a mudança mínima detectável (MMD) dos testes de força muscular (FM) e DF (DF) em mulheres com mais de 40 anos. Métodos: Um total de 113 mulheres, com idade média de 58 ± 11 anos (82% pós-menopausa), e índice de massa corporal de $28,5 \pm 5,8$ kg/m², participaram deste estudo. Foram realizados testes de uma repetição máxima (1RM) para sete exercícios de força: leg press a 45°, supino, extensão de pernas, pec deck, pull-down lateral com pegada larga, flexão de pernas e remada sentada com cabo, com reteste após 48 horas. O Timed-Up-and-Go, caminhada de seis minutos e teste de 30 segundos para levantar e sentar foram realizados no mesmo dia e retestados após 72 horas. Os testes e retestes foram administrados pelos mesmos avaliadores. A confiabilidade relativa (consistência da posição do participante entre teste e reteste) foi avaliada usando o coeficiente de correlação intraclass para consistência e concordância, e a confiabilidade absoluta (precisão da pontuação) foi avaliada usando a MDM com base no erro padrão de predição. Os testes de 1RM e de desempenho apresentaram excelente confiabilidade: leg press a 45° (consistência = 0,99, concordância = 0,98), supino (consistência = 0,96, concordância = 0,96), extensão de pernas (consistência = 0,93, concordância = 0,91), pec deck (consistência = 0,90, concordância = 0,88), pull-down lateral com pegada larga (consistência = 0,91, concordância = 0,89), flexão de pernas (consistência = 0,84, concordância = 0,83), remada sentada com cabo (consistência = 0,92, concordância = 0,94), Timed-Up-and-Go (consistência = 0,87, concordância = 0,87), caminhada de seis minutos (consistência = 0,96, concordância = 0,95), e teste de 30 segundos para levantar e sentar (consistência = 0,85,

concordância = 0,80). Esses testes mostraram valores significativos de MDM, especialmente para o teste de 1RM realizado na máquina e o teste de 30 segundos para levantar e sentar: leg press a 45° = 22,2 kg (15%), supino = 4,9 kg (19%), extensão de pernas = 9,2 kg (35%), pec deck = 8,8 kg (44%), pull-down lateral com pegada larga = 9,4 kg (31%), flexão de pernas = 9,2 kg (51%), remada sentada com cabo = 11,4 kg (29%), Timed-Up-and-Go = 1,4 s (20%), caminhada de seis minutos = 50,9 m (10%), e teste de 30 segundos para levantar e sentar = 4,6 repetições (30%). Embora os testes de FM e DF demonstrem excelente confiabilidade relativa em mulheres com mais de 40 anos, eles apresentam baixa confiabilidade absoluta, especialmente o teste de 1RM realizado na máquina e o teste de 30 segundos para levantar e sentar. Portanto, enquanto esses testes mostram boa consistência na posição das mulheres dentro do grupo após medidas repetidas, sua medida de precisão é relativamente baixa (alta flutuação). Mudanças nestes testes menores que a MDM pode não indicar alterações reais em mulheres de meia-idade e mais velhas.

Palavras-chave: Função Física; Confiabilidade Teste-Retest; Uma Repetição Máxima; Mulheres.

ABSTRACT

This study investigated the test-retest reliability and minimal detectable change (MDC) of muscle strength (MS) and physical performance (PP) tests in women aged over 40 years. A total of 113 women, aged 58 ± 11 years (82% postmenopausal women), with a body mass index of 28.5 ± 5.8 kg/m², participated in this study. One repetition maximum (1RM) tests were conducted for seven strength exercises: 45° leg press, bench press, leg extension, pec deck, wide-grip lateral pull-down, leg curl, and seated cable row, with a retest after 48 hours. Timed-Up-and-Go, six-minute walk, and 30s-sit-to-stand tests were performed on the same day and retested after 72 hours. The tests and retests were administered by the same evaluators. Relative reliability (consistency of participant rank between test-retest) was assessed using the intraclass correlation coefficient for consistency and agreement, and absolute reliability (precision of score) was assessed using the MDC based on the standard error of prediction. The 1RM and performance tests exhibited excellent reliability: 45° leg press (consistency = 0.99, agreement = 0.98), bench press (consistency = 0.96, agreement = 0.96), leg extension (consistency = 0.93, agreement = 0.91), pec deck (consistency = 0.90, agreement = 0.88), wide-grip lateral pull-down (consistency = 0.91, agreement = 0.89), leg curl (consistency = 0.84, agreement = 0.83), seated cable row (consistency = 0.92, agreement = 0.94), timed-Up-and-Go (consistency = 0.87, agreement = 0.87), six-minute walk (consistency = 0.96, agreement = 0.95), and 30s-sit-to-stand (consistency = 0.85, agreement = 0.80). These tests showed meaningful MDC values, particularly for the 1RM test performed on the machine and the 30s-sit-to-stand test: 45° leg press = 22.2 kg (15%), bench press = 4.9 kg (19%), leg extension = 9.2 kg (35%), pec deck = 8.8 kg (44%), wide-grip lateral pull-down = 9.4 kg (31%), leg curl = 9.2 kg (51%), seated cable row = 11.4 kg (29%), timed-Up-and-Go = 1.4 s (20%), six-minute walk = 50.9 m (10%), and 30s-sit-to-stand = 4.6 reps (30%). Although the MS and PP tests demonstrate excellent relative reliability in women aged over 40 years, they exhibit low absolute reliability, particularly the 1RM test performed on the machine and the 30s-sit-to-stand test. Therefore, while these tests show good consistency in the positioning of women within the group after repeated measures, their precision measure is relatively low (high fluctuation). Changes in these tests that are smaller than the MDC may not indicate real changes in women at middle age and older.

Key Words: Physical Function; Test-Retest Reliability; One Repetition Maximum; Women.

INTRODUÇÃO

À medida que as pessoas envelhecem, é inevitável que experimentem uma significativa diminuição tanto na força muscular (FM) quanto no DF (DF). Essa redução na FM e DF acarreta diversos riscos, incluindo o desenvolvimento de condições como a sarcopenia, uma maior suscetibilidade a quedas e, em casos mais graves, taxas elevadas de mortalidade (PATRIZIO et al., 2021). Consequentemente, a avaliação precisa da FM e do DF torna-se crucial para identificar indivíduos em estágios iniciais de risco e criar intervenções direcionadas para mitigar esses resultados adversos (PATRIZIO et al., 2021). Além disso, a obtenção de medidas confiáveis capazes de detectar até mesmo mudanças sutis na FM e no DF desempenha um papel fundamental no diagnóstico preciso dessas condições, bem como na monitorização eficaz e avaliação do impacto de intervenções na prática clínica (WEIR, 2005; RIEMANN; LININGER, 2018).

A necessidade de estabelecer a confiabilidade dessas medições e determinar suas mudanças mínimas detectáveis (MMD) é particularmente convincente entre mulheres com 40 anos ou mais. Este grupo específico de idade enfrenta declínios hormonais, mudanças fisiológicas e diminuição nos níveis de vitamina D durante as fases pré e pós-menopausa, todos os quais podem contribuir para reduções na FM e no DF (TSENG et al., 2012; EL KHOUDARY et al., 2019). No entanto, a literatura existente carece de pesquisa abrangente sobre a confiabilidade desses testes em mulheres dentro dessa faixa etária.

Ao avaliar o DF, medidas práticas e funcionais como os testes Timed-Up-and-Go (TUG), 30 segundos sentar e levantar e caminhada de seis minutos (6MWT) são frequentemente utilizados em idosos (YUKSEL et al., 2017; MUÑOZ-BERMEJO et al., 2021; PATRIZIO et al., 2021). Esses testes ganharam popularidade devido à sua simplicidade, custo-efetividade e capacidade de refletir o desempenho em atividades da vida diária (SPAGNUOLO et al., 2010; SHOEMAKER et al., 2013; PATRIZIO et al., 2021). No entanto, há uma escassez de pesquisas abrangentes sobre a confiabilidade desses testes em mulheres com 40 anos ou mais.

Uma ferramenta de avaliação amplamente utilizada para avaliar a FM é o teste de uma repetição máxima (1RM) (GRGIC et al., 2020b). No entanto, apesar da revisão sistemática de Grgic et al. (2020) sobre a confiabilidade teste-reteste do 1RM, ainda

persistem lacunas na literatura. A maioria dos 32 estudos selecionados em sua revisão envolveu participantes jovens. Embora vários exercícios tenham sido avaliados, a maioria dos estudos se concentrou em uma gama limitada de exercícios, com poucos explorando exercícios em máquinas, como o pec deck. Além disso, variações nos protocolos de teste, a falta de consideração de fatores como mudanças sistêmicas e informações limitadas sobre mudanças mínimas detectáveis foram falhas comuns. Embora seja possível calcular o MMD em alguns estudos, a maioria das análises dependeu do erro padrão da medida (EPM), que pode não fornecer o mais alto grau de precisão (CHARTER, 1996; WEIR, 2005). O erro padrão de previsão (EPP) foi recomendado como uma medida alternativa (CHARTER, 1996; WEIR, 2005).

Para preencher a lacuna existente nesta área de pesquisa, nosso estudo teve como objetivo investigar a confiabilidade teste-reteste e o MMD de testes de FM e DF em mulheres com 40 anos ou mais. Ao focar especificamente nesse grupo demográfico, nosso objetivo era oferecer *insights* valiosos sobre a eficácia dessas ferramentas de avaliação. As contribuições deste empreendimento de pesquisa avançarão as técnicas de medição para profissionais em diversas áreas, permitindo-lhes avaliar e monitorar com mais precisão FM e DF dentro desta faixa etária específica.

MÉTODOS

Desenho do Estudo

Este estudo teve uma duração de 5 semanas, iniciando em 19 de setembro e terminando em 21 de outubro de 2022. Inicialmente, uma entrevista foi conduzida para selecionar participantes com base nos critérios de inclusão, resultando em um total de 113 mulheres incluídas no estudo. Cada participante compareceu a uma visita para a administração dos testes de DF, seguida por outra visita para os retestes de DF após um intervalo de 72 horas (semana 1). Além disso, houve um período de familiarização com três visitas das semanas 2 a 4, durante o qual os participantes se familiarizaram com exercícios de resistência. Posteriormente, uma visita foi dedicada à realização dos testes de 1RM, com um reteste realizado após 48 horas (semana 5). Para minimizar a influência de variações biológicas diurnas, todas as avaliações foram consistentemente realizadas no mesmo horário do dia. Além disso, os testes de FM e

DF foram padronizados e supervisionados de perto por pesquisadores experientes para garantir a técnica adequada, confiabilidade e a segurança e bem-estar dos participantes.

Mulheres

As participantes deste estudo forneceram consentimento informado e foram recrutadas por meio de canais de mídia na cidade de Uberaba-MG, Brasil. O estudo recebeu aprovação do comitê de ética da Universidade Federal do Triângulo Mineiro (UFTM), onde foi especificamente conduzido dentro do programa de pós-graduação em Educação Física. As participantes foram recrutadas em bairros residenciais e aquelas que manifestaram interesse em participar passaram por uma entrevista abrangente. A entrevista envolveu a coleta de informações sobre idade, situação de emprego, indicadores de saúde, histórico médico passado e presente, intervenções terapêuticas e atividades físicas. Inicialmente, 140 mulheres passaram por triagem de elegibilidade. Dentre essas, um total de 113 participantes atendeu aos critérios de inclusão no estudo. Os critérios de inclusão incluíram: (i) ser pré-menopausal ou pós-menopausal (caracterizada por amenorreia espontânea por pelo menos 12 meses); (ii) não usar terapia hormonal ou suplementos alimentares; (iii) ter pressão arterial e níveis de glicose no sangue controlados; (iv) não ter distúrbios musculares, tromboembólicos ou gastrointestinais; (v) não ter histórico de doenças cardiovasculares ou infecciosas; (vi) ser não consumidoras de álcool (nenhum consumo de álcool em sua dieta); (vii) ser não fumantes; (viii) ter experiência prévia com exercício resistido, mas não se envolver em atividade física regular mais de uma vez por semana nos últimos 6 meses. No entanto, devido a problemas articulares específicos, algumas participantes só puderam realizar um subconjunto dos testes. A distribuição de participantes para cada teste pode ser encontrada na Tabela 2.

Teste de uma repetição máxima (1RM)

Durante o período de familiarização, os participantes completaram três sessões de treinamento por semana, às segundas, quartas e sextas-feiras. As sessões de treinamento visavam familiarizar os participantes com as técnicas dos exercícios.

Cada sessão ocorreu em dias alternados, e os participantes treinaram com cargas mais leves, evitando a falha concêntrica voluntária. Os exercícios incluídos no período de familiarização foram legpress 45°, supino, cadeira extensora, pec deck, puxada frontal, mesa flexora e remada baixa.

Após o período de familiarização, foram conduzidos testes de 1RM para todos os exercícios no mesmo dia, com um intervalo de 5 minutos entre cada exercício. Antes da sessão de teste, os participantes realizaram um aquecimento geral caminhando por 5 minutos, seguido por um aquecimento específico usando cargas subjetivas determinadas durante o período de familiarização. O aquecimento específico envolveu a realização de 12 repetições com cargas mais leves. Após um período de descanso de 90 segundos, a intensidade da carga foi aumentada em aproximadamente 20%, e os participantes realizaram entre três e cinco repetições. Após um período de descanso de 3 a 5 minutos, a intensidade da carga foi significativamente aumentada, e os participantes foram incentivados a fazer um esforço máximo e usar a amplitude total de movimento. Se a intensidade da carga foi superestimada ou subestimada, os participantes descansaram por 3 a 5 minutos antes de tentar uma nova repetição com uma carga mais baixa ou mais alta, respectivamente (aproximadamente 5% a 10% de diferença). Esse processo foi repetido por duas a cinco tentativas até que a intensidade da carga de 1RM equivalente fosse determinada².

A intensidade da carga na qual os participantes não conseguiam realizar mais do que uma repetição foi considerada como sua 1RM para cada exercício. Nos casos em que não foi possível estabelecer a carga para uma única repetição, a fórmula de Brzycki [$1RM = \text{peso levantado} / (1.0278 - 0.0278 * \text{repetições executadas})$] foi usada como alternativa.

A sessão de reteste seguiu o mesmo protocolo do teste inicial, com um intervalo mínimo de 48 horas entre as sessões de teste e reteste.

Teste *Timed-Up-and-Go* (TUG)

Para o teste TUG, os participantes estavam sentados em uma cadeira com as costas apoiadas por uma parede, os pés totalmente no chão e as mãos apoiadas nas

coxas. Eles foram instruídos a levantar-se da cadeira e caminhar uma distância de 3 metros em um piso plano. Eles contornaram um cone e retornaram à cadeira, onde se sentaram o mais rapidamente possível. Os participantes foram instruídos a completar o percurso sem correr, focando na velocidade. O tempo necessário para toda a sequência, desde o movimento inicial de levantar-se da cadeira até retornar a sentar-se novamente, foi registrado com precisão de 0,01 segundos usando um cronômetro⁹. Cada participante realizou o teste uma vez como ensaio de familiarização, e duas tentativas foram registradas para análise. A média das duas tentativas foi calculada como o valor representativo. Para o reteste, o mesmo protocolo foi seguido após um intervalo de 72 horas.

Teste de 30 segundos sentar e levantar (*30s-sit-to-stand*)

Durante o teste de *30s-sit-to-stand*, os participantes estavam sentados em uma cadeira com altura de 47 cm, posicionada contra uma parede. Eles garantiram que suas costas estivessem em contato com o encosto, seus pés estivessem completamente apoiados no chão e suas mãos estivessem cruzadas, apoiadas nos ombros opostos. O teste envolveu uma pessoa registrando o tempo e outra pessoa contando o número de repetições realizadas pelo participante. Cada participante teve de 1 a 2 tentativas com um período de descanso apropriado antes do teste final. Após um intervalo de 72 horas, o reteste foi conduzido seguindo o mesmo protocolo.

Teste de Caminhada de Seis Minutos (6MWT)

O 6MWT foi realizado em uma superfície plana. O percurso de caminhada foi demarcado por dois cones, criando uma distância total de 20 metros. Os participantes foram instruídos a caminhar de um cone para o outro (cobrindo uma distância de 20 metros), contornar um cone e retornar ao cone inicial, completando o percurso. Durante o 6MWT, os participantes foram encorajados a caminhar o mais rápido possível, sem pausas ou corridas, por uma duração de 6 minutos. Incentivos verbais foram fornecidos ao longo do teste para motivar os participantes, utilizando frases como "você está indo muito bem" e "caminhe o mais rápido que puder". A distância percorrida por cada participante foi registrada ao final do 6MWT. Após um intervalo de

72 horas, o reteste foi realizado seguindo o mesmo protocolo do teste inicial para garantir condições de teste consistentes e avaliar a confiabilidade dos resultados.

Análise Estatística

Todas as análises estatísticas foram conduzidas utilizando o software JAMOV[®]. Avaliamos a distribuição das diferenças entre as medições com o teste de Shapiro-Wilk, que indicou uma distribuição normal. Para avaliar a confiabilidade teste-reteste, utilizamos o pacote SympleAgree do jamovi.

Medimos a confiabilidade relativa, que avalia a consistência das classificações dos participantes entre o teste e o reteste, utilizando o coeficiente de correlação intraclass (ICC) tanto para consistência quanto para concordância. Além disso, avaliamos a confiabilidade absoluta, que mede a precisão das pontuações, utilizando a Mudança Mínima Detectável (MMD) com base no erro padrão de predição.

A análise estatística envolveu o cálculo do ICC por meio de uma ANOVA de dois fatores, considerando tanto a consistência (dois fatores fixos) quanto a concordância (dois fatores aleatórios) com uma unidade simples. Os valores de ICC foram categorizados da seguinte forma: excelente ($>0,90$), bom ($0,75-0,90$), moderado ($0,50 - 0,75$) ou fraco ($<0,50$), com base em critérios predefinidos (KOO; LI, 2016).

O MMD foi determinado utilizando a fórmula $SEP (1,96 \times SEP)$ em vez de SEM ao quadrado (SEM^2). Avaliamos a associação entre os valores médios de teste-reteste e o viés ($\Delta = \text{reteste} - \text{teste}$) usando o coeficiente de correlação de Pearson.

Para investigar a influência da menopausa na confiabilidade teste-reteste, conduzimos a análise de ICC excluindo mulheres pré-menopausadas. Além disso, para examinar o impacto da idade na confiabilidade teste-reteste, correlacionamos os deltas de teste com as idades dos participantes.

RESULTADOS

As características das participantes estão resumidas na Tabela 1.

Tabela 1 – Características gerais das mulheres (n=113).

	Média	DP
Idade, anos	57.6	10.0
IMC, kg/m²	28.5	5.8
Gordura corporal, %	43.5	6.5
IMM, kg/m²	6.85	1.21
Nível de atividade física, min/week	947.0	817.0
Problemas articulares, n (%)	29 (26%)	
Tempo educacional, yrs	10.0	7.0
Mulheres na pré-menopausa, n (%)	20 (18%)	

IMC = Índice de massa corporal, IMM = Índice de massa muscular, DP = Desvio padrão

Análise de Consistência e Concordância

A Tabela 2 apresenta os valores de consistência e concordância para os testes de FM e DF. Todos os testes de 1RM e DF demonstraram consistência e concordância excelentes ou boas (ou seja, > 0,75) entre o teste inicial e o reteste, independentemente da presença de mulheres pré-menopáusicas

Tabela 2 - Coeficiente de correlação intraclass para consistência e concordância dos testes de força muscular e desempenho físico.

Testes	Todas as mulheres (n=113)					Mulheres na pós menopausa (n=93)				
	n	ICC ¹	IC95%	ICC ²	IC95%	n	ICC ¹	IC95%	ICC ²	IC95%
Leg press, kg	79	0.99	(0.98-0.99)	0.98	(0.91-0.99)	65	0.99	(0.98-0.99)	0.98	(0.94-0.99)
Supino reto, kg	79	0.96	(0.95-0.98)	0.96	(0.91-0.97)	63	0.96	(0.95-0.98)	0.96	(0.92-0.97)
Cadeira extensora, kg	92	0.93	(0.91-0.95)	0.91	(0.83-0.95)	77	0.94	(0.91-0.96)	0.92	(0.91-0.96)
Pec deck, kg	92	0.90	(0.86-0.92)	0.88	(0.80-0.92)	75	0.89	(0.84-0.92)	0.87	(0.79-0.92)
Puxada frontal, kg	110	0.91	(0.87-0.93)	0.89	(0.83-0.93)	77	0.89	(0.85-0.92)	0.87	(0.78-0.91)
Mesa flexora, kg	92	0.84	(0.79-0.89)	0.83	(0.77-0.88)	75	0.87	(0.81-0.91)	0.85	(0.77-0.90)
Remada baixa, kg	85	0.92	(0.88-0.94)	0.90	(0.82-0.94)	72	0.93	(0.90-0.95)	0.92	(0.86-0.95)
TUG, s	113	0.87	(0.83-0.91)	0.87	(0.83-0.90)	86	0.87	(0.82-0.91)	0.87	(0.82-0.90)
6MWT, m	110	0.96	(0.94-0.97)	0.95	(0.93-0.97)	85	0.96	(0.95-0.97)	0.96	(0.93-0.97)
SL, reps	105	0.85	(0.80-0.89)	0.80	(0.61-0.89)	82	0.86	(0.80-0.90)	0.81	(0.63-0.89)

1 = consistência, 2 = concordância, ICC = Coeficiente de Correlação Intraclass, IC95% = Intervalo de Confiança de 95%, TUG = Timed up and go, 6MWT = Teste de caminhada de seis minutos, SL = Teste sentar e levantar.

MMD (Mudança Mínima Detectável)

A Tabela 3 apresenta o viés (viés sistemático) entre o teste inicial e o reteste, bem como o MMD (Mudança Mínima Detectável). No caso dos testes de 1RM, o viés variou de 5% a 8% em relação à média do teste e reteste. A diferença mínima necessária para considerar uma mudança real (MMD) variou de 15% a 51% em relação à média do teste e reteste. Notavelmente, os exercícios leg press e supino apresentaram a menor variação, enquanto os exercícios mesa flexora e pec deck exibiram a maior variação.

Em relação aos testes de DF, o MDC variou de 10% a 30%. O teste de caminhada de 6 minutos (6MWT) mostrou a menor variação (10%), enquanto o teste de sentar e levantar em 30 segundos apresentou a maior variação (30%).

No contexto dos testes de 1RM, foi observado um viés positivo ($r = 0,23$, $P = 0,027$) para o exercício de cadeira extensora. No entanto, não houve associação significativa entre o valor médio e a diferença entre o teste inicial e o reteste para os outros exercícios e testes de DF.

Tabela 3 - Valores de concordância (viés) entre teste e reteste e limites de concordância.

Testes	n	Test (A)	Re-test (B)	Média (A and B)	Viés	DP de Viés.	SEM	SEP	MMD	r	P-value (r)
Leg press, kg	79	146.3±49.6	153.3±51.5	149.8	7.0 (5%)	8.0	5.6	11.3	22.2 (15.0%)	0.23	0.038
Supino reto, kg	79	25.6±6.5	26.5±6.7	26.0	0.9 (4%)	1.9	1.3	2.5	4.9 (19%)	0.11	0.318
Cadeira extensora, kg	92	25.0±9.1	26.8±9.4	25.9	1.9 (7%)	3.4	2.4	4.7	9.2 (35%)	0.08	0.445
Pec deck, kg	92	19.1±6.9	20.6±7.4	19.9	1.5 (8%)	3.3	2.3	4.5	8.8 (44%)	0.16	0.126
Puxada frontal, kg	110	29.5±14.1	30.7±14.2	30.1	1.3 (4%)	3.2	2.4	4.7	9.2 (31%)	0.09	0.312
Mesa flexora, kg	92	17.8±6.1	18.8±6.1	18.3	1.0 (6%)	3.4	2.4	4.8	9.4 (51%)	0.02	0.818
Remada baixa, kg	85	38.6±10.4	40.7±10.3	39.6	2.1 (5%)	4.2	3.0	5.8	11.4 (29%)	-0.02	0.829
TUG, s	113	7.0±1.0	6.9±1.0	7.0	-0.1 (-1%)	0.5	0.4	0.7	1.4 (20%)	0.02	0.823
6MWT, m	110	503.0±66.9	509.0±64.5	506.0	6.2 (1%)	19.9	13.4	26.5	51.9 (10%)	-0.13	0.184
SL, reps	105	14.8±3.2	15.9±3.0	15.4	1.1 (7%)	1.7	1.2	2.3	4.6 (30%)	-0.16	0.093

TUG = Timed up and go, SL = Sentar e levantar 30'', 6MWT = Teste de caminhada de seis minutos, viés = Diferença entre A e B, DP = Desvio padrão, SEM = Erro padrão de medição, SEP = Erro padrão de previsão, MMD = Mudança mínima detectável

Análise de Correlação

A análise de correlação não revelou associação significativa entre os deltas (Δ = reteste - teste) de cada teste e a idade. Os seguintes valores de correlação foram observados: TUG ($P = 0,479$, $R = -0,073$), 6MWT ($P = 0,924$, $R = 0,010$), teste de sentar e levantar em 30 segundos ($P = 0,342$, $R = 0,099$), 1RM leg press 45° ($P = 0,739$, $R = -0,037$), 1RM supino ($P = 0,469$, $R = -0,082$), 1RM pec deck ($P = 0,516$, $R = 0,072$), 1RM mesa flexora ($P = 0,149$, $R = 0,160$), 1RM puxada frontal ($P = 0,385$, $R = 0,095$), 1RM remada baixa ($P = 0,769$, $R = -0,033$) e 1RM cadeira extensora ($P = 0,871$, $R = -0,018$).

DISCUSSÃO

O objetivo deste estudo foi avaliar a confiabilidade teste-reteste de testes de Força Muscular (FM) e DF em mulheres com 40 anos ou mais. Embora esses testes demonstrem excelente confiabilidade relativa, indicada por fortes correlações teste-reteste, entre mulheres nessa faixa etária, eles também apresentam baixa confiabilidade absoluta, caracterizada por uma significativa variabilidade nas medidas repetidas. Essa variabilidade é especialmente proeminente no teste de 1RM (uma repetição máxima) realizado em máquina e no teste de 30s-sit-to-stand. Esses testes consistentemente produzem resultados semelhantes ao avaliar o desempenho de mulheres na mesma faixa etária em várias medidas. No entanto, a precisão dessas medições é relativamente baixa, resultando em flutuações perceptíveis nos resultados quando os testes são repetidos.

Os Coeficientes de Correlação Intraclass (ICC) para os testes de 1RM variaram de 0,80 a 0,99, alinhando-se com os achados de outro estudo conduzido em uma população semelhante (BARBALHO et al., 2018; GRGIC et al., 2020a). Destaca-se que os exercícios de supino e prensa de pernas a 45° exibiram valores de ICC mais elevados em comparação com os outros exercícios. Exercícios que envolvem a manipulação de cargas específicas usando máquinas, como um aumento predefinido de 5 kg e a estimativa de 1RM usando fórmulas preditivas, podem ter contribuído para ICCs mais baixos. Isso sugere que o uso de pesos livres, que permite uma manipulação mais fácil da carga, pode ter contribuído para esses resultados favoráveis (BARBALHO et al., 2018; GRGIC et al., 2020a). Apesar dessas variações,

é importante enfatizar que todos os exercícios demonstraram excelente ou boa confiabilidade relativa, com ICCs superiores a 0,80. Isso implica que os testes de 1RM exibem consistência e concordância nas classificações dos participantes entre o teste inicial e o reteste, independentemente de serem realizados em máquina ou com pesos livres (RIEMANN; LININGER, 2018). Consequentemente, isso sugere que os testes de 1RM são uma ferramenta adequada para diagnosticar e classificar a FM, distinguindo entre níveis mais altos e mais baixos de FM, em mulheres com 40 anos ou mais.

Testes de DF desempenham um papel crucial na avaliação do risco de quedas e mortalidade, tornando-os ferramentas valiosas tanto na prática clínica quanto na pesquisa. Estudos anteriores utilizaram essa ferramenta para diagnosticar baixo DF em mulheres mais velhas (CHAN; PIN, 2019), mas é importante observar que declínios no DF são observados em mulheres mais jovens durante a transição menopáusica (TSENG et al., 2012; EL KHOUDARY et al., 2019). Assim, essa ferramenta pode ser útil para o diagnóstico precoce de baixo DF em mulheres. Em nosso estudo, os testes de DF demonstraram excelente confiabilidade relativa (ICC > 0,82), em linha com descobertas de estudos anteriores conduzidos com idosos (Boshnjaku A, Bahtiri A, Feka K, Krasniqi E, Tschan., 2021). Os valores de confiabilidade relativa relatados nesses estudos estão alinhados com os observados na avaliação de ferramentas de DF incluídas em nosso estudo, como os testes de levantar e sentar 30'' (MUÑOZ-BERMEJO et al., 2021), 6MWT (CHAN; PIN, 2019) e TUG (YUKSEL et al., 2017). Isso implica que os testes de DF exibem consistência e concordância nas classificações dos participantes entre o teste inicial e o reteste. Consequentemente, isso sugere que os testes de DF são uma ferramenta adequada para distinguir níveis mais altos e mais baixos de DF em mulheres com 40 anos ou mais.

Curiosamente, idade e status menopausal não pareceram impactar as mudanças e posições das medidas. Essa falta de influência provavelmente é atribuída à facilidade de entendimento e aplicabilidade dos testes. Consequentemente, esses testes demonstram uma forte confiabilidade relativa em diferentes faixas etárias e condições fisiológicas, estabelecendo-os firmemente como excelentes ferramentas para diagnosticar FM, além de DF, em mulheres pré e pós-menopausadas.

No contexto do planejamento e interpretação de intervenções ou monitoramento longitudinal de pacientes, é crucial ter uma compreensão abrangente de erros sistemáticos e limites de erros (MDC) associados a testes comumente utilizados, pois esses podem ocorrer (WEIR, 2005; RIEMANN; LININGER, 2018). Em nossa análise atual, foram observados erros sistemáticos em todos os testes de funcionalidade física, contribuindo para a determinação de um MDC. O MDC desempenha um papel fundamental em ajudar os profissionais a interpretar uma mudança significativa entre as condições pré e pós-intervenção. Para ilustrar esse conceito, tomemos o exemplo do exercício de leg press 45°. Em nosso estudo, o MDC para esse exercício foi determinado como 22,2 kg, o que corresponde a aproximadamente 15% do valor médio obtido pela amostra no teste. Isso implica que mudanças que excedem esse limite são consideradas mudanças verdadeiras, além do erro de medição, e ajudam a considerar erros instrumentais (WEIR, 2005; RIEMANN; LININGER, 2018). Portanto, quaisquer mudanças observadas no exercício de leg press 45° que ultrapassem esse limite podem ser consideradas significativas (mudanças verdadeiras) e não devem ser atribuídas exclusivamente à variabilidade de medição (WEIR, 2005; RIEMANN; LININGER, 2018). Detectar o MDC é de extrema importância em ensaios clínicos, pois ajuda a determinar a significância clínica da intervenção. Ao quantificar a mudança necessária na variável avaliada que excede possíveis erros de medição, os pesquisadores podem interpretar os resultados com confiança e avaliar o verdadeiro impacto da intervenção.

Os valores de MMD obtidos em nosso estudo são consistentes com os relatados em outros estudos que utilizaram testes para avaliar DF (YUKSEL et al., 2017; MURDOCH et al., 2023). No entanto, diferem da literatura limitada disponível sobre o teste de 1RM (BARBALHO et al., 2018; INFANTE et al., 2021) e os testes de DF (Chan WLS, Pin TW., 2019; MUÑOZ-BERMEJO et al., 2021). Uma possível explicação para essa discrepância pode ser atribuída à fórmula de MDC. A maioria dos cálculos dependeu do erro padrão da medida (SEM), que pode não fornecer o mais alto grau de precisão (CHARTER, 1996; WEIR, 2005). Conforme a recomendação (CHARTER, 1996; WEIR, 2005), calculamos o MMD considerando o erro padrão da predição. Outra possível explicação para essa discrepância pode ser atribuída aos diferentes equipamentos usados em nossa pesquisa. Alguns equipamentos empregados em nosso estudo tinham características biomecânicas distintas em comparação com os

usados na literatura existente. É importante notar que, em ambientes profissionais e laboratoriais para TF, existe uma ampla variedade de equipamentos, cada um com suas propriedades biomecânicas únicas. Portanto, estudos futuros devem investigar o MMD dos testes de 1RM usando máquinas com diversas características biomecânicas para fornecer uma compreensão mais abrangente de seu desempenho.

Em nosso estudo, não observamos mudanças sistemáticas significativas (viés), mas identificamos um alto valor de limite de erro (MMD). Tanto nos testes de 1RM quanto no teste de sentar e levantar em 30 segundos, o limite de erro excedeu 30% entre medidas repetidas. Esse limite de erro pode ultrapassar os aumentos típicos na FM e DF comumente observados em várias intervenções como nutricionais (NASIMI et al., 2023) ou treinamento de força (CUIYUL-VÁSQUEZ et al., 2023) ou influenciados por fatores como idade ou condições de saúde específicas, incluindo doenças incapacitantes (PICCA et al., 2022). Essa observação é relevante porque a aplicação principal dos testes de 1RM e de DF muitas vezes envolve a avaliação de mudanças na FM após um programa de treinamento específico ou o monitoramento longitudinal de pacientes.

É importante reconhecer algumas limitações deste estudo. Em primeiro lugar, nossa pesquisa foi conduzida por um único avaliador, o que significa que não podemos identificar erros que podem ocorrer entre diferentes avaliadores. Em segundo lugar, as diferenças observadas na confiabilidade relativa e absoluta são específicas de exercícios e testes particulares do estudo. É importante destacar que nosso estudo carecia de poder estatístico para avaliar as medidas teste-reteste em mulheres pré-menopausadas, pois menos de 20% da amostra se enquadrava nessa categoria. Finalmente, este estudo focou em uma população específica de mulheres com 40 anos ou mais. Portanto, deve-se ter cautela ao generalizar as descobertas para outras populações com características físicas e fisiológicas diferentes. Por outro lado, uma fortaleza deste estudo é a inclusão de um grande número de mulheres cobrindo uma ampla faixa etária.

CONCLUSÃO

Embora os testes de força muscular e DF demonstrem excelente confiabilidade relativa (correlação teste-reteste) em mulheres com mais de 40 anos, apresentam

baixa confiabilidade absoluta (variabilidade de medidas repetidas), especialmente o teste de 1RM realizado em máquina e o teste de sentar e levantar em 30 segundos. Como resultado, esses testes são ferramentas eficazes para identificar ou diagnosticar mulheres com FM e DF elevados ou baixos. No entanto, eles podem ter limitações ao avaliar intervenções que produzem pequenos efeitos nesses testes. Portanto, alterações nos resultados desses testes que ficam abaixo da Mudança Mínima Detectável podem indicar que mulheres de meia idade e mais velhas não experimentaram uma mudança real na FM ou DF.

REFERÊNCIAS

BARBALHO, M.; GENTIL, P.; RAIOL, R.; DEL VECCHIO, F.B.; RAMIREZ-CAMPILLO, R.; COSWIG, V.S. High 1RM Tests Reproducibility and Validity are not Dependent on Training Experience, Muscle Group Tested or Strength Level in Older Women. *Sports (Basel)*, v. 6, n. 4, 2018.

BOSHNIJAKU, A.; BAHTIRI, A.; FEKA, K.; KRASNIQI, E.; TSCHAN, H.; WESSNER, B. Test-retest reliability data of functional performance, strength, peak torque and body composition assessments in two different age groups of Kosovan adults. *Data in Brief*, v. 36, p. 106988, 2021.

CHAN, W.L.S.; PIN, T.W. Reliability, validity and minimal detectable change of 2-minute walk test, 6-minute walk test and 10-meter walk test in frail older adults with dementia. *Experimental Gerontology*, [Validation Study], v. 115, p. 9-18, 2019.

CHARTER, R.A. Revisiting the standard errors of measurement, estimate, and prediction and their application to test scores. *Perceptual and Motor Skills*, v. 82, n. 3_suppl, p. 1139-1144, 1996.

CUYUL-VASQUEZ, I.; PEZO-NAVARRETE, J.; VARGAS-ARRIAGADA, C. et al. Effectiveness of Whey Protein Supplementation during Resistance Exercise Training on Skeletal Muscle Mass and Strength in Older People with Sarcopenia: A Systematic Review and

EL KHOUDARY, S.R.; GREENDALE, G.; CRAWFORD, S.L. et al. The menopause transition and women's health at midlife: a progress report from the Study of

Women's Health Across the Nation (SWAN). Menopause, [Research Support, N.I.H., Extramural Review], v. 26, n. 10, p. 1213-1227, 2019.

GRGIC, J.; LAZINICA, B.; SCHOENFELD, B.J.; PEDISIC, Z. Test-Retest Reliability of the One-Repetition Maximum (1RM) Strength Assessment: a Systematic Review. Sports Medicine - Open, [Review], v. 6, n. 1, p. 31, 2020.

INFANTE, M.A.; HARRELL, G.M.; STRAND, K.L.; PARRINO, R.L.; SIGNORILE, J.F. One Repetition Maximum Test-Retest Reliability and Safety Using Keiser Pneumatic Resistance Training Machines With Older Women. Journal of Strength and Conditioning Research, v. 35, n. 12, p. 3513-3517, 2021.

KOO, T.K.; LI, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine, v. 15, n. 2, p. 155-163, 2016.

MUÑOZ-BERMEJO, L.; ADSUAR, J.C.; MENDOZA-MUÑOZ, M. et al. Test-Retest Reliability of Five Times Sit to Stand Test (FTSST) in Adults: A Systematic Review and Meta-Analysis. Biology, [Review], v. 10, n. 6, 2021.

MURDOCH, M.; WINDOW, P.; MORTON, C.; O'DONOHUE, R.; BALLARD, E.; CLAUS, A. People at a persistent pain service can walk it, but some struggle to talk about it: Reliability, detectable difference and clinically important difference of the six-minute walk test. Musculoskeletal Care, v. 21, n. 1, p. 221-231, 2023.

NASIMI, N.; SOHRABI, Z.; NUNES, E.A. et al. Whey Protein Supplementation with or without Vitamin D on Sarcopenia-Related Measures: A Systematic Review and Meta-Analysis. Adv Nutr, [Meta-Analysis Review Systematic Review], v. 14, n. 4, p. 762-773, 2023.

PATRIZIO, E.; CALVANI, R.; MARZETTI, E.; CESARI, M. Physical Functional Assessment in Older Adults. The Journal of Frailty & Aging, v. 10, n. 2, p. 141-149, 2021.

Picca A, Coelho-Junior HJ, Calvani R, Marzetti E, Vetrano DL. Biomarkers shared by frailty and sarcopenia in older adults: A systematic review and meta-analysis. Ageing Research Reviews. 2022;73:101530.

RIEMANN, B.L.; LININGER, M.R. Statistical Primer for Athletic Trainers: The Essentials of Understanding Measures of Reliability and Minimal Important Change. *Journal of Athletic Training*, [Review], v. 53, n. 1, p. 98-103, 2018.

SHOEMAKER, M.J.; CURTIS, A.B.; VANGSNES, E.; DICKINSON, M.G. Clinically meaningful change estimates for the six-minute walk test and daily activity in individuals with chronic heart failure. *Cardiopulmonary Physical Therapy Journal*, v. 24, n. 3, p. 21-29, 2013.

SPAGNUOLO, D.L.; JURGENSEN, S.P.; IWAMA, A.M.; DOURADO, V.Z. Walking for the assessment of balance in healthy subjects older than 40 years. *Gerontology*, [Comparative Study Research Support, Non-U.S. Gov't Validation Study], v. 56, n. 5, p. 467-473, 2010.

TSENG, L.A.; EL KHOUDARY, S.R.; YOUNG, E.A. et al. The association of menopause status with physical function: the Study of Women's Health Across the Nation. *Menopause*, [Multicenter Study Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't], v. 19, n. 11, p. 1186-1192, 2012.

WEIR, J.P. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, [Review], v. 19, n. 1, p. 231-240, 2005.

YUKSEL, E.; KALKAN, S.; CEKMECE, S.; UNVER, B.; KARATOSUN, V. Assessing Minimal Detectable Changes and Test-Retest Reliability of the Timed Up and Go Test and the 2-Minute Walk Test in Patients With Total Knee Arthroplasty. *The Journal of Arthroplasty*, v. 32, n. 2, p. 426-430, 2017.

2.2 ARTIGO 2

EFEITO DO TREINAMENTO DE FORÇA COM CARGA AUTOSSELECIONADA SOBRE O DESEMPENHO FÍSICO, FORÇA E MASSA MUSCULAR

RESUMO

Com o passar dos anos inevitavelmente ocorre o declínio da massa muscular, força e DF, associado a riscos como maior suscetibilidade a quedas, sarcopenia e mortalidade. O TF (TF) é uma alternativa para atenuar esses efeitos, particularmente relevante para mulheres acima dos 40 anos, faixa etária em que já se observa redução de força e massa muscular. A intensidade do TF, é uma variável chave, com recomendações do American College of Sports Medicine (ACSM) entre 60% e 80% da repetição máxima (1RM). No entanto, evidências na literatura sugerem ganhos significativos com baixas intensidades. Apesar dos benefícios substanciais do TF, a adesão ao TF nessa população é desafiadora. A autosseleção da intensidade surge como uma alternativa para melhorar a aderência. Portanto, o objetivo desse estudo foi verificar se o protocolo com intensidade autosselecionada promove resultados equivalentes na massa muscular, força e DF quando comparado ao protocolo de treinamento com cargas estabelecidas pela recomendação da ACSM. Ensaio clínico randomizado conduzido com 60 mulheres acima de 40 anos submetidas a dois protocolos de TF (TF) por 12 semanas. No início, foram aplicados questionários sobre função cognitiva (MOCA), atividade física (IPAC), qualidade de vida (SF36) e aspectos sociodemográficos. Antes e após a intervenção, foram realizados testes de DF (TUG, LS, 6MWT), avaliação da força muscular (1RM) e composição corporal. Após a avaliação inicial, as participantes foram randomizadas em dois grupos: o grupo tradicional (GT), que treinou com intensidade imposta de 50-80% de 1RM, e o grupo de carga autosselecionada (GA), com autonomia para escolher a intensidade. Ao longo da intervenção foram coletadas a escala de repetição de reserva (RR). A seleção da intensidade pelo GA foi substancialmente igual ao GT. No entanto, foram observadas diferenças nos exercícios supino ($p=0,017$), cadeira extensora ($p=0,011$) e remada baixa ($p=0,018$), nos quais o GT demonstrou uma superioridade. O volume apresentou diferenças apenas no leg press ($p=0.008$) e no volume total ($P=0.050$). No efeito tempo todas as variáveis tiveram resultados significativos exceto a força no exercício de mesa flexora ($p=0.183$), TUG ($p=0.169$) e na massa muscular dos braços ($p=0.530$). Não houve efeito grupo nem interação. Não foi possível estabelecer uma equivalência total dos protocolos em todas as variáveis sendo o GA ligeiramente inferior na maioria das variáveis. Notou-se que a intensidade teve relevância na contribuição do ganho de força em quase todos os exercícios, já o volume foi relevante

para respostas na massa muscular. Além disso, a intensidade e o volume mostraram-se relevantes para resultados além da alteração mínima detectável. O protocolo com intensidade autosselecionada demonstrou ser eficaz para ganho de força, massa muscular e melhora no DF. Contudo, não foi possível estabelecer equivalência como o protocolo tradicionalmente recomendado pelo ACSM. A intensidade apresentou-se como variável importante para o ganho de força muscular, adicionalmente o volume foi relevante para respostas na massa muscular. Além disso, as respostas além da AMD mostraram-se influenciadas tanto pela intensidade quanto pelo volume.

Palavras-chave: Envelhecimento; Treinamento com Peso; Menopausa

ABSTRACT

In the aging process, there is a natural decline in muscle mass, strength, and physical performance, associated with risks such as increased susceptibility to falls, sarcopenia, and mortality. Strength training (ST) emerges as an alternative to mitigate these effects, particularly relevant for women over 40, an age group where a reduction in strength and muscle mass is already observed. The intensity of ST is a key variable, with recommendations from the American College of Sports Medicine (ACSM) ranging between 60% and 80% of one-repetition maximum (1RM). However, literature evidence suggests significant gains with lower intensities. Despite the substantial benefits of ST, adherence in this population is challenging. Intensity self-selection arises as an alternative to improve adherence. Objective to determine whether a protocol with self-selected intensity yields equivalent results in muscle mass, strength, and physical performance compared to a protocol with loads established by ACSM recommendations. A randomized clinical trial involving 60 women over 40 undergoing two 12-week strength training (ST) protocols. At the beginning, questionnaires on cognitive function (MOCA), physical activity (IPAC), quality of life (SF36), and sociodemographic aspects were administered. Before and after the intervention, physical performance tests (TUG, LS, 6MWT), muscular strength assessment (1RM), and body composition were conducted. After the initial assessment, participants were randomized into two groups: the traditional group (TG), which trained with an imposed intensity of 50-80% of 1RM, and the self-selected intensity group (SG), with autonomy

to choose the intensity. The repetition reserve (RR) scale was collected throughout the intervention. The intensity selected by SG was substantially similar to TG. However, differences were observed in the bench press ($p=0.017$), leg extension ($p=0.011$), and low row ($p=0.018$) exercises, where TG demonstrated superiority. Volume showed differences only in leg press ($p=0.008$) and total volume ($P=0.050$). In the time effect, all variables had significant results except strength in the leg curl exercise ($p=0.183$), TUG ($p=0.169$), and arm muscle mass ($p=0.530$). There was no group effect or interaction. It was not possible to establish total equivalence of the protocols in all variables, with SG slightly inferior in most variables. It was noted that intensity was relevant for strength gain in almost all exercises, while volume was relevant for muscle mass responses. Additionally, both intensity and volume were relevant for results beyond the minimum detectable change. The protocol with self-selected intensity proved effective for gaining strength, muscle mass, and improving physical performance. However, equivalence with the traditionally recommended ACSM protocol was not established. Intensity emerged as an important variable for muscle strength gain, and volume was relevant for muscle mass responses. Furthermore, responses beyond the minimum detectable change were influenced by both intensity and volume.

Keywords: Aging; Resistance Training; Menopause

INTRODUÇÃO

No processo de envelhecimento, ocorre inevitavelmente a diminuição da massa muscular, da força e do DF funcional (CRUZ-JENTOFT; SAYER, 2019). A redução dessas variáveis está correlacionada a condições de vulnerabilidade, tais como a predisposição a quedas, o desenvolvimento de sarcopenia até mesmo aumento nas taxas de mortalidade. (ROSENBERG, 1997). Frente a isto, o TF é uma alternativa para atenuar os efeitos do envelhecimento em marcadores de DF, força e massa muscular. Nesse sentido, a mulher a partir dos 40 anos pode se beneficiar substancialmente dos efeitos do TF, uma vez que essa faixa etária compreende o período denominado transição menopausal, onde observa-se a redução em média de 0,06kg de massa muscular por ano (GREENDALE et al., [s.d.]). Além disso, a força muscular e o DF na mulher apresenta mudanças deletérias cerca de quase 50% mais acentuadas que nos homens na mesma faixa etária de idade (RHODES et al., 1999; SAMSON et al., 2000b).

Nesse intuito, estudos e intervenções na prática clínica profissional são desenvolvidos com o TF a fim de maximizar o efeito do mesmo sobre marcadores de força e DF. Para tanto, os testes de força e de DF são utilizados para verificar o efeito desse tipo de intervenção, visto que os mesmos são ferramentas acessíveis e eficazes para monitorar os impactos de ensaios clínicos com TF. Contudo, é relevante verificar efeitos que extrapolam os vieses de medidas advindas dessas ferramentas, considerando que alterações nos resultados desses testes que ficam abaixo da Mudança Mínima Detectável (MMD) podem indicar que mulheres de meia idade e mais velhas não experimentem uma mudança real das variáveis avaliadas (ASSIS SILVA et al., 2024). Essas medidas fornecem resultados fidedignos sobre marcadores de força e DF, dados esses, que ainda não são contemplados nas pesquisas e na prática clínica, porém relevantes para estabelecer melhores modelos para prescrição segura e eficaz do TF principalmente quando se trata da manipulação de variáveis como a intensidade ideal para idosos e mulheres na transição menopausal e na pós menopausa.

Assim, a intensidade do TF é uma das variáveis que é manipulada para promover maiores benefícios à saúde, como aumento da força muscular e do DF. Afim de fornecer parâmetros de prescrição segura e eficiente para melhoras significativas em marcadores de força, massa muscular e DF, o American College of Sports Medicine (ACSM) propõe diretrizes que preconizam uma intensidade situada entre 60% e 80% da repetição máxima (1RM) (AMERICAN COLLEGE OF SPORTS MEDICINE, 2009). Essa janela de intensidade já demonstrou resultados relevantes para marcadores de força e composição corporal em idosos, sobretudo em mulheres na pós menopausa (ORSATTI et al., 2008; CAVALCANTE et al., 2023).

Embora a recomendação do ACSM seja pertinente e benéfica para o aumento da massa muscular, força, densidade mineral óssea e DF (SINAKI; OFFORD, 1988; ORSATTI et al., 2008; SANDERS et al., 2020; CARNEIRO et al., 2021), a adesão a programas de TF na população de mulheres acima de 40 anos ainda é desafiadora, o que pode ser explicado por diversos fatores como a fragilidade física, problemas de saúde e respostas sensíveis, especialmente relacionadas à percepção de desconforto resultante da alta intensidade do exercício (ARIKAWA; O'DOUGHERTY; SCHMITZ, 2011), uma vez que as intensidades no TF recomendadas pelo ACSM abrange uma janela de alta intensidade, o que pode representar um baixo nível de adesão (GARCIA-HERMOSO et al., 2023). Desta forma, se faz necessário investigar protocolos com intensidade mais baixa com o intuito de melhorar a tolerância ao TF e, com isso, aumentar a aderência considerando que já existem evidências associadas a ganhos substanciais na força e massa muscular em protocolos com baixa intensidade (KAK et al., 2013; NASCIMENTO DE OLIVEIRA-JÚNIOR et al., 2022).

Uma alternativa estabelecida a fim de reduzir a percepção de desconforto e aumentar a aderência ao TF é a autosseleção da intensidade, pois realizar atividades em intensidade autosselecionada parece melhorar a aderência ao treinamento por fatores como o respeito à individualidade biológica, maior percepção de prazer e menor percepção de esforço (PARFITT; HUGHES, 2009). Além disso, este protocolo já demonstrou melhorias na composição corporal, no metabolismo e no DF (ALVES et al., 2022). Assim, com o objetivo de promover maior adesão ao TF, têm-se explorado e implementado estratégias que envolvam a autosseleção da intensidade.

Contudo, ainda não estão claras as respostas deste protocolo quanto aos efeitos crônicos comparados com a recomendação do ACSM sobre marcadores de massa muscular, força e DF. Nota-se ainda que não há evidências dos efeitos de ambos os protocolos além das medidas de erros sistemáticos, denominado alteração mínima detectável.

Portanto, o objetivo deste estudo foi verificar se o protocolo com intensidade autosselecionada promove resultados equivalentes sobre marcadores de massa muscular, força e DF quando comparado ao protocolo de treinamento com intensidades estabelecidas pelo ACSM, bem como verificar a associação da intensidade, volume e frequência sobre os efeitos independentemente do protocolo, além de identificar respondedores além das medidas de MMD.

METODOLOGIA

Ensaio clínico, randomizado, sobre o número CAAE: 14551619.3.0000.5154 parecer número: 5.654.381, desenvolvido com mulheres acima de 40 anos submetidas a dois protocolos de TF por 12 semanas. Um dos protocolos foi constituído por intensidade imposta e outro por intensidade autosselecionada pelas voluntárias. Ao início das avaliações, foram aplicados os seguintes questionários: questionários de função cognitiva Montreal Cognitive Assessment (MOCA), Índice de Prática de Atividade Física (IPAC), qualidade de vida pelo Short Form 36 Health Survey (SF36) e aspectos sociodemográficos. Antes e após a intervenção, foram realizadas as seguintes avaliações: testes de DF: Timed-Up-and-Go (TUG), levantar e sentar (LS) e teste de caminhada de 6 minutos (6MWT). Para mensurar a força muscular, foi realizado o teste de uma repetição máxima (1RM). Houve avaliação da composição corporal. Após período de avaliação houve a randomização em dois grupos: grupo tradicional (GT) treinou com uma intensidade imposta de 50-80% de 1RM e grupo de carga autosselecionada (GA), que tiveram autonomia para escolher a intensidade, desde que seguisse a seguinte orientação: “Escolher uma intensidade que fosse confortável, mas que lhe permitisse sentir um mínimo de esforço”.

Participantes

As participantes foram convidadas por meio de veículos midiáticos no município de Uberaba-MG. Foi aplicada uma entrevista que envolveu a coleta de informações para caracterização, incluindo idade, nível sociodemográfico, indicadores de saúde, histórico médico passado e presente, intervenções terapêuticas e atividades físicas.

No total, 91 voluntárias foram randomizadas de acordo com os seguintes critérios de inclusão: I- mulheres acima de 40 anos, II- não fazer uso de terapia hormonal ou suplementos dietéticos, III- pressão arterial e glicemia controladas, IV- não apresentar distúrbios musculares, tromboembólicos ou gastrointestinais, V- não ter histórico de doenças cardiovasculares ou infecciosas, VI- não consumir álcool- VII- não ser tabagista, VIII- ter experiência prévia com exercícios de força, mas não estar envolvida em atividades físicas regulares mais de uma vez por semana nos últimos 6 meses. Os critérios de exclusão foram I- Idade inferior a 40 anos II- uso de terapia hormonal ou suplemento dietético, III- problemas cardíacos IV- tabagista e/ou etilista V- ter se envolvido com TF nos últimos seis meses.

Teste de uma repetição máxima (1RM)

Durante o período de familiarização, os participantes realizaram três sessões de treinamento por semana, nas segundas-feiras, quartas-feiras e sextas-feiras. As sessões de treinamento tinham como objetivo familiarizar os participantes com as técnicas dos exercícios. As sessões ocorreram em dias alternados, e as participantes treinaram com cargas leves, evitando a falha concêntrica voluntária. Os exercícios incluídos no período de familiarização foram: leg press 45°, supino reto, cadeira extensora, pec deck, puxada ampla pela lateral, mesa flexora e remada baixa.

Após o período de familiarização, foram realizados testes de 1RM para todos os exercícios no mesmo dia, com um intervalo de 5 minutos entre cada exercício. Antes do teste, as participantes realizaram um aquecimento geral, caminhando por 5 minutos, seguido por um aquecimento específico usando cargas subjetivas, selecionadas e determinadas durante o período de familiarização. O aquecimento específico envolveu a realização de uma série de 12 repetições com cargas leves. Após um período de descanso de 90 segundos, a carga foi aumentada em cerca de 20% e as participantes realizaram entre três e cinco repetições. Após um período de descanso de 3 a 5 minutos, a carga foi significativamente aumentada e as

participantes foram encorajadas a fazer o máximo esforço e utilizar toda a amplitude de movimento. Se a carga foi superestimada ou subestimada, as participantes descansavam por 3 a 5 minutos antes de tentar uma nova repetição com uma carga mais baixa ou mais alta, respectivamente (aproximadamente uma diferença de 5% a 10%). Esse processo foi repetido de duas a cinco vezes até que a carga equivalente a 1RM fosse determinada.

A carga na qual as participantes não puderam realizar mais do que uma repetição foi considerada como sua 1RM para cada exercício. Caso não tenha sido possível estabelecer a carga para uma única repetição, a fórmula de Brzycki [$1RM = \text{peso levantado} / (1.0278 - 0.0278 * \text{repetições executadas})$] foi usada como alternativa.

A sessão de reteste seguiu o mesmo protocolo do teste inicial, com um intervalo mínimo de 48 horas entre as sessões de teste e reteste.

Teste Timed-Up-and-Go (TUG)

Para realizar o teste TUG, a voluntária foi posicionada em uma cadeira com as costas apoiadas, os pés totalmente apoiados no chão e as mãos descansando nas coxas. Partindo da posição sentada, a voluntária devia levantar-se da cadeira, caminhar por uma distância de 3 metros em um piso plano, contornar um cone e retornar à cadeira, devendo ser realizado o mais rápido possível sem correr e sem trotar. O tempo decorrido durante toda a sequência, desde o movimento inicial de levantar-se da cadeira até o retorno e sentar novamente foi registrado com precisão de 0,01 segundos usando um cronômetro. Cada participante realizou o teste uma vez como um ensaio de familiarização e duas tentativas foram registradas para análise. A média das duas tentativas foi calculada como o valor representativo. Para o reteste, o mesmo protocolo foi aplicado após um intervalo de 72 horas.

Teste levantar e sentar de 30 segundos (LS)

O teste consiste em sentar e levantar por 30 segundos. Para isso, foi utilizada uma cadeira com altura de 47 cm, de onde a voluntária devia iniciar o teste, partindo da posição sentada. Foi assegurado que suas costas estivessem em contato com o encosto da cadeira, seus pés totalmente apoiados no chão e suas mãos cruzadas,

descansando nos ombros opostos. O teste envolveu uma pessoa registrando o tempo e outra pessoa contando o número de repetições realizadas pela participante. Cada participante teve de 1 a 2 tentativas com um período de descanso apropriado antes do teste final. Após um intervalo de 72 horas, o reteste foi conduzido seguindo o mesmo protocolo.

Teste de caminhada de seis minutos (6MWT)

Para avaliar a velocidade de marcha, foi utilizado o teste de caminhada de 6 minutos (6MWT), realizado em uma superfície plana. A rota de caminhada foi demarcada por dois cones, criando uma distância total de 20 metros. As participantes receberam instruções para caminhar de um cone para o outro, completando um percurso total de 40 metros, de ida e volta, até completar o tempo de teste de 6 minutos. Durante o 6MWT, as participantes foram encorajadas a caminhar o mais rapidamente possível, sem fazer pausas ou correr, e receberam encorajamento verbal ao longo do teste, utilizando frases como "você está se saindo muito bem" e "caminhe o mais rápido que puder". A distância percorrida por cada participante foi registrada ao final do 6MWT. Após um intervalo de 72 horas, o reteste foi conduzido utilizando o mesmo protocolo do teste inicial para garantir condições de teste consistentes e avaliar a confiabilidade dos resultados.

Escala de repetições de reserva (RR)

As repetições de reserva (RR), foram uma tentativa subjetiva de diagnosticar o quão perto a voluntária chegou da falha concêntrica. A aplicação foi verbal seguida pela pergunta "Quantas repetições você supostamente conseguiria realizar além das oitos executadas, com esta mesma carga?". Os valores respondidos pela voluntária foram anotados em cada série de cada exercício uma vez na semana.

Protocolo de intensidade imposta

A intervenção teve duração de 12 semanas e consistiu em TF, com carga imposta entre 50-80% de 1RM. As voluntárias do GT receberam orientações de profissionais

em cada sessão de treinamento. O protocolo consistiu em 3 sessões semanais, em dias alternados, com a realização de 3 séries de 8 repetições para os 7 exercícios (leg press 45°, supino reto, cadeira extensora, pec deck, puxador frontal, mesa flexora e remada baixa). Durante o protocolo, foi coletada a escala Repetições de Reserva (RR), após cada série em todos os exercícios, uma vez por semana.

Protocolo de intensidade autosselecionada

A intervenção teve duração de 12 semanas e consistiu em TF, com carga autosselecionada. As voluntárias do grupo GA receberam orientações de profissionais em cada sessão de treinamento. O protocolo consistiu em 3 sessões por semana, em dias alternados, com a realização de 3 séries de 8 repetições para os 7 exercícios (leg press 45°, supino reto, cadeira extensora, pec deck, puxador frontal, mesa flexora e remada baixa). Durante o protocolo, foi coletada a escala Repetições de Reserva (RR), após cada série em todos os exercícios, uma vez por semana.

Composição Corporal

Para avaliar a composição corporal, foi utilizada a densitometria de dupla emissão de raios X (DEXA), utilizando a técnica recomendada pelo fabricante, por meio da qual foi possível avaliar a densidade óssea, distribuição de gordura e massa magra corporal. A voluntária foi instruída a usar roupas leves, sem objetos metálicos, jóias ou acessórios que pudessem interferir no exame. Em seguida, a voluntária foi posicionada em decúbito dorsal na mesa de exame. O avaliador certificou-se que a voluntária estivesse corretamente alinhada e imóvel durante o procedimento.

Análise Estatística

Os dados foram submetidos ao teste de normalidade Shapiro-Wilk. O teste t de amostra independente foi utilizado para caracterizar a amostra, bem como para analisar a carga e o volume empregados ao longo da intervenção. Na ausência de normalidade, optou-se pelo teste não paramétrico U-Mann-Whitney.

Para avaliação do efeito do tempo, grupo e interação tempo pelo grupo, recorreu-se a uma ANOVA de medidas repetidas. Em seguida, a análise de covariância (ANCOVA) foi empregada para avaliar a interação, ajustada pela idade, frequência, MOCA e valores do momento pré. O teste TOSTER foi usado para testar a equivalência entre grupos.

Foi conduzida correlação de Spearman, controlada pela idade, para verificar a associação entre intensidade, volume, RR e frequência com os deltas das variáveis de desfecho. Em seguida, as voluntárias foram classificadas como respondedoras e não respondedoras baseadas na alteração mínima detectável (AMD). A técnica de AMD está apresentada no artigo “Assessing the robustness of muscle strength and physical performance measures in women over 40: a test-retest reliability study” (anexo 01). Para verificar a associação de intensidade, volume, RR e frequência com a responsividade do TF foi realizada uma regressão logística binomial. Para significância estatística foi considerado $p < 0.05$ e as análises foram conduzidas no software Jamovi®.

RESULTADOS

Caracterização

Na tabela 1 são apresentados dados de variáveis para a caracterização da amostra com os valores de média e desvio padrão.

Tabela 1 - Caracterização da amostra.

Caracterização	GT(33)	GA(27)	p
Idade, anos	60.7±10.3	59.5±10.1	0.657
IMC, kg/m ²	27.8±6.7	29.3±6.9	0.240
Tempo de Menopausa, anos [*]	13.0±9.8	10.9±10.9	0.343
MOCA	18.0±4.64	20.5±4.44	0.064
Dinamometria [*]	25.5±4.69	26.1±7.06	0.388
Doenças	GT(33)	GA(27)	p
Reumatismo	4(12.1%)	4(14.8%)	1.000
Artrite/Osteoartrite	11(33.3%)	10(37.0%)	0.792
Esclerose Múltipla	0	1(3.7%)	0.450

Hipotiriodismo	7(21.2%)	5(18,5%)	1.000
Diabetes tipo 2	5(15.1%)	6(22.2%)	0.511
Problemas de Coluna	9(27.2%)	10(37%)	0.410
Tumores Malignos	4(12.1%)	3(11,1%)	1.000
Covid19	14(42.4%)	17(63.0%)	0.116

IMC = Índice de massa corporal, (*) = U-Mann-Whitney, MOCA = Montreal Cognitive Assessment, GT = Grupo Tradicional, GA = Grupo autosselecionado

A seleção da intensidade pelo GA ao longo das 12 semanas de intervenção mostrou-se estatisticamente semelhante ao GT com carga imposta entre 50-80% de 1RM na maioria dos exercícios. No entanto, foram observadas diferenças nos exercícios supino ($p=0,017$), cadeira extensora ($p=0,011$) e remada baixa ($p=0,018$), nos quais o GT demonstrou uma superioridade nas cargas de treinamento ao longo da intervenção.

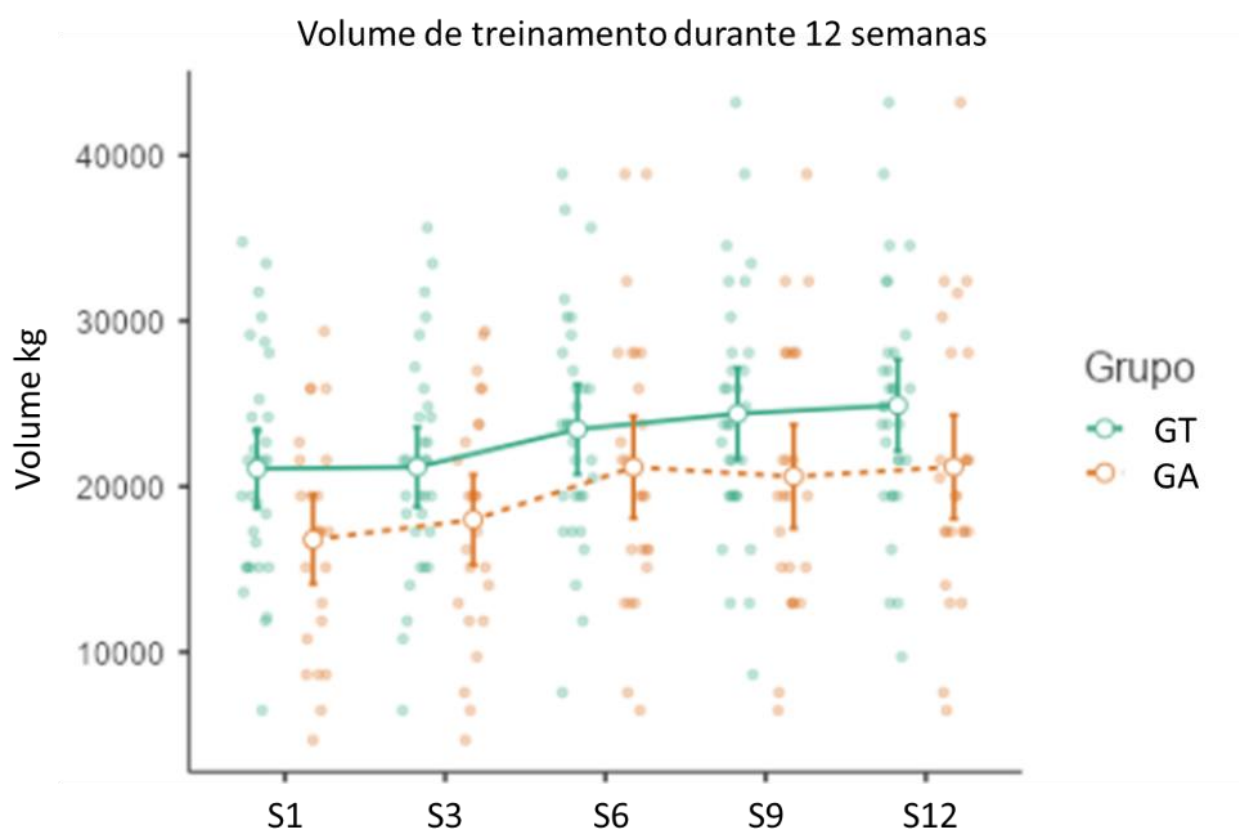
Adicionalmente, o volume total de treinamento calculado com base na carga e na quantidade de repetições ao longo das 12 semanas apresentou diferença apenas no exercício leg press ($p=0,008$). No entanto, o volume total de treinamento, considerando todos os exercícios, mostrou-se significativamente superior no GT ($p=0,050$). Essas diferenças podem ser atribuídas ao percentual de frequência, que também foi maior no GT ($p=0,030$).

Tabela 2: Dados de intensidade e volume de treinamento ao longo da intervenção.

Intensidade %1RM			
	GT	GA	p
Leg press	63±11.2	62.2±20.0	0.500*
Supino reto	70.7±11.0	63.1±11.7	0.017
Cadeira extensora	73.8±24.3	59.8±15.3	0.011*
Mesa flexora	61.1±16.7	61.4±15.6	0.954
Puxada frontal	65.1±14.4	62.7±12.0	0.528
Remada baixa	57.1±15.5	47.3±15.0	0.018*
Pec deck	77.5±24.2	79.3±22.4	0.771*
Média percentual de todos exercícios	65.6±10.3	62.5±11.5	0.308
Volume kg/1000			
Leg press	238.8±82.9	181.5±70.5	0.008
Supino reto	41.6±13.0	37.3±10.6	0.180

Cadeira extensora	36.6±14.9	32.0±11.4	0.281*
Mesa flexora	25.9±10.3	23.5±10.8	0.405
Puxada frontal	51.7±13.7	48.6±14.7	0.413
Remada baixa	52.5±15.8	47.4±15.6	0.233
Pec deck	30.3±11.4	32.2±10.9	0.532
Média de todos exercícios, Kg	477.4±146.6	402.5±131.5	0.050
Frequência%	77.0±17.6	68.9±12.8	0.030*

Nota: (*) = U-Mann-Whitney: 1RM = Teste de uma repetição máxima, GT = Grupo Tradicional, GA = Grupo autosselecionado



Nota: S1= Semana 1; S3= Semana 3; S6= Semana 6; S9= Semana 9; S12= Semana 12; GT= Grupo Tradicional; GA= Grupo Autosselecionado.

Resultado do efeito tempo, grupo e interação

Embora a intensidade e volume de treino tenham apresentado diferenças relevantes, a tabela 3 expõe dados relativos ao efeito do tempo, grupo e interação.

Observa-se que houve resultados significativos para o efeito do tempo nas variáveis de força, DF e massa muscular em ambos os grupos, com valores de $p < 0,05$. Contudo, a variável de força no exercício mesa flexora não apresentou efeito do tempo ($p = 0,183$). Para desfechos de DF, o teste TUG foi o único que não apresentou efeito do tempo ($p = 0,169$). Por fim, a Massa Muscular de braços (MMB) não demonstrou efeito significativo de tempo ($p = 0,530$).

No efeito entre grupos e interação, não foram encontradas diferenças significativas em nenhuma variável relacionada à força muscular, DF e massa muscular. Além disso, os resultados não demonstraram distinções entre os protocolos em nenhuma variável, mesmo após ajustes considerando o valor do momento pré, idade, frequência e avaliação cognitiva pelo MOCA sugerindo que, independentemente da massa muscular, força e DF no início do estudo, o efeito ao longo do tempo entre os protocolos foi semelhante, assim como a idade, frequência e o possível declínio cognitivo rastreado pelo MOCA.

Tabela 3: Valores do momento pré e pós de ambos os grupos das variáveis de força, desempenho físico e composição corporal.

	GT-GA	GT		GA		P _t	P _g	P _i	P ancova
Massa Muscular	n	Pré	Pós	Pré	Pós				
MMP, kg	33-27	13.1 (12.0-14.3)	13.6 (12.5-14.7)	13.9 (12.6-15.2)	14.3 (13.0-15.5)	<.001	0.388	0.446	0.628
MMB, kg	33-27	4.02 (3.66-4.38)	3.93 (3.59-4.26)	4.26 (3.86-4.66)	4.29 (3.92-4.66)	0.530	0.245	0.279	0.489
IMM, kg	33-27	6.73 (6.25-7.22)	6.90 (6.43-7.36)	7.12 (6.59-7.65)	7.23 (6.72-7.74)	0.002	0.308	0.535	0.808
MMT, kg	33-27	37.8 (35.3-40.3)	39.9 (37.1-42.7)	38.1 (35.7-40.5)	40.1 (37.5-42.8)	0.145	0.265	0.994	0.971
Força Máxima	n	Pré	Pós	Pré	Pós				
1RMsupino, kg	27-22	26.6 (24.1-29.1)	28.0 (25.1-30.8)	28.2 (25.4-30.9)	30.0 (26.9-33.2)	0.016	0.381	0.734	0.352
1RMLeg, kg	30-23	149 (129-168)	175 (156-194)	153 (131-175)	173 (151-194)	<.001	0.944	0.457	0.282
1RMFlexiora, kg	29-22	18.3 (15.4-21.2)	19.9 (17.5-22.4)	19.5 (16.2-22.9)	20.0 (17.2-22.7)	0.183	0.735	0.415	0.609
1RMRemada, kg	30-23	38.2 (34.0-42.3)	41.9 (38.4-45.4)	41.5 (36.8-46.3)	42.8 (38.8-46.8)	0.036	0.427	0.296	0.331
1RMPec, kg	30-23	18.3 (15.3-21.3)	26.1 (22.4-29.9)	21.5 (18.1-24.9)	26.5 (22.2-30.7)	<.001	0.464	0.111	0.203
1RMExtensora, kg	31-23	23.9 (20.5-27.3)	29.4 (25.9-32.9)	27.4 (23.4-31.3)	30.5 (26.4-34.6)	<.001	0.366	0.137	0.297
1RMPuxada, kg	30-23	32.8 (29.5-36.2)	38.4 (35.0-41.7)	35.4 (31.6-39.2)	38.8 (35.0-42.7)	<.001	0.527	0.182	0.068
1RM total, kg	27-19	319 (283-355)	367 (331-402)	343 (300-387)	376 (33-419)	<.001	0.609	0.355	0.453
Desempenho Físico	n	Pré	Pós	Pré	Pós				
6MWT, m	33-25	508 (485-532)	520 (494-545)	504 (476-531)	519 (489-549)	0.023	0.887	0.714	0.534
TUG, seg	33-27	7.04 (6.48-7.60)	6.94 (6.35-7.54)	7.67 (7.06-8.29)	7.54 (6.88-8.20)	0.169	0.151	0.824	0.361
SL, rep	33-27	15.9 (14.7-17.0)	18.3 (16.6-19.9)	15.6 (14.3-16.8)	17.2 (15.4-19.0)	<.001	0.480	0.390	0.279

Nota: MMP= Massa muscular de pernas; MMB= Massas muscular de braços; IMM= Índice de massa muscular; MMT= Massa muscular total; 1RM= Teste de uma repetição máxima; 6MWT= Teste de caminhada de 6 minutos; TUG= Teste Timed-Up-and-Go; SL= Teste de levantar e sentar, GT = Grupo Tradicional, GA = Grupo autosselecionado

Equivalência

A Tabela 4 apresenta estimativas referentes ao efeito de tempo de cada variável do grupo GA em comparação com o grupo GT. Não houve superioridade de um protocolo sobre o outro. No entanto, destaca-se que o protocolo com carga autosselecionada exibe valores negativos quando comparados aos resultados do GT. Adicionalmente, em quase todas as variáveis relacionadas à massa muscular, força e DF, os limites inferiores do intervalo de confiança de 90% ultrapassam a janela de equivalência, sugerindo que, embora pequena, existe uma inferioridade do GA em comparação com o GT. Sendo assim, não há equivalência total entre os protocolos.

É relevante observar que as variáveis Massa Muscular do Braço (MMB), TUG e 1RM no legpress apresentaram resultados positivos e os valores do limite superior do intervalo de confiança de 90% ultrapassam a janela de equivalência. Portanto, mesmo em pequena escala, há uma superioridade do GA sobre o GT nessas variáveis específicas.

Tabela 4 - Valores de estimativa com IC90% do teste t-testes de equivalência

	Estimativa de efeito Cohen's d (IC90%)	Estimativa de efeito Raw (IC90%)	p. t- teste	p. teste lower	p. teste upper
Massa Muscular					
MMP, kg	-0.335 (-0.796; 0.105)	-0.211 (-0.496; 0.0739)	0.220	0.002	0.271
MMB, kg	0.211 (-0.231; 0.665)	0.0895 (-0.0988; 0.278)	0.430	0.140	0.005
IMM, kg	-0.305 (-0.768; 0.138)	-0.0920 (-0.229; 0.0452)	0.266	0.002	0.238
MMT, kg	-0.121 (-0.573; 0.323)	-0.131 (-0.627; 0.364)	0.658	0.013	0.084
Desempenho Físico					
TUG	0.0676 (-0.377; 0.516)	-0.0430 (-0.242; 0.328)	0.801	0.056	0.019
SL	-0.208 (-0.664; 0.233)	-0.729 (-2.31; 0.856)	0.444	0.006	0.143
6MWT	-0.147 (-0.606; 0.298)	-8.30 (-34.7-18.1)	0.599	0.013	0.106
Força					
Leg press	0.0275 (-0.453; 0.510)	0.922 (-15.4; 17.3)	0.925	0.055	0.038
Supino reto	-0.0799 (-0.571; 0.406)	-0.348 (2.46; 1.77)	0.783	0.025	0.076
Cadeira extensora	-0.322 (-0.813; 0.147)	-1.78 (-4.37; 0.814)	0.256	0.003	0.264
Mesa flexora	-0.349 (-0.852; 0.131)	-1.81 (-4.31; 0.693)	0.231	0.002	0.300
Puxada frontal	-0.360 (-0.858; 0.113)	-2.61 (-4.40; 0.646)	0.218	0.002	0.314

Remada baixa	-0.205 (-0.697; 0.270)	-1.68 (-5.70; 2.34)	0.485	0.010	0.159
Pec deck	-0.455 (-0.959; 0.0167)	-2.91 (-6.03; 0.198)	0.123	<.001	0.438
Carga total	-0.228 (-0.790; 0.306)	-12.0 (-42.2; 18.2)	0.502	0.020	0.212

Tabela 4 1RM: Teste de uma repetição máxima; 6MWT: Teste de caminhada de 6 minutos; TUG: Teste Timed-Up-and-Go; SL: Teste de levantar e sentar, MMP = Massa muscular pernas, MMB = Massa muscular braços, IMM = Índice de massa muscular, MMT = Massa muscular total

Correlação

A Tabela 5 avança na análise dos resultados, relacionando a intensidade, volume, RR e a frequência ajustada pela idade, a fim de estabelecer quais dessas variáveis apresentam influência nas respostas do ganho de força, DF e massa muscular. Assim, é possível observar que a intensidade teve relevância na resposta da força, conforme estabelecido pelo 1RM em vários exercícios, como leg press ($\rho=0,552$, $p<0,001$), cadeira extensora ($\rho=0,311$, $p=0,028$), puxada ($\rho=0,313$, $p=0,032$), remada baixa ($\rho=0,296$, $p=0,039$) e pec deck ($\rho=0,362$, $p=0,011$). Além disso, o exercício pec deck também mostrou que o volume ($\rho=0,454$, $p=0,001$) e a frequência ($\rho=0,298$, $p=0,037$) desempenharam um papel relevante no delta da força nesse exercício. No entanto, o exercício supino e mesa flexora não tiveram a intensidade como uma variável relevante no efeito do tempo, assim como nenhuma outra variável incluída no modelo mostrou-se relevante para a resposta no ganho de força nesses exercícios.

Entretanto, o volume de treino, em relação à massa muscular, mostrou-se relevante no efeito ao longo do tempo na massa muscular de pernas (MMP) ($\rho=0,274$, $p=0,041$), massa muscular total (MMT) ($\rho=0,270$, $p=0,044$) e índice de massa muscular (IMM) ($\rho=0,303$, $p=0,025$).

Além disso, o delta obtido nos testes de DF não apontou nenhuma variável como relevante no desfecho, exceto o teste LS, que mostrou uma relação inversamente proporcional com a variável de RR ($\rho= -0,272$, $p=0,042$).

Tabela 5- Valores da correlação de Pearson das variáveis para intensidade, volume, repetições de reserva e frequência.

Força	%1RM	Volume	RR	Frequência
-------	------	--------	----	------------

	0.552	0.091	0.067	0.230
Leg press	p<.001	p=0.533	p=0.652	p=0.112
	0.280	0.236	-0.004	0.033
Supino reto	p=0.060	p=0.110	p=0.980	p=0.824
	0.311	0.140	0.064	0.080
Cadeira Extensora	p=0.028	p=0.332	p=0.661	p=0.580
	0.225	0.157	0.005	0.204
Mesa Flexora	p=0.128	p=0.286	p=0.978	p=0.164
	0.313	0.267	-0.029	0.042
Puxada Frontal	p=0.032	p=0.063	p=0.842	p=0.776
	0.296	0.005	0.098	0.162
Remada Baixa	p=0.039	p=0.974	p=0.503	p=0.266
	0.362	0.454	-0.104	0.298
Pec deck	p=0.011	p=0.001	p=0.478	p=0.037
Desempenho físico	%1RM	Volume	RR	Frequência
	0.031	0.019	0.087	0.105
TUG	p=0.828	p=0.888	p=0.523	p=0.439
	-0.167	-0.011	0.122	-0.173
6MWT, m	p=0.242	p=0.935	p=0.369	p=0.202
	0.184	0.049	-0.272	-0.163
SL	p=0.196	p=0.721	p=0.042	p=0.231
Massa Muscular	%1RM	Volume	RR	Frequência
	0.059	0.274	-0.022	0.158
MMP	p=0.680	p=0.041	p=0.872	p=0.244
	-0.091	0.133	0.067	0.158
MMB	0.527	p=0.329	p=0.623	p=0.245
	0.044	0.303	-0.021	0.228
IMM	p=0.761	p=0.025	p=0.877	p=0.094
	0.025	0.270	-0.017	0.226
MMT	p=0.861	p=0.044	p=0.898	p=0.095

Nota: 1RM= Teste de uma repetição máxima; 6MWT= Teste de caminhada de 6 minutos; TUG= Teste Timed-Up-and-Go; SL= Teste de levantar e sentar MMP = Massa muscular pernas, MMB = Massa muscular braços, IMM = Índice de massa muscular, MMT = Massa muscular total, RR = Repetições de reserva

Respondedoras além da medida de Alteração Mínima Detectável

A partir da análise da Alteração Mínima Detectável (AMD) realizada na amostra deste ensaio clínico randomizado, conduzido em um estudo de confiabilidade teste-reteste intitulado "Avaliando a robustez das medidas de força muscular e DF em mulheres com mais de 40 anos: um estudo de confiabilidade teste-reteste" (anexo 01), determinamos a quantidade de respondedoras que excederam as medidas de AMD em todos os testes de força e DF.

A análise das respondedoras está representada na Tabela 6 com valores absolutos e relativos da quantidade de respondedoras além da AMD. Observa-se que não houve superioridade de um protocolo sobre o outro em relação às respondedoras além da AMD. No entanto, foi possível verificar que no teste TUG não houve respondedoras além da AMD e, nos testes de força, especificamente os testes de 1RM na mesa flexora e na remada, não foram apresentadas respondedoras, ambos no GA. Não foram encontradas diferenças significativas entre os protocolos devido ao reduzido número de respondedoras encontradas no GT. Esse resultado pode ser explicado pela baixa intensidade média aplicada nos exercícios de mesa flexora e remada baixa em cada grupo ao longo da intervenção, conforme observado na Tabela 2.

Tabela 6- Quantidade absoluta e percentual de respondedoras além da MMD.

Força	GT	GA	GT+GA	P
Leg press	18(60.0%)	12(52.2%)	30(56.6%)	0.772
Supino reto	7(25.9%)	7(31.8%)	14(28.6)	0.892
Cadeira extensora	10(32.3%)	2(8.7%)	12(22.2%)	0.084
Mesa flexora	1(3.4%)	0(0.0%)	1(2.0%)	1.000
Puxada frontal	9(30.0%)	4(17.4%)	13(24.5%)	0.462
Remada baixa	5(16.7%)	0(0.0%)	5(9.4%)	0.113
Pec deck	12(40.0%)	6(26.1%)	18(34.0%)	0.443
Desempenho Físico	GT	GA	GT+GA	P
TUG, seg	1(3.0%)	0(0.0%)	1(1.7%)	1.000

6MWT, m	4(12.1%)	4(14.8%)	8(13.3%)	1.000
SL, rep	6(18.2%)	4(14.8%)	10(16.7%)	1.000

Nota: 1RM = Teste de uma repetição máxima; 6MWT = Teste de caminhada de 6 minutos; TUG = Teste Timed-Up-and-Go; SL= Teste de levantar e sentar, GT = Grupo tradicional, GA = Grupo autosselecionado

Após a identificação das respondedoras além da MMD, foi conduzida análise de regressão logística para identificar possíveis contribuintes para respostas superiores aos valores de MMD. A Tabela 7 apresenta as contribuições da intensidade, mostrando-se relevante nos exercícios leg press (OR=1,08, p=0,005) e puxada (OR=1,07, p=0,033) para alcançar resultados superiores à MMD. Além disso, observou-se que o volume teve papel significativo nas respostas que ultrapassaram a MMD no exercício de supino (OR=1,12, p=0,009), pec deck (OR=1,08, p=0,018) e puxada frontal (OR=1,06, p=0,050). No entanto, as variáveis relacionadas às RR, bem como a frequência, não demonstraram poder estatístico significativo na contribuição para resultados superiores à AM

Tabela 7- Regressão logística de cada variável com intensidade, volume, repetições de reserva e frequência.

	%1RM			VOLUME			RR			Frequência		
	OR	IC	P	OR	IC	P	OR	IC	P	OR	IC	P
Leg press	1.082	1.02-1.14	0.005	0.999	0.99-1.01	0.829	0.973	0.57-1.65	0.919	1.020	0.97-1.06	0.346
Supino reto	1.020	0.95-1.09	0.524	112.857	1.03-1.24	0.009	1.092	0.62-1.90	0.753	1.012	0.96-1.06	0.628
Cadeira extensora	1.010	0.98-1.04	0.501	1.027	0.97-1.08	0.312	0.683	0.34-1.34	0.270	1.030	0.97-1.08	0.267
Mesa flexora	0.971	0.84-1.11	0.678	1.436	0.72-2.85	0.302	0.501	0.06-3.59	0.491	0.957	0.82-1.11	0.572
Puxada Frontal	1.075	1.00-1.15	0.033	10.629	1.00-1.13	0.050	1.031	0.57-1.84	0.918	0.996	0.94-1.05	0.863
Remada baixa	1.060	0.99-1.13	0.081	1.10	0.99-1.22	0.062	0.677	0.26-1.71	0.410	1.153	0.96-1.38	0.119
Pec deck	1.018	0.99-1.05	0.189	108.529	1.01-1.161	0.018	0.784	0.43-1.41	0.415	10.401	0.99-1.09	0.101
6MWT	0.991	0.91-1.078	0.831	1.007	0.99-1.015	0.099	1.738	0.74-0.96	0.185	0.982	0.92-1.04	0.546
SL	1.001	0.99-1.00	0.596	0.999	0.99-1.00	0.819	0.525	0.21-1.25	0.147	0.984	0.94-1.03	0.468

Nota: 6MWT= Caminhada de 6 minutos; TUG= Teste Timed-Up-and-Go; SL= Teste de levantar e sentar 1RM= Teste de uma repetição máxima, RR = Repetições de reserva

DISCUSSÃO

O objetivo deste estudo foi verificar a equivalência do protocolo autosselecionado com as recomendações do ACSM. Embora essa pesquisa não resultou em uma equivalência total entre os protocolos, ela mostrou que não há diferença estatisticamente significativa entre os protocolos. Colocando a intensidade autosselecionada como mais uma estratégia no âmbito clínico e profissional. Já que anteriormente foi possível observar que a autosseleção de intensidade por mulheres na pós menopausa é compatível com a intensidade recomendada pelo ACSM (GLASS; AHMAD; GABLER, 2020). Mesmo havendo controvérsias, visto que um estudo com mulheres idosas não demonstrou a autosseleção de intensidades próximas da recomendada (ELSANGEDY et al., 2013). Em nosso estudo, a média de intensidade selecionada pelo grupo GA ocorreu dentro da faixa de intensidade recomendada, superior a 60% de 1RM em quase todos os exercícios, exceto na remada baixa ($47,3 \pm 15.0$ p=0,018). Apesar dos demais exercícios apresentarem uma média de autosseleção entre 60-79% de 1RM, exceto a remada baixa, houve diferenças significativas entre grupos nos exercícios de supino reto p=0,017 e cadeira flexora p=0,011. Essa diferença entre grupos, nesses exercícios, deve-se ao ajuste monitorado da intensidade durante quatro semanas no grupo GT, o que elevou a média de intensidade usada ao longo da intervenção para o grupo GT.

Além da intensidade, o volume, calculado por meio da fórmula carga*reps*frequência, assim como a frequência foram monitoradas durante as doze semanas. Notou-se que o grupo GT apresentou um volume total mais elevado que o grupo GA (p=0,050). Essa diferença se explica pela intensidade usada pelo grupo GT resultando em um volume consideravelmente maior. Todavia, o principal responsável por essa diferença se explica pela frequência, uma vez que o GT se mostrou mais frequente em comparação ao grupo GA (p=0,030), contrapondo-se aos achados de Ekkekakis; Lind; Vazou, 2010 e da hipótese inicial sugerida neste estudo, em que o protocolo autosselecionado seria mais aderente; consequentemente, apresentaria valores de frequência maiores que o protocolo tradicional.

Os resultados apontaram que o protocolo com intensidade autosselecionada demonstrou efeito de tempo sobre marcadores de DF, força e massa muscular, como já visto em outro estudo com intensidade autosselecionada, evidenciando efeito de

tempo sobre esses mesmos marcadores (ALVES et al., 2022). Apesar de não encontrar efeito da interação do tempo pelo grupo em nenhum marcador de DF, força e massa muscular, nem mesmo quando ajustado por valores pré, idade, MOCA e frequência, ainda assim, o protocolo autosselecionado não estabelece equivalência com o protocolo de intensidade recomendada pelo ACSM. Embora a equivalência não tenha se estabelecido, não há evidências para afirmar diferença estatisticamente significativa entre protocolos. Portanto, essa pesquisa avançou em análise investigando o papel da intensidade, volume, repetições de reserva e frequência sobre os efeitos encontrados independentemente do protocolo.

Sendo assim, a correlação destacou a intensidade como variável relevante para respostas significativas na força muscular independentemente do protocolo, o que condiz com achados de um estudo envolvendo a população adulta/idosa (BORDE; HORTOBÁGYI; GRANACHER, 2015). Um estudo do nosso grupo de pesquisas, também conduzido com mulheres na pós menopausa mostrou que a alta intensidade gerou ganhos substanciais na força muscular. Entretanto, também foi possível notar ganhos de força com baixa intensidade, uma vez que o volume deste ensaio clínico para o grupo baixa intensidade era significativamente maior que para o grupo alta intensidade (CARNEIRO et al., 2021).

Ademais, o volume foi observado como variável relevante no ganho de massa muscular. Outros estudos que avaliaram o ganho de massa muscular na população de mulheres na pós menopausa também revelaram o volume como melhor variável para resposta sobre o efeito na massa muscular (CAVALCANTE et al., 2023) (NASCIMENTO DE OLIVEIRA-JÚNIOR et al., 2022), corroborando com os resultados observados nesta pesquisa, apresentados na tabela 5, onde os valores de correlação são significativos para o volume em marcadores de massa muscular.

A frequência foi significativamente diferente entre os grupos: o grupo de intensidade imposta apresentou-se como mais frequente em relação ao grupo de intensidade autosselecionada. Embora esse resultado seja relevante para medir a aderência dos protocolos, o mesmo não foi relevante frente aos efeitos no DF, na força e massa muscular, corroborando estudo de Orsatti et al., (2014) que evidenciou que o TF aumenta a força e a massa muscular independentemente da frequência.

Os efeitos além das medidas de MMD revelaram que, nos testes de força, os exercícios leg press e supino reto foram os exercícios com melhores respondedores além da MMD. Associadamente, esses mesmos exercícios são os que apresentam menores valores de MMD (ASSIS SILVA et al., 2023). A possibilidade de manipulação da intensidade com maior precisão nesses exercícios é um fator explicativo dos mesmos serem mais sensíveis a medidas além da MMD no teste de força de 1RM. Nossos achados destacam o supino reto e o leg press, exercícios como melhores preditores de efeito na força além da medida de MMD, para membros inferiores (leg press) e de membro superiores (supino reto).

Em contrapartida, os exercícios de mesa flexora e pec deck são os exercícios que apresentam as maiores medidas de MMD para o teste de 1RM, 9,2kg (51%); 8,8 kg (44%) respectivamente (ASSIS SILVA et al., 2023). A mesa flexora, por ser um exercício de maior expressividade em medidas de MMD, foi a que mostrou menor expressividade na quantidade de respondedores. Embora o pec deck no grupo de exercício de membros superiores tenha apresentado maiores valores de MMD, não apresentou menor expressividade de respondedores. A menor expressividade de respondedores ocorreu no exercício remada baixa, ainda que a remada baixa em relação ao pec deck seja a que expresse o menor valor MMD. (ASSIS SILVA et al., 2023). Uma explicação para isso pode ser em razão da baixa tolerância a intensidade especificamente no exercício de remada baixa. Visto que a remada baixa teve a menor expressividade na intensidade usada ao longo da intervenção, 47% 1RM no grupo GA e 57% 1RM no grupo GT, embora o grupo GT fosse imposta intensidades acima de 70% 1RM, durante a intervenção notamos a dificuldade de impor essa intensidade especificamente nesse exercício e adotamos a flexibilização em razão de possível evasão a amostra. O treinamento com baixa intensidade na remada baixa em ambos os grupos pode ter levado ao baixo número de respondedores, uma vez que já observa a relação da intensidade com a resposta na força muscular (BORDE; HORTOBÁGYI; GRANACHER, 2015) (CARNEIRO et al., 2021).

Nas avaliações de DF, as medidas de MMD foram de 51,9 m (10%), 1,4 seg (20%), 4,6 reps (30%) para os testes de caminhada de 6 minutos, TUG e teste de sentar e levantar de 30`, respectivamente (ASSIS SILVA et al., 2023). Embora o teste de sentar e levantar de 30 segundos seja o teste com maior medida de MMD, o mesmo

apresentou-se com maior número de respondedores: 16,7% da amostra total. Já o TUG como medida de DF foi o que apresentou menor quantidade de respondedores: 1,7% da amostra total. Essas informações destacam o teste de sentar e levantar de 30 segundos, assim como o teste de caminhada de 6 minutos como melhores medidas no processo de elucidar respondedores além das medidas de MMD.

Ao averiguar o papel da intensidade, volume, frequência e RR como mediadores nos efeitos além das medidas de MMD, a intensidade mostrou-se relevante quanto aos efeitos para força no leg press e puxada frontal. Já o volume foi uma variável significativa para respostas além da MMD no supino reto, puxada frontal e pec deck. Entretanto, nenhuma das variáveis se apresentou como relevante na contribuição de efeitos que extrapole as medidas de MMD nos testes de DF.

Esse estudo apresenta limitações relevantes, entre elas, a análise de regressão logística utilizada não é uma estatística sensível à uma amostra relativamente pequena. Embora tenhamos recrutado uma amostra maior, o número de voluntárias que desistiram não permitiu que terminássemos o estudo com uma amostra sensível a uma regressão logística. Mesmo que esse estudo tenha uma amostra consideravelmente grande e viável para o desfecho principal. Além disso, trata-se de uma população específica, mulheres com mais de 40 anos, portanto, deve-se ter cautela ao generalizar esses dados para as demais populações. Um ponto forte desse estudo foi a verificação do efeito da intervenção além das medidas de MMD, extraídas da própria amostra, trazendo mais fidedignidade para esse dado relevante, porém pouco explorado em ensaios clínicos.

Essa pesquisa vislumbra beneficiar a área clínica e profissional elucidando os reais benefícios da intensidade autosselecionada em contraposto com as recomendações de intensidade no TF já bem fundamentadas do ACSM. Colocando em evidencia mais essa estratégia para a área clínica e profissional. Além disso, essa pesquisa torna-se inovadora elucidando a necessidade dos ensaios clínicos realizarem análise de respondedores além das medidas de MMD uma vez que esse dado é relevante para comunidade científica, a visto de verificar quanto um método ou estratégia de treinamento pode ser mais respondente do que outro em medidas de força e DF.

CONCLUSÃO

O protocolo com intensidade autosselecionada demonstrou ser eficaz para ganho de força, massa muscular e melhora no DF. Contudo, não foi possível estabelecer equivalência com o protocolo tradicionalmente recomendado pelo ACSM. A intensidade apresentou-se como variável importante para o ganho de força muscular. Adicionalmente, o volume foi relevante para respostas na massa muscular. Além disso, as respostas além da MMD foram estatisticamente sem diferença entre protocolos e influenciadas tanto pela intensidade quanto pelo volume em medidas de força.

REFERÊNCIAS

ALVES, R. C. et al. Effect of Different Training Programs at Self-Selected Intensity on Body Composition, Perceptual Responses and Fitness Outcomes in Obese Women. **International Journal of Exercise Science**, v. 15, n. 4, p. 373–385, 1 fev. 2022.

AMERICAN COLLEGE OF SPORTS MEDICINE. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. **Medicine and Science in Sports and Exercise**, v. 41, n. 3, p. 687–708, mar. 2009.

ARIKAWA, A. Y.; O'DOUGHERTY, M.; SCHMITZ, K. H. Adherence to a strength training intervention in adult women. **Journal of Physical Activity & Health**, v. 8, n. 1, p. 111–118, jan. 2011.

ASSIS SILVA, S. H. et al. Assessing the robustness of muscle strength and physical performance measures in women older than 40 years: a test-retest reliability study. **Menopause (New York, N.Y.)**, 12 dez. 2024.

BARBALHO, M. et al. High 1RM tests reproducibility and validity are not dependent on training experience, muscle group tested or strength level in older women. **Sports**, v. 6, n. 4, p. 171, 2018.

BORDE, R.; HORTOBÁGYI, T.; GRANACHER, U. Dose–Response Relationships of Resistance Training in Healthy Old Adults: A Systematic Review and Meta-Analysis. **Sports Medicine (Auckland, N.z.)**, v. 45, n. 12, p. 1693–1720, 2015.

Boshnjaku A, Bahtiri A, Feka K, Krasniqi E, Tschan... - Google Acadêmico.

Disponível em: <https://scholar.google.com.br/scholar?hl=pt-BR&as_sdt=0%2C5&q=Boshnjaku+A%2C+Bahtiri+A%2C+Feka+K%2C+Krasniqi+E%2C+Tschan+H%2C+Wessner+B.+Test-retest+reliability+data+of+functional+performance%2C+strength%2C+peak+torque+and+body+composition+assessments+in+two+different+age+groups+of+Kosovan+adults.+Data+in+brief.+2021%3B36%3A106988&btnG=>>. Acesso em: 1 mar. 2024.

CARNEIRO, M. A. et al. Effect of whole-body resistance training at different load intensities on circulating inflammatory biomarkers, body fat, muscular strength, and physical performance in postmenopausal women. **Applied Physiology, Nutrition, and Metabolism**, v. 46, n. 8, p. 925–933, ago. 2021.

CHAN, W. L. S.; PIN, T. W. Reliability, validity and minimal detectable change of 2-minute walk test, 6-minute walk test and 10-meter walk test in frail older adults with dementia. **Experimental Gerontology**, v. 115, p. 9–18, jan. 2019.

Chan WLS, Pin TW. Reliability, validity and minimal... - Google Acadêmico.

Disponível em: <https://scholar.google.com.br/scholar?hl=pt-BR&as_sdt=0%2C5&q=Chan+WLS%2C+Pin+TW.+Reliability%2C+validity+and+minimal+detectable+change+of+2-minute+walk+test%2C+6-minute+walk+test+and+10-meter+walk+test+in+frail+older+adults+with+dementia.+Experimental+gerontology.+%5BValidation+Study%5D.+2019%3B115%3A9-18&btnG=>>. Acesso em: 1 mar. 2024.

CHARTER, R. A. Revisiting the Standard Errors of Measurement, Estimate, and Prediction and Their Application to Test Scores. **Perceptual and Motor Skills**, v. 82, n. 3_suppl, p. 1139–1144, jun. 1996.

CUYUL-VÁSQUEZ, I. et al. Effectiveness of whey protein supplementation during resistance exercise training on skeletal muscle mass and strength in older people with sarcopenia: a systematic review and meta-analysis. **Nutrients**, v. 15, n. 15, p. 3424, 2023.

EKKEKAKIS, P.; LIND, E.; VAZOU, S. Affective responses to increasing levels of exercise intensity in normal-weight, overweight, and obese middle-aged women. **Obesity (Silver Spring, Md.)**, v. 18, n. 1, p. 79–85, jan. 2010.

EL KHOUDARY, S. R. et al. The menopause transition and women's health at midlife: a progress report from the Study of Women's Health Across the Nation (SWAN). **Menopause (New York, N.y.)**, v. 26, n. 10, p. 1213–1227, 23 set. 2019.

ELSANGEDY, H. M. et al. Is the Self-Selected Resistance Exercise Intensity by Older Women Consistent With the American College of Sports Medicine Guidelines to Improve Muscular Fitness? **The Journal of Strength & Conditioning Research**, v. 27, n. 7, p. 1877, jul. 2013.

ELSANGEDY, H. M. et al. Effects of Self-selected Resistance Training on Physical Fitness and Psychophysiological Responses in Physically Inactive Older Women: A Randomized Controlled Study. **Perceptual and Motor Skills**, v. 128, n. 1, p. 467–491, 1 fev. 2021.

GLASS, S. C.; AHMAD, S.; GABLER, T. Effectiveness of a 2-Week Strength Training Learning Intervention on Self-selected Weight-Training Intensity. **Journal of Strength and Conditioning Research**, v. 34, n. 9, p. 2443–2448, set. 2020.

GREENDALE, G. A. et al. Changes in body composition and weight during the menopause transition. **JCI Insight**, v. 4, n. 5, p. e124865, [s.d.].

GRGIC, J. et al. Test-Retest Reliability of the One-Repetition Maximum (1RM) Strength Assessment: a Systematic Review. **Sports Medicine - Open**, v. 6, n. 1, p. 31, 17 jul. 2020a.

GRGIC, J. et al. Test–Retest Reliability of the One-Repetition Maximum (1RM) Strength Assessment: a Systematic Review. **Sports Medicine - Open**, v. 6, n. 1, p. 31, dez. 2020b.

HAILE, L. et al. Affective and metabolic responses to self-selected intensity cycle exercise in young men. **Physiology & Behavior**, v. 205, p. 9–14, 1 jun. 2019.

INFANTE, M. A. et al. One repetition maximum test-retest reliability and safety using Keiser pneumatic resistance training machines with older women. **Journal of strength and conditioning research**, v. 35, n. 12, p. 3513–3517, 2021.

KOO, T. K.; LI, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. **Journal of chiropractic medicine**, v. 15, n. 2, p. 155–163, 2016.

MUÑOZ-BERMEJO, L. et al. Test-retest reliability of five times sit to stand test (FTSST) in adults: A systematic review and meta-analysis. **Biology**, v. 10, n. 6, p. 510, 2021.

MURDOCH, M. et al. People at a persistent pain service can walk it, but some struggle to talk about it: Reliability, detectable difference and clinically important difference of the six-minute walk test. **Musculoskeletal Care**, v. 21, n. 1, p. 221–231, mar. 2023.

NASIMI, N. et al. Whey protein supplementation with or without vitamin D on sarcopenia-related measures: A systematic review and meta-analysis. **Advances in Nutrition**, 2023.

ORSATTI, F. L. et al. Plasma hormones, muscle mass and strength in resistance-trained postmenopausal women. **Maturitas**, v. 59, n. 4, p. 394–404, 20 abr. 2008.

PARFITT, G.; HUGHES, S. The Exercise Intensity–Affect Relationship: Evidence and Implications for Exercise Behavior. **Journal of Exercise Science & Fitness**, v. 7, n. 2, Supplement, p. S34–S41, 1 jan. 2009.

PATRIZIO, E. et al. Physical Functional Assessment in Older Adults. **The Journal of Frailty & Aging**, v. 10, n. 2, p. 141–149, 2021.

PICCA, A. et al. Biomarkers shared by frailty and sarcopenia in older adults: A systematic review and meta-analysis. **Ageing research reviews**, v. 73, p. 101530, 2022.

RIEMANN, B. L.; LININGER, M. R. Statistical Primer for Athletic Trainers: The Essentials of Understanding Measures of Reliability and Minimal Important Change. **Journal of Athletic Training**, v. 53, n. 1, p. 98–103, 1 jan. 2018.

SAMSON, M. M. et al. Relationships between physical performance measures, age, height and body weight in healthy adults. **Age and Ageing**, v. 29, n. 3, p. 235–242, maio 2000a.

SAMSON, M. M. et al. Relationships between physical performance measures, age, height and body weight in healthy adults. **Age and Ageing**, v. 29, n. 3, p. 235–242, 1 maio 2000b.

SHOEMAKER, M. J. et al. Clinically Meaningful Change Estimates for the Six-Minute Walk Test and Daily Activity in Individuals With Chronic Heart Failure. **Cardiopulmonary Physical Therapy Journal**, v. 24, n. 3, p. 21–29, set. 2013.

SPAGNUOLO, D. L. et al. Walking for the Assessment of Balance in Healthy Subjects Older than 40 Years. **Gerontology**, v. 56, n. 5, p. 467–473, 13 jan. 2010.

TSENG, L. A. et al. The association of menopause status with physical function: the Study of Women's Health Across the Nation. **Menopause (New York, N.Y.)**, v. 19, n. 11, p. 1186–1192, nov. 2012.

VIKBERG, S. et al. Effects of Resistance Training on Functional Strength and Muscle Mass in 70-Year-Old Individuals With Pre-sarcopenia: A Randomized Controlled Trial.

Journal of the American Medical Directors Association, v. 20, n. 1, p. 28–34, jan. 2019.

WEIR, J. P. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. **Journal of Strength and Conditioning Research**, v. 19, n. 1, p. 231–240, fev. 2005.

YUKSEL, E. et al. Assessing minimal detectable changes and test-retest reliability of the timed up and go test and the 2-minute walk test in patients with total knee arthroplasty. **The Journal of arthroplasty**, v. 32, n. 2, p. 426–430, 2017.

3 CONSIDERAÇÕES FINAIS

Embora os testes de força muscular e DF demonstrem excelente confiabilidade relativa (correlação teste-reteste) em mulheres com mais de 40 anos, também apresentam baixa confiabilidade absoluta (variabilidade de medidas repetidas), especialmente o teste de uma repetição máxima (1RM) realizado em máquina e o teste de sentar e levantar em 30 segundos. Como resultado, esses testes são ferramentas eficazes para identificar ou diagnosticar mulheres com força muscular e DF. No entanto, eles podem ter limitações ao avaliar intervenções que produzem pequenos efeitos nesses testes. Portanto, alterações nos resultados desses testes que ficam abaixo da Mudança Mínima Detectável podem indicar que mulheres de meia idade e mais velhas não experimentaram uma mudança real na FM ou DF.

Além disso, mesmo que não sejam equivalentes, independentemente dos protocolos analisados, ambos foram eficazes no ganho de força, massa muscular e no DF. Tornando a intensidade autosselecionada mais uma ferramenta no âmbito clínico e profissional. A intensidade foi relevante para o ganho de força muscular e o volume se apresentou com relevante para respostas no ganho de massa muscular independentemente do protocolo. As respostas além da MMD foram estatisticamente sem diferença entre protocolos e influenciadas tanto pela intensidade quanto pelo volume em medidas de força.

REFERÊNCIAS

ALVES, R. C. et al. Effect of Different Training Programs at Self-Selected Intensity on Body Composition, Perceptual Responses and Fitness Outcomes in Obese Women. **International Journal of Exercise Science**, v. 15, n. 4, p. 373–385, 1 fev. 2022.

AMERICAN COLLEGE OF SPORTS MEDICINE. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. **Medicine and Science in Sports and Exercise**, v. 41, n. 3, p. 687–708, mar. 2009.

ARIKAWA, A. Y.; O'DOUGHERTY, M.; SCHMITZ, K. H. Adherence to a strength training intervention in adult women. **Journal of Physical Activity & Health**, v. 8, n. 1, p. 111–118, jan. 2011.

ASSIS SILVA, S. H. et al. Assessing the robustness of muscle strength and physical performance measures in women older than 40 years: a test-retest reliability study. **Menopause (New York, N.Y.)**, 12 dez. 2024.

BARBALHO, M. et al. High 1RM tests reproducibility and validity are not dependent on training experience, muscle group tested or strength level in older women. **Sports**, v. 6, n. 4, p. 171, 2018.

BORDE, R.; HORTOBÁGYI, T.; GRANACHER, U. Dose–Response Relationships of Resistance Training in Healthy Old Adults: A Systematic Review and Meta-Analysis. **Sports Medicine (Auckland, N.z.)**, v. 45, n. 12, p. 1693–1720, 2015.

Boshnjaku A, Bahtiri A, Feka K, Krasniqi E, Tschan... - Google Acadêmico. Disponível em: <https://scholar.google.com.br/scholar?hl=pt-BR&as_sdt=0%2C5&q=Boshnjaku+A%2C+Bahtiri+A%2C+Feka+K%2C+Krasniqi+E%2C+Tschan+H%2C+Wessner+B.+Test-retest+reliability+data+of+functional+performance%2C+strength%2C+peak+torque+and+body+composition+assessments+in+two+different+age+groups+of+Kosovan+adults.+Data+in+brief.+2021%3B36%3A106988&btnG=>>. Acesso em: 1 mar. 2024.

CARNEIRO, M. A. et al. Effect of whole-body resistance training at different load intensities on circulating inflammatory biomarkers, body fat, muscular strength, and physical performance in postmenopausal women. **Applied Physiology, Nutrition, and Metabolism**, v. 46, n. 8, p. 925–933, ago. 2021.

CHAN, W. L. S.; PIN, T. W. Reliability, validity and minimal detectable change of 2-minute walk test, 6-minute walk test and 10-meter walk test in frail older adults with dementia. **Experimental Gerontology**, v. 115, p. 9–18, jan. 2019.

Chan WLS, Pin TW. Reliability, validity and minimal... - Google Acadêmico. Disponível em: <https://scholar.google.com.br/scholar?hl=pt-BR&as_sdt=0%2C5&q=Chan+WLS%2C+Pin+TW.+Reliability%2C+validity+and+minimal+detectable+change+of+2-minute+walk+test%2C+6-minute+walk+test+and+10-meter+walk+test+in+frail+older+adults+with+dementia.+Experimental+gerontology.+%5BValidation+Study%5D.+2019%3B115%3A9-18&btnG=>>. Acesso em: 1 mar. 2024.

CHARTER, R. A. Revisiting the Standard Errors of Measurement, Estimate, and Prediction and Their Application to Test Scores. **Perceptual and Motor Skills**, v. 82, n. 3_suppl, p. 1139–1144, jun. 1996.

CUYUL-VÁSQUEZ, I. et al. Effectiveness of whey protein supplementation during resistance exercise training on skeletal muscle mass and strength in older people with sarcopenia: a systematic review and meta-analysis. **Nutrients**, v. 15, n. 15, p. 3424, 2023.

EKKEKAKIS, P.; LIND, E.; VAZOU, S. Affective responses to increasing levels of exercise intensity in normal-weight, overweight, and obese middle-aged women. **Obesity (Silver Spring, Md.)**, v. 18, n. 1, p. 79–85, jan. 2010.

EL KHOUDARY, S. R. et al. The menopause transition and women's health at midlife: a progress report from the Study of Women's Health Across the Nation (SWAN). **Menopause (New York, N.y.)**, v. 26, n. 10, p. 1213–1227, 23 set. 2019.

ELSANGEDY, H. M. et al. Is the Self-Selected Resistance Exercise Intensity by Older Women Consistent With the American College of Sports Medicine Guidelines to Improve Muscular Fitness? **The Journal of Strength & Conditioning Research**, v. 27, n. 7, p. 1877, jul. 2013.

ELSANGEDY, H. M. et al. Effects of Self-selected Resistance Training on Physical Fitness and Psychophysiological Responses in Physically Inactive Older Women: A Randomized Controlled Study. **Perceptual and Motor Skills**, v. 128, n. 1, p. 467–491, 1 fev. 2021.

GLASS, S. C.; AHMAD, S.; GABLER, T. Effectiveness of a 2-Week Strength Training Learning Intervention on Self-selected Weight-Training Intensity. **Journal of Strength and Conditioning Research**, v. 34, n. 9, p. 2443–2448, set. 2020.

GREENDALE, G. A. et al. Changes in body composition and weight during the menopause transition. **JCI Insight**, v. 4, n. 5, p. e124865, [s.d.].

GRGIC, J. et al. Test-Retest Reliability of the One-Repetition Maximum (1RM) Strength Assessment: a Systematic Review. **Sports Medicine - Open**, v. 6, n. 1, p. 31, 17 jul. 2020a.

GRGIC, J. et al. Test–Retest Reliability of the One-Repetition Maximum (1RM) Strength Assessment: a Systematic Review. **Sports Medicine - Open**, v. 6, n. 1, p. 31, dez. 2020b.

HAILE, L. et al. Affective and metabolic responses to self-selected intensity cycle exercise in young men. **Physiology & Behavior**, v. 205, p. 9–14, 1 jun. 2019.

INFANTE, M. A. et al. One repetition maximum test-retest reliability and safety using Keiser pneumatic resistance training machines with older women. **Journal of strength and conditioning research**, v. 35, n. 12, p. 3513–3517, 2021.

KOO, T. K.; LI, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. **Journal of chiropractic medicine**, v. 15, n. 2, p. 155–163, 2016.

MUÑOZ-BERMEJO, L. et al. Test-retest reliability of five times sit to stand test (FTSST) in adults: A systematic review and meta-analysis. **Biology**, v. 10, n. 6, p. 510, 2021.

MURDOCH, M. et al. People at a persistent pain service can walk it, but some struggle to talk about it: Reliability, detectable difference and clinically important difference of the six-minute walk test. **Musculoskeletal Care**, v. 21, n. 1, p. 221–231, mar. 2023.

NASIMI, N. et al. Whey protein supplementation with or without vitamin D on sarcopenia-related measures: A systematic review and meta-analysis. **Advances in Nutrition**, 2023.

ORSATTI, F. L. et al. Plasma hormones, muscle mass and strength in resistance-trained postmenopausal women. **Maturitas**, v. 59, n. 4, p. 394–404, 20 abr. 2008.

PARFITT, G.; HUGHES, S. The Exercise Intensity–Affect Relationship: Evidence and Implications for Exercise Behavior. **Journal of Exercise Science & Fitness**, v. 7, n. 2, Supplement, p. S34–S41, 1 jan. 2009.

PATRIZIO, E. et al. Physical Functional Assessment in Older Adults. **The Journal of Frailty & Aging**, v. 10, n. 2, p. 141–149, 2021.

PICCA, A. et al. Biomarkers shared by frailty and sarcopenia in older adults: A systematic review and meta-analysis. **Ageing research reviews**, v. 73, p. 101530, 2022.

RIEMANN, B. L.; LININGER, M. R. Statistical Primer for Athletic Trainers: The Essentials of Understanding Measures of Reliability and Minimal Important Change. **Journal of Athletic Training**, v. 53, n. 1, p. 98–103, 1 jan. 2018.

SAMSON, M. M. et al. Relationships between physical performance measures, age, height and body weight in healthy adults. **Age and Ageing**, v. 29, n. 3, p. 235–242, maio 2000a.

SAMSON, M. M. et al. Relationships between physical performance measures, age, height and body weight in healthy adults. **Age and Ageing**, v. 29, n. 3, p. 235–242, 1 maio 2000b.

SHOEMAKER, M. J. et al. Clinically Meaningful Change Estimates for the Six-Minute Walk Test and Daily Activity in Individuals With Chronic Heart Failure. **Cardiopulmonary Physical Therapy Journal**, v. 24, n. 3, p. 21–29, set. 2013.

SPAGNUOLO, D. L. et al. Walking for the Assessment of Balance in Healthy Subjects Older than 40 Years. **Gerontology**, v. 56, n. 5, p. 467–473, 13 jan. 2010.

TSENG, L. A. et al. The association of menopause status with physical function: the Study of Women's Health Across the Nation. **Menopause (New York, N.Y.)**, v. 19, n. 11, p. 1186–1192, nov. 2012.

VIKBERG, S. et al. Effects of Resistance Training on Functional Strength and Muscle Mass in 70-Year-Old Individuals With Pre-sarcopenia: A Randomized Controlled

Trial. **Journal of the American Medical Directors Association**, v. 20, n. 1, p. 28–34, jan. 2019.

WEIR, J. P. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. **Journal of Strength and Conditioning Research**, v. 19, n. 1, p. 231–240, fev. 2005.

YUKSEL, E. et al. Assessing minimal detectable changes and test-retest reliability of the timed up and go test and the 2-minute walk test in patients with total knee arthroplasty. **The Journal of arthroplasty**, v. 32, n. 2, p. 426–430, 2017.

ANEXO 01

Abstract

Objectives: This study investigated the test-retest reliability and minimal detectable change (MDC) of muscle strength (MS) and physical performance (PP) tests in women aged over 40 years.

Methods: A total of 113 women, aged 58 ± 11 years (82% postmenopausal women), with a body mass index of 28.5 ± 5.8 kg/m², participated in this study. One repetition maximum (1RM) tests were conducted for seven strength exercises: 45° leg press, bench press, leg extension, pec deck, wide-grip lateral pull-down, leg curl, and seated cable row, with a retest after 48 hours. Timed-Up-and-Go, six-minute walk, and 30s-sit-to-stand tests were performed on the same day and retested after 72 hours. The tests and retests were administered by the same evaluators. Relative reliability (consistency of participant rank between test-retest) was assessed using the intraclass correlation coefficient for consistency and agreement, and absolute reliability (precision of score) was assessed using the MDC based on the standard error of prediction.

Results: The 1RM and performance tests exhibited excellent reliability: 45° leg press (consistency = 0.99, agreement = 0.98), bench press (consistency = 0.96, agreement = 0.96), leg extension (consistency = 0.93, agreement = 0.91), pec deck (consistency = 0.90, agreement = 0.88), wide-grip lateral pull-down (consistency = 0.91, agreement = 0.89), leg curl (consistency = 0.84, agreement = 0.83), seated cable row (consistency = 0.92, agreement = 0.94), timed-Up-and-Go (consistency = 0.87, agreement = 0.87), six-minute walk (consistency = 0.96, agreement = 0.95), and 30s-sit-to-stand (consistency = 0.85, agreement = 0.80). These tests showed meaningful MDC values, particularly for the 1RM test performed on the machine and the 30s-sit-to-stand test: 45° leg press = 22.2 kg (15%), bench press = 4.9 kg (19%), leg extension = 9.2 kg (35%), pec deck = 8.8 kg (44%), wide-grip lateral pull-down = 9.4 kg (31%), leg curl = 9.2 kg (51%), seated cable row = 11.4 kg (29%), timed-Up-and-Go = 1.4 s (20%), six-minute walk = 50.9 m (10%), and 30s-sit-to-stand = 4.6 reps (30%).

Conclusions: Although the MS and PP tests demonstrate excellent relative reliability in women aged over 40 years, they exhibit low absolute reliability, particularly the 1RM

test performed on the machine and the 30s-sit-to-stand test. Therefore, while these tests show good consistency in the positioning of women within the group after repeated measures, their precision measure is relatively low (high fluctuation). Changes in these tests that are smaller than the MDC may not indicate real changes in women at middle age and older.

Key Words: Physical Function – Test-Retest Reliability, One Repetition Maximum, Women.

link to access the magazine document: [23-00175_GME.pdf](#)

INTRODUCTION

As individuals age, they inevitably experience a significant decline in both muscle strength (MS) and physical performance (PP). This decline in MS and PP poses various risks, including the development of conditions such as sarcopenia, an increased susceptibility to falls, and, in severe cases, elevated mortality rates¹. Consequently, the precise assessment of MS and PP becomes crucial in identifying individuals at an early stage of risk and crafting targeted interventions to mitigate these adverse outcomes¹. Furthermore, acquiring dependable measurements capable of detecting even subtle changes in MS and PP plays a pivotal role in accurately diagnosing these conditions and effectively monitoring and evaluating the impact of interventions in clinical practice^{2, 3}.

The need to establish the reliability of these measurements and determine their minimal detectable changes (MDC) is particularly compelling among women aged 40 years and above. This specific age group experiences hormonal declines, physiological changes, and a decrease in vitamin D levels during the pre and postmenopausal phases, all of which may contribute to reductions in MS and PP^{4, 5}. Nevertheless, the existing body of literature lacks comprehensive research on the reliability of these tests in women within this age cohort.

In assessing PP, practical and functional measures such as the timed-up-and-go (TUG), 30-second sit-to-stand, and six-minute walk (6MWT) tests are frequently employed among older adults ^{1, 6-8}. These tests have gained popularity due to their simplicity, cost-effectiveness, and ability to reflect performance in activities of daily living^{1, 9, 10}. However, there remains a dearth of comprehensive research concerning the reliability of these tests in women aged 40 years and above.

One widely used assessment tool for evaluating MS is the one-repetition maximum test (1RM)¹¹. However, despite Grgic et al.'s (2020) systematic review on the test-retest reliability of 1RM, certain gaps in the literature persist ¹¹. Most of the 32 selected studies in their review involved young participants. While numerous exercises were assessed, most studies focused on a limited range of exercises, with few exploring machine exercises, such as the pec deck. Additionally, variations in testing protocols, the failure to account for factors like systemic changes, and limited information on minimal detectable changes were common shortcomings. While it is possible to calculate the MDC in some studies, most calculations relied on the standard error of measurement (SEM), which may not provide the highest degree of precision^{2, 12}. The standard error of prediction (SEP) has been recommended as an alternative measure^{2, 12}.

To bridge the existing research gap in this domain, our study aimed to investigate the test-retest reliability and MDC of MS and PP tests in women aged 40 years and above. By focusing specifically on this demographic, our objective was to offer valuable insights into the effectiveness of these assessment tools. The contributions of this research endeavor will advance measurement techniques for professionals across various fields, enabling them to more accurately assess and monitor MS and PP within this specific age group.

METHODS

Study design

This study lasted 5 weeks starting from september 19th to October 21st, 2022.

Initially, an interview was conducted to select participants based on the inclusion criteria, resulting in a total of 113 women being included in the study. Each participant attended one visit for the administration of the PP tests, followed by another visit for the PP retests after a 72-hour interval (week 1). In addition, there was a familiarization period consisting of three visits from weeks 2 to 4, during which participants were acquainted with resistance exercises. Subsequently, one visit was dedicated to conducting the 1RM tests, with a retest performed after 48 hours (week 5). To minimize the influence of diurnal biological variations, all assessments were consistently conducted at the same time of day. Moreover, the MS and PP tests were standardized and closely supervised by experienced researchers to ensure proper technique, reliability, and the safety and well-being of the participants.

Women

The participants in this study provided informed consent and were recruited through media channels in the city of Uberaba-MG, Brazil. The study received approval from the ethics committee of the Federal University of Triangulo Mineiro (UFTM), where it was specifically conducted within the postgraduate program in Physical Education.

Participants were recruited from residential neighborhoods, and those who expressed interest in participating underwent a comprehensive interview. The interview involved collecting information on age, employment status, health indicators, past and present medical history, therapeutic interventions, and physical activities. Initially, 140 women underwent screening for eligibility. Out of these, a total of 113 participants met the inclusion criteria for the study. The inclusion criteria included: (i) being either premenopausal or postmenopausal (characterized by experiencing spontaneous amenorrhea for at least 12 months); (ii) not using hormone therapy or dietary supplements; (iii) having controlled blood pressure and blood sugar levels; (iv) not having muscle, thromboembolic, or gastrointestinal disorders; (v) no history of cardiovascular or infectious diseases; (vi) being non-consumers of alcohol (no alcohol intake whatsoever in their diet); (vii) being non-smokers; (viii) having previous experience with resistance exercise, but not engaging in regular physical activity more than once per week in the last 6 months. However, due to specific joint problems, some

participants were only able to perform a subset of the tests. The distribution of participants for each test can be found in Table 2.

One repetition maximum test (1RM)

During the familiarization period, participants completed three training sessions per week on Monday, Wednesday, and Friday. The training sessions aimed to familiarize participants with the exercise techniques. Each session occurred on alternate days, and participants trained with lower loads, avoiding voluntary concentric failure. The exercises included in the familiarization period were 45° leg press, bench press, leg extension, pec deck, wide grip lateral pull-down, leg curl, and seated cable row.

After the familiarization period, 1RM tests were conducted for all exercises on the same day, with a 5-minute interval between each exercise. Prior to the test session, participants performed a general warm-up by walking for 5 minutes, followed by a specific warm-up using subjective loads determined during the familiarization period. The specific warm-up involved performing 12 repetitions with lower loads. After a rest period of 90 seconds, the load intensity was increased by approximately 20%, and participants performed between three to five repetitions. Following a rest period of 3 to 5 minutes, the load intensity was significantly increased, and participants were encouraged to exert maximum effort and utilize the full range of motion. If the load intensity was overestimated or underestimated, participants rested for 3 to 5 minutes before attempting a new lift with a lower or higher load intensity, respectively (approximately 5% to 10% difference). This process was repeated for two to five attempts until the equivalent 1RM load intensity was determined².

The load intensity at which participants could perform no more than one repetition was considered as their 1RM for each exercise. In cases where it was not possible to establish the load for a single repetition, the Brzycki formula [$1RM = \text{weight lifted} / (1.0278 - 0.0278 * \text{executed repetitions})$] was used as an alternative.

The retest session followed the same protocol as the initial test, with a minimum interval of 48 hours between the test and retest sessions.

Timed-Up-and-Go TUG

For the TUG test, participants were seated on a chair with their back supported by a wall, their feet fully on the floor, and their hands resting on their thighs. They were instructed to stand up from the chair and walk a distance of 3 meters on a flat floor. They then went around a cone and returned to the chair, where they sat down as quickly as possible. Participants were instructed to complete the path without running, focusing on speed. The time taken for the entire sequence, from the initial movement to rise from the chair until returning to sit down again, was recorded to the nearest 0.01 seconds using a stopwatch⁹. Each participant performed the test once as a familiarization trial, and two attempts were recorded for analysis. The average of the two attempts was calculated as the representative value. For the retest, the same protocol was followed after a 72-hour interval.

30s-sit-to-stand

During the 30s-sit-to-stand test, participants were seated on a chair with a height of 47 cm, positioned against a wall. They ensured their back was in contact with the backrest, their feet were fully supported on the floor, and their hands were crossed, resting on their opposite shoulders. The test involved one person recording the time and another person counting the number of repetitions performed by the participant. Each participant was given 1 to 2 attempts with an appropriate rest period before the final test. After a 72-hour interval, the retest was conducted following the same protocol.

6MWT

The 6MWT was conducted on a flat surface. The walking route was marked by two cones, creating a total distance of 20 meters. Participants were instructed to walk from one cone to the other (covering a distance of 20 meters), then go around a cone and return to the starting cone, completing the course. During the 6MWT, participants were encouraged to walk as quickly as possible without any breaks or running for a duration of 6 minutes. Verbal encouragement was provided throughout the test to

motivate participants, using phrases such as "you are doing great" and "walk as fast as you can." The distance covered by each participant was recorded at the end of the 6MWT⁷. Following a 72-hour interval, the retest was conducted using the same protocol as the initial test to ensure consistent testing conditions and evaluate the reliability of the results.

Statistical Analysis

All statistical analyses were conducted using JAMOV[®] software. We assessed the distribution of differences between measurements with the Shapiro-Wilk test, which indicated a normal distribution. To evaluate test-retest reliability, we utilized the SympleAgree package from jamovi.

We measured relative reliability, which assesses the consistency of participant rankings between test and retest, using the intraclass correlation coefficient (ICC) for both consistency and agreement. Additionally, we evaluated absolute reliability, which measures the precision of the scores, using the Minimum Detectable Change (MDC) based on the standard error of prediction.

The statistical analysis involved calculating the ICC through a two-way ANOVA, considering both consistency (two-way fixed) and agreement (two-way random) with a simple unit. The ICC values were categorized as follows: excellent (>0.90), good ($0.75-0.90$), moderate ($0.50 - 0.75$), or poor (<0.50), based on predefined criteria ¹³.

The MDC was determined using the formula $SEP (1.96 \times SEP)$ rather than SEM squared (SEM^2). We assessed the association between the average retest-test values and the bias ($\Delta = \text{retest} - \text{test}$) using the Pearson correlation coefficient.

To investigate the influence of menopause on test-retest reliability, we conducted the ICC analysis after excluding premenopausal women. Furthermore, to examine the impact of age on test-retest reliability, we correlated test deltas with participants' ages.

RESULTS

The characteristics of women are summarized in Table 1.

Table 1 - The general characteristics of women (n=113).

	Mean	SD
Age, yrs	57.6	10.0
BMI, kg/m²	28.5	5.8
Body fat, %	43.5	6.5
MMI, kg/m²	6.85	1.21
Physical activity level, min/week	947.0	817.0
Joint problems, n (%)	29 (26%)	
Educational time, yrs	10.0	7.0
Premenopausal women, n (%)	20 (18%)	

BMI = Body mass index, *MMI* = muscle mass index, SD = standard deviation

Analysis of Consistency and Agreement

Table 2 presents the consistency and agreement values for the MS and PP tests. All 1RM and PP tests demonstrated excellent or good consistency and agreement (i.e., > 0.75) between the initial test and the retest, regardless of the presence of premenopausal women.

Table 2 - Intraclass correlation coefficient for consistency and agreement of muscle strength and physical performance tests

Testes	All women (n=113)					Postmenopausal women (n=93)				
	n	ICC¹	IC95%	ICC²	IC95%	n	ICC¹	IC95%	ICC²	IC95%
<i>Leg press, kg</i>	79	0.99	(0.98-0.99)	0.98	(0.91-0.99)	65	0.99	(0.98-0.99)	0.98	(0.94-0.99)
<i>Bench press, kg</i>	79	0.96	(0.95-0.98)	0.96	(0.91-0.97)	63	0.96	(0.95-0.98)	0.96	(0.92-0.97)
<i>Leg extension, kg</i>	92	0.93	(0.91-0.95)	0.91	(0.83-0.95)	77	0.94	(0.91-0.96)	0.92	(0.91-0.96)
<i>Pec deck, kg</i>	92	0.90	(0.86-0.92)	0.88	(0.80-0.92)	75	0.89	(0.84-0.92)	0.87	(0.79-0.92)
<i>Wide grip lateral pull-down, kg</i>	110	0.91	(0.87-0.93)	0.89	(0.83-0.93)	77	0.89	(0.85-0.92)	0.87	(0.78-0.91)
<i>Leg curl, kg</i>	92	0.84	(0.79-0.89)	0.83	(0.77-0.88)	75	0.87	(0.81-0.91)	0.85	(0.77-0.90)
<i>Seated cable row, kg</i>	85	0.92	(0.88-0.94)	0.90	(0.82-0.94)	72	0.93	(0.90-0.95)	0.92	(0.86-0.95)
TUG, s	113	0.87	(0.83-0.91)	0.87	(0.83-0.90)	86	0.87	(0.82-0.91)	0.87	(0.82-0.90)
6MWT, m	110	0.96	(0.94-0.97)	0.95	(0.93-0.97)	85	0.96	(0.95-0.97)	0.96	(0.93-0.97)
30s-sit-to-stand, reps	105	0.85	(0.80-0.89)	0.80	(0.61-0.89)	82	0.86	(0.80-0.90)	0.81	(0.63-0.89)

Table 2 - Intraclass correlation coefficient for consistency and agreement of muscle strength and physical performance tests. 1 = consistency, 2 = agreement, ICC = Intraclass Correlation Coefficient, CI95% = Confidence Interval of 95%, TUG = Timed up and go, 6MWT = Six Minute Tests

MDC (Minimum Detectable Change)

Table 3 presents the bias (systematic bias) between the initial test and the retest, as well as the MDC. In the case of the 1RM tests, the bias ranged from 5% to 8% concerning the average of the test and retest. The minimum difference necessary to consider a real change (MDC) varied from 15% to 51% in relation to the average of the test and retest. Notably, the leg press and bench press exercises displayed the least variation, while the leg curl and pec deck exercises exhibited the most variation.

Regarding the PP tests, the MDC ranged from 10% to 30%. The 6MWT showed the least variation (10%), while the 30s-sit-to-stand test exhibited the most variation (30%).

In the context of the 1RM tests, a positive bias ($r = 0.23$, $P = 0.027$) was observed for the leg extension exercise. However, there was no significant association between the mean value and the difference between the initial test and the retest for the other exercises and PP tests.

Table 3: Values of agreement (bias) between test and retest and limits of agreement.

Tests	n	Test (A)	Re-test (B)	Mean	Bias	SD of bias.	SEM	SEP	MDC	r	P-value
<i>Leg press, kg</i>	79	146.3±49.6	153.3±51.5	149.8	7.0 (5%)	8.0	5.6	11.3	22.2 (15.0%)	0.23	0.038
<i>Bench press, kg</i>	79	25.6±6.5	26.5±6.7	26.0	0.9 (4%)	1.9	1.3	2.5	4.9 (19%)	0.11	0.318
<i>Leg extension, kg</i>	92	25.0±9.1	26.8±9.4	25.9	1.9 (7%)	3.4	2.4	4.7	9.2 (35%)	0.08	0.445
<i>Pec deck, kg</i>	92	19.1±6.9	20.6±7.4	19.9	1.5 (8%)	3.3	2.3	4.5	8.8 (44%)	0.16	0.126
<i>Wide grip lateral pull-down, kg</i>	110	29.5±14.1	30.7±14.2	30.1	1.3 (4%)	3.2	2.4	4.7	9.2 (31%)	0.09	0.312
<i>Leg curl, kg</i>	92	17.8±6.1	18.8±6.1	18.3	1.0 (6%)	3.4	2.4	4.8	9.4 (51%)	0.02	0.818
<i>Seated cable row, kg</i>	85	38.6±10.4	40.7±10.3	39.6	2.1 (5%)	4.2	3.0	5.8	11.4 (29%)	-0.02	0.829
TUG, s	113	7.0±1.0	6.9±1.0	7.0	-0.1 (-1%)	0.5	0.4	0.7	1.4 (20%)	0.02	0.823
6MWT, m	110	503.0±66.9	509.0±64.5	506.0	6.2 (1%)	19.9	13.4	26.5	51.9 (10%)	-0.13	0.184
30s-sit-to-stand, reps	105	14.8±3.2	15.9±3.0	15.4	1.1 (7%)	1.7	1.2	2.3	4.6 (30%)	-0.16	0.093

TUG = Timed up and go, 6MWT = Six Minute Tests, Bias = difference between A and B, SD = standard deviation, SEM = standard error of measurement, SEP = standard error of prediction, MDC = minimal detectable change

Correlation Analysis

The correlation analysis revealed no significant association between the deltas (Δ = retest - test) of each test and age. The following correlation values were observed: TUG (P = 0.479, R = -0.073), 6MWT (P = 0.924, R = 0.010), 30s-sit-to-stand (P = 0.342, R = 0.099), 1RM 45° leg press (P = 0.739, R = -0.037), 1RM bench press (P = 0.469, R = -0.082), 1RM pec deck (P = 0.516, R = 0.072), 1RM leg curl (P = 0.149, R = 0.160), 1RM wide-grip lateral pull-down (P = 0.385, R = 0.095), 1RM seated cable row (P = 0.769, R = -0.033), and 1RM leg extension (P = 0.871, R = -0.018).

DISCUSSION

The objective of this study was to assess the test-retest reliability of MS and PP tests in women aged 40 years and older. While these tests display excellent relative reliability, indicated by strong test-retest correlations, among women in this age group, they also exhibit low absolute reliability, characterized by significant repeated-measurement variability (fluctuation). This variability is particularly prominent in the case of the 1RM (one-repetition maximum) test conducted on a machine and the 30s-sit-to-stand test. In essence, these tests consistently produce similar results when assessing the performance of women within the same age group across multiple measurements. However, the precision of these measurements is relatively low, leading to noticeable fluctuations in results when the tests are repeated.

The Intraclass Correlation Coefficients (ICCs) for the 1RM tests ranged from 0.80 to 0.99, which aligns with the findings of another study conducted in a similar population^{11, 14}. It is noteworthy that the bench press and 45° leg press exercises exhibited higher ICCs values compared to the other exercises. Exercises that involve the manipulation of specific loads using machines, such as a pre-defined increase of 5 kg and estimation of 1RM using predictive formulas, may have contributed to lower ICCs. This suggests that the use of free weights, which allows for easier manipulation of load, may have contributed to these favorable results^{11, 14}. Despite these variations, it is important to emphasize that all exercises demonstrated excellent or good relative reliability, with ICCs exceeding 0.80. This implies that 1RM tests exhibit consistent and agreement in participants' rankings between the initial test and the retest, irrespective

of whether they are performed on a machine or with free weights ³. Consequently, this suggests that 1RM tests are a suitable tool for diagnosing and classifying MS, distinguishing between higher and lower MS, in women aged 40 years and older.

PP tests play a crucial role in assessing the risk of falls and mortality, making them valuable tools in both clinical practice and research. Previous studies have utilized this tool to diagnose poor PP in older women⁷, but it's important to note that declines in PP are observed in younger women during the menopausal transition^{4, 5}. Thus, this tool can be useful for the early diagnosis of low PP in women. In our study, the PP tests demonstrated excellent relative reliability (ICCs > 0.82), consistent with findings from previous studies conducted with elderly individuals¹⁵. The relative reliability values reported in these studies align with those observed in the evaluation of PP tools included in our study, such as the 30s-sit-to-stand⁶, 6MWT⁷, and TUG⁸ tests. This implies that PP tests exhibit consistency and agreement in participants' rankings between the initial test and the retest. Consequently, this suggests that PP tests are a suitable tool for distinguishing higher and lower PP in women aged 40 years and older.

Interestingly, age and menopausal status did not seem to have an impact on the changes and positions of the measurements. This lack of influence is likely attributed to the ease of understanding and applicability of the tests. Consequently, these tests demonstrate strong relative reliability across different age groups and physiological conditions, firmly establishing them as excellent tools for diagnosing high and low PP, as well as MS, in both pre- and post-menopausal women.

In the context of planning and interpreting interventions or longitudinally monitoring patients, it is crucial to have a comprehensive understanding of systematic errors and the limits of errors (MDC) associated with commonly used tests, as these can occur^{2, 3}. In our current research analysis, systematic errors were observed in all physical functional tests, contributing to the determination of an MDC. The MDC plays a pivotal role in helping professionals interpret a significant change between pre-intervention and post-intervention conditions. To illustrate this concept, let's take the example of the 45° leg press exercise. In our study, the MDC for the 45° leg press exercise was determined as 22.2 kg, which corresponds to approximately 15% of the average value obtained by the sample in the test. This implies that changes exceeding

this threshold are considered true changes, beyond measurement error, and help account for instrumental errors^{2, 3}. Therefore, any observed changes in the 45° leg press exercise that surpass this threshold can be considered meaningful (true changes) and should not be solely attributed to measurement variability^{2, 3}. Detecting the MDC is of utmost importance in clinical trials as it aids in determining the clinical significance of the intervention. By quantifying the necessary change in the evaluated variable that exceeds possible measurement errors, researchers can confidently interpret the results and assess the true impact of the intervention.

The MDC values obtained in our study are consistent with those reported in other studies that utilized tests to evaluate PP^{8, 16}. However, they differ from the limited literature available on the 1RM test^{14, 17} and the PP tests^{6, 7}. One possible explanation for this discrepancy could be attributed to the formula of MDC. Most calculations relied on the standard error of measurement (SEM), which may not provide the highest degree of precision^{2, 12}. As per the recommendation^{2, 12}, we calculate the MDC by considering the standard error of the prediction. Another possible explanation for this discrepancy could be attributed to the different equipment used in our research. Some equipment employed in our study had distinct biomechanical characteristics compared to those used in the existing literature. It is important to note that in professional and laboratory settings for strength training, there is a wide variety of equipment, each with its unique biomechanical properties. Therefore, future studies should investigate the MDC of 1RM tests using machinery with diverse biomechanical characteristics to provide a more comprehensive understanding of their performance.

In our study, we did not observe significant systematic changes (bias), but we did identify a high error limit value (MDC). For both the 1RM tests and the 30s sit-and-stand test, the error limit exceeded 30% between repeated measurements. This error limit may surpass the typical increases in MS and PP commonly seen in various interventions (such as nutritional¹⁸ or strength training¹⁹) or influenced by factors like age or specific health conditions, including disabling diseases²⁰. This observation is noteworthy because the primary application of the 1RM and physical performance tests often involves assessing changes in MS following a particular training program or longitudinally monitoring patients.

It is important to acknowledge certain limitations of this study. Firstly, our

research was conducted by a single evaluator, which means that we cannot identify errors that may occur between different evaluators. Secondly, the observed differences in relative and absolute reliability are specific to particular exercises and tests in the study. It is important to highlight that our study lacked statistical power to evaluate the test-retest measures in pre-menopausal women, as less than 20% of the sample fell under this category. Finally, this study focused on a specific target population of women aged 40 and over. Therefore, caution should be exercised when generalizing the findings to other populations with different physical and physiological characteristics. On the other hand, a strength of this study is the inclusion of a large number of women covering a wide age range, from the menopausal transition to the postmenopausal phase.

CONCLUSION

While the MS and PP tests demonstrate excellent relative reliability (test-retest correlation) in women over 40 years of age, they exhibit low absolute reliability (variability of repeated measures), particularly the 1RM test performed on the machine and the 30s-sit-to-stand test. As a result, these tests are effective tools for identifying or diagnosing women with high or low MS and PP. However, they may have limitations when assessing interventions that produce small effects on these tests. Therefore, changes in these tests that fall below the MDC may indicate that middle-aged and older women have not experienced a true change in their MS or PP.

REFERENCES

1. Patrizio E, Calvani R, Marzetti E, Cesari M. Physical Functional Assessment in Older Adults. *The Journal of frailty & aging*. 2021;10(2):141-9.
2. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of strength and conditioning research*. [Review]. 2005;19(1):231-40.
3. Riemann BL, Lininger MR. Statistical Primer for Athletic Trainers: The Essentials of Understanding Measures of Reliability and Minimal Important Change. *Journal of athletic training*. [Review]. 2018;53(1):98-103.
4. El Khoudary SR, Greendale G, Crawford SL, et al. The menopause transition and women's health at midlife: a progress report from the Study of Women's Health Across the Nation (SWAN). *Menopause*. [Research Support, N.I.H., Extramural Review]. 2019;26(10):1213-27.
5. Tseng LA, El Khoudary SR, Young EA, et al. The association of menopause status with physical function: the Study of Women's Health Across the Nation. *Menopause*. [Multicenter Study Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. 2012;19(11):1186-92.
6. Munoz-Bermejo L, Adsuar JC, Mendoza-Munoz M, et al. Test-Retest Reliability of Five Times Sit to Stand Test (FTSST) in Adults: A Systematic Review and Meta-Analysis. *Biology*. [Review]. 2021;10(6).
7. Chan WLS, Pin TW. Reliability, validity and minimal detectable change of 2-minute walk test, 6-minute walk test and 10-meter walk test in frail older adults with dementia. *Experimental gerontology*. [Validation Study]. 2019;115:9-18.
8. Yuksel E, Kalkan S, Cekmece S, Unver B, Karatosun V. Assessing Minimal Detectable Changes and Test-Retest Reliability of the Timed Up and Go Test and the 2-Minute Walk Test in Patients With Total Knee Arthroplasty. *The Journal of arthroplasty*. 2017;32(2):426-30.
9. Spagnuolo DL, Jurgensen SP, Iwama AM, Dourado VZ. Walking for the

assessment of balance in healthy subjects older than 40 years. *Gerontology*. [Comparative Study Research Support, Non-U.S. Gov't Validation Study]. 2010;56(5):467-73.

10. Shoemaker MJ, Curtis AB, Vangsnes E, Dickinson MG. Clinically meaningful change estimates for the six-minute walk test and daily activity in individuals with chronic heart failure. *Cardiopulmonary physical therapy journal*. 2013;24(3):21-9.

11. Grgic J, Lazinica B, Schoenfeld BJ, Pedisic Z. Test-Retest Reliability of the One-Repetition Maximum (1RM) Strength Assessment: a Systematic Review. *Sports medicine - open*. [Review]. 2020;6(1):31.

12. Charter RA. Revisiting the standard errors of measurement, estimate, and prediction and their application to test scores. *Perceptual and Motor Skills*. 1996;82(3_suppl):1139-44.

13. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*. 2016;15(2):155-63.

14. Barbalho M, Gentil P, Raiol R, Del Vecchio FB, Ramirez-Campillo R, Coswig VS. High 1RM Tests Reproducibility and Validity are not Dependent on Training Experience, Muscle Group Tested or Strength Level in Older Women. *Sports (Basel)*. 2018;6(4).

15. Boshnjaku A, Bahtiri A, Feka K, Krasniqi E, Tschan H, Wessner B. Test-retest reliability data of functional performance, strength, peak torque and body composition assessments in two different age groups of Kosovan adults. *Data in brief*. 2021;36:106988.

16. Murdoch M, Window P, Morton C, O'Donohue R, Ballard E, Claus A. People at a persistent pain service can walk it, but some struggle to talk about it: Reliability, detectable difference and clinically important difference of the six-minute walk test. *Musculoskeletal care*. [Research Support, Non-U.S. Gov't]. 2023;21(1):221-31.

17. Infante MA, Harrell GM, Strand KL, Parrino RL, Signorile JF. One Repetition Maximum Test-Retest Reliability and Safety Using Keiser Pneumatic Resistance

Training Machines With Older Women. *Journal of strength and conditioning research*. 2021;35(12):3513-7.

18. Nasimi N, Sohrabi Z, Nunes EA, et al. Whey Protein Supplementation with or without Vitamin D on Sarcopenia-Related Measures: A Systematic Review and Meta-Analysis. *Adv Nutr. [Meta-Analysis Review Systematic Review]*. 2023;14(4):762-73.

19. Cuyul-Vasquez I, Pezo-Navarrete J, Vargas-Arriagada C, et al. Effectiveness of Whey Protein Supplementation during Resistance Exercise Training on Skeletal Muscle Mass and Strength in Older People with Sarcopenia: A Systematic Review and Meta-Analysis. *Nutrients. [Meta-Analysis Review Systematic Review]*. 2023;15(15).

20. Picca A, Coelho-Junior HJ, Calvani R, Marzetti E, Vetrano DL. Biomarkers shared by frailty and sarcopenia in older adults: A systematic review and meta-analysis. *Ageing research reviews. [Meta-Analysis Research Support, Non-U.S. Gov't Review Systematic Review]*. 2022;73:101530.