

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE EDUCAÇÃO FÍSICA, FISIOTERAPIA E DANÇA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DO MOVIMENTO HUMANO

Rafael Paiva Ribeiro

**VALIDADE, REPRODUTIBILIDADE E ACURÁCIA DIAGNÓSTICA DE TESTES
SACROILÍACOS: TESTE DE FLEXÃO EM PÉ (TFP), TESTE DE FLEXÃO
SENTADO (TFS) E TESTE DE DOWNING**

Porto Alegre

2020

Rafael Paiva Ribeiro

**VALIDADE, REPRODUTIBILIDADE E ACURÁCIA DIAGNÓSTICA DE TESTES
SACROILÍACOS: TESTE DE FLEXÃO EM PÉ (TFP), TESTE DE FLEXÃO
SENTADO (TFS) E TESTE DE DOWNING**

Dissertação de mestrado submetida ao Programa de Pós-Graduação em Ciências do Movimento Humano, da Escola de Educação Física, Fisioterapia e Dança, da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Mestre em Ciências do Movimento Humano.

Orientadora: Prof^ª. Dr^ª. Cláudia Tarragô Candotti

Porto Alegre

2020

CIP - Catalogação na Publicação

Ribeiro, Rafael Paiva

VALIDADE, REPRODUTIBILIDADE E ACURÁCIA DIAGNÓSTICA DE TESTES SACROILÍACOS: TESTE DE FLEXÃO EM PÉ (TFP), TESTE DE FLEXÃO SENTADO (TFS) E TESTE DE DOWNING / Rafael Paiva Ribeiro. -- 2020.

119 f.

Orientadora: Cláudia Tarragô Candotti.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Escola de Educação Física, Programa de Pós-Graduação em Ciências do Movimento Humano, Porto Alegre, BR-RS, 2020.

1. Articulação Sacroilíaca. 2. Confiabilidade e Validade. 3. Acurácia Diagnóstica. 4. Fisioterapia. 5. Osteopatia. I. Candotti, Cláudia Tarragô, orient. II. Título.

Rafael Paiva Ribeiro

**VALIDADE, REPRODUTIBILIDADE E ACURÁCIA DIAGNÓSTICA DE TESTES
SACROILÍACOS: TESTE DE FLEXÃO EM PÉ (TFP), TESTE DE FLEXÃO
SENTADO (TFS) E TESTE DE DOWNING**

Conceito final: _____

Aprovado em dede.....

BANCA EXAMINADORA

Prof^ª. Dr^a Clarice Sperotto dos Santos Rocha – UFRGS

Prof. Dr. Jefferson Fagundes Loss – UFRGS

Prof. Dr. Sandro da Silva Groisman – IBO

Orientadora - Prof^ª. Dr^a. Cláudia Tarragô Candotti – UFRGS

AGRADECIMENTOS

Primeiramente agradeço por ter saúde e por ter a possibilidade em finalizar esta Dissertação em plena pandemia por covid-19. Tenho a consciência de que isso é um privilégio para poucos e, de posse disso, busco e luto para que acabe. Por um dia em que as oportunidades sejam as mesmas para todos, e que as desigualdades não sejam tão influentes.

Agradeço à minha Família, por serem a minha base e estrutura, por permitirem a concretização dos meus objetivos e metas. Por serem meu início, meio e fim.

Agradeço à minha companheira, Ana, por permitir dividir a vida, apoiando-me nos momentos mais difíceis e partilhando os momentos mais felizes.

Agradeço à minha orientadora, Professora Cláudia, por confiar em mim e se aventurar em um projeto de pesquisa desafiador. Agradeço pelos 7 anos de ensinamentos, desde 2013 quando tinha 18 anos, ensina-me a fazer ciência, acampanhou e influenciou de perto todo o meu desenvolvimento profissional. A Professora mais importante da minha vida, que sempre me lembrarei.

Agradeço à minha equipe de avaliação, Filipe, Eduardo, Luiza e Mateus, por dedicarem seus tempos em me ajudar. Sem vocês esse trabalho não seria possível.

Agradeço ao Professor Jefferson, por todas as contribuições, pela assistência com os protocolos de avaliação e todos os pilotos.

Agradeços aos componentes da banca, Professora Clarice e Professor Sandro, pela prontidão em contribuir com esse estudo. É uma honra para mim.

Agradeço ao Grupo Biomec, por todas as trocas e trabalhos em conjunto durante esses anos, um grupo de pesquisa que trabalha unido, e faz ciência como ela tem que ser, com ética, seriedade e qualidade.

Por fim, agradeço à CAPES, pela contribuição com a bolsa, foi de extrema importância e um significativo estímulo durante essa jornada.

RESUMO

A região pélvica, estrutura anatômica que envolve a articulação sacroilíaca (ASI), é importante fonte de dores e disfunções, além de ser uma das principais regiões responsáveis pelo alinhamento e estabilidade corporal. A disfunção sacroilíaca pode ser definida por um estado de hipomobilidade da amplitude de movimento articular, seguida de uma alteração posicional entre a relação do sacro e o ilíaco. Dentre as possibilidades de avaliação da ASI, estão o Teste de Flexão em Pé (TFP), o Teste de Flexão Sentado (TFS) e o teste de Downing, contudo, esses testes clínicos palpatórios carecem de informações acerca das suas validades concorrentes, reprodutibilidades e acurácias diagnósticas. Desse modo, para melhor organização, esta dissertação foi dividida em três capítulos. **CAPÍTULO 1:** foi realizada uma revisão sistemática com o objetivo primário de investigar quais são os valores de reprodutibilidade e validade de testes clínicos palpatórios de mobilidade sacroilíaca, e objetivo secundário de investigar quais são os testes clínicos palpatórios de mobilidade sacroilíaca existentes na literatura. Essa revisão sistemática seguiu a metodologia PRISMA, as bases de dados PubMed, Embase, Scopus, Medline e PEDro foram pesquisadas utilizando as seguintes palavras chave, na língua inglesa: “Sacroiliac Joint”, “Range of Motion, Articular”, “Physical Examination” and “Palpation”, juntamente com seus termos livres. Não houve restrição quanto ao tipo de estudo e participantes. Foram incluídos 28 estudos, onde 15 testes clínicos palpatórios de mobilidade sacroilíaca foram descritos: (1) Teste de Flexão em Pé (TFP), (2) Teste de Gillet, (3) Teste de Flexão Sentado (TFS), (4) Teste de Flexão de joelho em prono, (5) Teste de Stork, (6) Supine Long Sitting Test, (7) Teste de Downing, (8) Inferior test, (9) The Initiation of the Squat Motion test, (10) Bourdillon's method, (11) Mitchell and Moran, (12) Joint Play (Menell) Test, (13) Spine Test, (14) Lateral Flexion e (15) Overtake Phenomenon. Em relação a reprodutibilidade, 14 estudos realizaram a avaliação interavaliador, cuja concordância pela estatística kappa variou de -0,05 a 0,77, e 8 estudos também realizaram a avaliação intra-avaliador, variando o valor de kappa de 0,08 a 0,73. Nenhum estudo incluído nessa revisão sistemática verificou os valores de validade desses testes. A metanálise para a reprodutibilidade intra-avaliador mostrou moderada a boa concordância para o Teste de Gillet ($k=0,46$), TFP ($k=0,61$) e TFS ($k=0,68$). **CAPÍTULO 2:** o objetivo desse capítulo foi avaliar a validade concorrente, a reprodutibilidade intra e interavaliadores e a acurácia diagnóstica do Teste de Flexão em Pé (TFP) e do Teste de Flexão Sentado (TFS). Estudo prospectivo, onde a amostra foi composta por 30 indivíduos de ambos os sexos. A equipe de avaliação foi composta por 5 Fisioterapeutas (A, B, C, D e E). O padrão referência utilizado para a validade concorrente foi a medição tridimensional (cinemetria 3D), por meio do sistema BTS Smart-DX. Análise estatística: foi utilizado o percentual de concordância (%C) e a estatística kappa (k) para as análises de concordância, e a curva ROC (AUC) para determinar a capacidade discriminativa, os valores de sensibilidade e especificidade, além dos valores preditivos positivos (VPP) e negativos (VPN). Os resultados foram, para a validade concorrente: TFP (%C=66,7%; $k=0,42$; $p<0,01$) e TFS (%C=46,7%; $k=0,12$; $p=0,25$). Curva ROC: TFP (AUC=0,700; $p<0,05$; sensibilidade=80%; especificidade=60%; VPP=85%; VPN=78%) e TFS (AUC=0,602; $p=0,52$; sensibilidade=75%; especificidade=46%; VPP=20%; VPN=91%). Reprodutibilidade intra-avaliador: TFP (%C=66,3%; $k=0,43$; $p<0,01$) e TFS (%C=56,7%; $k=0,38$; $p<0,01$). Reprodutibilidade interavaliador: TFP (%C=10%; $k=-0,02$; $p=0,825$) e TFS (%C=13,3%; $k=0,01$; $p=0,836$). Portanto, o Teste de Flexão em Pé (TFP) é válido, reprodutível entre o mesmo avaliador e não acurado. O Teste de Flexão Sentado (TFS) não é válido, não é reprodutível e nem acurado. **CAPÍTULO 3:** o objetivo desse capítulo foi avaliar a validade concorrente, a reprodutibilidade

intra e interavaliadores e a acurácia diagnóstica do teste de Downing. Estudo prospectivo, onde a amostra foi composta por 30 indivíduos de ambos os sexos. A equipe de avaliação foi composta por 5 Fisioterapeutas (A, B, C, D e E). O padrão referência utilizado para a validade concorrente foi a medição tridimensional (cinemetria 3D), por meio do sistema BTS Smart-DX. Análise estatística: foi utilizado o percentual de concordância (%C) e a estatística kappa (k) para as análises de concordância, e a curva ROC (AUC) para determinar a capacidade discriminativa, os valores de sensibilidade e especificidade, além dos valores preditivos positivos (VPP) e negativos (VPN). Os resultados foram: validade concorrente (%C=54,5%; k=0,43; p<0,01), curva ROC para disfunção posterior (AUC=0,889; p≤0,05; sensibilidade=78%; especificidade=100%; VPP=100%; VPN=60%), curva ROC para disfunção anterior (AUC=0,657; p=0,37; sensibilidade=71%; especificidade=60%; VPP=71%; VPN=60%), reprodutibilidade intra-avaliador (%C=66,7%; k=0,43; p<0,01) e reprodutibilidade interavaliador (%C=16,7%; k=0,15; p<0,05). Portanto, o teste de Downing é válido, reprodutível somente entre o mesmo avaliador e acurado para disfunção posterior.

Palavras-chave: Terapia Manual; Fisioterapia; Amplitude de Movimento Articular; Articulação Sacroilíaca; Confiabilidade e Validade; Revisão Sistemática; Metanálise; Exame Físico.

ABSTRACT

The pelvic region, an anatomical structure that involves a sacroiliac joint (SIJ), is an important source of pain and dysfunction, in addition to being one of the main priorities for body alignment and stability. Sacroiliac dysfunction can be defined by a state of hypomobility in the range of motion of the joint, followed by a positional change between the relationship between the sacrum and the iliac. Among the possibilities of SIJ evaluation are the Standing Flexion Test (STFT), the Sitting Flexion Test (SIFT) and the Downing test, however, these palpatory clinical tests lack information about their competing validities, reproducibilities and accuracy. diagnostics. Thus, for better organization, this dissertation was divided into three chapters. **CHAPTER 1:** a systematic review was carried out with the primary objective is to investigate what are the values of reliability and validity of palpatory clinical tests of sacroiliac mobility, and secondary objective is to investigate which palpatory clinical tests of sacroiliac mobility exist in the literature. This systematic review followed the PRISMA methodology, the PubMed, Embase, Scopus, Medline and PEDro databases were searched using the following keywords: "Sacroiliac Joint", "Range of Motion, Articular", "Physical Examination" and "Palpation", along with its free terms. There was no restriction on the type of study and participants. 28 studies were included, where 15 palpatory clinical tests of sacroiliac mobility were described: (1) Standing Flexion Test (STFT), (2) Gillet Test, (3) Sitting Flexion Test (SIFT), (4) Prone Knee Flexion, (5) Stork Test, (6) Supine Long Sitting Test, (7) Downing Test, (8) Inferior test, (9) The Initiation of the Squat Motion test, (10) Bourdillon's method, (11) Mitchell and Moran, (12) Joint Play (Menell) Test, (13) Spine Test, (14) Lateral Flexion and (15) Overtake Phenomenon. Regarding reliability, 14 studies performed the inter-examiner evaluation, whose agreement by the kappa statistic ranged from -0.05 to 0.77, and 8 studies also performed the intra-examiner evaluation, ranged the kappa value from 0.08 to 0.73. No study included in this systematic review has verified the validity values of these tests. The meta-analysis for intra-examiner reliability showed moderate to good agreement for the Gillet Test ($k = 0.46$), STFT ($k = 0.61$) and SIFT ($k = 0.68$). **CHAPTER 2:** the objective of this chapter was to evaluate concurrent validity, intra- and inter-examiner reliability and the diagnostic accuracy of the Standing Flexion Test (STFT) and the Sitting Flexion Test (SIFT). Prospective study, where the sample was composed of 30 individuals of both sexes. The evaluation team was composed of 5 Physiotherapists (A, B, C, D and E). The reference standard used for concurrent validity was three-dimensional measurement (3D kinematics), using the BTS Smart-DX system. Statistical analysis: the percentage of agreement (%C) and the kappa statistic (k) were used for the analysis of agreement, and the analysis of the ROC curve (AUC) to determine the discriminative capacity, the sensitivity and specificity values, in addition to the positive and negative predictive values (PPV and NPV). The results were, for concurrent validity: STFT (%C=66.7%; $k=0.42$; $p<0.01$) and SIFT (%C=46.7%; $k=0.12$; $p=0.25$). ROC curve: STFT (AUC=0.700; $p<0.05$; sensitivity=80%; specificity=60%; PPV=85%; NPV=78%) and SIFT (AUC=0.602; $p=0.52$; sensitivity=75%; specificity=46%; PPV= 20%; NPV=91%). Intra-examiner reliability: STFT (%C= 66.3%; $k=0.43$; $p<0.01$) and SIFT (%C=56.7%; $k=0.38$; $p<0.01$). Inter-examiner reliability: STFT (%C=10%; $k=-0.02$; $p=0.825$) and SIFT (%C=13.3%; $k=0.01$; $p=0.836$). Therefore, the Standing Flexion Test (STFT) is valid, reliable between the same examiner and not accurate. The Seated Flexion Test (SIFT) is not valid, it is not reproducible or accurate. **CHAPTER 3:** the objective of this chapter was to evaluate concurrent validity, intra- and inter-examiner reliability and the diagnostic accuracy of the Downing test. Prospective study, where the sample was composed of 30 individuals of both sexes. The evaluation team was composed of 5 Physiotherapists (A, B, C, D and E). The reference standard used for concurrent validity was three-dimensional measurement (3D kinematics), using the BTS Smart-DX system. Statistical analysis: the percentage of agreement

(%C) and the kappa statistic (k) were used for the analysis of agreement, and the analysis of the ROC curve (AUC) to determine the discriminative capacity, the sensitivity and specificity values, in addition to the positive and negative predictive values (PPV and NPV). The results were: concurrent validity (%C=54.5%; k=0.43; p<0.01), ROC curve for posterior dysfunction (AUC=0.889; p<0.05; sensitivity=78%; specificity=100%; PPV=100%; NPV=60%), ROC curve for previous dysfunction (AUC=0.657; p=0.37; sensitivity=71%; specificity=60%; PPV=71%; NPV=60%), intra-examiner reliability (%C=66.7%; k=0.43; p<0.01) and inter-examiner reliability (%C=16.7%; k=0.15; p<0.05). Therefore, the Downing test is valid, reproducible between the same evaluator and accurate for posterior dysfunction.

Keywords: Manual Therapy; Physiotherapy; Articular Range of Motion; Sacroiliac joint; Reliability and Validity; Systematic review; Meta-analysis; Physical examination.

LISTA DE TABELAS, FIGURAS E APÊNDICES

Capítulo 1

Figura 1 - Fluxogram of included studies according to PRISMA. ¹⁴	23
Tabela 1 - Results of the methodological quality assessment, according to the validation and reliability critical appraisal scale. ¹⁹	25
Tabela 2 – Palpatory clinical tests of sacroiliac mobility found in studies included in the systematic review.	26
Tabela 3 - Data from studies reporting inferential statistical analysis.....	31
Tabela 4 – Results of statistical analyses of reliability studies.....	33
Figura 2 - Intra-examiner reliability meta-analysis (k); Random effect was used.....	38

Capítulo 2

Figura 1 - Desenho do 1º dia de avaliação.....	50
Figura 2 - Desenho do 2º dia de avaliação.....	51
Figura 3a - Posição inicial do TFP. Avaliador realizando a palpação das EIPs.....	52
Figura 3b - Posição final do TFP. Avaliador observa a movimentação das EIPs.....	52
Figura 4a - Posição inicial do TFS. Avaliador realizando a palpação das EIPs.....	53
Figura 4b - Posição final do TFS. Avaliador observa a movimentação das EIPs.....	53
Figura 5 - Marcador refletivo colado no polegar do avaliador A (I) e palpação da Espinha Ilíaca Pósterior Superior (EIPS) (II).....	54
Figura 6 - SCL criado utilizando uma faixa com quatro marcadores refletivos acima dos olhos.....	55
Figura 7a – Diagrama de fluxo referente ao Teste de Flexão em Pé (TFP), realizado pelo avaliador A, e a medição tridimensional (cinemetria 3D), para o ponto de corte de 3 mm.....	58
Figura 7b - Diagrama de fluxo referente ao Teste de Flexão Sentado (TFS), realizado pelo avaliador A, e a medição tridimensional (cinemetria 3D), para o ponto de corte de 7 mm.....	59
Tabela 1 – Informações descritivas das avaliações com a medição tridimensional (cinemetria 3D), do Teste de Flexão em Pé (TFP) e do Teste de Flexão Sentado (TFS), referente ao 1º dia de avaliação.....	60
Tabela 2 - Validade concorrente entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão em Pé (TFP) para os 5 pontos de corte.....	61
Tabela 3 - Validade concorrente entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão Sentado (TFS), para os 5 pontos de corte.....	62
Tabela 4 - Análise da curva ROC entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão em Pé (TFP) para os 5 pontos de corte.....	64

Tabela 5 - Análise da curva ROC entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão Sentado (TFS) para os 5 pontos de corte.....	66
Tabela 6 – Reprodutibilidade intra-avaliador (BxB) para o o Teste de Flexão em Pé (TFP) e Teste de Flexão Sentado (TFS).....	67
Tabela 7 – Reprodutibilidade interavaliador (AxBxC; AxB; AxC; BxC) para o Teste de Flexão em Pé (TFP) e Teste de Flexão Sentado (TFS).....	68
Apêndice 1 - Quadro de anotação do Teste de Flexão em Pé (TFP) e Teste de Flexão Sentado (TFS).....	78
Capítulo 3	
Figura 1 - Desenho do 1º dia de avaliação.....	83
Figura 2 - Desenho do 2º dia de avaliação.....	84
Figura 3 - Execução do teste de Downing (ponte, manobra de alongamento e manobra de encurtamento)	86
Figura 4 - Marcadores refletivos colados nos polegares do avaliador A, palpando a EIAS (I) e MM (II).....	87
Figura 5 - SCL criado utilizando uma faixa com quatro marcadores refletivos acima dos olhos.....	88
Figura 6 - Diagrama de fluxo referente ao teste de Downing (avaliador A) e a medição tridimensional (cinemetria 3D), com base no ponto de corte de 3 milímetros, considerando as avaliações pelos MMs, para a conclusão do teste.....	91
Tabela 1 – Informações descritivas das avaliações com a medição tridimensional (cinemetria 3D), com base nos MMs e nas EIASs, referente ao 1º dia de avaliação.....	93
Tabela 2 – Validade concorrente entre avaliador A e a medição tridimensional (cinemetria 3D) do teste de Downing, com base nas mensurações dos MMs e das EIASs.....	94
Tabela 3 - Concordância entre as avaliações feitas com base nos MMs e nas EIASs para a medição tridimensional (cinemetria 3D), feitas pelo avaliador A.....	97
Tabela 4 – Análise da curva ROC entre o avaliador A e a medição tridimensional (cinemetria 3D), com base no MMs, para o ponto de corte de 3 mm.....	99
Tabela 5 – Reprodutibilidade intra-avaliador (BxB) do teste de Downing, com base nas mensurações dos MMs e das EIASs.....	101
Tabela 6 – Reprodutibilidade interavaliador (AxBxC) do teste de Downing, com base nas mensurações dos MMs e das EIASs.....	103
Tabela 7 – Reprodutibilidade interavaliador (AxB; AxC; BxC) do teste de Downing, com base nas mensurações dos MMs e das EIASs.....	105
Apêndice 1 - Quadro de anotação do teste de Downing.....	115

LISTA DE ABREVIATURAS E SIGLAS

SIJ: Sacroiliac Joint

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

STFT: Standing Flexion Test

SIFT: Sitting Flexion Test

CDR: Clinical Diagnostic Rule

PSIS: Postero Superior Iliac Spine

PPV: Positive Predictive Value

NPV: Negative Predictive Value

Na: Not applicable

Se: Standard error

Sd: Standard deviation

CI: Confidence Interval

ASI: Articulação Sacroilíaca

TFP: Teste de Flexão em Pé

TFS: Teste de Flexão Sentado

LAPEX: Laboratório de Pesquisa do Exercício

ESEFID: Escola de Educação Física, Fisioterapia e Dança

UFRGS: Universidade Federal do Rio Grande do Sul

JMPT: Journal of Manipulative and Physiological Therapeutics

GRRAS: Guidelines for Reporting Reliability and Agreement Studies

STARD: Standards for Reporting of Diagnostic Accuracy

IMC: Índice de massa corporal

MsIs: Membros Inferiores

CAAE: Certificado de Apresentação de Apreciação Ética

TCLE: Termo de Consentimento Livre e Esclarecido

EIAS: Espinha Ilíaca Antero Superior

EIASs: Espinhas Ilíacas Antero Superiores

MM: Maléolo Medial

MMs: Maléolos Mediais

EIPS: Espinha Ilíaca Pósterio Superior

EIPSS: Espinhas Ilíacas Pósterio Superiores
SCL: Sistema de Coordenada Local
SPSS: Statistical Package for the Social Sciences
AUC: Área sob a curva
VPP: Valores Preditivos Positivos
VPN: Valores Preditivos Negativos
mm: Milímetros
SEM: Erro padrão de medida
p: Nível de significância
*: p estatisticamente significativo
n: Amostra
k: kappa
%C: Percentual de concordância
PI: Posição Inicial
PF: Posição Final
C: Conclusão do teste
IC: Intervalo de Confiança
ROC: Curva característica de operação do receptor
RSA: Análise Radiostereométrica
n: Paramétricos
nb: Não Paramétricos
+: Positivo
-: Negativo
D: Direita
E: Esquerda

SUMÁRIO

1 APRESENTAÇÃO	16
2 INTRODUÇÃO	17
3 CAPÍTULO 1	19
Validity and reliability of palpatory clinical tests of sacroiliac joint mobility: a systematic review and meta-analysis	19
ABSTRACT	19
INTRODUCTION	20
METHOD	21
Identification and selection of studies	21
Methodological quality assessment	22
Data extraction and analysis	22
RESULTS	23
The flow of studies through the review	23
Characteristics of included studies	24
DISCUSSION	40
CONCLUSION	43
REFERENCES	44
4 CAPÍTULO 2	47
Validade, reprodutibilidade e acurácia diagnóstica do Teste de Flexão em Pé (TFP) e do Teste de Flexão Sentado (TFS)	47
RESUMO	47
INTRODUÇÃO	48
MÉTODOS	49
DESENHO DO ESTUDO E PARTICIPANTES	49
PROTOCOLO DE AVALIAÇÃO	50
(1º dia) - Reprodutibilidade interavaliador e validade concorrente	50
(2º dia) - Reprodutibilidade intra-avaliador	52
ANAMNESE	52
TESTE DE FLEXÃO EM PÉ (TFP)	53
TESTE DE FLEXÃO SENTADO (TFS)	54
MEDIÇÃO TRIDIMENSIONAL	55
ANÁLISE ESTATÍSTICA	56
RESULTADOS	58
DISCUSSÃO	70

CONCLUSÃO	75
REFERÊNCIAS	76
APÊNDICE 1 – Quadro de anotação do Teste de Flexão em Pé (TFP) e Teste de Flexão Sentado (TFS).	79
5 CAPÍTULO 3	80
Validade, reprodutibilidade e acurácia diagnóstica do teste de Downing.....	80
RESUMO	80
INTRODUÇÃO.....	81
MÉTODOS	82
DESENHO DO ESTUDO E PARTICIPANTES	82
PROTOCOLO DE AVALIAÇÃO	83
(1º dia) - Reprodutibilidade interavaliador e validade concorrente	83
(2º dia) - Reprodutibilidade intra-avaliador	85
ANAMNESE	85
TESTE DE DOWNING	86
MEDIÇÃO TRIDIMENSIONAL (CINEMETRIA 3D)	88
ANÁLISE ESTATÍSTICA.....	89
RESULTADOS	91
DISCUSSÃO.....	109
CONCLUSÃO	114
REFERÊNCIAS.....	114
APÊNDICE 1 – Quadro de anotação do teste de Downing.....	116
6 CONSIDERAÇÕES FINAIS E PERSPECTIVAS.....	117
7 REFERÊNCIAS BIBLIOGRÁFICAS DA INTRODUÇÃO	118

1 APRESENTAÇÃO

O atual estudo foi realizado com o objetivo de investigar as propriedades de medida (validade concorrente, reprodutibilidade e acurácia diagnóstica) de três testes clínicos palpatórios de mobilidade sacroilíaca: Teste de Flexão em Pé (TFP), Teste de Flexão Sentado (TFS) e teste de Downing. Esses testes clínicos são descritos para auxiliar o profissional Fisioterapeuta e Osteopata na avaliação da Articulação Sacroilíaca (ASI), com o intuito de diagnosticar disfunções de mobilidade dessa articulação.

Para avaliar a validade concorrente dos testes, foi utilizado o sistema BTS Smart-DX (BTS Bioengineering, EUA), por meio de medições tridimensionais (cinemetria 3D). A reprodutibilidade, intra e interavaliador, foi mensurada a partir de três pesquisadores, nesse estudo chamados de A, B e C, Fisioterapeutas, um Osteopata mais experiente e dois estudantes de Osteopatia, treinados, independentes e cegados. Além disso, a acurácia diagnóstica foi investigada utilizando as análises da medição tridimensional e do pesquisador A. As coletas foram realizadas no Laboratório de Biomecânica, do Laboratório de Pesquisa do Exercício (LAPEX), da Escola de Educação Física, Fisioterapia e Dança (ESEFID), da Universidade Federal do Rio Grande do Sul (UFRGS).

Para dar conta desses objetivos, essa dissertação foi dividida em três capítulos: (1) revisão sistemática com metanálise, intitulada “Validity and reliability of palpatory clinical tests of sacroiliac joint mobility: a systematic review and meta-analysis”, cujo objetivo primário foi investigar quais são os valores de reprodutibilidade e validade de testes clínicos palpatórios de mobilidade sacroilíaca, e objetivo secundário foi investigar quais são os testes clínicos palpatórios de mobilidade sacroilíaca existentes na literatura. Esse estudo foi submetido na revista *Journal of Manipulative and Physiological Therapeutics* (JMPT) e já passou por duas rodadas de avaliação. No momento, encontra-se no estágio: “aguardando decisão do editor”; (2) capítulo sobre o Teste de Flexão em Pé (TFP) e o Teste de Flexão Sentado (TFS), intitulado “Validade, reprodutibilidade e acurácia diagnóstica do Teste de Flexão em Pé (TFP) e do Teste de Flexão Sentado (TFS)”, cujo objetivo foi avaliar a validade concorrente, a reprodutibilidade intra e interavaliadores e a acurácia diagnóstica do TFP e do TFS; e (3) capítulo sobre o teste de Downing, intitulado “Validade, reprodutibilidade e acurácia diagnóstica do teste de Downing”, cujo objetivo foi avaliar a validade concorrente, a reprodutibilidade intra e interavaliadores e a acurácia diagnóstica do teste de Downing. Os dois últimos capítulos seguem as diretrizes *Guidelines for Reporting Reliability and Agreement Studies* (GRRAS) e *Standards*

for Reporting of Diagnostic Accuracy (STARD) e serão organizados para submissão nas revistas *Journal of Manipulative and Physiological Therapeutics (JMPT)* e *International Journal of Osteopathic Medicine (IJOM)*.

2 INTRODUÇÃO

A região pélvica, estrutura anatômica que envolve a Articulação Sacroilíaca (ASI), é considerada uma base central e por ela são transmitidas forças direta e indiretamente para o restante do corpo. É uma das principais regiões responsáveis pelo alinhamento e estabilidade corporal (PRATHER, 2000), suportando a parte superior do corpo e amortecendo o impacto durante a deambulação (HARRISON, HARRISON, TROYANOVICH, 1997). As articulações sacroilíacas são multiplanares, realizam deslocamentos ao longo de três eixos de movimento (X, Y, Z), e sua amplitude pode chegar até 8° para a rotação e 8 milímetros (mm) para a translação (GOODE *et al.*, 2008). Os principais movimentos sacrais são descritos como rotação e translação (VLEEMING *et al.*, 1992), e o principal movimento do íliaco é a rotação (BEMIS, DANIEL, 1987). Dessa maneira, o sacro move-se em relação ao íliaco e o contrário também acontece e, quando a movimentação da articulação está prejudicada ou anormal, o termo disfunção é dado (BEMIS, DANIEL, 1987). A disfunção sacroilíaca pode ser definida por um estado de hipomobilidade da amplitude de movimento da articulação seguida de uma alteração posicional entre a relação do sacro e o íliaco (DREYFUSS *et al.*, 1994) e, segundo Mitchell (1958), o que mantém as disfunções são os músculos, fáscias, ligamentos e outras estruturas.

A ASI é também considerada uma importante fonte de dores e disfunções. Em relação à coluna lombar, a ASI é capaz de suportar cerca de seis a sete vezes mais carga (BERNARD, CASSIDY, 1991). Ainda, alguns autores compartilham da opinião de que a ASI é uma considerável fonte de dor lombar e dor nas nádegas (MAIGNE *et al.*, 1996; SHAW, 1992; STURESSON *et al.*, 1989; POTTER; ROTHSTEIN, 2006; SLIPMAN *et al.*, 2001), podendo irradiar para a parte posterior da coxa e/ou anterior na região inguinal (PRATHER, 2000). Já foi reportada uma prevalência de 13 a 30% de dor na ASI em pacientes com dor lombar (SCHWARZER, APRIL, BOGDUK, 1995; MAIGNE, AIVALIKLIS, PFEFER, 1996). Esse dado requer atenção por parte dos profissionais da área, que deveriam avaliar constantemente em seus pacientes a ASI, uma vez que a dor lombar é o principal problema de saúde pública em todo o mundo, com uma prevalência maior que 84% durante a vida (HOY *et al.*, 2010; AIRAKSINEN *et al.*, 2006).

Grande parte dos artigos e trabalhos científicos envolvendo técnicas manuais para a ASI são conduzidos para identificar os efeitos a curto ou longo prazo de manipulações ou técnicas de tratamento (AL-SUBAHI *et al.*, 2017). No entanto, para chegar a um diagnóstico clínico que resulte na necessidade de uma intervenção, é necessária a realização de uma avaliação minuciosa anteriormente. Na prática clínica, existem diversas técnicas de avaliação possíveis de serem realizadas pelo terapeuta manual, seja ele, o Fisioterapeuta, o Osteopata ou o Quiropraxista, já bem aceitas pela comunidade da área. Contudo, são raras as técnicas de avaliação disponíveis que possuem as suas propriedades de medidas como, validade, reprodutibilidade e acurácia diagnóstica, confirmadas (VAN DER WURFF, MEYNE, HAGMEIJER, 2000a, 2000b). Ainda, fora do contexto clínico, no ambiente da pesquisa científica, em relação a ASI, nota-se uma carência na literatura de estudos que investiguem os testes clínicos palpatórios avaliativos, sendo a última revisão sistemática sobre o assunto datada há 20 anos (VAN DER WURFF, MEYNE, HAGMEIJER, 2000a, 2000b). Nesse contexto, parece razoável assumir que tanto os profissionais clínicos quanto os pesquisadores da área mantêm suas atenções nos efeitos das intervenções, negligenciando a etapa anterior, que é crucial para a tomada de decisão de qual técnica de tratamento utilizar.

Na prática clínica existem alguns testes para investigar as disfunções ilíacas, como o Teste de Flexão em Pé (TFP) e o teste de Downing (VINCENT-SMITH, GIBBONS, 1999; MITCHEL, 1958), e outros para investigar as disfunções sacrais, como o Teste de Flexão Sentado (TFS) (POTTER, ROTHSTEIN, 2006). O TFP e o TFS são testes globais e estão entre os mais utilizados por Osteopatas americanos (FRYER; MORSE; JOHNSON, 2009). No entanto, as informações sobre a validade e acurácia diagnóstica desses três testes são desconhecidas, e os valores de reprodutibilidade são heterogêneos, não havendo um consenso (VAN DER WURFF, MEYNE, HAGMEIJER, 2000a, 2000b; VINCENT-SMITH; GIBBONS, 1999; ÅSTRÖM; GUMMESSON, 2014; ARAB *et al.*, 2009; RIDDLE, FREBURGER; NORTH AMERICAN ORTHOPAEDIC REHABILITATION RESEARCH NETWORK, 2002; BOWMAN, GRIBBLE, 1995; TONG *et al.*, 2006; PAYDAR, THIEL, GEMMELL, 1994; FRYER, MCPHERSON, O'KEEFE, 2005; LIMA *et al.*, 2019). Sendo assim, entende-se necessária a condução de estudos que busquem investigar as propriedades de medida desses testes palpatórios avaliativos tão corriqueiros na prática clínica de terapeutas manuais, tais como o Teste de Flexão em Pé (TFP), o Teste de Flexão Sentado (TFS) e o teste de Downing.

3 CAPÍTULO 1

Validity and reliability of palpatory clinical tests of sacroiliac joint mobility: a systematic review and meta-analysis

Palpatory clinical tests of sacroiliac joint mobility: a systematic review

ABSTRACT

Objective: The primary objective is to investigate what are the values of reliability and validity of palpatory clinical tests of sacroiliac mobility. The secondary is to investigate which palpatory clinical tests of sacroiliac mobility exist in the literature.

Methods: The PubMed, Embase, Scopus, Medline and PEDro were searched. There was no restriction on the study design and participants. The data extracted from each study were: sample size, study design and clinical test used. If there was information on reliability values, the number of examiners, concurrent validity values, the gold standard used and the inferential statistical test used was also extracted. For Intra-examiner reliability, the data were expressed as kappa values that were meta-analyzed using random effects.

Results: Fifteen palpatory clinical tests of sacroiliac mobility from 28 studies, fourteen studies performed inferential statistical analysis, of which all performed their inter-examiner reliability analysis, where kappa values ranged from -0.05 to 0.77. Intra-examiner reliability analysis was performed in eight studies, and kappa ranged from 0.08 to 0.73. No study included in this systematic review verified the concurrent validity of the tests. The meta-analysis of intra-examiner reliability showed moderate to good agreement results for the Gillet Test ($k= 0.46$), Standing Flexion Test (STFT) ($k= 0.61$) and Sitting Flexion Test (SIFT) ($k= 0.68$).

Conclusion: We found 15 palpatory clinical tests of sacroiliac mobility in this systematic review. According to the meta-analysis, only the SIFT obtained a good and statistically significant agreement. Further studies are necessary to evaluate the reliability and validity of these tests.

Key Indexing Terms: *Sacroiliac Joint; Systematic Review; Meta-Analysis; Reliability and Validity; Physical Examination.*

INTRODUCTION

It has been well described in the literature that the sacroiliac joint (SIJ) is a considerable source of lower back pain and buttock pain^{1,2,3,4} and may irradiate to lower extremity.⁵ Regarding the clinical practice, the SIJ pain complaints are constant. Complaints that have a prevalence of 13 to 30% reported in patients with low back pain^{1,6} which is the main public health problem worldwide.⁷

Sacral movements are described as rotation and translation,⁸ and in the iliac, the main movement is rotation.⁹ A previous systematic review showed that the SIJ's rotation movement can be up to 8° and the translation movement up to 8mm.¹⁰ Besides, the sacrum moves relative to the iliac and the opposite also occurs, however when joint movement is impaired or abnormal, the term dysfunction is given.¹¹

Manual therapists often use clinical tests to investigate SIJ, classified into 3 categories: (1) positional tests, aiming to verify asymmetries of the involved bone landmarks; (2) provocative pain tests to reproduce the patient's symptoms; and (3) mobility tests to verify movement deficits in the joint.¹¹ For osteopaths, joint evaluation through palpatory clinical tests of mobility is routinely performed in the diagnosis of dysfunctions.¹¹

Despite the routine use of these clinical tests, the latest systematic reviews on clinical evaluation of sacroiliac mobility^{12,13} has shown that reliability results are discouraging and that there are no adequate data on the validity of these tests. In addition, these systematic reviews date back more than 20 years. During that time, new tests may have been introduced or old tests may have been adapted or modified. Thus, we believe that it is necessary to carry out a new systematic review, updating the scientific community.

Therefore, the research questions were:

1. What are the reliability and validity values of palpatory clinical tests of sacroiliac mobility?
2. What are the palpatory clinical tests of sacroiliac mobility in the literature?

Therefore, the main objective of this study is to investigate what are the values of reliability and validity of palpatory clinical tests of sacroiliac mobility. The secondary objective is to investigate what are the palpatory clinical tests of sacroiliac mobility in the literature. Our

hypothesis is that there are several palpatory clinical tests for sacroiliac mobility, but few tests with reliability and validity.

METHOD

This review follows the methodology recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.¹⁴ Has registration in: PROSPERO CRD42019122969.

Identification and selection of studies

A search was conducted in five databases: Pubmed, Embase, Scopus, Medline and PEDro, on August 27 and 28, 2019. The search strategy used was the mesh terms: “Sacroiliac Joint”, “Range of Motion, Articular”, “Physical Examination” and “Palpation”, separated by the Boolean operator “AND”, together with their respective free terms: “Joint, Sacroiliac”, “Joints, Sacroiliac”, “Sacroiliac joints”, “Joint Range of Motion”, “Joint Flexibility”, “Flexibility, Joint”, “Range of Motion”, “Passive Range of Motion”, “Examination, Physical”, “Examinations, Physical”, “Physical Examinations”, “Physical Examinations and Diagnoses” And “palpations”, separated by the Boolean operator “OR”. According to the example in the Pubmed database: (“Joint, sacroiliac”[tw] OR “Joints, sacroiliac”[tw] OR “Sacroiliac joints”[tw] OR “sacroiliac joint”[MeSH]) AND (“Joint Range of Motion”[tw] OR “Joint Flexibility”[tw] OR “Flexibility, Joint”[tw] OR “Range of Motion”[tw] OR “Passive Range of Motion”[tw] OR “Range of Motion, Articular”[MeSH]) AND (“Examination, Physical”[tw] OR “Examinations, Physical”[tw] OR “Physical Examinations”[tw] OR “Physical Examinations and Diagnoses”[tw] OR “Physical Examination”[MeSH]) AND (palpations[tw] OR palpation[MeSH]).

There were no limits applied in the research for the population of interest (e.g. gender, age or health condition) and year of publication of the studies. In order to contemplate the secondary objective of this review, there were also no limits applied for the design of study. To include all studies, contact was made by e-mail with the main authors of studies that were not available in full-text, and manual search through the reference lists of the studies obtained and Google Scholar. There were no studies published in languages other than English and Portuguese, nor summary and unpublished studies.

Two independent authors (R.P.R and F.G.G), both Physiotherapists with at least three years' experience in palpatory clinical tests of sacroiliac mobility, read the titles and abstracts according to the inclusion criteria. If the eligibility criteria were not clarified only through titles and abstracts, the article was included for full reading by the authors independently. At this stage, in case of disagreement, the decision was taken by consensus.

The inclusion criteria for the main objective were to present a palpatory clinical test of sacroiliac mobility and its reliability and/or validity values. To contemplate the secondary objective of this study, it was only necessary that the studies cite a palpatory clinical test of sacroiliac mobility in its title or abstract for the study to be classified for the next stage. Thus, there was no restriction on the study design.

The exclusion criteria adopted were not citing or naming the palpatory clinical test of sacroiliac mobility, evaluating sacroiliac mobility not with a clinical test and evaluating SIJ with a provocative pain test.

Methodological quality assessment

The authors (R.P.R and F.G.G) independently performed the methodological quality assessment using the full content of the included studies, according to Brink and Louw (2012),¹⁵ only in studies that analyzed the reliability of the tests. In case of disagreement, the third author was requested (E.N.C).

The scale has a 13-item checklist, whose answers range from “yes”, “no” or “not applicable”. Questions 4, 5, 6 and 8 refer to reliability studies, questions 3, 7, 9 and 11 to concurrent validity studies and the remaining questions to both studies. As the scale does not have a cutoff point to classify the studies,¹⁶ a 75% cutoff was adopted, where studies that scored 75% or higher were classified as having a high methodological quality.

Data extraction and analysis

Data extraction was independently conducted by the authors (R.P.R and F.G.G) using a standard form in Microsoft Excel software. In case of disagreement, the opinion of the third reviewer (E.N.C) was decisive. The data extracted from each study were: sample size, study design and clinical test used. If there was information on reliability values, the number of examiners, concurrent validity values, the gold standard used and the inferential statistical test used, was also extracted.

For the synthesis of the meta-analysis of reliability values, a kappa cutoff > 0.40 was stipulated, only studies with minimally moderate kappa values were included.¹⁷ Therefore, for the clinical test to be included in the meta-analysis, it should behave at least two studies with a kappa value > 0.40 , assuming that any value less than 0.40 is considered unacceptable.¹⁸ Data were analyzed and calculated using Microsoft Excel software, according to Sun (2011),¹⁹ the random effect was used and heterogeneity (I^2) was calculated according to Higgins et al. (2003).²⁰ It was assumed throughout the study for the percentage of agreement, the classification proposed by Janse et al. (2004):²¹ poor agreement (<0.30), weak (0.31-0.50), moderate (0.51-0.70), good (0.71-0.90) and excellent (0.91-1.00). For kappa values, the classification proposed by Altman (1990)¹⁷ was assumed: k values ≤ 0.20 = poor; 0.21-0.40 = mild; 0.41-0.60 = moderate; 0.61-0.80 = good and 0.81-1 = very good.

RESULTS

The flow of studies through the review

Of the 74 potentially relevant studies found in the databases and study references, 13 were removed by duplicates, and 17 were deleted after reading titles and abstracts. After, 44 studies were analyzed by reading in full and then 28 met the inclusion criteria (Figure 1).^{3,9,11,22-}

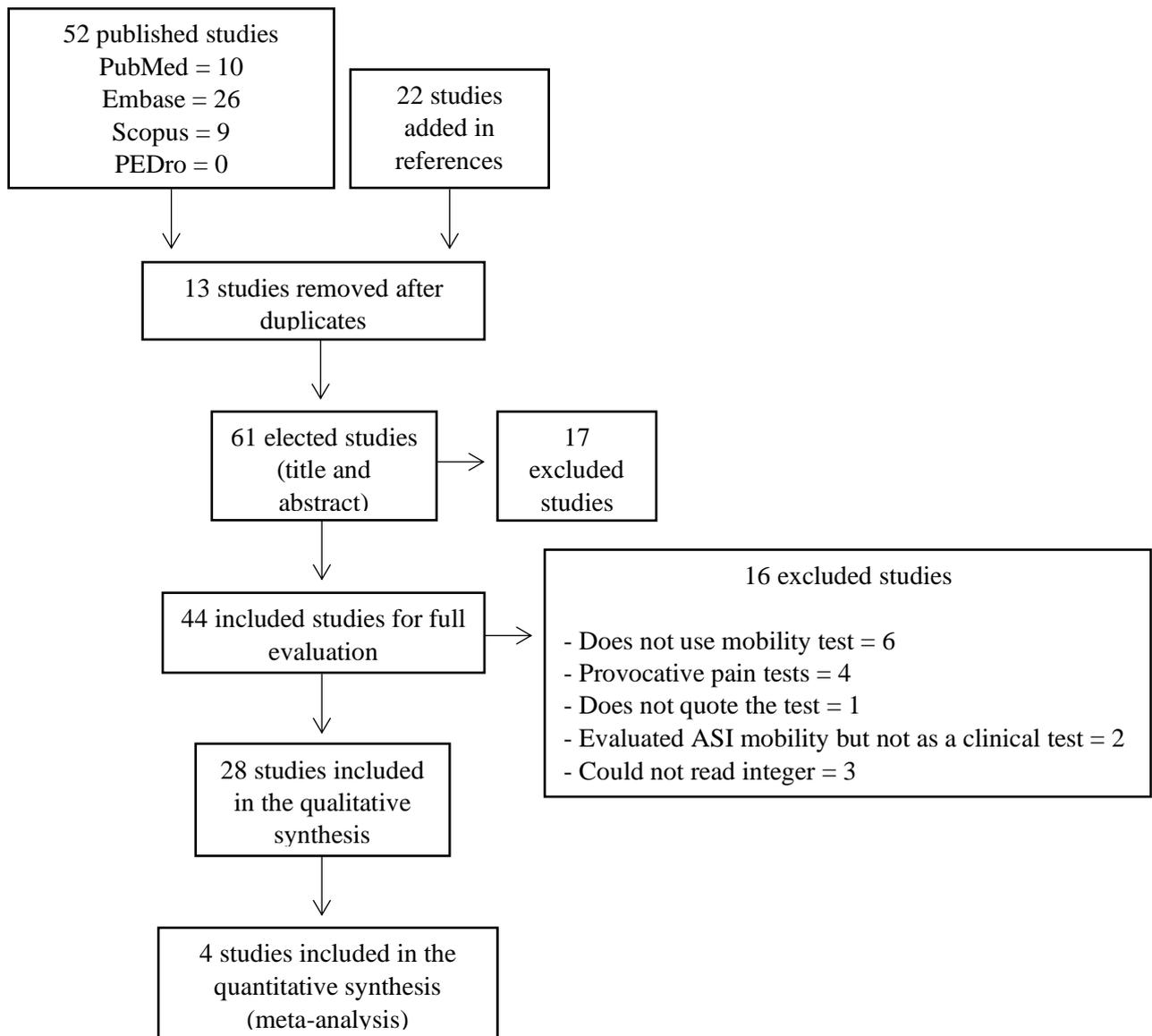


Figure 1 - Fluxogram of included studies according to PRISMA.¹⁴

Characteristics of included studies

Of the studies that performed reliability analysis ($n = 14$), only three^{26,28,30} had a low methodological quality (Table 1), i.e. they scored less than 75% on the scale evaluation criteria. No article has evaluated the concurrent validity of the tests. Six studies^{3,28,29,30,31,35} were classified as “not applicable” in question 5, as they performed only the inter-examiner reliability analysis, and this question deals with intra-examiner reliability. Only in question 13, which deals with the statistical method used, all studies were classified as “yes”. In addition, questions

6 and 8 are the ones in which the studies have the most negative answers, they are about the randomization of examinations and the time interval between repeated measures.

Table 1 - Results of the methodological quality assessment, according to the validation and reliability critical appraisal scale.¹⁹

1 st author (year)	1	2	3	4	5	6	7	8	9	10	11	12	13	Number of ✓	% of ✓	Methodological quality
Arab (2009)	✓	✓	na	✓	✓	✓	na	X	na	✓	na	✓	✓	8	88,9	High
Åström (2014)	✓	✓	na	✓	✓	X	na	X	na	✓	na	✓	✓	7	77,8	High
Bowman (1995)	X	✓	na	✓	na	X	na	✓	na	✓	na	✓	✓	6	75,0	High
Fryer (2005)	✓	✓	na	✓	✓	✓	na	X	na	✓	na	✓	✓	8	88,9	High
Herzog (1989)	X	✓	na	✓	✓	X	na	X	na	✓	na	✓	✓	6	66,7	Low
Hungerford (2007)	✓	✓	na	X	na	✓	na	X	na	✓	na	✓	✓	6	75,0	High
Lima (2019)	✓	✓	na	✓	✓	✓	na	✓	na	✓	na	✓	✓	9	100	High
Meijne (1999)	✓	✓	na	✓	✓	X	na	✓	na	✓	na	✓	✓	8	88,9	High
Paydar (1994)	✓	✓	na	✓	✓	✓	na	X	na	✓	na	X	✓	7	77,8	High
Potter (1985)	✓	✓	na	✓	na	✓	na	✓	na	✓	na	✓	✓	8	100	High
Riddle (2002)	✓	✓	na	✓	na	X	na	✓	na	✓	na	✓	✓	7	87,5	High
Tong (2006)	✓	X	na	✓	na	X	na	X	na	✓	na	✓	✓	5	62,5	Low
Vincent- Smith (1999)	✓	✓	na	✓	✓	✓	na	X	na	✓	na	✓	✓	8	88,9	High
Wiles (1980)	✓	✓	na	✓	na	X	na	X	na	X	na	✓	✓	5	62,5	Low

CRITERIA: 1) Did the authors give a detailed description of the sample? 2) Did the authors clarify the qualification or competence of the rater(s)? 3) Was the reference standard explained? 4) If inter-rater reliability was tested, were raters blinded to the findings of other raters? 5) If intra-rater reliability was tested, were raters blinded to their own prior findings? 6) Was the order of examination randomized? 7) Was the time period between the reference standard and the evaluated test short enough to be reasonably sure that the target condition did not change between the two tests? 8) Was the stability of the variable being measured taken into account when determining the suitability of the time interval between repeated measures? 9) Was the reference standard independent of the evaluated test? 10) Was the execution of the evaluated test described in sufficient detail to permit replication of the test? 11) Was the execution of the reference standard described in sufficient detail to permit its replication? 12) Were withdrawals from the study explained? 13) Were the statistical methods appropriate for the purpose of the study? ANSWERS TO THE CRITERIA: ✓ = Yes; X = No, na = Not applicable.

Table 2 shows all palpatory sacroiliac mobility clinical tests found in this review, described in the 28 studies. In all there were 15 different tests, being the Standing Flexion Test (STFT),^{3,22,23,24,28,29,33-41} the Sitting Flexion Test (SIFT)^{3,11,27,28,34,36,38-42} and the Gillet Test,^{3,22,24-26,30,34,36,38-42} the most frequent.

Table 2 – Palpatory clinical tests of sacroiliac mobility found in studies included in the systematic review.

Test name and description (number of studies)	Author (year)
<p>Standing Flexion Test (STFT) (n=16)</p> <p>The STFT is performed by palpating the posterior superior iliac spine (PSIS) while the subject is bending forward from standing position. The test is negative if PSISs appear to move equally and symmetrically and positive on the side in which the PSIS moves cranially and ventrally more than other side. A positive result in a STFT indicates limited movement of the ilium on the sacrum, and therefore limited SIJ motion on the side of the superior PSIS.^{3,29,33}</p>	<p>Arab (2009) Åström (2014) Bemis (1987) Bennis (2017) Bowman (1995) Cibulka (1999) Dreyfuss (1994) Egan (1996) Levangie (1999) Potter (1985) Riddle (2002) Sipko (2010) Soleimanifar (2017) Tong (2006) Vincent-Smith (1999) Wurff (2006)</p>
<p>Gillet Test (n=13)</p> <p>The Gillet test is performed by palpating the PSISs in the standing position. The subject is asked to stand on one leg while flex the opposite knee pulling up to chest. The test is repeated with other legs. On the normal side, the PSIS move inferiorly. If the PSIS on the side on which the knee is flexed and pulled to chest remains at the level of other PSISs or moves down minimally or even paradoxically moves superiorly, it indicates a positive test.^{3,25,36}</p>	<p>Arab (2009) Åström (2014) Bennis (2017) Dreyfuss (1994) Herzog (1989) Levangie (1999) Meijne (1999) Petrides (2014) Potter (1985) Sipko (2010) Soleimanifar (2017) Wiles (1980) Wurff (2006)</p>

Table 2 – Palpatory clinical tests of sacroiliac mobility found in studies included in the systematic review (continuation).

Test name and description (number of studies)	Author (year)
<p>Sitting Flexion Test (SIFT) (n=13)</p> <p>Also done to the STFT but the subject starts from the sitting position. The test is negative if PSISs appear to move equally and symmetrically and positive on the side in which the PSIS moves cranially and ventrally more than other side. However, a positive result in a SIFT indicates limited movement of the sacrum on the ilium, and therefore limited SIJ motion on the side of the superior PSIS. ^{3,29,33}</p>	<p>Arab (2009) Bemis (1987) Bennis (2017) Dreyfuss (1994) Fryer (2005) Levangie (1999) Paydar (1994) Petrides (2014) Potter (1985) Sipko (2010) Soleimanifar (2017) Tong (2006) Wurff (2006)</p>
<p>Prone Knee Flexion Test (n=6)</p> <p>The subject's position is prone. While the examiner holds both heels, the patient's knees are passively flexed to 90°. The leg lengths are compared by examining the left and right soles of the heel in the prone and prone knees flexed position. The test is negative if no relative change in leg lengths between two positions occurred. If one leg appears shorter than other in the prone knee extended position, apparent lengthening of the short leg in prone knees flexed position implies a hypothesized posterior innominate rotation. ^{3,29,33}</p>	<p>Arab (2009) Bemis (1987) Cibulka (1999) Potter (1985) Riddle (2002) Soleimanifar (2017)</p>
<p>Stork Test (n=6)</p> <p>The Stork test is performed by palpating the posterior superior iliac spine (PSIS) and palpating the midline of the sacrum at the same level, while the subject is in standing position. The subject then flexes the contralateral hip and knee to a minimum of 90 degrees. The sacroiliac joint motion is considered normal if the thumb on the PSISs moves caudal and abnormal if the thumb on the PSISs does not move or if it rises. The test is then repeated on the other side. The dysfunction is recorded as bilateral, left, or right. ^{28,31}</p>	<p>Bennis (2017) Hungerford (2007) Lebouef-Yde (2002) Tong (2006) Kirkaldy-willis (1979) Wiles (1980)</p>

Table 2 – Palpatory clinical tests of sacroiliac mobility found in studies included in the systematic review (continuation).

Test name and description (number of studies)	Author (year)
<p style="text-align: center;">Downing's Test (n=4)</p> <p>The test is done with the subject placed in the supine position. First the patient is instructed to bend his knees up and, keeping his feet on the couch, raise his hips off the couch. After that, the movements are all passive, the stretching movement (external rotation, adduction and hip extension) and shortening (internal rotation, adduction and hip extension) of both legs are performed. If there is no stretching or shortening of any of the legs, the test is called positive for that side. The measurement is made by the medial malleoli in each phase. At the end of the test, the possible dysfunctions are: negative test, anterior dysfunction (D or E) and posterior dysfunction (D or E).⁴⁶</p>	<p style="text-align: center;">Bowman (1995) Lima (2019) Peres (2011) Duffield (1940)</p>
<p style="text-align: center;">Inferior Test (n=2)</p> <p>To test left side: place thumb of right hand over spinous process S2 and thumb of left hand over the posterior superior spine. Instruct patient to flex the left hip and knee to 90° and observe movement of the thumb. In the normal joint the thumb will move caudally. In the abnormally fixed joint, the thumb will move cephalad. The test is also described: Place right thumb over the last sacral spinous process and left thumb over ischial tuberosity. Again, instruct subject to flex the hip and knee to 90°. In the normal joint the left thumb will move laterally, while in the abnormally fixed joint the thumb will remain stationary or move cephalad.⁴³</p>	<p style="text-align: center;">Kirkaldy-willis (1979) Wiles (1980)</p>

Table 2 – Palpatory clinical tests of sacroiliac mobility found in studies included in the systematic review (continuation).

Test name and description (number of studies)	Author (year)
<p>The initiation of the squat motion test (n=1)</p> <p>The examiner palpated the right and left PSIS with a loose grip holding their thumbs horizontally. The subject was asked to begin to squat. The examiner followed the right and left PSIS through the initiation of the squat motion and analyzed whether there was asymmetry in the pelvic motion around a transverse axis. The PSIS was expected to move in an inferior direction. If one of the PSIS moved less inferiorly than the other, that side was reported less mobile. If both sides moved equally, the test was reported symmetric.²⁴</p>	<p>Åström (2014)</p>
<p>Bourdillon's Method (n=1)</p> <p>The test is done similarly to the Downing's test.</p> <p>a) Shortening- flex knee and hip. Abduct and externally rotate the leg. Fully flex the leg in the neutral position. Adduct and internally rotate. Straighten.</p> <p>b) Lengthening- flex knee and hip. Adduct and internally rotate. Full flexion in neutral. Abduct and externally rotate.</p> <p>The leg length is compared by the examiner placing his thumbs on the distal borders of the medial malleoli and seeing whether the thumbs are level or if one is above the other. The tests are then performed on each leg and changes in leg length noted.³⁵</p>	<p>Bowman (1995)</p>

Table 2 – Palpatory clinical tests of sacroiliac mobility found in studies included in the systematic review (continuation).

Test name and description (number of studies)	Author (year)
<p>Mitchell and Moran (n=1)</p> <p>The test is done similarly to the Bourdillon’s method.</p> <p>a) Shortening - flex knee and hip to 90°. Abduct and externally rotate. Maintain external rotation until straight.</p> <p>b) Lengthening - flex knee and hip to 90°. Adduct and internally rotate. Maintain internal rotation until straight.</p> <p>The leg length is compared by the examiner placing his thumbs on the distal borders of the medial malleoli and seeing whether the thumbs are level or if one is above the other. The tests are then performed on each leg and changes in leg length noted.³⁵</p>	Bowman (1995)
<p>Joint Play (Menell) test (n=1)</p> <p>No description was presented</p>	Petrides (2014)
<p>Spine Test (n=1)</p> <p>No description was presented</p>	Wurff (2006)
<p>Lateral – Flexion (n=1)</p> <p>No description was presented</p>	Wurff (2006)
<p>Overtake Phenomenon (n=1)</p> <p>No description was presented</p>	Wurff (2006)

Table 3 refers to data (author, sample, clinical test, number of examiners, study design and statistical analysis) from studies reporting inferential statistical analysis.^{3,11,22-32,35}

Table 3 - Data from studies reporting inferential statistical analysis.

1° author (year)	Sample size (n)	Clinical test	Number of examiners	Study Design	Statistical analyses
Arab (2009)	Men and women between 20 and 65 years old (25) with low back pain	- STFT - Gillet - SIFT - Prone Knee Flexion	2	Intra and inter-examiner reliability	kappa
Åström (2014)	Healthy men and women between 18 and 65 years old (34)	- STFT - Gillet - The initiation of the squat motion test	4	Intra and inter-examiner reliability	% agreement and kappa
Bowman (1995)	(16) With or without low back pain	- STFT	3	Inter-examiner reliability	% agreement and kappa
Fryer (2005)	Asymptomatic women between 22 and 24 years old (10)	- SIFT	10	Intra and inter-examiner reliability	% agreement and kappa
Herzog (1989)	Adults (11) with sacroiliac dysfunction	- Gillet	10	Intra and inter-examiner reliability	% agreement
Hungerford (2007)	Men and women average 36 years (33) with lower back or leg pain	- Stork Test	3	Inter-examiner reliability	% agreement and kappa
Lima (2019)	Men and women average 21 years (108) symptomatic and asymptomatic	Downing's test	2	Intra and inter-examiner reliability	kappa
Meijne (1999)	Men between 21 e 25 years old (38) symptomatic and asymptomatic	- Gillet	2	Intra and inter-examiner reliability	% agreement and kappa

Table 3 - Data from studies reporting inferential statistical analysis (continuation).

1° author (year)	Sample size (n)	Clinical test	Number of examiners	Study Design	Statistical analyses
Paydar (1994)	Asymptomatic men and women between 18 and 50 years old (32)	- SIFT	2	Intra and inter- examiner reliability	% agreement and kappa
Potter (1985)	Men and women between 24 and 58 years old (17) with low back pain	- STFT - Gillet - SIFT - Prone Knee Flexion - Supine Long Sitting	8	Inter- examiner reliability	% agreement
Riddle (2002)	Men and women between 18 and 81 years old (65) with low back pain	- STFT - Prone Knee Flexion - Supine Long Sitting	34	Inter- examiner reliability	% agreement and kappa
Tong (2006)	Men and women with a mean age of 68.3 years (24) with low back pain	- STFT - SIFT - Stork Test	2	Inter- examiner reliability	kappa
Vincent- Smith (1999)	Healthy men and women between 20 and 25 years old (14)	- STFT	9	Intra and inter- examiner reliability	% agreement and kappa
Wiles (1980)	Men and women between 20 and 25 years old (46)	- Gillet - Stork Test - Inferior Test	2	Inter- examiner reliability	% agreement

Table 4 presents the results of the statistical analyses of the studies for Intra- and inter-examiner reliabilities,^{3,11,22-32,35} according to each palpatory sacroiliac mobility clinical test.

Table 4 – Results of statistical analyses of reliability studies.

	Author (year)	Inter-examiner reliability			Intra-examiner reliability		
		% agreement	kappa	95% CI	% agreement	kappa	95% CI
Standing Flexion Test (STFT)	Arab (2009)	-	0.51	0.08-0.95	-	0.68	0.35-1.01
	Åström (2014)	59	0.10-0.20	-	87	0.70	-
	Bowman (1995)	52	0.23	-	-	-	-
	Potter (1985)	44	-	-	-	-	-
	Riddle (2002)	55	0.32	-	-	-	-
	Tong (2006)	-	0.14	-	-	-	-
	Vincent-Smith (1999)	43	0.05	-	68	0.46	-

Table 4 – Results of statistical analyses of reliability studies (continuation).

	Author (year)	Inter-examiner reliability			Intra-examiner reliability		
		% agreement	kappa	95% CI	% agreement	kappa	95% CI
Gillet Test	Arab (2009)	-	0.41	0.03-0.87	-	0.42	-0.01-0.87
	Åström (2014)	59-65	0.20-0.30	-	74	0.50	-
	Herzog (1989)	61	-	-	72	-	-
	Meijne (1999)	77	-0.05	-0.06-0.12	83,5	0.08	0.01-0.14
	Potter (1985)	47	-	-	-	-	-
	Wiles (1980)	64	-	-	-	-	-
Sitting Flexion Test (SIFT)	Arab (2009)	-	0.75	0.42-1.08	-	0.73	0.45-1.01
	Fryer (2005)	71	0.14	-	-	0.41	-
	Paydar (1994)	34	0.09	-	58,1	0.29	-
	Potter (1985)	50	-	-	-	-	-
	Tong (2006)	-	0.06	-	-	-	-

Table 4 – Results of statistical analyses of reliability studies (continuation).

	Author (year)	Inter-examiner reliability			Intra-examiner reliability		
		% agreement	kappa	95% CI	% agreement	kappa	95% CI
Prone Knee Flexion Test	Arab (2009)	-	0.58	0.25-0.91	-	0.41	0.07-0.78
	Potter (1985)	23	-	-	-	-	-
	Riddle (2002)	60	0.26	-	-	-	-
Stork Test	Hungerford (2007)	90	0.77	-	-	-	-
	Tong (2006)	-	0.27	-	-	-	-
	Wiles (1980)	47	-	-	-	-	-
Supine Long Sitting Test	Potter (1985)	40	-	-	-	-	-
	Riddle (2002)	45	0.19	-	-	-	-
Downing's test	Lima (2019)	-	0.18	0.09-0.27	-	0.12	0.03-0.22

Table 4 – Results of statistical analyses of reliability studies (continuation).

	Author (year)	Inter-examiner reliability			Intra-examiner reliability		
		% agreement	kappa	95% CI	% agreement	kappa	95% CI
The initiation of the squat motion test	Åström (2014)	65-71	0.20-0.40	-	74	0.40	-
Inferior Test	Wiles (1980)	62	-	-	-	-	-

se= standard error; sd= standard deviation; CI= confidence interval.

Figure 2 presents the meta-analysis results for intra-examiner reliability for studies with kappa value >0.40 . Due to the cutoff point, the clinical tests analyzed were STFT,²²⁻²⁴ Gillet Test^{22,24} and SIFT.^{11,22}

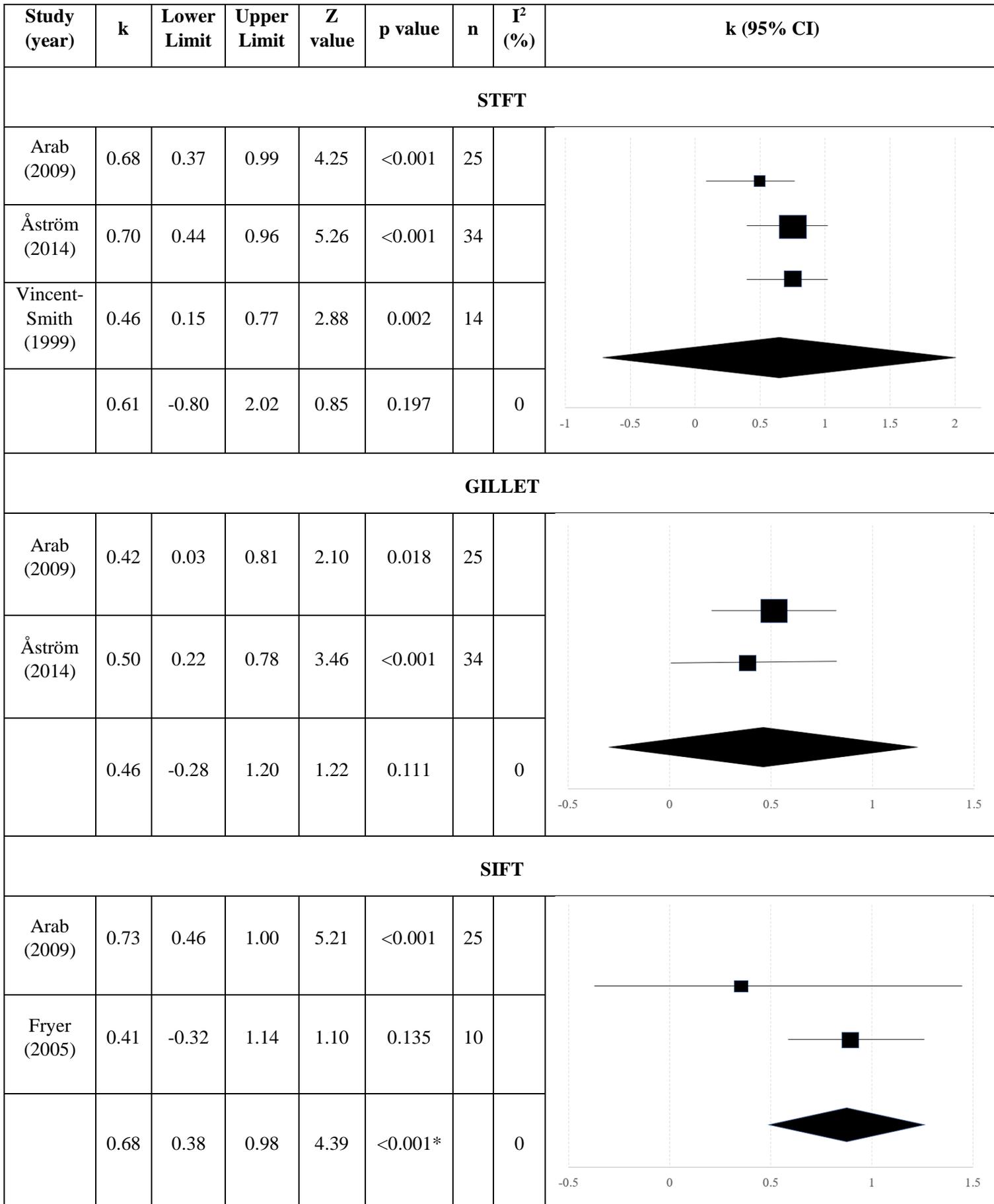


Figure 2 - Intra-examiner reliability meta-analysis (k); Random effect was used.

DISCUSSION

This systematic review has as its main objective is to investigate what are the values of reliability and validity of palpatory clinical tests of sacroiliac mobility. The secondary objective is to investigate what are the palpatory clinical tests of sacroiliac mobility in the literature. In the 28 included studies, fifteen palpatory mobility clinical tests were described: (1) standing flexion test (STFT), (2) Gillet test, (3) sitting flexion test (SIFT), (4) prone knee flexion test, (5) stork test, (6) supine long sitting test, (7) downing's test, (8) inferior test, (9) the initiation of the squat motion test, (10) bourdillon's method, (11) mitchell and moran, (12) joint play (Menell) test, (13) spine test, (14) lateral flexion and (15) overtake phenomenon. The most recurrent tests were the standing flexion test (STFT), the Gillet test, and the sitting flexion test (SIFT), which were found in 16,^{3,22,23,24,28,29,33-41} 13^{3,22,24-26,30,34,36,38-42} and 13 studies,^{3,11,27,28,34,36,38-42} respectively (Table 2).

In investigating the reliability and validity of these palpatory clinical mobility tests, 14 studies^{3,11,22-32,35} performed inter-examiner reliability analysis and 8^{11,22-27,32} also analyzed intra-examiner reliability (Table 3). No study included in this systematic review has verified the concurrent validity of clinical tests. The most recurrent tests were STFT, Gillet test and SIFT. Were analyzed for intra-examiner reliability in 3,²²⁻²⁴ 4,^{22,24-26} and 3 studies,^{11,22,27} and in 7,^{3,22-24,28,29,35} 6,^{3,22,24-26,30} and 5 studies^{3,11,22,27,28} for inter- examiner reliability, respectively (Table 4).

According to Kottner et al. (2011)⁴⁷ inter-examiner reliability is the agreement between different examiners, using the same scale, classification, instrument or procedure, evaluating the same subjects or objects. Intra-examiner reliability, also known as test and retest, is the agreement between the same examiner using the same scale, rating, instrument or procedure, evaluating the same subject or object at different times. In musculoskeletal research and practice, there is often a need to determine the reliability of measurements,¹⁸ since a test or instrument is of no use if it is not reliable.⁴⁸

Regarding STFT, for inter-examiner reliability, the percentage of agreement ranged from 42.7%²³ to 59%²⁴ and the kappa value ranged from 0.05²³ to 0.51,²² resulting in a percentage agreement of weak to moderate and a kappa value from poor to moderate. For intra-examiner reliability, the agreement ranged from 68%²³ to 87%²⁴ and the kappa value ranged from 0.46²³ to 0.70²⁴ resulting in a concordance percentage and a kappa value from moderate to good. Thus, it is possible to notice that intra-examiner reliability has better agreement values

in relation to inter-examiner. Therefore, when the objective is the evaluation and re-evaluation of patients, according to these studies, it is preferable that the same clinician makes measurements with the STFT at different times, instead of two clinicians comparing their measurements.

For the Gillet test, concerning inter-examiner reliability, the agreement percentage ranged from 46.7%³ to 76.8%²⁵ and the kappa value from -0.05²⁵ to 0.41,²² ie, a percentage agreement from weak to good and a kappa value from poor to moderate. For intra-examiner reliability, the agreement percentage ranged from 72%²⁶ to 83.5%²⁵ and the kappa value ranged from 0.08²⁵ to 0.50,²⁴ which represents a good agreement percentage and a kappa value from poor to moderate. Taking into account the results of the agreement percentages, for the Gillet test, intra-examiner reliability has better results compared to inter-examiner, generating a trend of thinking similar to STFT. However, this situation is not repeated for the kappa values, making the results still heterogeneous for the Gillet test.

Regarding SIFT, for inter-examiner reliability, the agreement percentage ranged from poor (34.4%)²⁷ to good (71%),¹¹ and the kappa value from poor (0.06),²⁸ to good (0,75).²² However, the study by Tong et al. (2006)²⁸ was assessed as having a low methodological quality, in other words, a high risk of methodological bias (Table 1). For intra-examiner reliability, only the study by Paydar (1994)²⁷ evaluated the agreement percentage and obtained a result of 58.1%, being a moderate agreement. The kappa values ranged from 0.29²⁷ to 0.73,²² resulting in a mild to good kappa. So, in relation to the SIFT, the intra-examiner reliability presented better agreement values through the kappa statistics. However, the same cannot be said for the percentage of agreement, considering that only one study performed this assessment for intra-examiner reliability and obtained moderate results. Therefore, further studies are needed to carry out this assessment.

The Prone knee flexion test was found in 3 studies analyzing inter- examiner reliability, and percentage of agreement values ranged from 23.5%³ to 60%²⁹ and kappa values ranged from 0.26²⁹ to 0.58,²² showing a poor to moderate agreement percentage and mild to moderate kappa. Only one study evaluated intra-examiner reliability for the Prone knee flexion test²² and obtained a moderate kappa of 0.41. Concerning the Stork test, inter- examiner reliability was analyzed in 3 studies, where the agreement percentage ranged from 47%³⁰ to 89.9%³¹ and the kappa values ranged from 0.27²⁸ to 0.77,³¹ poor to a good agreement and a mild to good kappa.

It should be considered that the study by Tong et al. (2006)²⁸ and Wiles (1980)³⁰ were assessed as having a high risk of bias (Table 1). Intra-examiner reliability has not been evaluated.

The Supine long sitting test was described in 2 studies, which examined their inter-examiner reliability and the values found for the agreement percentage were 40%³ and 44.6%,²⁹ being weak, and the kappa value of 0.19,²⁹ denominated poor. The studie of Lima et al. (2019)³² aimed to measure Downing's test intra and inter-examiner reliabilities, finding, respectively, kappa of 0.12 and 0.18, both considered poor. Still, Åström and Gummesson (2014)²⁴ also evaluated reliabilities, but from The initiation of squat motion test, they obtained a good agreement percentage of 74% and a mild kappa of 0.4 for intra-examiner reliability. For inter-examiner reliability, they found 65% and 71%, moderate and good values, for the agreement percentage and 0.2 and 0.4 kappa values, being poor and mild. Still, Wiles (1980)³⁰ performed the same previous analysis, this time from the Inferior test, and obtained a moderate agreement percentage of 62%.

Based on previous analyses, the reliability results for palpatory clinical tests of sacroiliac mobility are very divergent. The Gillet test²⁴⁻²⁶ and The initiation of squat motion test²⁴ showed good results for the agreement percentage, although the poor to moderate kappa values. One test with good results for the agreement percentage and kappa was the SIFT, however, these findings were only in 2 studies.^{11,22} Regarding the STFT, there were good results for the agreement percentage and for kappa, however, only for intra-examiner reliability and also only in 2 studies.^{22,24} Accordingly, good results were also found for the Stork test, however, in only one study³¹ and only for inter-examiner reliability. In contrast, the Downing test and the Supine long sitting had poor agreement results, but only in 1³² and 2 studies,^{3,29} respectively.

According to the meta-analysis (Figure 2) results for intra-examiner reliability, in all included tests (STFT, Gillet and SIFT) a low heterogeneity was observed ($I^2 = 0.00$). For both STFT²²⁻²⁴ and SIFT,^{11,22} a good agreement was found between the same examiner, given by a kappa of 0.61 and 0.68, respectively. However, only for SIFT, the result was statistically significant ($p < 0.001$). Regarding the Gillet test,^{22,24} the agreement between the same examiner was moderate, given by a kappa of 0.46, not having statistical significance ($p = 0.111$). It is pertinent to note that this meta-analysis was performed with only three studies for STFT and two for SIFT and Gillet, however, the four included studies were identified with high methodological quality (Table 1).

There are studies in the literature dealing with the Clinical Diagnostic Rule (CDR) of tests, Petersen et al. (2017)⁴⁹ conducted a systematic review of systematic reviews on this subject. They report that certain CDRs have high specificity for low back pain from spondylolisthesis, disc herniation with nerve root involvement, and spinal stenosis. For SIJ, in relation to provocative pain tests, there is also a high specificity of CDR including at least 3 positive tests of 5,⁵⁰⁻⁵² according to the following tests: distraction, thigh thrust, Gaenslen's test, compression and sacral thrust. In addition, pain in the postero superior iliac spine (PSIS) may increase the specificity of the finding,⁵³ and the phenomena of centralization and peripheralization are also used as clinical reasoning for exclusion in the diagnosis of pain in the sacroiliac joint, since these are associated with discogenic pain.⁵⁴ Despite being tests that assess pain and non-mobility, like the current study, we understand the importance of CDRs which do not yet exist for SIJ mobility tests. As a suggestion, future studies could investigate a CDR of mobility tests, analyzing the STFT, SIFT and Gillet Test, which were the most recurrent tests and with the best agreement values in this systematic review.

We understand that performing the meta-analysis with three or two studies per test can be seen as a limitation of the current review, further studies on this subject yet are needed for conclusions with more evidence. In addition, another limitation may have been the realization of the methodological quality of only studies that reported statistical analysis and not of those that were used only to list and describe the clinical tests in the literature.

CONCLUSION

We found 15 palpatory clinical tests of sacroiliac mobility in this systematic review. No study has evaluated the concurrent validity of these tests. There is a divergence in the reliability studies of these clinical trials. According to the meta-analysis, only the Sitting Flexion Test (SIFT) obtained a good statistically significant agreement for intra-examiner reliability. Therefore, the results of this systematic review suggest that further studies should be conducted to investigate the reliability and validity of palpatory sacroiliac mobility clinical trials.

REFERENCES

1. Maigne J, Aivaliklis A, Pfefer F. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. *Spine*. 1996;21(16): 1889-1892.
2. Stuesson B, Selvik G, Uden A. Movements of the sacroiliac joints. A roentgen stereophotogrammetric analysis. *Spine*. 1989;14(2):162-165.
3. Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. *Phys Ther*. 1985;65(11):1671-1675.
4. Slipman CW, Whyte WS, Chow DW, Chou L, Lenrow D, Ellen M. Sacroiliac joint syndrome. *Pain Physician*. 2001;4(2):143-152.
5. Prather H. Pelvis and sacral dysfunction in sports and exercise. *Physical Medicine and Rehabilitation Clinics*. 2000;11(4):805-836.
6. Schwarzer AC, Aprill CN, Bogduk N. The sacroiliac joint in chronic low back pain. *Spine*. 1995;20(1):31-37.
7. Airaksinen O, Brox J, Cedraschi C, Hildebrandt J, Klaber-Moffett, Kovacs F, et al. Chapter 4 European guidelines for the management of chronic nonspecific low back pain. *European Spine Journal*. 2006;15(2):192.
8. Vleeming A, Van Wingerden JP, Dijkstra PF, Stoeckart R, Snijders CJ, Stijnen T. Mobility in the sacroiliac joints in the elderly: a kinematic and radiological study. *Clinical Biomechanics*. 1992;7(3):170-176.
9. Bemis T, Daniel M. Validation of the Long Sitting Test on Subjects with Iliosacral Dysfunction. *Journal of Orthopaedic & Sports Physical Therapy*. 1987;8(7):336-345.
10. Goode A, Hegedus EJ, Sizer P, Brismee JM, Linberg A, Cook CE. Three-Dimensional Movements of the Sacroiliac Joint: A Systematic Review of the Literature and Assessment of Clinical Utility. *Journal of Manual & Manipulative Therapy*. 2008;16(1):25-38.
11. Fryer G, McPherson HC, O'Keefe P. The effect of training on the inter-examiner and intra-examiner reliability of the seated flexion test and assessment of pelvic anatomical landmarks with palpation. *International Journal of Osteopathic Medicine*. 2005;8(4):131-138.
12. Van der Wurff P, Hagmeijer RH, Meyne W. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 1: Reliability. *Manual Therapy*. 2000;5(1):30-36.
13. Van der Wurff P, Meyne W, Hagmeijer RH. Clinical tests of the sacroiliac joint: a systematic methodological review. Part 2: validity. *Manual Therapy*. 2000;5(2):89-96.
14. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*. 2010;8(5):336-341.
15. Brink Y, Louw QA. Clinical instruments: reliability and validity critical appraisal. *Journal of Evaluation in Clinical Practice*. 2012;18(6):1126-1132.
16. Rondoni A, Rossetini G, Ristori D, Gallo F, Strobe M, Giaretta F, et al. Intrarater and inter-rater reliability of active cervical range of motion in patients with nonspecific neck pain measured with technological and common use devices: a systematic review with meta-regression. *Journal of Manipulative and Physiological Therapeutics*. 2017;40(8):597-608.
17. Altman D. Practical statistics for medical research. CRC press. 1990.
18. Sim J, Wright C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*. 2005;85(3):257-268.

19. Sun S. Meta-analysis of Cohen's kappa. *Health Services and Outcomes Research Methodology*. 2011;11(3-4):145-163.
20. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560.
21. Janse AJ, Gemke RBJ, Uiterwaal CSPM, Van der Tweel I, Kimpen JLL, Sinnema G. Quality of life: patients and doctors don't always agree: a meta-analysis. *Journal of Clinical Epidemiology*. 2004;57(7):653-661.
22. Arab AM, Abdollahi I, Joghataei MTMT, Golafshani Z, Kazemnejad A. Inter- and intra-examiner reliability of single and composites of selected motion palpation and pain provocation tests for sacroiliac joint. *Manual Therapy*. 2009;14(2):213-221.
23. Vincent-Smith B, Gibbons P. Inter-examiner and intra-examiner reliability of the standing flexion test. *Manual Therapy*. 1999;4(2):87-93.
24. Åström, M., & Gummesson C. Assessment of asymmetry in pelvic motion –An inter- and intra-examiner reliability study. *The European Journal of Physiotherapy*. 2014;16(2):76-81.
25. Meijne W, van Neerbos K, Aufdemkampe G, van der Wurff P. Intraexaminer and interexaminer reliability of the Gillet test. *Journal of Manipulative and Physiological Therapeutics*. 1999;22(1):4-9.
26. Herzog W, Read LJ, Conway PJ, Shaw LD, McEwen MC. Reliability of motion palpation procedures to detect sacroiliac joint fixations. *Journal of Manipulative and Physiological Therapeutics*. 1989;12(2):86-92.
27. Paydar D, Thiel H, Gemmell H. Intra- and Interexaminer Reliability of Certain Pelvic Palpatory Procedures and the Sitting Flexion Test for Sacroiliac Joint Mobility and Dysfunction. *J Neuromusculoskel Syst*. 1994;2(2):65-69.
28. Tong HC, Heyman OG, Lado DA, Isser MM. Interexaminer reliability of three methods of combining test results to determine side of sacral restriction, sacral base position, and innominate bone position. *Journal of the American Osteopathic Association*. 2006;106(8):464-468.
29. Riddle. Evaluation of the Presence of Sacroiliac Joint Region Dysfunction Using a Combination of Tests: A Multicenter Intertester Reliability Study. *Physical Therapy*. 2002;82(8):772-781.
30. Wiles MR. Reproducibility and interexaminer correlation of motion palpation findings of the sacroiliac joints. *The Journal of the Canadian chiropratic Association*. 1980;24:59-69.
31. Hungerford BA, Gilleard W, Moran M, Emmerson C. Evaluation of the Ability of Physical Therapists to Palpate Intrapelvic Motion With the Stork Test on the Support Side. *Physical Therapy*. 2007;87(7):879-887.
32. Lima PO, Melo WP, Bezerra MA, Almeida GP, Nunes ACL, Oliveira RR. Intraexaminer and Interexaminer Reproducibility of the Downing Test for Sacroiliac Joint Evaluation of Symptomatic and Asymptomatic Individuals. *Journal of Chiropractic Medicine*. 2019
33. Cibulka MT, Koldehoff R. Clinical Usefulness of a Cluster of Sacroiliac joint Tests in Patients With and Without Low Back Pain. *Journal of Orthopaedic & Sports Physical Therapy*. 1999;29(2):83-92.
34. Bennis S, Hwang S. Office Evaluation of Pelvic Pain. *Physical Medicine and Rehabilitation Clinics*. 2017;28(3):461-476.
35. Bowman, C., & Gribble R. The value of the forward flexion test and three tests of leg length changes in the clinical assessment of movement of the sacroiliac joint. *Journal of Orthopaedic Medicine*. 1995;17(2):66-67.
36. Dreyfuss P, Dryer S, Griffin J, Hoffman J, Walsh N. Positive sacroiliac screening tests in asymptomatic adults. *Spine*. 1994;19(10):1138-1143.

37. Egan D, Cole J TL. The standing forward flexion test: an inaccurate determinant of sacroiliac joint dysfunction. *Physiotherapy*. 1996;82(4):236-242.
38. Soleimanifar M, Karimi N, Arab AM. Association between composites of selected motion palpation and pain provocation tests for sacroiliac joint disorders. *Journal of Bodywork and Movement Therapies*. 2017;21(2):240-245.
39. Sipko T, Grygier D, Barczyk K, Eliaz G. The Occurrence of Strain Symptoms in the Lumbosacral Region and Pelvis During Pregnancy and After Childbirth. *Journal of Manipulative and Physiological Therapeutics*. 2010;33(5):370-377.
40. Van der Wurff P. Clinical diagnostic tests for the sacroiliac joint: Motion and palpation tests. *Australian Journal of Physiotherapy*. 2006;52(4):308.
41. Levangie PK. Four Clinical Tests of Sacroiliac Joint Dysfunction: The Association of Test Results With Innominate Torsion Among Patients With and Without Low Back Pain. *Physical Therapy*. 1999;79(11):1043-1057.
42. Petrides S. Non-inflammatory sacroiliac joint disorders. *Indian Journal of Rheumatology*. 2014;9:S54-S63.
43. Kirkaldy-Willis WH, Hill RJ. A more precise diagnosis for low-back pain. *Spine*. 1979;4(2):102-109.
44. Leboeuf-Yde C, van Dijk J, Franz C, et al. Motion palpation findings and self-reported low back pain in a population-based study sample. *Journal of Manipulative and Physiological Therapeutics*. 2002;25(2):80-87.
45. Peres C, Risso L, Oliveira LU. Efeitos da manipulação do íliaco na descarga de peso no retopé em indivíduos com disfunção sacro-ilíaca. *Rev Ter Man*. 2011;9(42):150-154.
46. Duffield HJ. Manipulation of the Sacro-Iliac Joint. *The Ulster Medical Journal*. 1940;9(2):108.
47. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *International Journal of Nursing Studies*. 2011;48(6):661-671.
48. Thomas J, Nelson J, Silverman S. Métodos de pesquisa em atividade física. Artmed Editora. 2009.
49. Petersen T, Laslett M, Juhl C. Clinical classification in low back pain: best-evidence diagnostic rules based on systematic reviews. *BMC Musculoskeletal Disorders*. 2017;18(188).
50. Laslett M, Young S, Aprill C, McDonald B. Diagnosing painful sacroiliac joints. A validity study of a McKenzie evaluation and sacroiliac provocation tests. *Aust J Physiother*. 2003;49:89-97.
51. Van der Wurff P, Buijs EJ, Groen GJ. A multitest regimen of pain provocation tests as an aid to reduce unnecessary minimally invasive sacroiliac joint procedures. *Arch Phys Med Rehabil*. 2006;87(1):10-4.
52. Ozcocmen S, Bozgeyik Z, Kalcik M, Yildirim A. The value of sacroiliac pain provocation tests in early active sacroiliitis. *Clin Rheumatol*. 2008;27(10):1275-82.
53. Van der Wurff P, Buijs EJ, Groen GJ. Intensity mapping of pain referral areas in sacroiliac joint pain patients. *J Manipulative Physiol Ther*. 2006;29(3):190-5.
54. Young S, Aprill C, Laslett M. Correlation of clinical examination characteristics with three sources of chronic low back pain. *Spine J*. 2003;3(6):460-5.

4 CAPÍTULO 2

Validade, reprodutibilidade e acurácia diagnóstica do Teste de Flexão em Pé (TFP) e do Teste de Flexão Sentado (TFS)

RESUMO

INTRODUÇÃO: A disfunção sacroilíaca é um estado de hipomobilidade da amplitude de movimento da articulação, seguida de uma alteração posicional entre a relação do sacro e o ilíaco. Em geral, os testes clínicos que avaliam a Articulação Sacroilíaca (ASI) e suas disfunções carecem de valores de validade, reprodutibilidade e acurácia diagnóstica. **OBJETIVO:** Avaliar a validade concorrente, a reprodutibilidade intra e interavaliadores e a acurácia diagnóstica do Teste de Flexão em Pé (TFP) e do Teste de Flexão Sentado (TFS). **MÉTODO:** Estudo prospectivo, onde a amostra foi composta por 30 indivíduos de ambos os sexos. A equipe de avaliação foi composta por 5 Fisioterapeutas (A, B, C, D e E). As avaliações aconteceram em dois dias diferentes: (1º dia) reprodutibilidade interavaliador e validade concorrente; (2º dia) reprodutibilidade intra-avaliador. O padrão referência utilizado para a validade concorrente foi a medição tridimensional (cinemetria 3D), por meio do sistema BTS Smart-DX. Análise estatística: foi utilizado o percentual de concordância (%C) e a estatística kappa (k) para as análises de concordância, e a curva ROC (AUC) para determinar a capacidade discriminativa, os valores de sensibilidade e especificidade, além dos valores preditivos positivos (VPP) e negativos (VPN). **RESULTADOS:** Validade concorrente: TFP (%C=66,7%; k=0,42; p<0,01) e TFS (%C=46,7%; k=0,12; p=0,25). Curva ROC: TFP (AUC=0,700; p<0,05; sensibilidade=80%; especificidade=60%; VPP=85%; VPN=78%) e TFS (AUC=0,602; p=0,52; sensibilidade=75%; especificidade=46%; VPP=20%; VPN=91%). Reprodutibilidade intra-avaliador: TFP (%C=66,3%; k=0,43; p<0,01) e TFS (%C=56,7%; k=0,38; p<0,01). Reprodutibilidade interavaliador: TFP (%C=10%; k=-0,02; p=0,825) e TFS (%C=13,3%; k=0,01; p=0,836). **CONCLUSÃO:** O Teste de Flexão em Pé (TFP) é válido, reprodutível entre o mesmo avaliador e não acurado. O Teste de Flexão Sentado (TFS) não é válido, não é reprodutível e nem acurado. Sugerimos mais estudos que investiguem as propriedades de medida de testes clínicos palpatórios de mobilidade da ASI.

Palavras-chave: Terapia Manual; Fisioterapia; Amplitude de Movimento Articular; Articulação Sacroilíaca; Confiabilidade e Validade.

INTRODUÇÃO

A articulação sacroilíaca (ASI) é uma considerável fonte de dor na região lombar e das nádegas (MAIGNE et al., 1996; SHAW, 1992; STURESSON et al., 1989; POTTER; ROTHSTEIN, 2006; SLIPMAN et al., 2001), podendo irradiar para a parte posterior das coxas e anterior da região inguinal (PRATHER, 2000). A ASI é uma articulação diartrodial, multiplanar, com cápsula articular e líquido sinovial e realiza movimentos ao longo de três eixos (SLIPMAN et al., 2001; GOODE et al., 2008). Os movimentos sacrais são descritos como rotação e translação, e o principal movimento dos ilíacos é a rotação (VLEEMING et al., 1992; BEMIS et al., 1987).

A disfunção sacroilíaca pode ser definida por um estado de hipomobilidade da amplitude de movimento da articulação seguida de uma alteração posicional entre a relação do sacro e o ilíaco (DREYFUSS et al., 1994). Dentre a gama de possibilidades de avaliação da ASI e suas disfunções (PRATHER, 2000), o Teste de Flexão em Pé (TFP) e o Teste de Flexão Sentado (TFS) estão entre as opções de mensuração da mobilidade articular (FRYER et al., 2009). O TFP é utilizado para investigar primordialmente o comportamento dos ilíacos, enquanto o TFS investiga o comportamento do sacro. Ambos os testes estão entre os mais utilizados por Osteopatas americanos. (VINCENT-SMITH; GIBBONS, 1999; POTTER; ROTHSTEIN, 2006; FRYER et al., 2009)

Apesar da relevância e grande utilização desses testes, os seus valores de reprodutibilidade intra e interavaliadores ainda são muito divergentes, variando de concordância pobre a boa (ARAB et al., 2009; ÅSTRÖM et al., 2014; BOWMAN et al., 1995; RIDDLE et al., 2002; TONG et al., 2006; VINCENT-SMITH; GIBBONS, 1999; FRYER et al., 2005; PAYDAR et al., 1994). Além disso, não há nenhum estudo na literatura científica reportando valores de validade e tampouco informações acerca da acurácia diagnóstica desses testes, como ponto de corte, sensibilidade e especificidade. Dessa maneira, o objetivo desse estudo é avaliar a validade concorrente, a reprodutibilidade intra e interavaliadores e a acurácia diagnóstica do Teste de Flexão em Pé (TFP) e do Teste de Flexão Sentado (TFS). Como hipóteses do estudo, acreditamos que o TFP e TFS terão validade e acurácia diagnóstica confirmadas, bem como serão reprodutíveis para avaliações realizadas por um mesmo avaliador e por diferentes avaliadores.

MÉTODOS

Esse estudo está descrito de acordo com o Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (KOTTNER et al., 2011), para as análises de reprodutibilidade, e de acordo com Standards for Reporting of Diagnostic Accuracy (STARD) (BOSSUYT et al., 2015) para a análise da validade concorrente e acurácia diagnóstica. Conforme as recomendações de Kottner et al. (2011), decidimos utilizar dois guias para reportar esse estudo, com a justificativa de que os estudos de acurácia diagnóstica diferem dos estudos de reprodutibilidade ou concordância na utilização de um padrão de referência/ouro.

Para verificar se o teste realmente avalia aquilo que ele se propõe, é necessário analisar a sua validade. Para isso, é preciso utilizar um padrão de referência clínica ou padrão ouro, ou seja, o melhor método disponível para estabelecer a presença ou ausência da condição de destino. Um padrão-ouro seria um padrão de referência sem erros (BOSSUYT et al., 2015). Em nosso estudo assumimos o sistema de medição tridimensional (cinemetria 3D) como o padrão referencia.

Segundo Kottner *et al.* (2011) a reprodutibilidade interavaliador é a concordância entre diferentes avaliadores, utilizando a mesma escala, classificação, instrumento ou procedimento, avaliando os mesmos assuntos ou objetos. A reprodutibilidade intra-avaliador, também conhecida como teste e re-teste, é a concordância entre o mesmo avaliador utilizando a mesma escala, classificação, instrumento ou procedimento, avaliando o mesmo sujeito ou objeto em diferentes momentos.

DESENHO DO ESTUDO E PARTICIPANTES

Esse é um estudo prospectivo, onde a coleta de dados foi planejada antes da realização das avaliações. A amostra foi composta por indivíduos de ambos os sexos. Os participantes da pesquisa foram voluntários, identificados por meio de convite verbal e virtual, entre os meses de setembro a novembro de 2019. As avaliações ocorreram no setor de Biomecânica do laboratório de pesquisa de uma universidade. Para as análises de reprodutibilidade e validade, a amostra foi definida com base em um teste bi-caudal, adotando um poder de 90%, assumindo uma hipótese nula com valor de $\kappa = 0,00$ e um κ detectável de 0,70 com uma proporção de classificações positivas de 0,50, resultando em 22 indivíduos (SIM; WRIGHT, 2005). Prevendo 30% de perda amostral, foi adotado uma quantidade alvo de 30 indivíduos.

Os critérios de inclusão foram: (1) indivíduos com idades entre 18 e 60 anos, (2) não obesos ($IMC < 30 \text{ kg/m}^2$) (WORLD HEALTH ORGANIZATION, 2000) e (3) sem cirurgias na lombar, pelve ou quadril. Os critérios de exclusão foram: (1) dor lombar ou no quadril no momento da coleta, (2) impossibilidade de realizar os testes do protocolo, ou seja, não conseguir flexionar o tronco e não conseguir permanecer sentado sem encosto, e (3) mais que 2 cm de diferença de comprimento entre os membros inferiores (MsIs) (GURNEY, 2002).

Com o objetivo de abordar todos os resultados possíveis dos testes, a amostra foi preenchida de forma consecutiva, de modo que três características diferentes dos participantes estivessem presentes na amostra: (1) TFP superior ao TFS, ou seja, indivíduos com TFP positivo [+] e TFS negativo [-] ou TFP positivo [++] e TFS positivo [+], indicando um TFP mais relevante que o TFS; (2) TFS superior ao TFP, ou seja, TFS [+] e TFP [-] ou TFS [++] e TFP [+]; e (3) TFP e TFS iguais, ou seja, TFP [-] e TFS [-] ou TFP [+] e TFS [+]. A informação de qual característica o avaliado possuía não foi divulgada nem aos participantes e tampouco a equipe de avaliação, garantindo o cegamento do estudo.

PROTOCOLO DE AVALIAÇÃO

A equipe de avaliação foi composta por 5 Fisioterapeutas (A, B, C, D e E), dentre eles, um Osteopata (A) e dois são estudantes de Osteopatia do 3º ano (B e C). O avaliador A foi considerado o mais experiente e referência, com 11 anos de prática clínica na realização dos testes do protocolo. Os avaliadores B e C, ambos com 2 anos de experiência na execução dos testes. O pesquisador D foi o responsável pela anamnese e randomização das avaliações, e o pesquisador E pela coleta de dados no sistema BTS Smart-DX (BTS Bioengineering, EUA). Os avaliadores A, B e C passaram por um treinamento de 20 horas, onde foi abordada a forma de realizar os testes e seus comandos, e testado o protocolo de avaliação completo.

Esse estudo foi aprovado pelo comitê de ética da universidade (CAAE: 15499219.4.0000.5347) e os participantes assinaram o termo de consentimento livre e esclarecido (TCLE). As avaliações ocorreram em dois dias diferentes: (1º dia) reprodutibilidade interavaliador e validade concorrente; (2º dia – após 24 às 72h) reprodutibilidade intra-avaliador. O padrão referência utilizado para a validade concorrente foi a medição tridimensional (cinemetria 3D), por meio do sistema BTS Smart-DX.

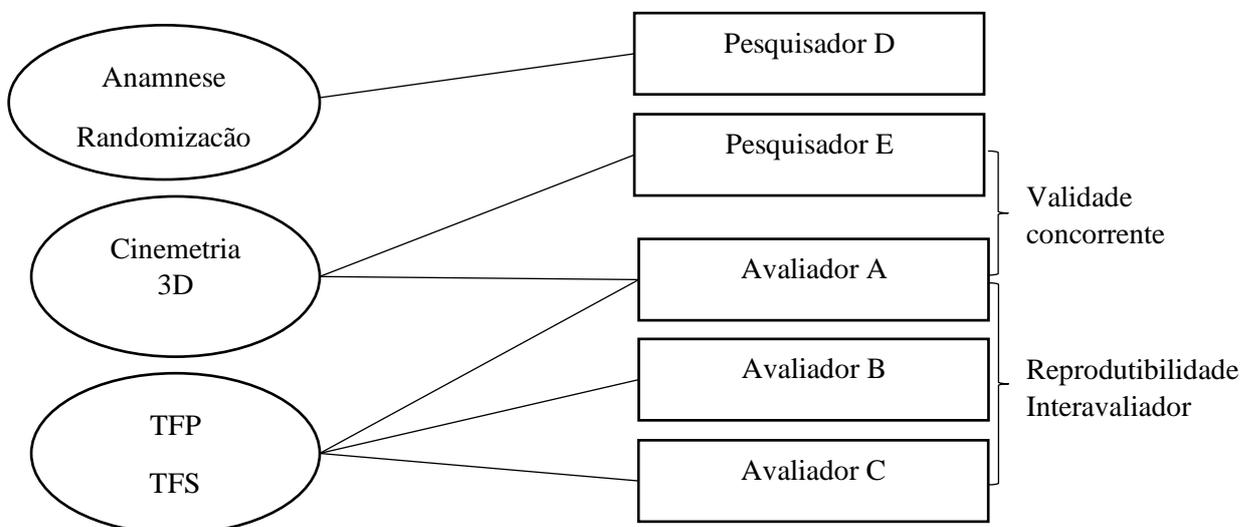
(1º dia) - Reprodutibilidade interavaliador e validade concorrente

Cada participante da pesquisa foi submetido ao seguinte protocolo de avaliação: (a) Anamnese e randomização da ordem dos testes e ordem dos avaliadores; (b) Realização do TFP e TFS pelos avaliadores A, B e C; (c) Medição tridimensional (cinemetria 3D) simultânea a realização dos testes. Nesse 1º dia de avaliação (Figura 1), todos os pesquisadores participaram (A, B, C, D e E). Os avaliadores A, B e C, independentes e cegados, realizaram a reprodutibilidade interavaliador.

A randomização foi feita pelo pesquisador D, o qual efetuou dois sorteios manuais: sorteio da ordem dos testes (TFP e TFS) e sorteio da ordem entre os avaliadores B e C. O avaliador A realizou sempre duas avaliações para cada teste, a primeira e a última, portanto, não randomizadas. A primeira avaliação foi utilizada para análise da reprodutibilidade e da validade e a última para controle, com o objetivo de assegurar que as características anatômicas iniciais não modificassem, após as repetições de cada teste. Caso o participante obtivesse resultados diferentes dos testes entre as mensurações feitas pelo avaliador A, a coleta de dados seria interrompida e o participante excluído do estudo.

Como o protocolo foi feito teste por teste, todos os avaliadores realizaram a mesma sequência sorteada quanto a ordem dos testes. Cada participante foi submetido a quatro repetições de cada teste, sucessivamente.

Figura 1 - Desenho do 1º dia de avaliação.

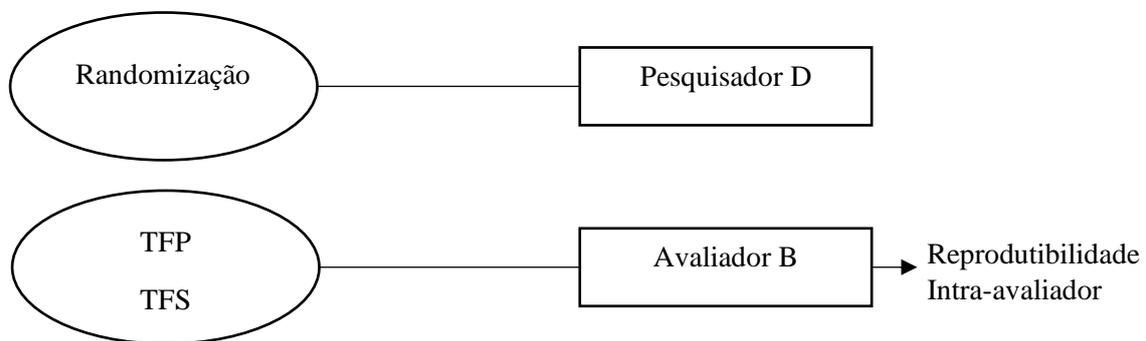


Após a anamnese, o protocolo de coleta foi explicado aos participantes, que foram submetidos aos testes com o objetivo de familiarização dos movimentos, sendo que correções poderiam ser feitas, caso necessário, como por exemplo: flexionar corretamente a coluna, realizar o movimento mais lentamente, não flexionar os joelhos, afastar ou aproximar os pés, entre outros. Ao final do 1º dia, o avaliado recebeu uma folha lembrete, contendo a data e horário do 2º dia de avaliação e orientações para não realizar atividade física, como por exemplo: academia, alongamentos, corrida; atividades que envolvam esforço físico, como por exemplo: faxina, carregar peso, relação sexual; e tratamentos terapêuticos, como: fisioterapia, osteopatia, quiropraxia.

(2º dia) - Reprodutibilidade intra-avaliador

Nesse 2º dia de avaliação (Figura 2) cada participante da pesquisa foi submetido ao seguinte protocolo de avaliação: (a) Randomização da ordem dos testes, pelo pesquisador D; (b) Realização do TFP e TFS pelo avaliador B. O 2º dia foi feito após um intervalo de 24 a 72h do 1º dia, foi optado esse tempo de intervalo para minimizar possíveis chances de mudanças nas características anatômicas dos participantes entre os dias de avaliação. O participante era questionado se estava com dor naquele momento e se tinha respeitado as orientações feitas no final do 1º dia.

Figura 2 - Desenho do 2º dia de avaliação.



ANAMNESE

A anamnese foi aplicada no 1º dia de avaliações, pelo pesquisador D. Inicialmente foram coletados dados como data de nascimento, estatura e massa corporal, a partir do auto relato do avaliado. Posteriormente foi feito o cálculo do índice de massa corporal (IMC), por meio da fórmula: estatura/massa corporal² (ANJOS, 1992). Foi questionado também se o indivíduo possuía dor lombar ou no quadril no momento da avaliação e se já fez cirurgia na

lombar, pelve ou quadril. Além disso, foi mensurado o comprimento real dos membros inferiores, conforme Maloney e Keeney (2004), onde, por meio de fita métrica, foi mensurada a distância entre a Espinha Ilíaca Antero Superior (EIAS) ao Maléolo Medial (MM) ipsilateral, tendo essa metodologia sua validação e reprodutibilidade confirmadas segundo Gogia e Braatz (1986). Caso o participante não se encaixasse nos critérios de elegibilidade feitos na anamnese, a coleta de dados seria cancelada automaticamente

TESTE DE FLEXÃO EM PÉ (TFP)

Os participantes foram orientados a permanecerem em posição ortostática, com os membros superiores ao longo do corpo, com os pés na linha dos quadris (GREENMAN, 2003), posicionados paralelamente sem nenhum ângulo de rotação. O avaliador se posicionava atrás do avaliado, em uma posição confortável para palpar as Espinhas Ilíacas Póstero Superiores (EIPSS) do avaliado. O avaliador colocava suas mãos bilateralmente e lateralmente nas cristas ilíacas do avaliado e deslizava com os polegares ao encontro das EIPSSs. As bases dos polegares se posicionavam na obliquidade inferior das EIPSSs (FIGURA 3a). Os participantes então eram orientados a realizar uma flexão máxima de coluna, lentamente, iniciando o movimento pela cervical, mantendo os joelhos estendidos (FIGURA 3b). O teste era considerado negativo se a movimentação das EIPSSs fosse simétrica ou teste positivo se um dos lados movimentasse mais que o outro em direção cefálica e/ou ventralmente. Somente um dos três resultados do teste era possível: teste negativo, positivo a direita (D) e positivo a esquerda (E) (VINCENT-SMITH; GIBBONS, 1999; POTTER e ROTHSTEIN, 1985; RIDDLE, 2002; CIBULKA e KOLDEHOFF, 1999). Os avaliadores A, B e C anotavam os achados em uma tabela desenvolvida para o estudo (APÊNDICE 1), nos dois dias de avaliação. Para melhor investigação do teste, as anotações foram divididas em três etapas: posição inicial, posição final e conclusão do teste.

Figura 3a - Posição inicial do TFP.

Avaliador realizando a palpação das EIPSSs.

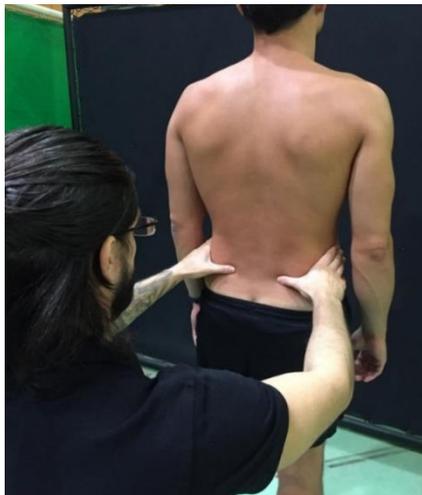
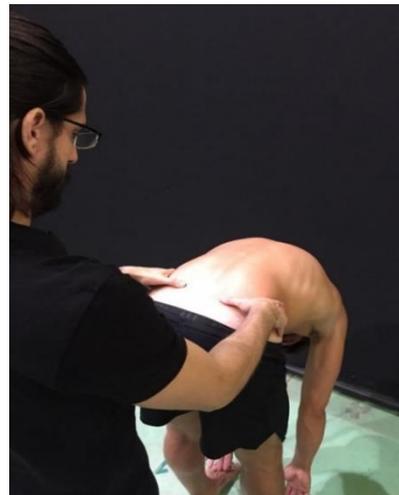


Figura 3b - Posição final do TFP.

Avaliador observa a movimentação das EIPSSs.



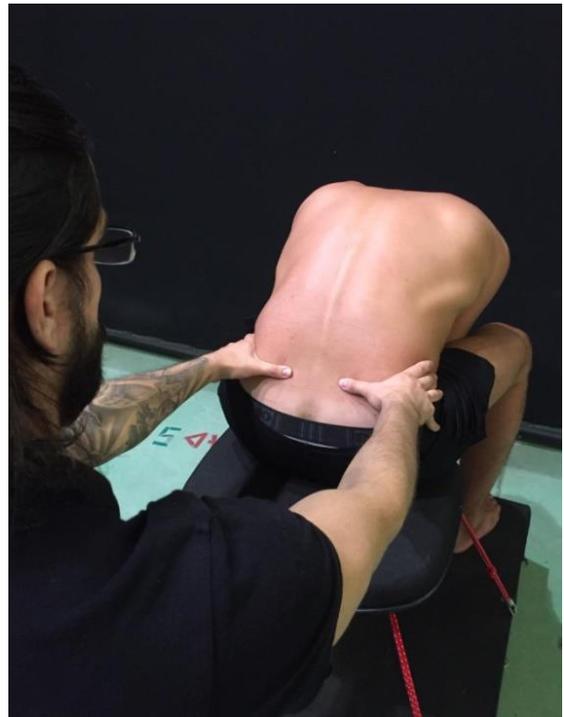
TESTE DE FLEXÃO SENTADO (TFS)

O TFS é feito semelhantemente ao TFP, porém os indivíduos partem da posição sentada. Os participantes eram orientados a sentarem em um assento regulável, com a coluna ereta, com os pés apoiados em uma superfície plana, posicionados paralelamente sem nenhum ângulo de rotação, os joelhos e quadris na largura dos ombros e aproximadamente a 90° de flexão. O posicionamento do avaliador era igual ao inferior para a palpção das EIPSSs. Os participantes então eram orientados a cruzar as mãos atrás da cabeça, aproximar os cotovelos e realizar uma flexão máxima de coluna, lentamente, iniciando o movimento pela cervical (FIGURA 4a e 4b) (PAYDAR; THIEL; GEMMELL, 1994; POTTER e ROTHSTEIN, 1985; RIDDLE, 2002; CIBULKA e KOLDEHOFF, 1999). Os apontamentos possíveis do teste eram os mesmos do TFP. Os avaliadores utilizavam a mesma tabela de anotação dos achados (APÊNDICE 1) e o mesmo critério de anotação.

Figura 4a - Posição inicial do TFS.
Avaliador realizando a palpção das EIPSSs.



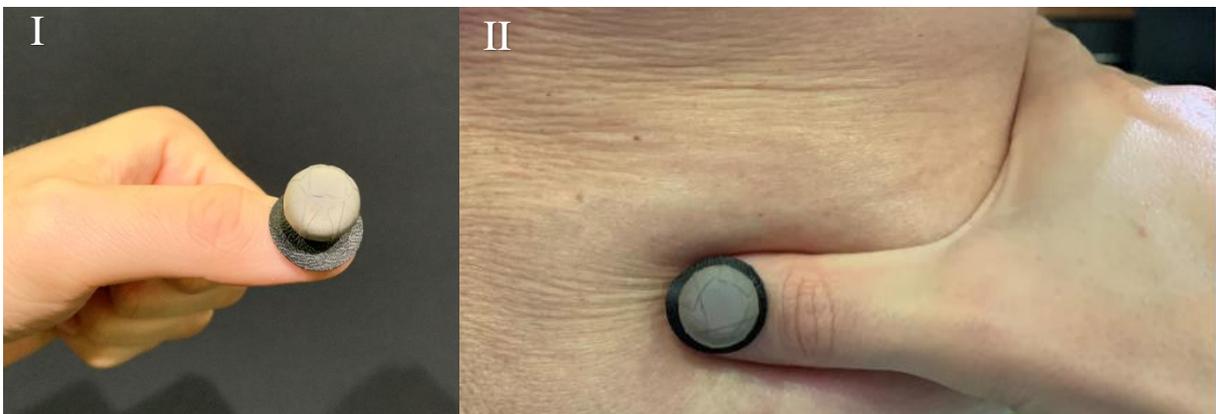
Figura 4b - Posição final do TFS.
Avaliador observa a movimentação das EIPSSs.



MEDIÇÃO TRIDIMENSIONAL

A validade concorrente dos testes foi realizada considerando a medição tridimensional (cinemetria 3D) por meio do sistema BTS Smart-DX com 10 câmeras como sendo o padrão referência. As coletas foram realizadas com uma taxa de amostragem de 100Hz, por meio do software BTS Smart Capture. Devido à variabilidade existente no processo de retirada e colocação de marcadores anatômicos, neste estudo, os pontos anatômicos de referência no avaliado foram identificados por meio dos polegares do avaliador durante a palpação. Ao longo do protocolo de coleta, que ocorreu no 1º dia de avaliação, o avaliador A permaneceu com os marcadores refletivos colados com fita dupla face nos polegares (FIGURA 5) para a realização dos testes.

Figura 5 - Marcador refletivo colado no polegar do avaliador A (I) e palpação da Espinha Ilíaca Pósterio Superior (EIPS) (II).



O protocolo criado para esse estudo é baseado no cálculo da diferença linear (em milímetros) da posição média das EIPs direita (D) e esquerda (E) no eixo Y (caudal-cranial do avaliado). Para que esse cálculo fosse possível, foi criado um Sistema de Coordenada Local (SCL) no avaliado - por um cluster de três pontos não colineares utilizando os marcadores dos polegares do avaliador e por mais um marcador anatômico na região de L3 do avaliado - e outro SCL no avaliador - por um cluster de quatro pontos utilizando uma faixa logo acima dos olhos com quatro marcadores refletivos (FIGURA 6). Um SCL é caracterizado por um conjunto de três coordenadas (x, y e z) que se desloca juntamente com o local em que é inserido. Durante a execução dos testes, o avaliador A foi orientado a acompanhar o movimento do avaliado com a cabeça, para que os SCLs se mantivessem em uma angulação semelhante.

Figura 6: SCL criado utilizando uma faixa com quatro marcadores refletivos acima dos olhos.



Para análise dos dados, foi utilizado um filtro Butterworth passa baixa, com uma frequência de corte de 1Hz. O trackeamento dos pontos anatômicos foi feito no software BTS Smart Tracker e a análise dos dados no software BTS Smart Analyzer com rotinas criadas especificamente para os testes. Foram feitas duas capturas estáticas por avaliado, com duração de 10 segundos cada: (1) posição inicial em pé (TFP) ou sentado (TFS) e (2) posição final em flexão máxima de coluna em pé (TFP) ou sentado (TFS). No momento da análise dos dados, o primeiro segundo inicial e o último segundo final das coletas foram descartados. A conclusão do teste foi obtida por meio da análise do deslocamento entre a posição inicial e final de cada avaliado.

ANÁLISE ESTATÍSTICA

Os dados da anamnese como idade, estatura, massa corporal e IMC foram submetidos à estatística descritiva, como média e desvio padrão. A análise estatística de reprodutibilidade intra-avaliador (BxB), reprodutibilidade interavaliador (AxBxC; AxB; AxC; BxC) e validade concorrente foram realizadas no software Statistical Package for the Social Sciences (IBM SPSS Statistics 21) e no software Microsoft Office Excel 2016.

Os dados de validade e reprodutibilidade foram analisados utilizando o percentual de concordância (%C) e a estatística kappa (k) (SIM; WRIGHT, 2005; FLEISS, 1971). Para os

valores de k , a classificação adotada foi de acordo com Altman (1990) onde valores de $k \leq 0,20$ = pobre; 0,21-0,40 = leve; 0,41-0,60 = moderado; 0,61-0,80 = bom e 0,81-1 = muito bom. Para os valores do %C, foi de acordo com por Janse et al. (2004), onde $\% C \leq 0,30$ = pobre; 0,31-0,50 = fraco; 0,51-0,70 = moderado; 0,71-0,90 = bom; e 0,91-1,00 = excelente. O nível de significância adotado foi $\leq 0,05$. Após todas as análises, adotamos como critério para considerar se os testes seriam válidos e reprodutíveis, um $\%C > 0,50$ (concordância moderada) e um $k > 0,40$ (concordância moderada).

Além disso, foi utilizado o método estatístico baseado na análise da curva ROC (curva característica de operação do receptor) (AKOBENG, 2007) com o objetivo de determinar a acurácia diagnóstica do TFP e TFS, como a área sob a curva (AUC) e os valores atingidos de sensibilidade e especificidade. Sensibilidade corresponde à proporção de resultados verdadeiro positivos de um teste, ou seja, identificar os participantes com a condição alvo/doença, e especificidade à proporção de resultados verdadeiro negativos, ou seja, identificar os participantes sem a condição alvo/doença, ambos variam de 0 a 1 (0 a 100%) (AKOBENG, 2007). Foi realizado também a análise dos valores preditivos positivos (VPP) e negativos (VPN), que correspondem a proporção de verdadeiros positivos entre todos os indivíduos com teste positivo e a proporção de verdadeiros negativos entre todos os indivíduos com teste negativo, respectivamente. O VPP e VPN expressam a probabilidade de um paciente com o teste positivo ter a doença, e um paciente com teste negativo não ter a doença, respectivamente (STEINBERG, FINE & CHAPPELL, 2009).

Foram atribuídos 5 pontos de corte de 3 a 7 milímetros (mm) para a categorização dos valores oriundos da medição tridimensional (cinemetria 3D). Então, os valores numéricos em mm da cinemetria 3D, foram categorizados a partir desses pontos de corte, em toda a amostra, para ambos os testes. Esses pontos de corte foram baseados na experiência de realização dos testes e também de acordo com a revisão sistemática de Goode et al. (2008) onde os autores relatam uma mobilidade de translação da articulação sacroilíaca de 4 a 5 mm, com um erro padrão de medida (SEM) de 1,3 mm, durante o movimento de flexão de quadril bilateral.

Os dados escalares em milímetros (mm) oriundos da medição tridimensional (cinemetria 3D) foram categorizados em: negativo, positivo a D e positivo a E, para a conclusão do teste; e D acima, E acima e simétricas, para a posição inicial e final. A tabela de anotação (APÊNDICE 1) ilustra as nomenclaturas utilizadas para a categorização dos dados. Foi adotado um intervalo de confiança de 95%. Para os valores de acurácia diagnóstica pela AUC foi utilizada a

classificação de acordo com Swets (1988), onde valores $> 0,9$ = alta acurácia; entre $0,7-0,9$ = moderada acurácia; $0,5-0,7$ = baixa acurácia e $<0,5$ = resultado casual. Para que os testes sejam considerados acurados deveriam apresentar $AUC >0,7$ (moderada acurácia).

RESULTADOS

A amostra foi composta por 30 indivíduos, 17 do sexo masculino e 13 do sexo feminino, com idade média de 30 (± 8) anos, estatura média de 171 (± 9) cm, massa média de 71 (± 12) kg e índice de massa corporal (IMC) médio de 24 (± 2) kg/cm². Não houve perda amostral, assim como nenhum indivíduo excluído por apresentar uma diferença no comprimento dos membros inferiores >2 cm, dor lombar ou no quadril no dia da coleta, impossibilidade de realizar os testes do protocolo ou ter resultados diferentes dos testes nas duas mensurações do avaliador A.

As Figuras 7a e 7b mostram o fluxo de participantes pelo estudo, para o TFP e o TFS, respectivamente. Não houve participantes excluídos e tampouco com resultados inconclusivos nos testes e na medição tridimensional (cinemetria 3D).

Figura 7a - Diagrama de fluxo referente ao Teste de Flexão em Pé (TFP), realizado pelo avaliador A, e a medição tridimensional (cinemetria 3D), para o ponto de corte de 3 mm.

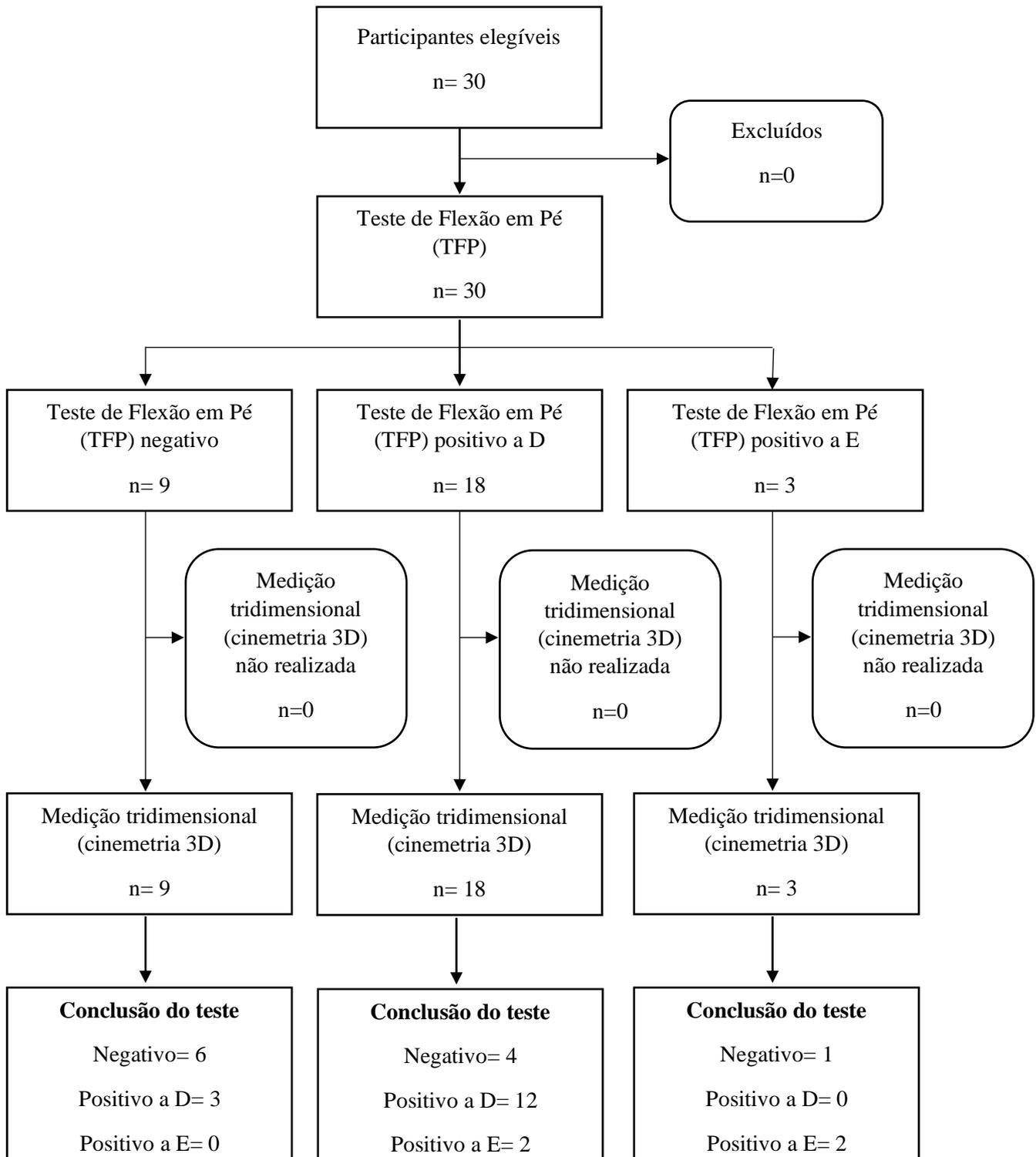
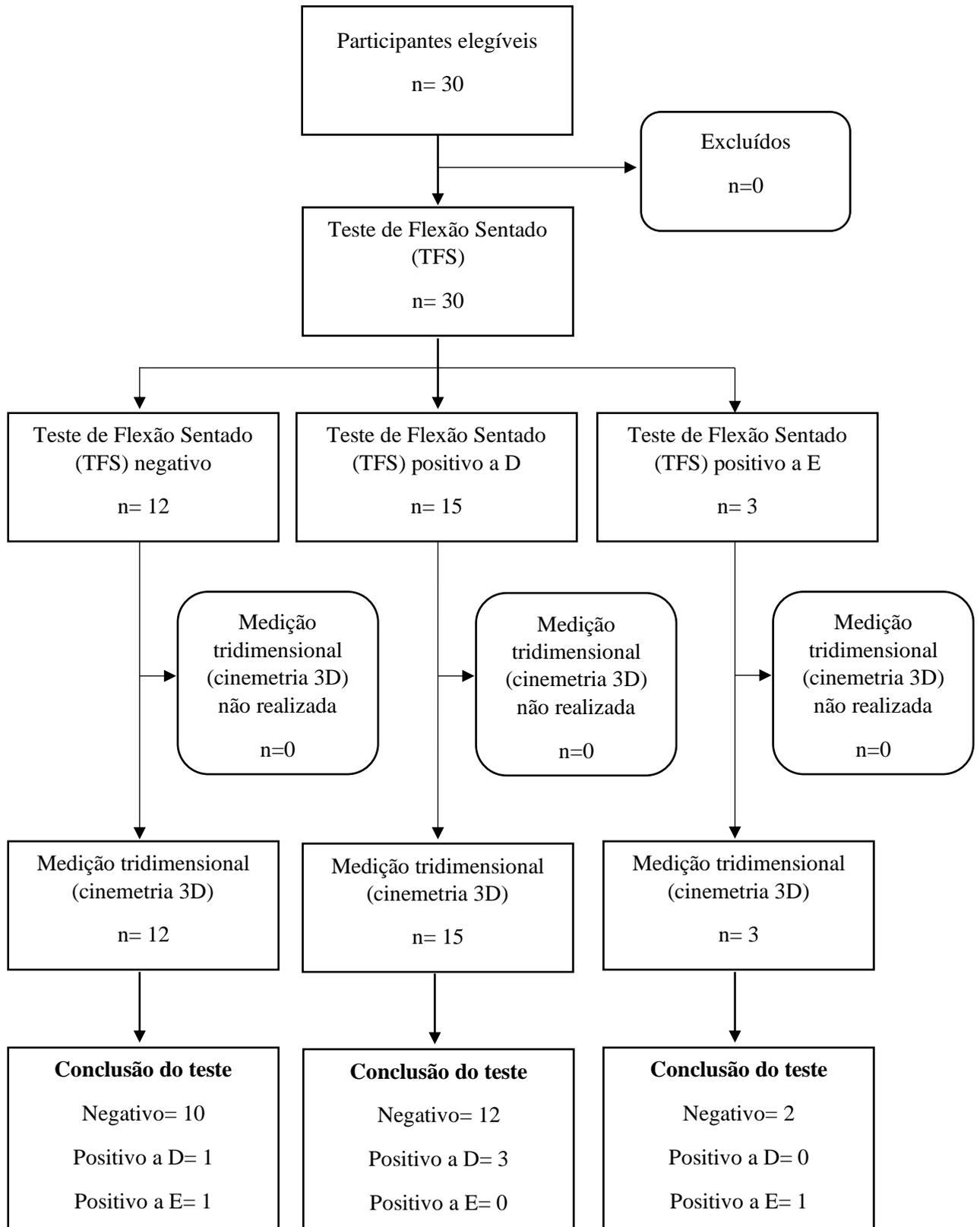


Figura 7b - Diagrama de fluxo referente ao Teste de Flexão Sentado (TFS), realizado pelo avaliador A, e a medição tridimensional (cinemetria 3D), para o ponto de corte de 7 mm.



A Tabela 1 trata sobre as informações descritivas (média, mediana, desvio padrão, valor mínimo e máximo) em milímetros (mm) das avaliações com a medição tridimensional (cinemetria 3D), do Teste de Flexão em Pé (TFP) e do Teste de Flexão Sentado (TFS) referente ao 1º dia de avaliação, de acordo com as etapas: posição inicial, posição final e conclusão do teste. Em geral, os valores de assimetria (posição inicial e final) e deslocamento (conclusão do teste) são similares, não havendo relevantes discrepâncias entre as etapas.

Tabela 1 – Informações descritivas das avaliações com a medição tridimensional (cinemetria 3D), do Teste de Flexão em Pé (TFP) e do Teste de Flexão Sentado (TFS), referente ao 1º dia de avaliação.

	Teste de Flexão em Pé (TFP)			Teste de Flexão Sentado (TFS)		
	Posição Inicial ^{nb} (mm)	Posição Final ^{nb} (mm)	Conclusão do teste ^{nb} (mm)	Posição Inicial ^b (mm)	Posição Final ^b (mm)	Conclusão do teste ^{nb} (mm)
Média	6	10	6	9	11	5
Mediana	5	10	7	9	12	5
DP	5	7	5	6	6	4
Mínimo	0	1	0	0	0	0
Máximo	24	31	21	25	21	14

^b: dados paramétricos; ^{nb}: dados não paramétricos; DP: desvio padrão; mm: milímetros;

A Tabela 2 mostra os resultados de validade concorrente entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão em Pé (TFP), para os 5 pontos de corte (3, 4, 5, 6 e 7 mm) e para toda as etapas (posição inicial, posição final e conclusão do teste), em um total de 30 indivíduos, referente ao 1º dia de avaliação. O ponto de corte que apresentou o conjunto de melhores valores de concordância (k e %C) foi o de 3 mm. Analisando todos os pontos de corte, os %C variaram de moderados a bom, e a concordância k variou de leve a moderada.

Tabela 2 - Validade concorrente entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão em Pé (TFP) para os 5 pontos de corte.

Teste de Flexão em Pé (TFP) - ênfase na mobilidade dos Ilíacos										
	3 mm		4 mm		5 mm		6 mm		7 mm	
	n (%C)	k (p)								
Posição Inicial (em pé)	24 (80%)	0,57 (<0,01)*	23 (76,7%)	0,53 (<0,01)*	22 (73,3%)	0,48 (<0,01)*	23 (76,7%)	0,55 (<0,01)*	17 (56,7%)	0,26 (<0,05)*
Posição Final (flexão máxima)	23 (76,7%)	0,53 (<0,01)*	23 (76,7%)	0,53 (<0,01)*	24 (80%)	0,62 (<0,01)*	24 (80%)	0,63 (<0,01)*	22 (73,3%)	0,51 (<0,01)*
Conclusão do teste (durante a flexão)	20 (66,7%)	0,42 (<0,01)*	19 (63,3%)	0,38 (<0,01)*	19 (63,3%)	0,40 (<0,01)*	19 (63,3%)	0,40 (<0,01)*	19 (63,3%)	0,41 (<0,01)*

p: nível de significância; *: $p < 0,05$; n: amostra; k: kappa; %C: percentual de concordância; mm: milímetros.

A Tabela 3 trata sobre os resultados de validade concorrente entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão Sentado (TFS), para os 5 pontos de corte (3, 4, 5, 6 e 7 mm) e para toda as etapas (posição inicial, posição final e conclusão do teste), em um total de 30 indivíduos, referente ao 1º dia de avaliação. O ponto de corte que apresentou o conjunto de melhores valores de concordância (k e %C) foi o de 7 mm. Analisando todos os pontos de corte, os %C variaram de moderados a bom, e a concordância k variou de leve a moderada, exceto para a conclusão do teste, onde os %C foram leves e os valores de k foram pobres.

Tabela 3 - Validade concorrente entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão Sentado (TFS), para os 5 pontos de corte.

Teste de Flexão Sentado (TFS) – ênfase na mobilidade do Sacro										
	3mm		4mm		5mm		6mm		7mm	
	n	k								
	(%C)	(p)								
Posição Inicial (sentado)	20 (66,7%)	0,36 (<0,05)*	20 (66,7%)	0,36 (<0,05)*	19 (63,3%)	0,30 (0,06)	23 (76,7%)	0,56 (<0,01)*	24 (80%)	0,62 (<0,01)*
Posição Final (flexão máxima)	20 (66,7%)	0,37 (<0,01)*	22 (73,3%)	0,50 (<0,01)*	22 (73,3%)	0,50 (<0,01)*	23 (76,7%)	0,55 (<0,01)*	23 (76,7%)	0,55 (<0,01)*
Conclusão do teste (durante a flexão)	15 (50%)	0,20 (0,14)	15 (50%)	0,20 (0,14)	14 (46,7%)	0,15 (0,23)	14 (46,7%)	0,11 (0,35)	14 (46,7%)	0,12 (0,25)

p: nível de significância; *: $p < 0,05$; n: amostra; k: kappa; %C: percentual de concordância; mm: milímetros.

A Tabela 4 trata sobre a análise da curva ROC entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão em Pé (TFP), para os 5 pontos de corte (3, 4, 5, 6 e 7 mm) e para toda as etapas (posição inicial, posição final e conclusão do teste). Nessa análise foram incluídos os indivíduos caracterizados em positivos a D e negativos, por serem a grande maioria da amostra e devido esse teste estatístico requerer variáveis dicotômicas. O ponto de corte que apresentou o conjunto de melhores valores de acurácia diagnóstica (AUC) foi o de 3 mm, sendo considerada uma acurácia baixa para a conclusão do teste e acurácia moderada para a posição inicial e final.

Tabela 4 - Análise da curva ROC entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão em Pé (TFP) para os 5 pontos de corte.

Teste de Flexão em Pé (TFP) - ênfase na mobilidade dos Íliacos															
	3mm			4mm			5mm			6mm			7mm		
	PI	PF	C												
	n= 29	n= 27	n= 25	n= 30	n= 27	n= 25	n= 30	n= 27	n= 25	n= 30	n= 27	n= 25	n= 30	n= 27	n= 26
AUC	0,797	0,786	0,700	0,767	0,786	0,666	0,746	0,796	0,686	0,794	0,791	0,686	0,717	0,756	0,700
Erro Padrão	0,097	0,118	0,112	0,091	0,118	0,113	0,092	0,103	0,109	0,083	0,099	0,109	0,095	0,102	0,104
p	0,01*	<0,05*	<0,05*	0,01*	<0,05*	0,16	<0,05*	<0,05*	0,12	<0,01*	0,01*	0,12	0,09	<0,05*	0,09
IC 95%	0,607- 0,988	0,556- 1,000	0,480- 0,920	0,589- 0,944	0,556- 1,000	0,444- 0,887	0,565- 0,926	0,594- 0,998	0,473- 0,899	0,630- 0,957	0,597- 0,985	0,473- 0,899	0,532- 0,903	0,556- 0,955	0,496- 0,904
Sensibilidade	90%	77%	80%	93%	77%	79%	93%	84%	83%	100%	88%	83%	100%	88%	90%
Especificidade	70%	80%	60%	60%	80%	55%	56%	75%	54%	59%	70%	54%	44%	64%	50%
VPP	85%	94%	75%	70%	94%	69%	65%	89%	63%	65%	83%	63%	65%	78%	53%
VPN	78%	44%	67%	90%	44%	67%	100%	67%	78%	100%	78%	78%	100%	78%	89%

PI: posição inicial; PF: posição final; C: conclusão do teste; AUC: área sob a curva; IC: intervalo de confiança; VPP: valor preditivo positivo; VPN: valor preditivo negativo; *: p<0,05

A Tabela 5 trata sobre a análise da curva ROC entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão Sentado (TFS), para os 5 pontos de corte (3, 4, 5, 6 e 7 mm) e para toda as etapas (posição inicial, posição final e conclusão do teste). Igualmente a análise anterior, também foram incluídos os indivíduos positivos a D e negativos. O ponto de corte que apresentou o conjunto de melhores valores de acurácia diagnóstica (AUC) foi o de 7 mm, sendo considerada uma acurácia moderada para a posição inicial e final e acurácia baixa para a conclusão do teste.

Tabela 5 - Análise da curva ROC entre o avaliador A e a medição tridimensional (cinemetria 3D) do Teste de Flexão Sentado (TFS) para os 5 pontos de corte.

Teste de Flexão Sentado (TFS) – ênfase na mobilidade do Sacro															
	3mm			4mm			5mm			6mm			7mm		
	PI	PF	C												
	n= 28	n= 28	n= 23	n= 28	n= 28	n= 23	n= 28	n= 28	n= 24	n= 28	n= 28	n= 26	n= 28	n= 28	n= 26
AUC	0,762	0,813	0,572	0,762	0,841	0,572	0,700	0,841	0,563	0,823	0,875	0,558	0,856	0,875	0,602
Erro Padrão	0,102	0,085	0,122	0,102	0,074	0,122	0,110	0,074	0,129	0,085	0,065	0,134	0,078	0,065	0,150
p	<0,05*	<0,05*	0,56	<0,05*	0,01*	0,56	0,10	0,01*	0,63	<0,01*	<0,01*	0,67	<0,01*	<0,01*	0,52
IC 95%	0,563- 0,961	0,646- 0,979	0,334- 0,810	0,563- 0,961	0,697- 0,985	0,334- 0,810	0,484- 0,916	0,697- 0,985	0,310- 0,816	0,656- 0,990	0,747- 1,000	0,295- 0,821	0,703- 1,000	0,747- 1,000	0,309- 0,896
Sensibilidade	67%	63%	73%	67%	68%	73%	65%	68%	71%	81%	75%	67%	87%	75%	75%
Especificidade	86%	100%	42%	86%	100%	42%	75%	100%	41%	83%	100%	45%	85%	100%	46%
VPP	93%	100%	53%	93%	100%	53%	87%	100%	33%	87%	100%	27%	87%	100%	20%
VPN	46%	31%	63%	46%	46%	63%	46%	46%	78%	77%	62%	82%	85%	62%	91%

PI: posição inicial; PF: posição final; C: conclusão do teste; AUC: área sob a curva; IC: intervalo de confiança; VPP: valor preditivo positivo; VPN: valor preditivo negativo; *: p<0,05

A Tabela 6 é sobre os resultados de reprodutibilidade intra-avaliador (BxB) para ambos os testes (TFP e TFS), em todas as etapas (posição inicial, posição final e conclusão do teste), referente ao 2º dia de avaliação, em um total de 30 indivíduos. Observamos um %C moderado para ambos os testes, em todas as etapas. Em relação a concordância k, para o TFP observamos uma concordância moderada na posição final e na conclusão do teste e, concordância leve para a posição inicial. Para o TFS, houve concordância leve em todas as etapas.

Tabela 6 – Reprodutibilidade intra-avaliador (BxB) para o o Teste de Flexão em Pé (TFP) e Teste de Flexão Sentado (TFS).

	Teste de Flexão em Pé (TFP)		Teste de Flexão Sentado (TFS)	
	Concordância n (%C)	kappa k (p)	Concordância n (%C)	kappa k (p)
Posição Inicial	20 (66,7%)	0,38 (<0,01)*	19 (66,3%)	0,39 (<0,01)*
Posição Final	19 (66,3%)	0,43 (<0,01)*	16 (53,3%)	0,31 (<0,01)*
Conclusão do teste	19 (66,3%)	0,43 (<0,01)*	17 (56,7%)	0,38 (<0,01)*

*: $p < 0,01$; p: nível de significância; n: amostra; k: kappa; %C: percentual de concordância.

A Tabela 7 apresenta os resultados de reprodutibilidade interavaliador (AxBxC; AxB; AxC; BxC) para ambos os testes (TFP e TFS), em todas as etapas (posição inicial, final e conclusão do teste), referente ao 1º dia de avaliação, em um total de 30 indivíduos. Observamos concordância pobre (%C e k) entre os 3 avaliadores (AxBxC), em ambos os testes e em todas as etapas.

Tabela 7 – Reprodutibilidade interavaliador (AxBxC; AxB; AxC; BxC) para o Teste de Flexão em Pé (TFP) e Teste de Flexão Sentado (TFS).

	Teste de Flexão em Pé (TFP)		Teste de Flexão Sentado (TFS)	
	Concordância	kappa	Concordância	kappa
	n (%C)	k (p)	n (%C)	k (p)
Posição Inicial AxBxC	8 (26,7%)	0,05 (0,56)	9 (30%)	0,07 (0,45)
Posição Inicial AxB	16 (53,3%)	0,14 (0,35)	17 (56,7%)	0,20 (0,21)
Posição Inicial AxC	12 (40%)	0,07 (0,55)	13 (43,3%)	-0,00 (0,98)
Posição Inicial BxC	12 (40%)	0,03 (0,79)	14 (46,7%)	0,05 (0,73)
Posição Final AxBxC	4 (13,3%)	0,02 (0,77)	4 (13,3%)	0,04 (0,63)
Posição Final AxB	12 (40%)	0,13 (0,22)	13 (43,3%)	0,13 (0,30)
Posição Final AxC	16 (53,3%)	0,20 (0,15)	10 (33,3%)	-0,14 (0,35)
Posição Final BxC	6 (20%)	-0,17 (0,15)	13 (43,3%)	0,14 (0,28)
Conclusão do teste AxBxC	3 (10%)	-0,02 (0,83)	4 (13,3%)	0,01 (0,84)
Conclusão do teste AxB	12 (40%)	0,13 (0,22)	13 (43,3%)	0,15 (0,19)
Conclusão do teste AxC	13 (43,3%)	0,02 (0,89)	9 (30%)	-0,04 (0,74)
Conclusão do teste BxC	7 (23,3%)	-0,111 (0,33)	10 (33,3%)	-0,015 (0,91)

p: nível de significância; n: amostra; k: kappa; %C: percentual de concordância.

DISCUSSÃO

O Teste de Flexão em Pé (TFP) e o Teste de Flexão Sentado (TFS), foram divididos em: posição inicial, posição final e conclusão. Para ambos os testes, a posição final ocorre em flexão máxima de coluna e a conclusão se dá de acordo com o deslocamento das Espinhas Ilíacas Pósteras Superiores (EIPSS) durante a flexão. O que difere entre os dois testes é a posição inicial, onde para o TFP é em pé, e para o TFS sentado. Não há na literatura estudos que dividam estes testes nestas etapas, sendo a análise feita somente de acordo com o deslocamento das EIPSS. A exploração descrita no atual estudo permite, não só, inferências sobre a conclusão do teste, ou seja, o seu resultado final, mas também sobre as posições estáticas dadas no início e final dos testes.

A Tabela 1 apresenta os resultados descritivos (média, mediana, desvio padrão, valor mínimo e máximo) da medição tridimensional (cinemetria 3D) para o TFP e TFS. Em geral, os testes apresentam os valores semelhantes tanto de assimetria, dada pela posição inicial e final, quanto de deslocamento, dado pela conclusão do teste. É possível observar uma diferença de 3 milímetros (mm) na posição inicial e 1 mm na posição final entre os testes, com o TFS mostrando uma maior assimetria média de 9 e 11 mm, respectivamente. Desse modo, no atual estudo o TFS apresentou maiores assimetrias das EIPSS, no entanto, em relação ao deslocamento o mesmo não acontece. Ainda, para os valores máximos, o TFP apresenta maiores valores, com uma diferença de 10 mm para a posição final e 7 mm para a conclusão do teste, em comparação ao TFS. Ambos os testes tendem a ter um deslocamento médio de 5 a 6 mm, estando esses valores de acordo com os pontos de corte utilizados nesse estudo.

De acordo com os resultados de validade concorrente do Teste de Flexão em Pé (TFP) (Tabela 2), é possível observar um percentual de concordância bom (%C=73,3-80%) para a posição inicial e final em todos os pontos de corte, exceto para o ponto de corte de 7 milímetros (mm) na posição inicial, cuja concordância foi moderada (%C= 56,7%). Para a conclusão do teste, o %C foi moderado (%C=63,3- 66,7%) em todos os pontos de corte. Referente aos valores de kappa (k), a concordância foi moderada (k= 0,48-0,57) para a posição inicial em todos os pontos de corte, exceto para o de 7 mm onde a concordância foi leve (k= 0,26). Para a posição final, a concordância foi moderada (k= 0,51 - 0,53) para os pontos de corte de 3, 4 e 7 mm e, concordância boa (k= 0,62-0,63) para os pontos de corte de 5 e 6 mm. Referente a conclusão do teste, a concordância foi moderada (k= 0,40-0,42), exceto para o ponto de corte de 4 mm, onde a concordância foi leve (k= 0,38). Portanto, esses achados confirmam a validade do TFP,

sendo o melhor ponto de corte o de 3 mm, onde a concordância foi moderada a boa (%C= 66,7–80%; $k= 0,42-0,57$) em todas as etapas.

No que tange a validade concorrente do Teste de Flexão Sentado (TFS) (Tabela 3), para a posição inicial, é possível observar um %C moderado (%C= 63,3-66,7%) para os pontos de corte de 3, 4 e 5 mm e %C bom (%C= 76,7-80%) para os pontos de corte de 6 e 7 mm. Referente a posição final, o %C foi bom (%C= 73,3-76,7%) para todos os pontos de corte, exceto para o de 3 mm onde foi moderado (%C= 66,7%). Para a conclusão do teste, o %C foi fraco (%C= 46,7 - 50%) em todos os pontos de corte. Referente aos resultados analisados pela estatística k , para a posição inicial, houve uma concordância leve ($k= 0,30-0,36$) para os pontos de corte de 3, 4 e 5 mm, concordância moderada ($k= 0,56$) em 6 mm e, concordância boa ($k= 0,62$) em 7 mm. Referente a posição final, houve concordância leve ($k= 0,37$) em 3 mm e, concordância moderada ($k= 0,50-0,55$) para o restante dos pontos de corte. Para a conclusão do teste, a concordância foi pobre ($k \leq 0,20$) em todos os pontos de corte. Assim sendo, o melhor ponto de corte para o TFS foi o de 7 mm, onde a concordância foi moderada e boa para a posição inicial e final (%C= 76,7–80%; $k= 0,55-0,62$), porém para a conclusão do teste a concordância foi pobre e fraca (%C= 46,7%; $k= 0,12$), confirmando a validade das posições estáticas, mas não da conclusão do teste. Esse achado parece indicar que, apesar do avaliador identificar corretamente a simetria das EIPs no início e final do teste, a interpretação dele, por meio do deslocamento dos referenciais anatômicos, foi um dificultador. É viável destacar que o deslocamento das EIPs é menor em comparação com a assimetria, sendo, em média, de 5 mm para o TFS.

Em relação as análises da curva ROC do Teste de Flexão em Pé (TFP) (Tabela 4), é possível observar acurácia moderada e significativa (AUC= 0,717-0,796; $p \leq 0,05$) para a posição inicial e final em todos os pontos de corte, exceto para a posição inicial no ponto de corte de 7 mm, onde o resultado não foi significativo ($p= 0,09$). Para a conclusão do teste, observa-se acurácia diagnóstica baixa e não significativa (AUC= 0,666 - 0,700; $p > 0,05$) em todos os pontos de corte, exceto para o ponto de corte de 3 mm, onde o resultado foi significativo ($p < 0,05$). Os valores de sensibilidade variaram de 77 a 100%, e de especificidade variaram de 44 a 80%. Os valores preditivos positivos e negativos variaram de 53 a 94% e 44 a 100%, respectivamente. Dessa maneira, para conclusão do TFP, utilizando o ponto de corte de 3 mm, obtém-se uma acurácia diagnóstica baixa (AUC= 0,700), com uma sensibilidade de 80%, especificidade de 60%, VPP de 75% e VPN de 67%.

Quanto as análises da curva ROC do Teste de Flexão Sentado (TFS) (Tabela 5), é possível observar moderada e significativa acurácia ($AUC= 0,762-0,875$; $p\leq 0,05$) para a posição inicial e final em todos os pontos de corte, exceto para o ponto de corte de 5 mm na posição inicial, onde a acurácia foi baixa e não significativa ($AUC= 0,700$; $p=0,104$). Referente a conclusão do teste, a acurácia diagnóstica foi baixa e não significativa ($AUC= 0,558 - 0,602$; $p> 0,05$) para todos os pontos de corte. Os valores de sensibilidade variaram de 63 a 87%, e de especificidade variaram de 41 a 100%. Os valores preditivos positivos e negativos variaram de 20 a 100% e 31 a 91%, respectivamente. Com base nesses achados, para o resultado do TFS, utilizando o ponto de corte de 7 mm, obtém-se uma acurácia diagnóstica baixa ($AUC= 0,602$), com uma sensibilidade de 75%, especificidade de 46%, VPP de 20% e VPN de 91%.

Os resultados do nosso estudo mostram que tanto o TFP quanto o TFS, são considerados testes não acurados ($AUC<0,700$). Contudo, o mesmo não acontece para a posição inicial e final, onde a acurácia diagnóstica é moderada. Desse modo, ambos os testes fornecem informações estáticas diagnósticas importantes de simetria das EIPSS, mas não da mobilidade da ASI durante o movimento de flexão de coluna. Por exemplo, de acordo com esse estudo, para o TFP na posição inicial, um indivíduo com assimetria (EIPS D acima) maior que 3 mm, será diagnosticado como assimétrico em 90% dos casos (sensibilidade) e ele terá uma proporção de ser verdadeiramente assimétrico em 85% dos casos (valor preditivo positivo). Para a posição final, a sensibilidade é de 77% e o VPP é de 94%. Realizando a mesma análise para o TFS na posição inicial, um indivíduo com assimetria maior que 7 mm, será diagnosticado como assimétrico em 87% dos casos e terá uma proporção de ser verdadeiramente assimétrico também em 87% das chances. Para a posição final, a sensibilidade é de 75% e o VPP é de 100%. Esses achados estáticos podem auxiliar os práticos clínicos durante a avaliação da ASI de pacientes, sendo uma forma disponível acurada de mensuração da posição das EIPSS.

Até onde se sabe, não há evidências na literatura sobre a realização de análises de validade concorrente e acurácia diagnóstica em testes clínicos de mobilidade para a ASI, e tampouco para o TFP e TFS. Alguns estudos realizaram a investigação da mobilidade da ASI, porém não utilizando um teste clínico. Bussey et al. (2004) utilizaram o exame de tomografia computadorizada e também um dispositivo de rastreamento magnético digitalizando as referências anatômicas e realizando o cálculo das medições com coordenadas 3D, durante o movimento de abdução e rotação externa do quadril em decúbito ventral. Sturesson et al. (1989) e Kibsgard et al. (2012) utilizaram a análise radiostereométrica (RSA) e a mobilidade da ASI também foi calculada utilizando coordenadas 3D com implante de marcadores na articulação.

Sturesson et al. (1989) avaliou a mobilidade da ASI durante os movimentos de decúbito dorsal para sentado, decúbito dorsal para em pé, e hiperextensão do quadril em decúbito ventral.

Goode et al. (2008) realizaram uma revisão sistemática sobre avaliação tridimensional dos movimentos sacroilíacos, os instrumentos utilizados pelos estudos foram: análise radiostereométrica (RSA), fotogrametria, tomografia computadorizada e goniometria. Os movimentos analisados foram divididos: (1) em translação da ASI durante movimentos dos membros inferiores, (2) rotação da ASI durante movimentos dos membros inferiores, (3) translação da ASI durante movimentos do tronco e (4) rotação da ASI durante movimentos do tronco. A partir dessa revisão, parece que ainda não há um consenso sobre o instrumento padrão ouro/referência para utilização na avaliação da mobilidade da ASI, porém, independente do instrumento, parece unânime a utilização de coordenadas 3D. Em nosso estudo, como padrão de referência para as análises de validade concorrente dos testes, realizamos a avaliação tridimensional por meio do sistema BTS Smart-DX.

Em relação a reprodutibilidade intra-avaliador do TFP (Tabela 6), em geral encontramos concordância moderada (%C= 66,3–66,7%; k= 0,43) em todas as etapas, exceto para o valor de k na posição inicial onde a concordância foi leve (k= 0,38). Quanto ao TFS, a reprodutibilidade intra-avaliador mostrou concordância leve (%C= 53,3–56,7%; k= 0,31–0,39) em todas as etapas, exceto para o %C na posição inicial onde a concordância foi moderada (%C= 66,3%). Referente a reprodutibilidade interavaliador do TFP e TFS (Tabela 7), a concordância foi pobre (%C<30%; k<0,20;) e fraca (%C< 50%), na maioria das etapas e em todas as combinações dos avaliadores (AxBxC; AxB; AxC; BxC), exceto entre os avaliadores AxB durante a posição inicial, para o TFP e TFS, onde a concordância foi moderada (%C= 53,3% e 56,7%, respectivamente) e entre os avaliadores AxC na posição final, somente para o TFP, a concordância também foi moderada (%C= 53,3%). Em vista disso, somente o TFP fornece resultados reprodutíveis (k>0,40) quando a avaliação é feita pelo mesmo avaliador, porém, o mesmo não ocorre quando os avaliadores são diferentes. Esses resultados trazem informações de que avaliações feitas com esse teste podem ser confiáveis para um profissional clínico que trabalha sozinho, mas não para uma equipe multiprofissional, comparando avaliações de um mesmo paciente entre diferentes clínicos.

Estudos anteriores sobre a reprodutibilidade intra-avaliador do TFP mostram concordância variando de moderada a boa, os valores variaram de 68 a 87% (VINCENT-SMITH et al., 1999; ÅSTRÖM et al., 2014) para o %C e de 0,46 a 0,70 (VINCENT-SMITH et

al., 1999; ÅSTRÖM et al., 2014; ARAB et al., 2009) para os valores de concordância k . Resultados esses que corroboram com o atual estudo, que obteve uma concordância moderada ($\%C = 66\%$; $k = 0,43$) para a conclusão do teste. Já para a reprodutibilidade interavaliador do TFP, nos estudos anteriores a concordância variou de pobre a moderada. Os valores variam de 42,7 a 59% (VINCENT-SMITH et al., 1999; ÅSTRÖM et al., 2014; RIDDLE et al., 2002; BOWMAN et al., 1995; POTTER et al., 1985) para o $\%C$ e de 0,05 a 0,51 (VINCENT-SMITH et al., 1999; ÅSTRÖM et al., 2014; ARAB et al., 2009; RIDDLE et al., 2002; BOWMAN et al., 1995) para os valores de k . Contudo, somente um estudo teve kappa acima de 0,40 (ARAB et al., 2009). Esses resultados corroboram em parte com os nossos, uma vez que obtivemos, em geral, concordância pobre ($\%C \leq 30\%$; $k \leq 0,20$) e leve ($\%C \leq 50\%$) entre os avaliadores.

Para o TFS, em relação aos estudos sobre a reprodutibilidade intra-avaliador, a concordância varia de leve a boa. No entanto, os resultados dos estudos anteriores são heterogêneos. Somente um estudo analisou o $\%C$ e obteve valor de 58% (PAYDAR et al., 1994), com valores de k variando entre 0,29 a 0,73 (FRYER et al., 2005; ARAB et al., 2009; PAYDAR et al., 1994). Nossos resultados corroboram em parte com os anteriores, pois obtivemos concordância leve ($k = 0,38$) e moderada ($\%C = 56,7\%$) para a conclusão do teste. Em relação a reprodutibilidade interavaliador do TFS, nos estudos anteriores a concordância variou de pobre a boa. Os valores variam de 34,4 a 71% (FRYER et al., 2005; PAYDAR et al., 1994; POTTER et al., 1985) para o $\%C$ e de 0,06 a 0,75 (FRYER et al., 2005; ARAB et al., 2009; PAYDAR et al., 1994; TONG et al., 2006) para os valores de k . Porém, desses quatro estudos (FRYER et al., 2005; ARAB et al., 2009; PAYDAR et al., 1994; TONG et al., 2006), somente um (ARAB et al., 2009) possui valor de k acima de 0,20. O nosso estudo corrobora com a maioria dos anteriores, pois mostrou, em geral, concordância pobre ($\%C \leq 30\%$; $k \leq 0,20$; $p > 0,05$) e leve ($\%C \leq 50\%$) entre os avaliadores.

Ao analisarmos as avaliações do avaliador A, considerado referência e experiente nesses testes, em relação aos avaliadores B e C, considerados com o mesmo nível de experiência e menos experientes em relação ao A, podemos observar que na maioria dos casos não houveram melhores resultados de concordância com a presença do avaliador experiente, e tampouco melhores resultados de concordância entre os avaliadores menos experientes, sendo grande parte dos resultados tendo um $k < 0,20$ e $\%C < 50\%$. Logo, no nosso estudo, a experiência dos avaliadores parece não ter interferido na reprodutibilidade dos testes. Segundo Fryer et al. (2005) existem muitos fatores que podem contribuir para a inconsistência interavaliador, como as expectativas e habilidades de diagnóstico clínico, fadiga, distração, grau de assimetria,

movimentos do avaliado, composição de gordura e espessura do tecido. Ainda de acordo com Fryer et al. (2005), no seu estudo, os avaliadores foram divididos em dois grupos: treinados e não treinados. Eles realizaram a avaliação do TFS e a palpação das EIPs, e, mesmo assim, ambos os grupos obtiveram resultados de reprodutibilidade interavaliador $<0,20$ para a concordância k . Em relação a reprodutibilidade intra-avaliador, para o TFS, o grupo treinado obteve um k de 0,41 e o não treinado de 0,02. Para a palpação das EIPs, o valor de k foi semelhante entre os grupos (0,54 e 0,49, respectivamente). Dessa forma, a relação da experiência do avaliador para a reprodutibilidade parece incerta, podendo a avaliação ser influenciada por múltiplos fatores.

Em suma, nossos resultados permitem que a hipótese do estudo seja parcialmente aceita, pois somente o TFP pode ser considerado válido e reprodutível para o mesmo avaliador. Além disso, ambos os testes não são acurados e não são reprodutíveis entre diferentes avaliadores. Como limitações do estudo, identificamos a não realização da reprodutibilidade intra-avaliador com o avaliador experiente, também a não realização da reprodutibilidade interavaliador entre o avaliador A e outro com o mesmo nível de experiência, e também a não realização da concordância entre os avaliadores menos experientes com o instrumento referência (BTS Smart-DX). Com essas análises poderíamos ser mais contundentes acerca das afirmações sobre a relação da experiência entre avaliadores e entre avaliador e a medição tridimensional (cinemetria 3D).

CONCLUSÃO

O Teste de Flexão em Pé (TFP) é válido e reprodutível para o mesmo avaliador, mas é não acurado. O Teste de Flexão Sentado (TFS) não é válido, não é reprodutível e nem acurado. Sugerimos mais estudos que investiguem as propriedades de medida de testes clínicos palpatórios de mobilidade da ASI.

REFERÊNCIAS

- AKOBENG, A.K. Understanding diagnostic tests 3: receiver operating characteristic curves. **Acta paediatrica**, v. 96, n. 5, p. 644-647, 2007.
- ALTMAN, Douglas G. **Practical statistics for medical research**. CRC press, 1990.
- ANJOS, Luiz A. Índice de massa corporal (massa corporal. estatura-2) como indicador do estado nutricional de adultos: revisão da literatura. **Revista de Saúde Pública**, v. 26, p. 431-436, 1992.
- ARAB, Amir Massoud; *et al.* A. Inter-and intra-examiner reliability of single and composites of selected motion palpation and pain provocation tests for sacroiliac joint. **Manual Therapy**, 14(2), 213-221. 2009.
- ÅSTRÖM, Malin; GUMMESSON, Christina. Assessment of asymmetry in pelvic motion—An inter-and intra-examiner reliability study. **The European Journal of Physiotherapy**, 1;16(2). 76-81. 2014.
- BEMIS, Terry; MONTE, Daniel. Validation of the long sitting test on subjects with iliosacral dysfunction. **Journal of Orthopaedic & Sports Physical Therapy**, v. 8, n. 7, p. 336-345, 1987.
- BOSSUYT, Patrick M. *et al.* STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. **Clinical chemistry**, v. 61, n. 12, p. 1446-1452, 2015.
- BOWMAN, Chris; GRIBBLE, Richard. The value of the forward flexion test and three tests of leg length changes in the clinical assessment of movement of the sacroiliac joint. **Journal of Orthopaedic Medicine**. 1, 17(2), p.66-7, 1995.
- BUSSEY, Melanie D; YANAI, Toshimasa; MILBURN, Peter. A non-invasive technique for assessing innominate bone motion. **Clinical Biomechanics**, 19, 85-90, 2004.
- CIBULKA, Michael T.; KOLDEHOFF, Rhonda. Clinical usefulness of a cluster of sacroiliac joint tests in patients with and without low back pain. **Journal of Orthopaedic & Sports Physical Therapy**, v. 29, n. 2, p. 83-92, 1999.
- DREYFUSS, Paul; *et al.* Positive sacroiliac screening tests in asymptomatic adults. **Spine**, v. 19, p. 1138-1143, 1994.
- FLEISS, Joseph L. Measuring nominal scale agreement among many raters. **Psychological bulletin**, v. 76, n. 5, p. 378, 1971.
- FLUSS, Ronen; FARAGGI, David; REISER, Benjamin. Estimation of the Youden Index and its associated cutoff point. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, v. 47, n. 4, p. 458-472, 2005.
- FRYER, Gary; MCPHERSON, Hayley Claire; O'KEEFE, Paul. The effect of training on the inter-examiner and intra-examiner reliability of the seated flexion test and assessment

of pelvic anatomical landmarks with palpation. **International Journal of Osteopathic Medicine**, 1;8(4):131-8: 2005.

FRYER, Gary; MORSE, Christopher M; JOHNSON, Jane C. Spinal and sacroiliac assessment and treatment techniques used by osteopathic physicians in the United States. **Osteopathic medicine and primary care**, 3(1), 4: 2009.

GOGIA, PP; Braatz, JH. Validity and reliability of leg length measurements. **Journal of Orthopaedic & Sports Physical Therapy**, V. 8, n. 4, p.185-188, 1986.

GOODE, Adam et al. Three-dimensional movements of the sacroiliac joint: a systematic review of the literature and assessment of clinical utility. **Journal of Manual & Manipulative Therapy**, v. 16, n. 1, p. 25-38, 2008.

GREENMAN, Ph E. **Principles of manual medicine**. Lippincott Williams & Wilkins, 2003.

GURNEY, Burke. Leg length discrepancy. **Gait & posture**, v. 15, n. 2, p. 195-206, 2002.

JANSE, A. J. et al. Quality of life: patients and doctors don't always agree: a meta-analysis. **Journal of clinical epidemiology**, v. 57, n. 7, p. 653-661, 2004.

KIBSGÅRD, Thomas J. et al. Precision and accuracy measurement of radiostereometric analysis applied to movement of the sacroiliac joint. **Clinical Orthopaedics and Related Research®**, v. 470, n. 11, p. 3187-3194, 2012.

KOTTNER Jan; *et al.* Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. **International journal of nursing studies**, v. 48, n. 6, p. 661-671, 2011.

MAIGNE, Jean-yves; AIVALIKLIS, Alain; PFEFER, Fabrice. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. **Spine**, v. 21, n. 16, p. 1889-1892, 1996.

MALONEY, William J; KEENEY, James A. Leg length discrepancy after total hip arthroplasty. **The Journal of arthroplasty**. Jun 1;19(4):108-10. 2004.

PAYDAR, Dylan; THIEL, H.; GEMMELL, H. Intra-and interexaminer reliability of certain pelvic palpatory procedures and the sitting flexion test for sacroiliac joint mobility and dysfunction. **J Neuromusculoskel Sys**, v. 2, n. 2, p. 65-69, 1994.

POTTER, Nancy A; ROTHSTEIN, Jules M. Intertester reliability for selected clinical tests of the sacroiliac joint. **Journal of Women's Health Physical Therapy**, v. 30, n. 1, p. 21-25, 2006.

PRATHER, Heidi. Pelvis and sacral dysfunction in sports and exercise. **Physical medicine and rehabilitation clinics of North America**, v. 11, n. 4, p. 805-836, 2000.

RIDDLE, Daniel L; FREBURGER, Janet K; NORTH AMERICAN ORTHOPAEDIC REHABILITATION RESEARCH NETWORK. Evaluation of the presence of sacroiliac

joint region dysfunction using a combination of tests: a multicenter intertester reliability study. **Physical Therapy**, 1;82(8):772-81: 2002.

SHAW, J. L. The role of the sacroiliac joint as a cause of low back pain and dysfunction. In: **The First Interdisciplinary World Congress on Low Back Pain and its Relation to the Sacroiliac Joint, San Diego, CA**. 1992. p. 5-6.

SIM, Julius; WRIGHT, Chris C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. **Physical therapy**, v. 85, n. 3, p. 257-268, 2005.

SLIPMAN, Curtis W; *et al.* M. Sacroiliac joint syndrome. **Pain Physician**, 4(2), 143-152, 2001.

STEINBERG, David M.; FINE, Jason; CHAPPELL, Rick. Sample size for positive and negative predictive value in diagnostic research using case-control designs. **Biostatistics**, v. 10, n. 1, p. 94-105, 2009.

STURESSON, Bengt; SELVIK, Göran; UDEN, A. Movements of the sacroiliac joints: a rontgen stereophotogrammetric analysis. **Spine**, v. 14, n. 2, p. 162-165, 1989.

SWETS, John A. Measuring the accuracy of diagnostic systems. **Science**, v. 240, n. 4857, p. 1285-1293, 1988.

TONG, Henry C; *et al.* Interexaminer reliability of three methods of combining test results to determine side of sacral restriction, sacral base position, and innominate bone position. **Journal of the American Osteopathic Association**, 1;106(8):464: 2006.

VINCENT-SMITH, B; GIBBONS, P. Inter-examiner and intra-examiner reliability of the standing flexio test. **Manual Therapy**, 4(2), 87-93: 1999.

VLEEMING, Andry; *et al.* Mobility in the sacroiliac joints in the elderly: a kinematic and radiological study. **Clinical Biomechanics**, v. 7, n. 3, p. 170-176, 1992.

WORLD HEALTH ORGANIZATION. **Obesity: preventing and managing the global epidemic**. World Health Organization, 2000.

APÊNDICE 1 – Quadro de anotação do Teste de Flexão em Pé (TFP) e Teste de Flexão Sentado (TFS).

Teste de Flexão em Pé (TFP) e Teste de Flexão Sentado (TFS)						
CÓDIGO	Posição Inicial TFP	Posição Final TFP	CONCLUSÃO	Posição Inicial TFS	Posição Final TFS	CONCLUSÃO
	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo
	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo
	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo
	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo
	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo
	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo
	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo
	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo	EIPS () D acima () E acima () simétricas	EIPS () D acima () E acima () simétricas	() Positivo a D () Positivo a E () Negativo

5 CAPÍTULO 3

Validade, reprodutibilidade e acurácia diagnóstica do teste de Downing

RESUMO

INTRODUÇÃO: A dor lombar é a principal causa de incapacidade global, sendo que de 13 a 30% dos pacientes com dor lombar referem dor na articulação sacroilíaca (ASI). Um dos testes clínicos disponíveis para a avaliação da mobilidade articular da ASI, é o teste de Downing. Porém, informações acerca da reprodutibilidade, validade e acurácia diagnóstica desse teste são escassas ou desconhecidas na literatura científica. **OBJETIVO:** Avaliar a validade concorrente, a reprodutibilidade intra e interavaliadores e a acurácia diagnóstica do teste de Downing. **METODOLOGIA:** Estudo prospectivo, onde a amostra foi composta por 30 indivíduos de ambos os sexos. A equipe de avaliação foi composta por 5 Fisioterapeutas. As avaliações aconteceram em dois dias diferentes: (1º dia) reprodutibilidade interavaliador e validade concorrente; (2º dia) reprodutibilidade intra-avaliador. O padrão referência utilizado para a validade concorrente foi a medição tridimensional (cinemetria 3D), por meio do sistema BTS Smart-DX. Para análise estatística foi utilizado o percentual de concordância (%C), a estatística kappa (k), a curva ROC (AUC), os valores de sensibilidade e especificidade, além dos valores preditivos positivos (VPP) e negativos (VPN). **RESULTADOS:** Validade concorrente (%C=54,5%; k=0,43; p<0,01), curva ROC para disfunção posterior (AUC=0,889; p<0,05; sensibilidade=78%; especificidade=100%; VPP=100%; VPN=60%), curva ROC para disfunção anterior (AUC=0,657; p=0,37; sensibilidade=71%; especificidade=60%; VPP=71%; VPN=60%), reprodutibilidade intra-avaliador (%C=66,7%; k=0,43; p<0,01) e reprodutibilidade interavaliador (%C=16,7%; k=0,15; p<0,05) **CONCLUSÃO:** O teste de Downing é válido, reprodutível entre o mesmo avaliador e acurado para disfunção posterior. Sugerimos mais estudos que investiguem as propriedades de medida desse teste.

Palavras-chave: Terapia Manual; Fisioterapia; Amplitude de Movimento Articular; Articulação Sacroilíaca; Confiabilidade e Validade.

INTRODUÇÃO

A dor lombar é a principal causa de incapacidade global (GLOBAL BURDEN OF DISEASE, 2016) e é reportada uma prevalência de 13 a 30% de dor na articulação sacroilíaca (ASI) em pacientes com dor lombar (SCHWARZER; APRILL; BOGDUK, 1995; MAIGNE; AIVALIKLIS; PFEFER, 1996; BERNARD; KIRKALDY-WILLIS, 1987). O diagnóstico da ASI é desafiador, pois existem diversas maneiras pelas quais os sintomas podem ocorrer e, possivelmente por essa razão, apesar de a dor na ASI ser uma causa comum de dor lombar, ela costuma ser subvalorizada (COHEN; CHEN; NEUFELD, 2013).

Existem alguns fatores que podem predispor a dor na ASI, como por exemplo: discrepância verdadeira ou aparente do comprimento dos membros inferiores, anormalidades na marcha, exercício vigoroso e prolongado, escoliose, fusão óssea entre coluna lombar e sacro, cirurgia na coluna vertebral e gravidez (COHEN; CHEN; NEUFELD, 2013; COHEN, 2005). Os profissionais de saúde costumam avaliar a ASI de diferentes maneiras: por exames de imagens, como tomografia computadorizada, ressonância magnética e Raios-X, por injeção de bloqueador anestésico intra-articular, ou por testes clínicos de mobilidade articular, provocativos de dor ou posicionais (COHEN; CHEN; NEUFELD, 2013; ARAB et al., 2009).

Um dos testes clínicos disponíveis para a avaliação da mobilidade articular da ASI, é o teste de Downing (DUFFIELD; 1940; BOWMAN; GRIBBLE, 1995; LIMA et al., 2019). O teste tem como objetivo avaliar a mobilidade de rotação anterior ou posterior do osso ilíaco em relação ao sacro. Quando essa mobilidade está alterada pode gerar uma discrepância aparente dos membros inferiores. Essa mobilidade de rotação é explicada por meio das tensões ligamentares da articulação coxo-femoral provocadas por movimentos passivos específicos da mesma (DUFFIELD; 1940; BOWMAN; GRIBBLE, 1995; LIMA et al., 2019). Segundo o raciocínio Osteopático, restrições de mobilidade articular, também chamadas de disfunções articulares, podem acarretar dores e predispor a lesões por meio de compensações do corpo (DREYFUSS *et al.*, 1994; PRATHER, 2000)

Sendo um teste clínico, as investigações da reprodutibilidade, validade e acurácia diagnóstica do teste de Downing são essenciais para que sua aplicabilidade gere resultados confiáveis (KOTTNER et al., 2011; BOSSUYT et al., 2015). Até onde se sabe, a reprodutibilidade do teste de Downing é reportada como pobre (LIMA et al., 2019) ou não melhor que o acaso (BOWMAN; GRIBBLE, 1995). Do mesmo modo, as informações acerca da validade e acurácia diagnóstica desse teste são desconhecidas na literatura científica. Assim

sendo, o objetivo desse estudo foi avaliar a validade concorrente, a reprodutibilidade intra e interavaliadores e a acurácia diagnóstica do teste de Downing. Como hipóteses do estudo, acreditamos que o teste de Downing terá validade e acurácia diagnóstica confirmadas, bem como será reprodutível para avaliações realizadas por um mesmo avaliador e por diferentes avaliadores.

MÉTODOS

Esse estudo está descrito de acordo com o Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (KOTTNER et al., 2011), para as análises de reprodutibilidade, e de acordo com Standards for Reporting of Diagnostic Accuracy (STARD) (BOSSUYT et al., 2015) para a validade concorrente e acurácia diagnóstica. Conforme as recomendações de Kottner et al. (2011), decidimos utilizar dois guias para reportar esse estudo, com a justificativa de que os estudos de acurácia diagnóstica diferem dos estudos de reprodutibilidade ou concordância na utilização de um padrão de referência/ouro.

Para verificar se o teste realmente avalia aquilo que ele se propõe, é necessário analisar a sua validade. Para isso, é preciso utilizar um padrão de referência clínica ou padrão ouro, ou seja, o melhor método disponível para estabelecer a presença ou ausência da condição de destino. Um padrão-ouro seria um padrão de referência sem erros (BOSSUYT et al., 2015), que no nosso estudo assumimos como sendo a medição tridimensional (cinemetria 3D).

Assumimos também que a reprodutibilidade interavaliador é a concordância entre diferentes avaliadores, utilizando a mesma escala, classificação, instrumento ou procedimento, avaliando os mesmos assuntos ou objetos. A reprodutibilidade intra-avaliador, também conhecida como teste e re-teste, é a concordância entre o mesmo avaliador utilizando a mesma escala, classificação, instrumento ou procedimento, avaliando o mesmo sujeito ou objeto em diferentes momentos (Kottner *et al.*, 2011).

DESENHO DO ESTUDO E PARTICIPANTES

Esse é um estudo prospectivo, onde a coleta de dados foi planejada antes da realização das avaliações. A amostra foi composta por indivíduos de ambos os sexos. Os participantes da pesquisa foram voluntários, identificados por meio de convite verbal e virtual, entre os meses de setembro a novembro de 2019. As avaliações ocorreram no setor de Biomecânica do laboratório de pesquisa da Universidade onde o estudo foi realizado. Para as análises de

reprodutibilidade e validade, a amostra foi definida com base em um teste bi-caudal, adotando um poder de 90%, assumindo uma hipótese nula com valor de $kappa = 0,00$ e um $kappa$ detectável de 0,70 com uma proporção de classificações positivas de 0,50, resultando em um número necessário de 22 avaliados (SIM; WRIGHT, 2005). Prevendo 30% de perda amostral, foi adotado uma quantidade alvo de 30 indivíduos. Esse mesmo tamanho amostral também foi assumido para a análise da acurácia diagnóstica.

Os critérios de inclusão foram: (1) indivíduos com idades entre 18 e 60 anos, (2) não obesos ($IMC < 30 \text{ kg/m}^2$) (WORLD HEALTH ORGANIZATION, 2000) e (3) sem cirurgias na lombar, pelve ou quadril. Os critérios de exclusão foram: (1) dor lombar ou no quadril no momento da coleta e (2) mais que 2 cm de diferença de comprimento entre os membros inferiores (MsIs) (GURNEY, 2002).

PROTOCOLO DE AVALIAÇÃO

A equipe de avaliação foi composta por 5 Fisioterapeutas (A, B, C, D e E), dentre eles, um Osteopata (A) e dois estudantes de Osteopatia do 3º ano (B e C). O avaliador A foi considerado o avaliador mais experiente e referência, com 11 anos de prática clínica na realização dos testes do protocolo. Os avaliadores B e C, ambos com 2 anos de experiência na execução dos testes. O pesquisador D foi o responsável pela anamnese e randomização das avaliações, e o pesquisador E pela coleta de dados no sistema BTS Smart-DX. Os avaliadores A, B e C passaram por um treinamento de 20 horas, onde foi abordada a forma de realizar o teste e seus comandos, e testado o protocolo de avaliação completo.

Esse estudo foi aprovado pelo comitê de ética da universidade (CAAE: 15499219.4.0000.5347) e os participantes assinaram o termo de consentimento livre e esclarecido (TCLE). As avaliações ocorreram em dois dias diferentes: (1º dia) reprodutibilidade interavaliador e validade concorrente; (2º dia – após 24 às 72h) reprodutibilidade intra-avaliador. O padrão referência utilizado para a validade concorrente foi a medição tridimensional (cinemetria 3D), por meio do sistema BTS Smart-DX (BTS Bioengineering, EUA).

(1º dia) - Reprodutibilidade interavaliador e validade concorrente

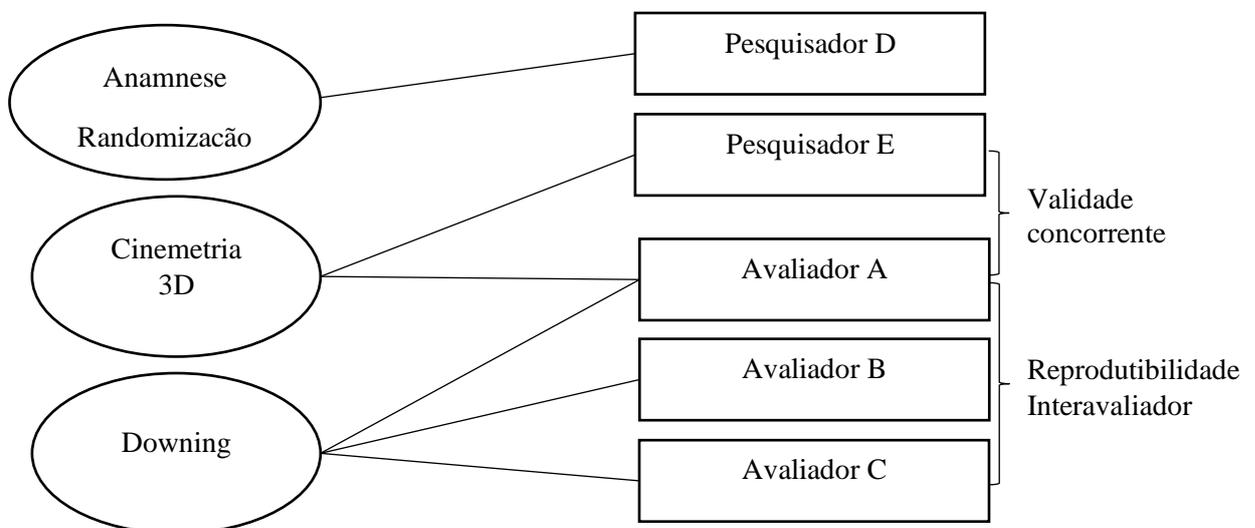
Cada participante da pesquisa foi submetido ao seguinte protocolo de avaliação: (a) Anamnese e randomização da ordem dos avaliadores; (b) Realização do teste de Downing pelos

avaliadores A, B e C (que independentes e cegados realizaram a reprodutibilidade interavaliador); (c) Medição tridimensional (cinemetria 3D) simultânea a realização do teste de Downing. Nesse 1º dia de avaliação (Figura 1), todos os pesquisadores participaram (A, B, C, D e E) da coleta de dados.

A randomização foi feita pelo pesquisador D, o qual efetuou dois sorteios manuais: (1) ordem de avaliação entre os avaliadores B e C e (2) ordem de qual etapa do teste começar (alongamento ou encurtamento) e em qual lado (direito ou esquerdo). O avaliador A realizou sempre duas mensurações, a primeira e a última, portanto, não randomizadas. A primeira foi utilizada para análise da reprodutibilidade e da validade e a última para controle. Na última avaliação, foi realizado somente a mensuração das características posicionais das Espinhas Ilíacas Ântero Superiores (EIASs) e dos Maléolos Mediais (MMs) na posição inicial, onde o avaliado permanecia em decúbito dorsal, com o objetivo de assegurar que as características anatómicas iniciais não modificassem, após as repetições do teste. Caso o participante obtivesse resultados diferentes do teste entre as mensurações feitas pelo avaliador A, a coleta de dados seria interrompida e o participante excluído do estudo.

Cada participante foi submetido a quatro repetições do teste e os avaliadores realizaram a mesma sequência sorteada para as etapas do teste.

Figura 1 - Desenho do 1º dia de avaliação.

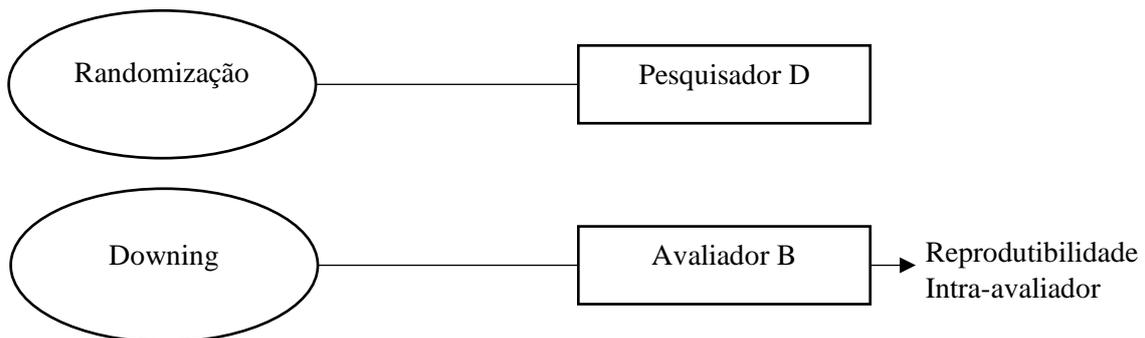


Durante a execução do teste, era orientado ao participante que não resistisse e não auxiliasse os movimentos passivos do avaliador. Ao final do 1º dia, o participante recebeu uma folha lembrete, contendo a data e horário do 2º dia de avaliação e orientações para não realizar atividade física, como por exemplo: academia, alongamentos, corrida; atividades que envolvam esforço físico, como por exemplo: faxina, carregar peso, relação sexual; e tratamentos terapêuticos, como: fisioterapia, osteopatia, quiropraxia.

(2º dia) - Reprodutibilidade intra-avaliador

Nesse 2º dia de avaliação (Figura 2), cada participante da pesquisa foi submetido ao seguinte protocolo de avaliação: (a) Randomização da ordem de qual etapa do teste começar (alongamento ou encurtamento) e em qual lado (direito ou esquerdo), pelo pesquisador D; (b) Realização do teste de Downing pelo avaliador B. O 2º dia de avaliação foi feito após um intervalo de 24 a 72h do 1º dia, foi optado esse tempo de intervalo para minimizar possíveis chances de mudanças nas características anatômicas dos participantes entre os dias de avaliação. O participante era questionado se estava com dor naquele momento e se tinha respeitado as orientações feitas no final do 1º dia.

Figura 2 - Desenho do 2º dia de avaliação.



ANAMNESE

A anamnese foi aplicada no 1º dia de avaliação, pelo pesquisador D. Inicialmente foram coletados dados como data de nascimento, estatura e massa corporal, a partir do auto relato do avaliado. Posteriormente foi feito o cálculo do índice de massa corporal (IMC), por meio da fórmula: $\text{estatura}/\text{massa corporal}^2$ (ANJOS, 1992). Foi questionado também se o indivíduo possuía dor lombar ou no quadril no momento da avaliação e se já fez cirurgia na lombar, pelve ou quadril. Além disso, foi mensurado o comprimento real dos membros inferiores, onde, por

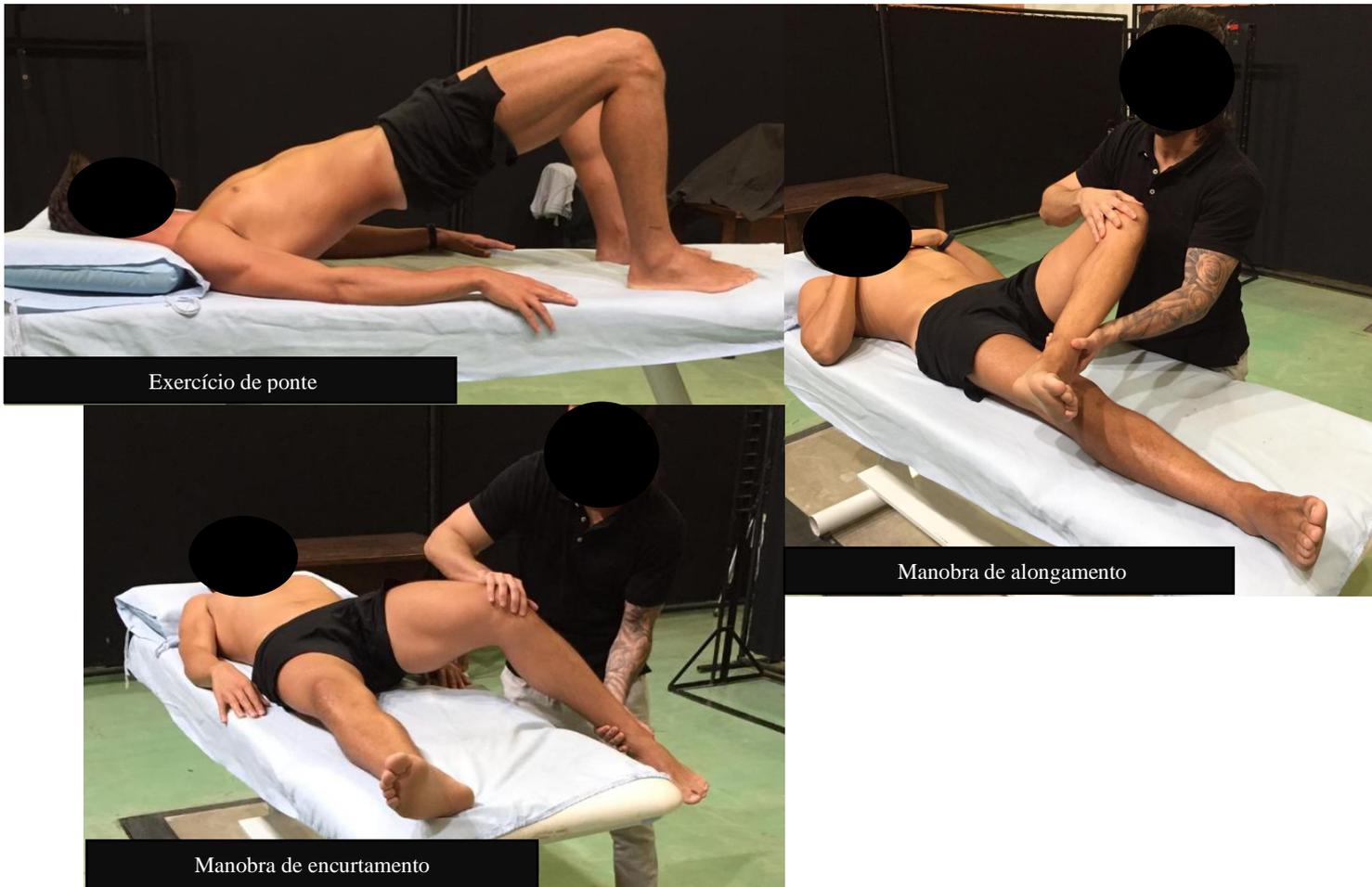
meio de fita métrica, foi mensurada a distância entre a Espinha Ilíaca Antero Superior (EIAS) ao Maléolo Medial (MM) ipsilateral (Maloney e Keeney, 2004). Essa metodologia de medição dos membros inferiores tem sua validação e reprodutibilidade confirmadas (Gogia e Braatz, 1986). Caso o participante não se encaixasse nos critérios de elegibilidade identificados na anamnese, a coleta de dados seria cancelada automaticamente.

TESTE DE DOWNING

A descrição do teste segue as seguintes etapas (DUFFIELD; 1940; BOWMAN; GRIBBLE, 1995; LIMA et al., 2019):

1. Normalização das tensões ligamentares por meio do exercício de ponte em decúbito dorsal (Figura 3a). A volta do movimento, dada pela extensão de quadril e joelho, é feita passivamente pelo avaliador. Por fim, é feita a mensuração qualitativa visual posicional das referências anatômicas.
2. Alongamento direito (D) e esquerdo (E) (Figura 3b): movimento de rotação externa de quadril, após, uma adução de quadril e em seguida uma extensão de quadril e joelho levando o membro inferior em direção à maca. Todos os são movimentos passivos e realizados pelo avaliador. Ao final de cada lado, é feita a mensuração qualitativa visual posicional das referências anatômicas para identificar a efetividade da manobra e a normalização das tensões ligamentares por meio de uma flexão passiva máxima de quadril e joelho.
3. Encurtamento direito (D) e esquerdo (E) (Figura 3): movimento de rotação interna de quadril, após, uma abdução de quadril e em seguida uma extensão de quadril e joelho levando o membro inferior em direção à maca. Os movimentos também são passivos e realizados pelo avaliador. Novamente, ao final de cada lado, é feita a mensuração qualitativa visual posicional das referências anatômicas e a normalização das tensões ligamentares por meio de uma flexão passiva máxima de quadril e joelho.

Figura 3 - Execução do teste de Downing (ponte, manobra de alongamento e manobra de encurtamento).



As referências anatômicas geralmente utilizadas são os MMs (DUFFIELD; 1940; BOWMAN; GRIBBLE, 1995) ou então marcações em linhas na perna (LIMA et al., 2019). Em nosso estudo utilizamos os MMs e, como inovação, as EIAs, com o objetivo de identificar se a pelve possui o mesmo comportamento dos membros inferiores.

A realização do teste de Downing foi dividida em seis etapas, não necessariamente nessa ordem, devido à randomização: (1) posição inicial; (2) alongamento D; (3) encurtamento D; (4) alongamento E; (5) encurtamento E; (6) posição final. A posição inicial é logo após o exercício de ponte, e a posição final é após as manobras de alongamento e encurtamento, e após todas as normalizações das tensões ligamentares. Em ambas as posições, o participante estava em decúbito dorsal, com os membros inferiores estendidos.

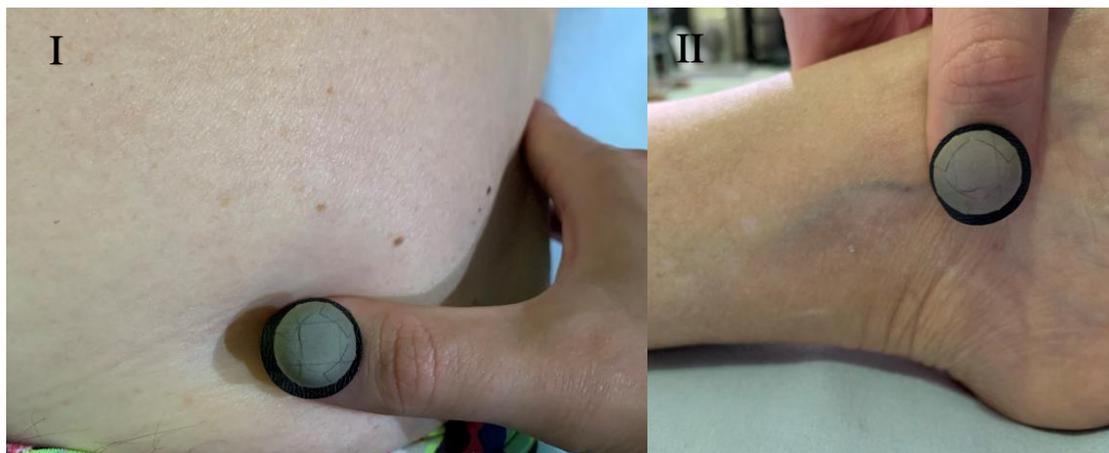
Ao final do teste, somente uma das seguintes conclusões é possível: (1) teste negativo, (2) disfunção anterior do osso Íliaco a D ou a E, e (3) disfunção posterior do osso Íliaco a D ou a

E. Todos os avaliadores preenchem a ficha de anotação desenvolvida para o estudo com as informações de cada etapa e a conclusão final do teste, de cada participante (APÊNDICE 1).

MEDIÇÃO TRIDIMENSIONAL (CINEMETRIA 3D)

Para a validade concorrente do teste de Downing foi assumido como padrão referência a medição tridimensional (cinemetria 3D) com 10 câmeras do sistema BTS Smart-DX. As coletas foram realizadas com uma taxa de amostragem de 100Hz, utilizando o software BTS Smart Capture. Devido à variabilidade existente no processo de retirada e colocação de marcadores anatômicos, neste estudo, os pontos anatômicos de referência no avaliado (EIASs e MMs) foram identificados por meio dos polegares do avaliador durante a palpação. Ao longo do protocolo de coleta, que ocorreu no 1º dia de avaliação, o avaliador A permaneceu com os marcadores refletivos colados com fita dupla face nos polegares (FIGURA 4) para a realização dos testes.

Figura 4 - Marcadores refletivos colados nos polegares do avaliador A, palpando a EIAS (I) e MM (II).



O protocolo criado para esse estudo é baseado no cálculo da diferença linear (em milímetros) da posição média das EIASs direita (D) e esquerda (E) no eixo Z do sistema de coordenadas cartesianas 3D. O eixo Z corresponde ao vetor na direção caudal-cranial do avaliado em decúbito dorsal. Para que esse cálculo fosse possível, foi criado um Sistema de Coordenada Local (SCL) no avaliado, por um cluster de três pontos não colineares utilizando os marcadores dos polegares do avaliador e a referência do chão, tanto para as EIASs quanto para os MMs. Também foi criado outro SCL no avaliador, por um cluster de quatro pontos utilizando uma faixa logo acima dos olhos com quatro marcadores refletivos, conforme

ilustrado na Figura 5. Um SCL é caracterizado por um conjunto de três coordenadas (x, y e z) que se desloca juntamente com o local em que é inserido. Durante a coleta de dados, o avaliador era orientado a manter a angulação da faixa paralela aos segmentos avaliados, para que os SCLs se mantivessem em uma angulação semelhante.

Figura 5 - SCL criado utilizando uma faixa com quatro marcadores refletivos acima dos olhos.



Para análise dos dados, foi utilizado um filtro Butterworth passa baixa, com uma frequência de corte de 1Hz. O trackeamento dos pontos anatômicos foi feito no software BTS Smart Tracker e a análise dos dados no software BTS Smart Analyzer com rotinas criadas especificamente para o teste. Foram feitas doze capturas estáticas por avaliado, seis para as EIAs e seis para os MMs, com duração de 10 segundos cada, conforme as etapas: (1) posição inicial; (2) alongamento D; (3) encurtamento D; (4) alongamento E; (5) encurtamento E; (6) posição final. No momento da análise dos dados, o primeiro segundo inicial e o último segundo final das coletas foram descartados. O resultado do teste de Downing, que será chamado de conclusão do teste, foi obtido por meio da análise de cada etapa em relação a posição inicial.

ANÁLISE ESTATÍSTICA

Os dados da anamnese como idade, estatura, massa corporal e IMC foram submetidos à estatística descritiva. A análise estatística de reprodutibilidade intra-avaliador (BxB), reprodutibilidade interavaliador (AxBxC; AxB; AxC; BxC) e validade concorrente foram

realizadas no software Statistical Package for the Social Sciences (IBM SPSS Statistics 21) e no software Microsoft Office Excel 2016. O nível de significância adotado em todos os testes foi $\leq 0,05$.

Os dados de validade e reprodutibilidade foram analisados utilizando o percentual de concordância (%C) e a estatística kappa (k) (SIM; WRIGHT, 2005; FLEISS, 1971). Para os valores de k, a classificação adotada foi de acordo com Altman (1994) onde valores de $k \leq 0,20$ = pobre; 0,21-0,40 = leve; 0,41-0,60 = moderado; 0,61-0,80 = bom e 0,81-1 = muito bom. Para os valores do %C, foi de acordo com por Janse et al. (2004), onde $\% C \leq 0,30$ = pobre; 0,31-0,50 = fraco; 0,51-0,70 = moderado; 0,71-0,90 = bom; e 0,91-1,00 = excelente. Após todas as análises, adotamos como critério para considerar se os testes seriam válidos e reprodutíveis, um $\%C > 0,50$ (concordância moderada) e um $k > 0,40$ (concordância moderada).

Para determinar a acurácia diagnóstica do teste de Downing foi realizada a análise da curva ROC (curva característica de operação do receptor) e calculada a área sob a curva (AUC) e os valores atingidos de sensibilidade e especificidade (AKOBENG, 2007). Sensibilidade corresponde à proporção de resultados verdadeiro positivos de um teste, ou seja, identificar os participantes com a condição alvo/doença, e especificidade à proporção de resultados verdadeiro negativos, ou seja, identificar os participantes sem a condição alvo/doença, ambos variam de 0 a 1 (0 a 100%) (AKOBENG, 2007). Foi realizado também a análise dos valores preditivos positivos (VPP) e negativos (VPN), que correspondem a proporção de verdadeiros positivos entre todos os indivíduos com teste positivo e a proporção de verdadeiros negativos entre todos os indivíduos com teste negativo, respectivamente. O VPP e VPN expressam a probabilidade de um paciente com o teste positivo ter a doença, e um paciente com teste negativo não ter a doença, respectivamente (STEINBERG, FINE & CHAPPELL, 2009).

Foram atribuídos 5 pontos de corte de 3 a 7 milímetros (mm) para a categorização dos valores oriundos da medição tridimensional (cinemetria 3D). Então, os valores numéricos em mm da cinemetria 3D, foram categorizados a partir desses pontos de corte, em toda a amostra e em todas as etapas do teste de Downing. Esses pontos de corte foram baseados na experiência de realização dos testes e também de acordo com a revisão sistemática de Goode et al. (2008) onde os autores relatam uma mobilidade de translação da articulação sacroilíaca de 4 a 5 mm, com um erro padrão de medida (SEM) de 1,3 mm, durante o movimento de flexão de quadril bilateral.

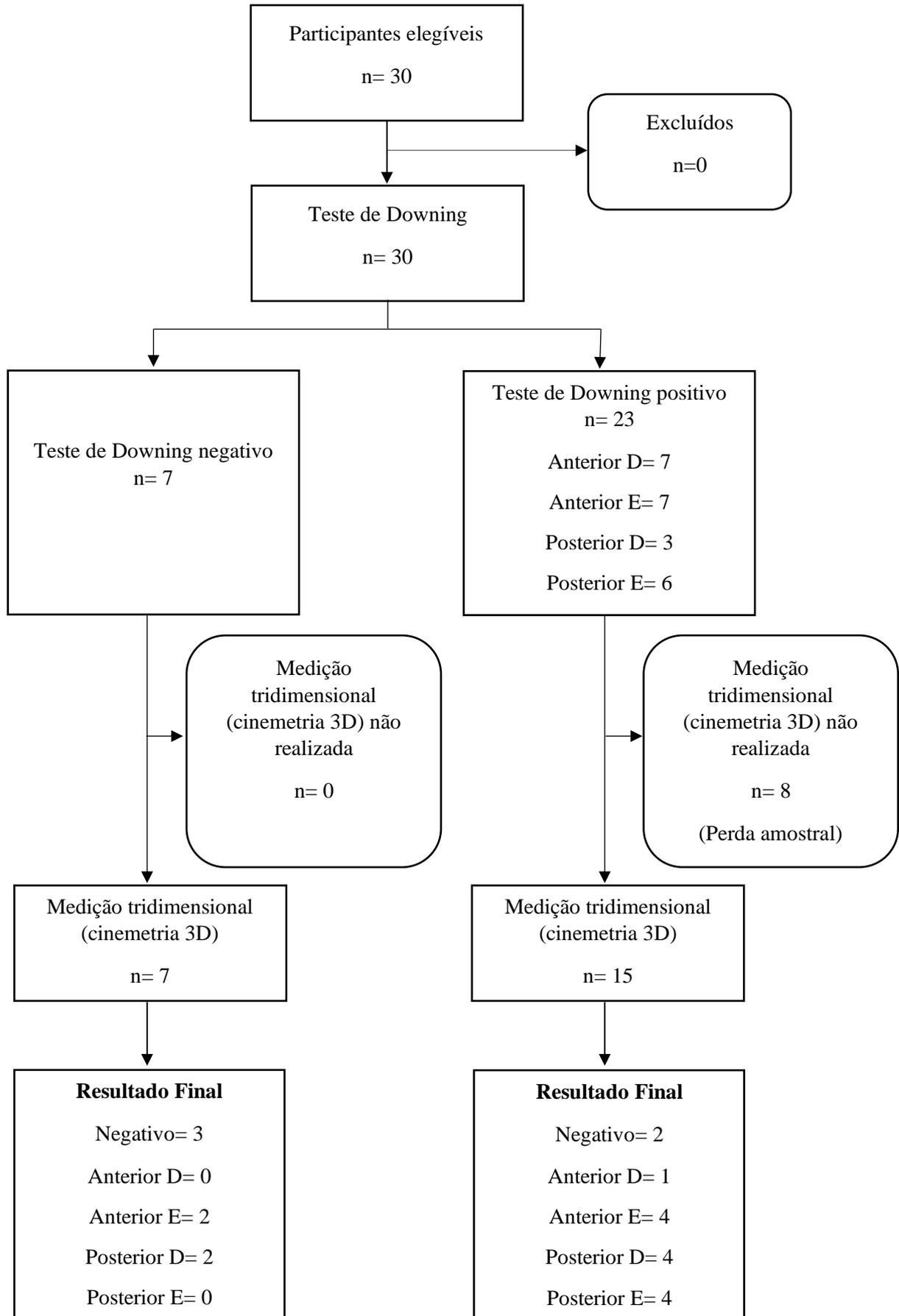
Os dados escalares em milímetros (mm) oriundos da medição tridimensional (cinemetria 3D) foram categorizados em: (1) para a conclusão do teste: negativo, disfunção anterior D ou E, e disfunção posterior D ou E; (2) para a posição inicial e final do teste: EIAS D/E ou MM D/E superior, inferior e simétrico; e (3) para as etapas de alongamento e encurtamento: EIAS D/E ou MM D/E superior, inferior e imóvel. A tabela de anotação (APÊNDICE 1) ilustra as nomenclaturas utilizadas para a categorização dos dados. Para os valores de acurácia diagnóstica pela AUC foi utilizada a classificação de acordo com Swets (1988), onde valores $> 0,9$ = alta acurácia; entre $0,7-0,9$ = moderada acurácia; $0,5-0,7$ = baixa acurácia e $< 0,5$ = resultado casual. Em nosso estudo, assumimos que para que o teste seja considerado acurado, deveria apresentar $AUC > 0,7$ (moderada acurácia).

RESULTADOS

A amostra foi composta por 30 indivíduos, 17 do sexo masculino e 13 do sexo feminino, com idade média de $30 (\pm 8)$ anos, estatura média de $171 (\pm 9)$ cm, massa média de $71 (\pm 12)$ kg e índice de massa corporal (IMC) médio de $24 (\pm 2)$ kg/cm². Não houve nenhum indivíduo excluído por apresentar uma diferença no comprimento dos membros inferiores >2 cm, por dor no momento da coleta ou por modificar as características anatômicas na mensuração de controle do avaliador A.

Não houve resultados do teste de Downing inconclusivos, mas houve perda amostral de 8 avaliados na medição tridimensional (cinemetria 3D), devido a falha do sinal durante a análise dos dados. Esses avaliados foram excluídos das análises de validação. A Figura 6 mostra o fluxo de avaliados pelo estudo.

Figura 6 - Diagrama de fluxo referente ao teste de Downning (avaliador A) e a medição tridimensional (cinemetria 3D), com base no ponto de corte de 3 milímetros, considerando as avaliações pelos MMs, para a conclusão do teste.



A Tabela 1 trata sobre as informações descritivas (média, mediana, desvio padrão, valor mínimo e máximo) em milímetros (mm) das avaliações com a medição tridimensional (cinemetria 3D), com base nos MMs e nas EIAs, referente ao 1º dia de avaliação, de acordo com as etapas: posição inicial, alongamento a D, encurtamento a D, alongamento a E, encurtamento a E e posição final. É possível observar que as EIAs possuem uma maior amplitude em mm de assimetria (posição inicial e final) e de deslocamento (alongamento e encurtamento) em comparação aos MMs.

Tabela 1 – Informações descritivas das avaliações com a medição tridimensional (cinemetria 3D), com base nos MMs e nas EIAs, referente ao 1º dia de avaliação.

MMs						
	Posição inicial^b	Alongamento D^b	Encurtamento D^b	Alongamento E^{nb}	Encurtamento E^b	Posição Final^b
	(mm)	(mm)	(mm)	(mm)	(mm)	(mm)
Média	5	3	5	5	4	6
Mediana	5	3	3	3	2	4
DP	3	2	5	5	4	7
Mínimo	0	0	0	0	0	0
Máximo	12	9	20	16	12	25

EIAs						
	Posição inicial^b	Alongamento D^b	Encurtamento D^{nb}	Alongamento E^{nb}	Encurtamento E^b	Posição Final^{nb}
	(mm)	(mm)	(mm)	(mm)	(mm)	(mm)
Média	9	9	11	13	8	8
Mediana	9	7	10	11	7	6
DP	8	9	8	7	7	7
Mínimo	0	0	0	1	0	1
Máximo	34	36	27	31	31	35

^b: dados paramétricos; ^{nb}: dados não paramétricos; DP: desvio padrão; mm: milímetros; EIAs: espinhas ilíacas ântero superiores; MMs: maléolos mediais; D: direita; E: esquerda.

A Tabela 2 retrata os resultados de validade concorrente entre o avaliador A e a medição tridimensional (cinemetria 3D), para os 5 pontos de corte (3, 4, 5, 6 e 7 mm) e para toda as etapas (posição inicial, alongamento D, alongamento E, encurtamento D, encurtamento E, posição final e conclusão do teste), referente ao 1º dia de avaliação. Em geral, a concordância k variou de pobre a leve, e o %C variou de fraco a moderado. De acordo com as avaliações pelos MMs, para a conclusão do teste, a concordância (k e %C) foi moderada nos pontos de corte 3, 5 e 6 mm. Os melhores valores de concordância (k e %C) foi no ponto de corte de 3 mm pelas avaliações feitas com base no MMs.

Tabela 2 – Validade concorrente entre avaliador A e a medição tridimensional (cinemetria 3D) do teste de Downing, com base nas mensurações dos MMs e das EIAsSs.

		<i>Teste de Downing</i>						
		<i>MMs (n= 22)</i>						
		Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
3 mm	n (%C)	9 (40,9%)	9 (40,9%)	18 (81,8%)	14 (63,6%)	13 (59,1%)	13 (59,1%)	12 (54,5%)
	k (p)	0,11 (0,42)	0,06 (0,66)	0,64 (<0,01)*	0,30 (0,10)	0,29 (0,07)	0,31 (<0,05)*	0,43 (<0,01)*
4 mm	n (%C)	10 (45,5%)	8 (36,4%)	15 (68,2%)	12 (54,5%)	12 (54,5%)	15 (68,2%)	11 (50%)
	k (p)	0,11 (0,46)	0,06 (0,53)	0,38 (<0,05)*	0,19 (0,24)	0,23 (0,13)	0,41 (<0,01)*	0,36 (<0,01)*
5mm	n (%C)	12 (54,5%)	6 (27,3%)	14 (63,6%)	11 (50%)	12 (54,5%)	13 (59,1%)	12 (54,5%)
	k (p)	0,21 (0,17)	-0,01 (0,85)	0,30 (0,10)	0,14 (0,35)	0,23 (0,13)	0,20 (0,21)	0,41 (<0,01)*

6mm	n (%C)	11 (50%)	6 (27,3%)	14 (63,6%)	11 (50%)	13 (59,1%)	13 (59,1%)	12 (54,5%)
	k (p)	0,05 (0,75)	-0,01 (0,85)	0,30 (0,10)	0,14 (0,35)	0,27 (0,07)	0,20 (0,21)	0,41 (<0,01)*
7 mm	n (%C)	12 (54,5%)	5 (22,7%)	14 (63,6%)	11 (50%)	12 (54,5%)	12 (54,5%)	11 (50%)
	k (p)	0,08 (0,62)	-0,05 (0,10)	0,30 (0,10)	0,14 (0,35)	0,20 (0,12)	0,08 (0,62)	0,34 (<0,01)*

EIASs (n= 28)

		Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
3 mm	n (%C)	7 (25%)	14 (50%)	15 (53,6%)	19 (67,9%)	12 (42,9%)	10 (35,7%)	11 (39,3%)
	k (p)	-0,14 (0,27)	0,12 (0,39)	0,20 (0,16)	0,10 (0,51)	0,02 (0,88)	0,02 (0,85)	0,24 (<0,01)*
4 mm	n (%C)	9 (32,1%)	14 (50%)	17 (60,7%)	19 (67,9%)	14 (50%)	10 (35,7%)	13 (46,4%)
	k (p)	-0,07 (0,58)	0,11 (0,45)	0,32 (<0,05)*	0,10 (0,51)	0,17 (0,22)	-0,02 (0,90)	0,33 (<0,01)*

5mm	n (%C)	9 (32,1%)	14 (50%)	17 (60,7%)	18 (64,3%)	13 (46,4%)	12 (42,9%)	12 (42,9%)
	k (p)	-0,07 (0,58)	0,11 (0,45)	0,32 (<0,05)*	0,04 (0,80)	0,12 (0,38)	0,06 (0,66)	0,28 (<0,01)*
6mm	n (%C)	10 (35,7%)	13 (46,4%)	18 (64,3%)	18 (64,3%)	13 (46,4%)	10 (35,7%)	12 (42,9%)
	k (p)	-0,04 (0,78)	0,07 (0,61)	0,36 (<0,05)*	0,11 (0,51)	0,12 (0,38)	-0,11 (0,44)	0,28 (<0,01)*
7 mm	n (%C)	10 (35,7%)	14 (50%)	18 (64,3%)	19 (67,9%)	16 (57,1%)	11 (39,2)	14 (50%)
	k (p)	-0,04 (0,78)	0,02 (0,90)	0,36 (<0,05)*	0,23 (0,19)	0,22 (0,16)	-0,07 (0,62)	0,37 (<0,01)*

*: $p \leq 0,05$; p: nível de significância; n: número de participantes; k: kappa; %C: percentual de concordância; EIASs: espinhas ilíacas ântero superiores; MMs: maléolos mediais; D: direita; E: esquerda.

A Tabela 3 apresenta os resultados de concordância entre as avaliações feitas com base nos MMs e nas EIAs para a medição tridimensional (cinemetria 3D) pelo avaliador A, para os 5 pontos de corte (3, 4, 5, 6 e 7 mm) em todas as etapas (posição inicial, alongamento D, alongamento E, encurtamento D, encurtamento E, posição final e conclusão do teste), referente ao 1º dia de avaliação. Em geral a concordância (k e %C) variou de pobre a fraca, somente para a etapa de encurtamento a D a concordância foi moderada.

Tabela 3 - Concordância entre as avaliações feitas com base nos MMs e nas EIAs para a medição tridimensional (cinemetria 3D), feitas pelo avaliador A.

<i>Teste de Downing</i>								
<i>MMs x EIAs</i>								
		Posição inicial n= 25	Alongamento D n= 24	Encurtamento D n= 23	Alongamento E n= 23	Encurtamento E n= 23	Posição Final n= 24	Conclusão do teste n= 22
3 mm	n (%C)	10 (40%)	9 (37,5%)	13 (56,5%)	12 (52,2%)	9 (39,1%)	7 (29,2%)	7 (31,8%)
	k (p)	0,12 (0,35)	-0,01 (0,97)	0,23 (0,17)	0,07 (0,59)	0,08 (0,53)	-0,12 (0,40)	0,11 (0,33)
4 mm	n (%C)	10 (40%)	10 (41,7%)	16 (69,6%)	9 (39,1%)	10 (43,5%)	11 (45,8%)	4 (18,2%)

5mm	k (p)	0,08 (0,57)	-0,05 (0,70)	0,47 (<0,01)*	0,01 (0,96)	0,12 (0,39)	0,07 (0,64)	-0,01 (0,90)
	n (%C)	9 (36%)	10 (41,7%)	15 (65,2%)	9 (39,1%)	9 (39,1%)	13 (54,2%)	3 (13,6%)
6mm	k (p)	-0,02 (0,90)	-0,05 (0,57)	0,40 (<0,01)*	0,03 (0,80)	0,02 (0,86)	0,09 (0,60)	-0,07 (0,47)
	n (%C)	10 (40%)	11 (45,8%)	16 (69,6%)	11 (47,8%)	10 (43,5%)	15 (62,5%)	3 (13,6%)
7 mm	k (p)	-0,01 (0,93)	-0,04	0,44 (<0,01)*	0,12 (0,40)	0,07 (0,57)	0,20 (0,22)	-0,08 (0,40)
	n (%C)	12 (48%)	14 (58,3%)	16 (69,6%)	12 (52,2%)	11 (47,8%)	15 (62,5%)	3 (13,6%)
	k (p)	0,10 (0,37)	0,05 (0,34)	0,44 (<0,01)*	0,17 (0,27)	-0,05 (0,70)	0,12 (0,44)	-0,09 (0,37)

*: $p \leq 0,05$; p: nível de significância; n: número de participantes; k: kappa; %C: percentual de concordância; EIAs: espinhas ilíacas ântero superiores; MMs: maléolos mediais; D: direita; E: esquerda.

A Tabela 4 trata sobre a análise da curva ROC entre o avaliador A e a medição tridimensional (cinemetria 3D), com base no MMs, para o ponto de corte de 3 mm, o qual obteve os melhores valores de concordância, em toda as etapas (posição inicial, alongamento D, alongamento E, encurtamento D, encurtamento E, posição final e conclusão do teste). A conclusão do teste foi dividida em disfunção anterior e disfunção posterior, juntamente com os participantes classificados com teste negativo, pois essa análise estatística requerer variáveis dicotômicas. O teste de Downing é considerado acurado ($AUC > 0,700$) somente para a disfunção posterior.

Tabela 4 – Análise da curva ROC entre o avaliador A e a medição tridimensional (cinemetria 3D), com base no MMs, para o ponto de corte de 3 mm.

<i>Teste de Downing</i>		
3mm - MMs		
	Conclusão do teste Disfunção Anterior n= 12	Conclusão do teste Disfunção Posterior n= 12
AUC	0,657	0,889
Erro Padrão	0,168	0,097
p	0,37	$\leq 0,05^*$
IC 95%	0,328-0,986	0,698-1,000
Sensibilidade	71%	78%
Especificidade	60%	100%
VPP	71%	100%
VPN	60%	60%

*: $p \leq 0,05$; p: nível de significância; n: número de participantes; MMs: maléolos mediais; D: direita; E: esquerda; VPP: valor preditivo positivo; VPN: valor preditivo negativo.

A Tabela 5 apresenta os resultados de reprodutibilidade intra-avaliador (BxB), com base nas mensurações das EIAs e dos MMs, em todas as etapas (posição inicial, alongamento D, alongamento E, encurtamento D, encurtamento E, posição final e conclusão do teste), referente ao 2º dia de avaliação, em um total de 30 indivíduos. Em geral, é possível observar %C bom, tanto para a avaliação pelas EIAs como pelos MMs. Em relação a estatística k, é possível observar concordância moderada em relação aos MMs na maioria das etapas, e concordância variando de pobre a boa em relação as EIAs.

Tabela 5 – Reprodutibilidade intra-avaliador (BxB) do teste de Downing, com base nas mensurações dos MMs e das EIAs.

<i>Teste de Downing (BxB)</i>							
<i>MMs (n= 30)</i>							
	Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
n (%C)	23 (76,7%)	26 (86,7%)	25 (83,3%)	30 (100%)	25 (83,3%)	22 (73,3%)	20 (66,7 %)
k (p)	0,58 (<0,01)*	0,43 (<0,05)*	-0,06 (0,69)	1,00 (<0,01)*	0,56 (<0,01)*	0,51 (<0,01)*	0,43 (<0,01)*
<i>EIAs (n= 30)</i>							
	Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
n (%C)	24 (80%)	22 (73,3%)	19 (63,3%)	23 (76,7%)	20 (66,7%)	26 (86,7%)	12 (40%)
k (p)	0,56 (<0,01)*	0,40 (<0,05)*	0,20 (0,29)	0,09 (0,63)	0,32 (0,08)	0,68 (<0,01)*	0,16 (0,11)

*: $p \leq 0,05$; p: nível de significância; n: número de participantes; k: kappa; %C: percentual de concordância; EIAs: espinhas ilíacas ântero superiores; MMs: maléolos mediais; D: direita; E: esquerda.

A Tabela 6 apresenta os resultados de reprodutibilidade interavaliador (AxBxC), com base nas mensurações das EIAS e dos MMs, em todas as etapas (posição inicial, alongamento D, alongamento E, encurtamento D, encurtamento E, posição final e conclusão do teste), referente ao 1º dia de avaliação, em um total de 30 indivíduos. Em geral, a concordância (%C e k) é pobre na maioria das etapas, tanto para as avaliações feitas pelas EIASs como pelos MMs.

Tabela 6 – Reprodutibilidade interavaliador (AxBxC) do teste de Downing, com base nas mensurações dos MMs e das EIAsS.

<i>Teste de Downing (AxBxC)</i>							
<i>MMs (n= 30)</i>							
	Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
n (%C)	1 (3,3%)	18 (60%)	11 (36,7%)	21 (70%)	8 (26,7%)	2 (6,7%)	5 (16,7%)
k (p)	-0,04 (0,60)	0,13 (0,22)	-0,03 (0,79)	0,07 (0,52)	-0,04 (0,70)	-0,02 (0,81)	0,15 (<0,05)*
<i>EIASs (n= 30)</i>							
	Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
n (%C)	5 (16,7%)	12 (40%)	10 (33,3%)	19 (63,3%)	7 (23,3%)	5 (16,7%)	1 (3,3%)
k (p)	-0,04 (0,55)	-0,02 (0,83)	0,05 (0,60)	0,07 (0,51)	-0,04 (0,70)	-0,03 (0,68)	0,01 (0,81)

.*: $p \leq 0,05$; p: nível de significância; n: número de participantes; k: kappa; %C: percentual de concordância; EIAsS: espinhas ilíacas ântero superiores; MMs: maléolos mediais; D: direita; E: esquerda.

A Tabela 7 apresenta os resultados de reprodutibilidade interavaliador (AxB; AxC; BxC) do teste de Downing com base nas mensurações das EIAS e dos MMs, em todas as etapas (posição inicial, alongamento D, alongamento E, encurtamento D, encurtamento E, posição final e conclusão do teste), referente ao 1º dia de avaliação, em um total de 30 indivíduos.

Tanto para as avaliações pelas EIASs como pelos MMs, na maioria dos casos, observando pelo %C, a concordância variou de fraca a moderada entre os avaliadores AxB, e pobre a moderada entre os avaliadores AxC. Para ambas as combinações de avaliadores, em geral, a concordância foi pobre pela estatística k. O mesmo comportamento anterior segue para os resultados entre os avaliadores BxC, porém somente para as avaliações feitas pelas EIASs. Quando são feitas com base nos MMs, é possível observar %C variando de fraco a bom, e concordância k variando de pobre a moderada.

Tabela 7 – Reprodutibilidade interavaliador (AxB; AxC; BxC) do teste de Downing, com base nas mensurações dos MMs e das EIAsS.

<i>Teste de Downing (AxB)</i>							
<i>MMs (n= 30)</i>							
	Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
n (%C)	12 (40%)	20 (66,7%)	16 (53%)	22 (73,3%)	20 (66,7%)	12 (40%)	11 (36,7%)
k (p)	0,02 (0,90)	-0,03 (0,85)	0,11 (0,35)	0,15 (0,12)	0,28 (0,09)	-0,01 (0,97)	0,19 (<0,05)*
<i>EIASs (n= 30)</i>							
	Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
n (%C)	12 (40%)	15 (50%)	18 (60%)	21 (70%)	21 (70%)	14 (46,7%)	9 (30%)
k (p)	-0,06 (0,63)	-0,14 (0,42)	0,23 (0,18)	0,15 (0,35)	0,38 (<0,05)*	0,01 (0,92)	0,13 (0,15)

Teste de Downing (AxC)

MMs (n= 30)

	Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
n (%C)	10 (33,3%)	20 (66,7%)	12 (40%)	22 (73,3%)	9 (30%)	10 (33,3%)	7 (23,3%)
k (p)	-0,03 (0,82)	-0,03 (0,85)	-1,15 (0,27)	0,15 (0,12)	-0,41 (<0,05)*	-0,03 (0,82)	0,02 (0,79)

EIASs (n= 30)

	Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
n (%C)	9 (30%)	19 (63,3%)	13 (43,3%)	22 (73,3%)	10 (33,3%)	9 (30%)	6 (20%)
k (p)	-0,09 (0,49)	-0,07 (0,71)	-0,05 (0,71)	0,15 (0,12)	-0,35 (0,05)	-0,09 (0,49)	-0,02 (0,86)

Teste de Downing (BxC)

MMs (n= 30)

	Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
n (%C)	9 (30%)	26 (86,7%)	24 (80%)	28 (93,3%)	17 (56,7%)	11 (36,7%)	15 (50%)
k (p)	-0,03 (0,76)	0,52 (<0,01)*	0,29 (0,11)	-0,03 (0,85)	0,10 (0,53)	0,04 (0,77)	0,30 (<0,01)*

EIASs (n= 30)

	Posição inicial	Alongamento D	Encurtamento D	Alongamento E	Encurtamento E	Posição Final	Conclusão do teste
n (%C)	12 (40%)	20 (66,7%)	19 (63,3%)	25 (83,3%)	13 (43,3%)	12 (40%)	7 (23,3%)
k (p)	0,06 (0,63)	0,19 (0,24)	0,13 (0,45)	-0,06 (0,69)	-0,14 (0,43)	0,05 (0,68)	-0,03 (0,79)

*: $p \leq 0,05$; p: nível de significância; n: número de participantes; k: kappa; %C: percentual de concordância; EIASs: espinhas ilíacas ântero superiores; MMs: maléolos mediais; D: direita; E: esquerda.

DISCUSSÃO

O atual estudo dividiu o teste de Downing em seis etapas: posição inicial, alongamento a D, encurtamento a D, alongamento a E, encurtamento a E e posição final. De posse dessas etapas, foi possível obter a conclusão do teste, ou seja, o seu diagnóstico final sobre o comportamento do osso Ilíaco. Classicamente os referenciais anatômicos utilizados para a realização do teste de Downing são os Maléolos Mediais (MMs) (DUFFIELD; 1940; BOWMAN; GRIBBLE, 1995). Contudo, nesse estudo, acrescentamos as Espinhas Ilíacas Ântero Superiores (EIASs) como referenciais anatômicos adicionais, com a justificativa de que o teste de Downing é uma mensuração da mobilidade da ASI, e então as EIASs estariam mais próximas da articulação alvo, compondo o osso Ilíaco.

Para as informações descritivas acerca das avaliações com a medição tridimensional (cinemetria 3D) (Tabela 1), é possível observar que as avaliações feitas com base nas EIASs possuem uma maior assimetria média, dada pela posição inicial e final, e maior deslocamento médio, dado pelas etapas de alongamento e encurtamento. Além disso, a mediana, o desvio padrão e os valores máximos também são maiores nas avaliações pelas EIASs em comparação as mensurações pelos MMs. Esses dados indicam que as EIASs possuem uma maior amplitude de assimetria e deslocamento em relação aos MMs quando é feito o teste de Downing. Uma hipótese é de que os MMs estariam mais distantes da ASI, em comparação às EIASs, sofrendo influências da articulação do quadril, joelho e tornozelo durante a execução do teste.

A avaliação da validade concorrente (Tabela 2) entre o avaliador A, considerado experiente, e a medição tridimensional (cinemetria 3D), demonstra uma diferença entre as avaliações feitas com base nos MMs e nas EIASs, onde é possível constatar que as concordâncias são melhores, para todas as etapas do teste, quando os referenciais anatômicos utilizados são os MMs. Uma hipótese viável é que os avaliadores do atual estudo e, nesse caso, principalmente o avaliador A, estão habituados a realizarem o teste de Downing na prática clínica de acordo com os MMs, e não com as EIASs. Em relação aos resultados com base nos MMs, é possível observar um %C moderado (51-70%) a bom (71-90%) para a posição inicial em 5 e 7 mm, para o encurtamento a D em todos os pontos de corte, para o alongamento a E em 3 e 4 mm, para o encurtamento a E e posição final em todos os pontos de corte, e para a conclusão do teste em 3, 5 e 6 mm. Ainda em relação as avaliações pelos MMs, porém analisando os resultados de concordância pela estatística kappa (k), é possível observar concordância moderada ($k=0,41-0,60$) a boa ($k=0,61-0,80$) para o encurtamento a D em 3 mm,

para a posição final em 4 mm, e para a conclusão do teste em 3, 5 e 6 mm. Também para os resultados de validade concorrente (Tabela 2), porém em relação as avaliações pelas EIAs, é possível observar um %C moderado (51-70%) para o encurtamento a D e alongamento a E em todos os pontos de corte, e para o encurtamento a E em 7 mm. Analisando os resultados pela estatística k, não houve concordância melhor que leve ($k \leq 0,37$) para nenhuma etapa e em nenhum ponto de corte.

A partir dos resultados anteriores, considerando a conclusão do teste, a validade concorrente do teste de Downing foi confirmada somente para as avaliações com base nos MMs ($k > 0,40$ e $\%C > 50\%$). O ponto de corte que apresentou o conjunto de melhores valores de concordância foi o de 3 mm. As avaliações com base nas EIAs, não se mostraram válidas, obtendo uma concordância leve para a conclusão do teste. Ainda, as etapas que apresentaram os melhores valores de concordância foram as de encurtamento, principalmente a D. Não sabemos a explicação desse fato, mas temos como hipótese que a manobra de encurtamento, que envolve rotação interna, abdução e extensão de quadril, possa ser mais fácil de ser realizada em comparação a manobra de alongamento, a qual envolve rotação externa, adução e extensão de quadril. Isso porque, fisiologicamente, o movimento de rotação interna de quadril possui menor amplitude de movimento em comparação com a rotação externa e, dessa maneira, o avaliador tem mais facilidade para abranger todo o arco de movimento durante a manobra. Até onde sabemos, não há estudos que tenham analisado a validade do teste de Downing e, tampouco, que tenham dividido a análise do teste de Downing em etapas (posição inicial e final, alongamento e encurtamento a D e a E). Portanto, entendemos como necessária a condução de novos estudos para confrontar os atuais achados.

Ao analisar a concordância entre as avaliações feitas com base nos MMs e nas EIAs (Tabela 3), em geral o %C foi pobre ($\leq 30\%$) a moderado (51-70%) e a estatística k foi pobre ($k \leq 0,20$), exceto para a etapa de encurtamento a D onde a concordância foi leve ($k = 0,23-0,40$) em 3 e 5 mm, e moderada ($k = 0,44-0,47$) em 4, 6 e 7 mm. Assim sendo, as avaliações baseadas nos MMs e nas EIAs não possuem concordância aceitável durante a maioria das etapas do teste de Downing, bem como para a conclusão do teste. Esses achados indicam que o profissional, ao realizar o teste, não deve ter as EIAs como parâmetro para a avaliação e conclusão do diagnóstico. Até onde se tem conhecimento, não há na literatura outros estudos que fazem análise semelhante entre essas referências anatômicas com o teste de Downing ou então avaliando a mobilidade da ASI.

Para realizar a análise da acurácia diagnóstica do teste de Downing (Tabela 4), foi utilizado o ponto de corte com o conjunto de melhores valores de concordância e os referenciais anatômicos com os melhores valores de validade concorrente. Dessa maneira, foi realizada a análise de curva ROC com base nos MMs em 3 mm. Em vista disso, o teste de Downing é considerado acurado ($AUC > 0,700$) somente para a disfunção posterior ($AUC = 0,889$; $EP = 0,097$; $p \leq 0,05$; $IC\ 95\% = 0,698-1,000$), mostrando uma sensibilidade de 78%, especificidade de 100%, valor preditivo positivo de 100% e valor preditivo negativo de 60%. O mesmo não acontece para a disfunção anterior ($AUC = 0,657$; $EP = 0,168$; $p = 0,37$; $IC\ 95\% = 0,328-0,986$), tendo uma sensibilidade e valor preditivo positivo de 71%, e especificidade e valor preditivo negativo de 60%. O ponto de corte de 3 mm aponta que basta uma diferença sutil entre os MMs para o profissional discriminar entre um indivíduo com ou sem disfunção da ASI, indicando que durante a realização do teste de Downing, o executante deve estar atento para mínimos deslocamentos e assimetrias. Não há na literatura outros estudos que fizeram a análise da acurácia diagnóstica desse teste.

Referente a reprodutibilidade intra-avaliador (BxB) (Tabela 5), em geral o %C foi bom ($\%C = 71-90\%$) e a concordância foi moderada ($k = 0,41-0,60$) pela estatística k. Exceto, para as avaliações pelas EIASs, de acordo com os valores de k, onde a concordância variou de pobre ($k \leq 0,20$) a leve ($k \leq 0,40$) na maiorias das etapas. Desse modo, o teste de Downing se mostra reprodutível para avaliações feitas pelo mesmo avaliador, tendo como referencial anatômico somente os MMs. Lima *et al.* (2019) realizaram também a análise da reprodutibilidade intra-avaliador do teste de Downing por um avaliador fisioterapeuta, treinado e experiente na realização do teste, em indivíduos sintomáticos, com dor lombar ou sacroilíaca, e indivíduos assintomáticos. Os autores encontraram concordância pobre para a amostra total ($k = 0,12$) e para os indivíduos assintomáticos ($k = 0,18$), e concordância leve ($k = 0,28$) para os indivíduos sintomáticos. Uma diferença desse estudo para o nosso, é que os autores utilizaram uma marcação na perna como referência anatômica, e não os MMs. Alguns fatores podem dificultar a reprodutibilidade do teste de Downing, como a experiência do avaliador, o treinamento para a realização das manobras de encurtamento e alongamento corretamente, assim como a palpação e a percepção de pequenos deslocamentos após cada manobra das referências anatômicas. Para conclusões mais assertivas acerca disso, permanece então a necessidade de mais estudos investigando a reprodutibilidade intra-avaliador do teste de Downing.

De acordo com a reprodutibilidade interavaliador entre os três avaliadores (AxBxC) (Tabela 6), para as avaliações feitas com base nos MMs, em geral o %C foi pobre ($\%C \leq 30\%$),

exceto para o encurtamento a D, onde foi fraco (%C= 36,7%), e para o alongamento a D e a E, que foi moderado (%C= 60-70%). Com relação as avaliações pelas EIAs, os resultados são semelhantes aos anteriores, em geral o %C foi pobre, exceto para o alongamento e encurtamento a D, onde foi fraco (%C= 33,3-40%), e para o alongamento a E, onde foi moderado (63,3%). Em relação a estatística k, a concordância foi pobre para todas as etapas ($k \leq 0,20$), tanto para os MMs como para as EIAs. Bowman e Gribble (1995) realizaram também a análise da reprodutibilidade interavaliador, com um osteopata experiente e dois estudantes de osteopatia, do teste de Downing com base nos MMs. Os autores encontraram uma concordância não melhor que a esperada pelo acaso ($k = 0$). Os fatores dificultantes do atual teste citados anteriormente, se repetem para a reprodutibilidade interavaliador, contudo parecem ser intensificados quando o avaliador não é o mesmo. À vista disso, o teste de Downing parece não ser um teste reprodutível entre diferentes avaliadores, carecendo mais estudos que façam tal análise.

Os resultados são semelhantes para a reprodutibilidade interavaliador por pares (AxB) (Tabela 7). Em geral, o %C variou de fraco (31-50%) a moderado (51-70%) e referente a estatística k, a concordância foi pobre ($k \leq 0,20$). Os resultados são semelhantes para a reprodutibilidade interavaliador (AxC) (Tabela 7), na maioria dos casos o %C foi pobre ($\%C \leq 30\%$) a fraco (31-50%) e a concordância foi pobre ($k \leq 0,20$) para todas as etapas, de acordo com a estatística k. Em relação a reprodutibilidade interavaliador por pares (BxC) (Tabela 7), entre ambos avaliadores considerados menos experientes em relação ao avaliador A, para as avaliações utilizando as EIAs, os resultados foram semelhantes aos anteriores. Contudo, de acordo com as avaliações pelos MMs, em comparação as análises anteriores, a concordância foi melhor, havendo mais etapas com %C moderado (51-70%) a excelente (91-100%) e também mais etapas com concordância leve ($k = 0,21-0,40$), como por exemplo para a conclusão do teste, a moderada ($k = 0,41-0,60$), de acordo com o valor de k. Em vista disso, mesmo que os resultados não sejam suficientes para confirmar a reprodutibilidade interavaliador do teste de Downing, eles são melhores quando são feitas as análises entre os avaliadores menos experientes (BxC).

Isto posto, o teste de Downing parece não ser reprodutível entre dois ou três avaliadores, não sendo aconselhável então, durante a prática clínica, profissionais fisioterapeutas e osteopatas compararem os seus resultados, por meio desse teste, de um mesmo paciente. Os resultados do atual estudo corroboram com o estudo de Lima *et al.* (2019), onde os autores realizaram a reprodutibilidade interavaliador do teste de Downing entre dois fisioterapeutas treinados e experientes, e obtiveram concordância pobre ($k \leq 0,20$) para a amostra total e para os

indivíduos sintomáticos, e concordância leve ($k=0,22$) para os indivíduos assintomáticos. Para mais, a influência da experiência dos avaliadores na execução do teste de Downing é incerta, apesar dos resultados de reprodutibilidade interavaliador por pares (AxB, AxC e BxC) não apresentarem, em geral, valores aceitáveis de concordância, é possível observar que quando as análises são feitas entre os dois avaliadores (BxC) considerados menos experientes que o avaliador A, porém com o mesmo tempo de prática entre si, os resultados de concordância são melhores do que quando comparados com as combinações em que o avaliador experiente está presente (AxB e AxC). A partir dessa observação, nós hipotetizamos uma possível interferência da experiência do avaliador na execução do teste. Onde, nesse caso, avaliadores menos experientes teriam resultados melhores de concordância. É cabível pensar que avaliadores com menos tempo de prática podem ser mais detalhistas e cuidadosos na execução do teste, minimizando assim a influência da subjetividade dos resultados, por se tratar de distâncias em milímetros. Assim sendo, não está clara ainda a interferência da experiência do avaliador na execução do teste, se é um fator contribuinte ou prejudicial. Então, sugerimos mais estudos realizando a análise da reprodutibilidade interavaliador do teste de Downing, e, também, comparando avaliadores com diferentes e iguais níveis de experiência.

Em suma, nossos resultados permitem que a hipótese do estudo seja parcialmente aceita, pois o teste de Downing pode ser considerado válido, reprodutível para o mesmo avaliador e acurado para o diagnóstico de disfunção posterior. Porém, não é reprodutível entre diferentes avaliadores e não é acurado para o diagnóstico de disfunção anterior. Como limitações do estudo, é cabível salientar a não realização da validade concorrente com os avaliadores B e C, e a não realização da reprodutibilidade intra-avaliador com o avaliador experiente (A). Além disso, outra limitação do estudo é a não realização da reprodutibilidade interavaliador entre o avaliador A e outro de mesma experiência. Com essas informações anteriores, análises adicionais sobre a interferência da experiência na execução do teste de Downing, e nas suas propriedades de medida (reprodutibilidade, validade e acurácia diagnóstica) poderiam ser feitas.

CONCLUSÃO

O teste de Downing é válido para avaliações com base nos Maléolos Mediais (MMs), reprodutível para o mesmo avaliador e acurado para o diagnóstico do osso Íliaco em disfunção posterior, utilizando o ponto de corte de 3 milímetros, com uma sensibilidade de 78% e especificidade de 100%. Sugerimos mais estudos que investiguem as propriedades de medida desse teste.

REFERÊNCIAS

ANJOS, Luiz A. Índice de massa corporal (massa corporal. estatura-2) como indicador do estado nutricional de adultos: revisão da literatura. **Revista de Saúde Pública**, v. 26, p. 431-436, 1992.

ARAB, Amir Massoud; *et al.* A. Inter-and intra-examiner reliability of single and composites of selected motion palpation and pain provocation tests for sacroiliac joint. **Manual Therapy**, 14(2), 213-221. 2009.

BERNARD, Jr TN; KIRKALDY-WILLIS, William H. Recognizing specific characteristics of nonspecific low back pain. **Clinical orthopaedics and related research**, n. 217, p. 266-280, 1987.

BOSSUYT, Patrick M. *et al.* STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. **Clinical chemistry**, v. 61, n. 12, p. 1446-1452, 2015.

BOWMAN, Chris; GRIBBLE, Richard. The value of the forward flexion test and three tests of leg length changes in the clinical assessment of movement of the sacroiliac joint. **Journal of Orthopaedic Medicine**, v. 17, n. 2, p. 66-67, 1995.

COHEN, Steven P. Sacroiliac joint pain: a comprehensive review of anatomy, diagnosis, and treatment. **Anesthesia & Analgesia**, v. 101, n. 5, p. 1440-1453, 2005.

COHEN, Steven P.; CHEN, Yian; NEUFELD, Nathan J. Sacroiliac joint pain: a comprehensive review of epidemiology, diagnosis and treatment. **Expert review of neurotherapeutics**, v. 13, n. 1, p. 99-116, 2013.

DREYFUSS, Paul; *et al.* Positive sacroiliac screening tests in asymptomatic adults. **Spine**, v. 19, p. 1138-1143, 1994.

DUFFIELD, Hugh J. Manipulation of the Sacro-Iliac Joint. **The Ulster medical journal**, v. 9, n. 2, p. 108, 1940.

VOS, Theo *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. **The lancet**, v. 388, n. 10053, p. 1545-1602, 2016.

- GOGIA, PP; Braatz, JH. Validity and reliability of leg length measurements. **Journal of Orthopaedic & Sports Physical Therapy**, V. 8, n. 4, p.185-188, 1986.
- GURNEY, Burke. Leg length discrepancy. **Gait & posture**, v. 15, n. 2, p. 195-206, 2002.
- KOTTNER, Jan et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. **International journal of nursing studies**, v. 48, n. 6, p. 661-671, 2011.
- LIMA, Pedro OP et al. Intraexaminer and Interexaminer Reproducibility of the Downing Test for Sacroiliac Joint Evaluation of Symptomatic and Asymptomatic Individuals. **Journal of Chiropractic Medicine**, 2019.
- MAIGNE, Jean-yves; AIVALIKLIS, Alain; PFEFER, Fabrice. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. **Spine**, v. 21, n. 16, p. 1889-1892, 1996.
- MALONEY, William J; KEENEY, James A. Leg length discrepancy after total hip arthroplasty. **The Journal of arthroplasty**. Jun 1;19(4):108-10. 2004
- PRATHER, Heidi. Pelvis and sacral dysfunction in sports and exercise. **Physical medicine and rehabilitation clinics of North America**, v. 11, n. 4, p. 805-836, 2000.
- SCHWARZER, Antony C; APRILL, Charles N; BOGDUK, Nicolai. The sacroiliac joint in chronic low back pain. **Spine**, 20:31-37. 8, 1995.
- SIM, Julius; WRIGHT, Chris C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. **Physical therapy**, v. 85, n. 3, p. 257-268, 2005.
- STEINBERG, David M.; FINE, Jason; CHAPPELL, Rick. Sample size for positive and negative predictive value in diagnostic research using case-control designs. **Biostatistics**, v. 10, n. 1, p. 94-105, 2009.
- WORLD HEALTH ORGANIZATION. Obesity: preventing and managing the global epidemic. **World Health Organization**, 2000.

APÊNDICE 1 – Quadro de anotação do teste de Downing

DOWNING							
CÓDIGO	POSIÇÃO INICIAL	ALONG () ENC ()	POSIÇÃO FINAL	CONCLUSÃO			
	D () E ()	D () E ()	D () E ()	D () E ()	D () E ()	D () E ()	
	EIAS () Superior () Inferior () simétricas MM () Superior () Inferior () simétricos	EIAS () Superior () Inferior () Imóvel MM () Superior () Inferior () Imóvel	EIAS () Superior () Inferior () Imóvel MM () Superior () Inferior () Imóvel	EIAS () Superior () Inferior () Imóvel MM () Superior () Inferior () Imóvel	EIAS () Superior () Inferior () Imóvel MM () Superior () Inferior () Imóvel	EIAS () Superior () Inferior () simétricas MM () Superior () Inferior () simétricos	() Disfunção anterior a () () Disfunção posterior a () () Teste negativo
CÓDIGO	POSIÇÃO INICIAL	ALONG () ENC ()	POSIÇÃO FINAL	CONCLUSÃO			
	D () E ()	D () E ()	D () E ()	D () E ()	D () E ()	D () E ()	
	EIAS () Superior () Inferior () simétricas MM () Superior () Inferior () simétricos	EIAS () Superior () Inferior () Imóvel MM () Superior () Inferior () Imóvel	EIAS () Superior () Inferior () Imóvel MM () Superior () Inferior () Imóvel	EIAS () Superior () Inferior () Imóvel MM () Superior () Inferior () Imóvel	EIAS () Superior () Inferior () Imóvel MM () Superior () Inferior () Imóvel	EIAS () Superior () Inferior () simétricas MM () Superior () Inferior () simétricos	() Disfunção anterior a () () Disfunção posterior a () () Teste negativo

EIAS: espinha ilíaca ântero superior; MM: maléolo medial; D: direito; E: esquerdo; ALONG: alongamento; ENC: encurtamento

6 CONSIDERAÇÕES FINAIS E PERSPECTIVAS

De acordo com a revisão sistemática (Capítulo 1), encontramos 15 testes clínicos palpatórios de mobilidade sacroilíaca. Nenhum estudo realizou a análise de validade concorrente e os valores de reprodutibilidade dos testes são divergentes. Referente a metanálise, somente o Teste de Flexão Sentado (TFS) obteve concordância boa e significativa para a reprodutibilidade intra-avaliador. Conforme o Capítulo 2, o Teste de Flexão em Pé (TFP) é válido e reprodutível para o mesmo avaliador, mas é não acurado. O Teste de Flexão Sentado (TFS) não é válido, não é reprodutível e nem acurado. Para o Capítulo 3, constatamos que o teste de Downing é válido para avaliações com base nos Maléolos Mediais (MMs), reprodutível para o mesmo avaliador e acurado para o diagnóstico do osso Ilíaco em disfunção posterior, utilizando o ponto de corte de 3 milímetros, com uma sensibilidade de 78% e especificidade de 100%. Sugerimos mais estudos que investiguem as propriedades de medida de testes clínicos palpatórios de mobilidade sacroilíaca.

Realizaremos a divulgação dos achados dessa dissertação no meio científico, via congressos, simpósios, e principalmente, em periódicos científicos. Esperamos que o atual estudo contribua aos profissionais Fisioterapeutas, Osteopatas, e estudantes, e que os conhecimentos gerados possibilitem um melhor entendimento dos presentes testes clínicos da Articulação Sacroilíaca (ASI). Aspiramos que esse estudo sirva de incentivo para futuras pesquisas, mais investigações sobre as reprodutibilidades desses testes são necessárias, analisando também a interferência da experiência dos avaliadores. Além disso, estudos avaliando as validades concorrentes e as acurácias diagnósticas desses testes também são estimulados, tendo em vista que o atual foi pioneiro em realizar tais análises e não há outros achados semelhantes para confronto. Os testes clínicos palpatórios ainda são um desafio científico extremamente necessário, considerando as suas aplicabilidades e facilidades na prática clínica.

7 REFERÊNCIAS BIBLIOGRÁFICAS DA INTRODUÇÃO

AL-SUBAHI, Moayad et al. The effectiveness of physiotherapy interventions for sacroiliac joint dysfunction: a systematic review. **Journal of physical therapy science**, v. 29, n. 9, p.1689-1694, 2017.

AIRAKSINEN, Olavi; et al. Chapter 4 European guidelines for the management of chronic nonspecific low back pain. **European spine journal**, 11;15, p. 192-300. 2006.

ARAB, Amir Massoud; *et al.* A. Inter-and intra-examiner reliability of single and composites of selected motion palpation and pain provocation tests for sacroiliac joint. **Manual Therapy**, 14(2), 213-221. 2009.

ÅSTRÖM, Malin; GUMMESSON, Christina. Assessment of asymmetry in pelvic motion—An inter-and intra-examiner reliability study. **The European Journal of Physiotherapy**, 1;16(2). 76-81. 2014.

BEMIS, Terry; MONTE, Daniel. Validation of the long sitting test on subjects with iliosacral dysfunction. **Journal of Orthopaedic & Sports Physical Therapy**, v. 8, n. 7, p. 336-345, 1987.

BERNARD, T; CASSIDY, J. **The Sacroiliac Joint Syndrome: Pathophysiology, Diagnosis, and Management. The Adult Spine: Principles and Practice.** Raven Press Ltd, New York, NY, USA, 2107–2130, 1991.

BOWMAN, Chris; GRIBBLE, Richard. The value of the forward flexion test and three tests of leg length changes in the clinical assessment of movement of the sacroiliac joint. **Journal of Orthopaedic Medicine**. 1, 17(2), p.66-7, 1995.

DREYFUSS, Paul; *et al.* Positive sacroiliac screening tests in asymptomatic adults. **Spine**, v. 19, p. 1138-1143, 1994.

FRYER, Gary; MCPHERSON, Hayley Claire; O'KEEFE, Paul. The effect of training on the inter-examiner and intra-examiner reliability of the seated flexion test and assessment of pelvic anatomical landmarks with palpation. **International Journal of Osteopathic Medicine**, 1;8(4):131-8: 2005.

FRYER, Gary; MORSE, Christopher M; JOHNSON, Jane C. Spinal and sacroiliac assessment and treatment techniques used by osteopathic physicians in the United States. **Osteopathic medicine and primary care**, 3(1), 4: 2009.

GOODE, Adam et al. Three-dimensional movements of the sacroiliac joint: a systematic review of the literature and assessment of clinical utility. **Journal of Manual & Manipulative Therapy**, v. 16, n. 1, p. 25-38, 2008.

HARRISON, Deed E.; HARRISON, Donald D.; TROYANOVICH, Stephan J. The sacroiliac joint: a review of anatomy and biomechanics with clinical implications. **Journal of manipulative and physiological therapeutics**, v. 20, n. 9, p. 607-617, 1997.

HOY, Damian et al. Measuring the global burden of low back pain. **Best practice & research Clinical rheumatology**, v. 24, n. 2, p. 155-165, 2010.

LIMA, Pedro OP et al. Intraexaminer and Interexaminer Reproducibility of the Downing Test for Sacroiliac Joint Evaluation of Symptomatic and Asymptomatic Individuals. **Journal of Chiropractic Medicine**, 2019.

MAIGNE, Jean-yves; AIVALIKLIS, Alain; PFEFER, Fabrice. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. **Spine**, v. 21, n. 16, p. 1889-1892, 1996.

MITCHELL, Frederic L. Structural pelvic function. **Academy of Applied Osteopathy Yearbook**, p. 71-90, 1958.

PAYDAR, Dylan; THIEL, H.; GEMMELL, H. Intra- and interexaminer reliability of certain pelvic palpatory procedures and the sitting flexion test for sacroiliac joint mobility and dysfunction. **J Neuromusculoskel Sys**, v. 2, n. 2, p. 65-69, 1994.

POTTER, Nancy A; ROTHSTEIN, Jules M. Intertester reliability for selected clinical tests of the sacroiliac joint. **Journal of Women's Health Physical Therapy**, v. 30, n. 1, p. 21-25, 2006.

PRATHER, Heidi. Pelvis and sacral dysfunction in sports and exercise. **Physical medicine and rehabilitation clinics of North America**, v. 11, n. 4, p. 805-836, 2000.

RIDDLE, Daniel L; FREBURGER, Janet K; NORTH AMERICAN ORTHOPAEDIC REHABILITATION RESEARCH NETWORK. Evaluation of the presence of sacroiliac joint region dysfunction using a combination of tests: a multicenter intertester reliability study. **Physical Therapy**, 1;82(8):772-81: 2002.

SCHWARZER, Antony C; APRILL, Charles N; BOGDUK, Nicolai. The sacroiliac joint in chronic low back pain. **Spine**, 20:31-37. 8, 1995.

SHAW, J. L. The role of the sacroiliac joint as a cause of low back pain and dysfunction. In: **The First Interdisciplinary World Congress on Low Back Pain and its Relation to the Sacroiliac Joint, San Diego, CA**. 1992. p. 5-6.

SLIPMAN, Curtis W; *et al.* M. Sacroiliac joint syndrome. **Pain Physician**, 4(2), 143-152, 2001.

STURESSON, Bengt; SELVIK, Göran; UDEN, A. Movements of the sacroiliac joints: a rontgen stereophotogrammetric analysis. **Spine**, v. 14, n. 2, p. 162-165, 1989.

TONG, Henry C; *et al.* Interexaminer reliability of three methods of combining test results to determine side of sacral restriction, sacral base position, and innominate bone position. **Journal of the American Osteopathic Association**, 1;106(8):464: 2006.

VAN DER WURFF, Peter; MEYNE, Wilco; HAGMEIJER, Ruud H.M. Clinical tests of the sacroiliac joint: a systematic methodological review. Part 1: reliability. **Manual therapy**, v. 5, n. 1, p. 30-36, 2000a.

VAN DER WURFF, Peter; MEYNE, Wilco; HAGMEIJER, Ruud H.M. Clinical tests of the sacroiliac joint: a systematic methodological review. Part 2: validity. **Manual therapy**, v. 5, n. 2, p. 89-96, 2000b.

VINCENT-SMITH, B; GIBBONS, P. Inter-examiner and intra-examiner reliability of the standing flexio test. **Manual Therapy**, 4(2), 87-93: 1999.

VLEEMING, Andry; et al. Mobility in the sacroiliac joints in the elderly: a kinematic and radiological study. **Clinical Biomechanics**, v. 7, n. 3, p. 170-176, 1992.