

eman ta zabal zazu



Universidad del País Vasco Euskal Herriko Unibertsitatea

# FIABILIDAD DE DIFERENTES PRUEBAS QUE SE UTILIZAN PARA EVALUAR EL RIESGO DE LESIÓN DEL MIEMBRO INFERIOR EN MUJERES DEPORTISTAS

AUTORA.

ARGIA LANGARIKA ROCAFORT

Directores.

Dr. José Ignacio Empananza Knörr

Dr. Julio María Calleja González

TESIS DOCTORAL  
2014



Jarduera Fisikoaren eta  
Kirolaren Zientzien Fakultatea  
Facultad de Ciencias de la  
Actividad Física y del Deporte



Departamento de  
Educación Física y Deportiva  
Gorputz eta Kirol  
Hezkuntzako Saila

© Servicio Editorial de la Universidad del País Vasco (UPV/EHU)  
- Euskal Herriko Unibertsitateko Argitalpen Zerbitzua (UPV/EHU)  
- EHU Press (UPV/EHU)  
ISBN: 978-84-9860-978-3

---

Esta tesis ha sido desarrollada gracias a la beca concedida dentro del programa de Formación del Personal de Investigación de la Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU) con el código de referencia PIFA-B/01/06.

---



*Amari,*

*nire lehen maistra*

*A todas las maestras y maestros de mi vida,*

*por su fe y ternura*

*“Bihotza jartzen denean pausu guztietan, Ithakarako bideak betetzen du xedea”*



## **AGRADECIMIENTOS**

El camino para realizar esta tesis doctoral se inició en 2006, con un pequeño e ilusionante proyecto que ha resultado ser un apasionante aprendizaje, superando todas mis expectativas iniciales. El proceso ha sido largo, duro y muchas veces solitario, pero sin duda siento que ha merecido la pena. He sido afortunada por el apoyo institucional recibido, y por la ayuda de todas las personas que me han arropado a lo largo de estos años. A todas ellas quiero expresar mi profundo agradecimiento en estas líneas:

A la Universidad del País Vasco/Euskal Herriko Unibertsitatea, al Departamento de Educación Física y Deportiva y a la Facultad de Ciencias de la Actividad Física y del Deporte por la confianza en el proyecto y los medios recibidos para poder realizar la tesis.

A mis directores, José F. Aramendi, José I. Emparanza, Julio Calleja y Julen Castellano por guiarme en el desarrollo de la tesis, por su dedicación, esfuerzo y paciencia.

A Mari Mar Ubeda por enseñarme a manejar las bases de datos y su valiosa colaboración en esta tesis. También por su apoyo incondicional, confianza y cariño recibido.

A Josean Lekue por ayudarme en la elaboración del proyecto y su colaboración en la realización de este trabajo.

A la Unidad de Epidemiología Clínica (Jose, Iratxe, María Jose y Mayte) y a la biblioteca (Eukene, Koro, Marimar) del Hospital Universitario Donostia por las facilidades ofrecidas para realizar esta tesis. También por acogerme con los brazos abiertos y hacerme sentir como en casa, haciendo del trabajo una delicia.

A Osasunkiroi, Salud y Deporte (Josetxo, Tximist y Oihane) que me acogieron con cariño y me ayudaron a poner en marcha el proyecto.

A Oidui Usabiaga e Iñaki Llodio que colaboraron en la realización de las revisiones con afán y confianza. También por su cariño y empatía mostrada.

Al Comité de Ética para las Investigaciones con Seres Humanos, sus muestras y sus datos (CEISH) de la UPV/EHU, en especial Susi y Mabel que me sensibilizaron y formaron con esmero y afecto en los aspectos éticos de la investigación.

A Jurdan Mendiguchia por sus consejos y contribuciones en esta tesis.

A Getxoko Aixerrota Boleibol Taldea por permitirme realizar los estudios, y a las entrenadoras y entrenadores y jugadoras por las facilidades y compromiso mostrado con la investigación. En especial a Alberto Olalde.

Al instituto de Aixerrota BHI por cederme y facilitarme el uso de las instalaciones para realizar los estudios de campo.

A las personas de mi hogar; Ama, Oiane y Aitor, por vuestra comprensión y apoyo en el muchas veces duro día a día de mi vida como doctoranda. Gracias por ese respaldo y ese amor incondicional mostrado. Vuestra es también esta tesis.

A mi familia, por arroparme en todo momento; a Amama Juli, Izeko Nere, Sonsoles, Javi, Anetxu, Juana, Josean, Juanan, Nere y todo el resto de familiares que me habéis mimado y animado a seguir adelante.

A Irati, Nerea, Pilartxo y Lore, que sois mi familia elegida, y que a pesar de las distancia que nos separa siempre os siento cerca.

A mis amigas y amigos que habéis sufrido mis ausencias con comprensión y cariño, y me habéis apoyado en todo momento para poder sacar adelante esta tesis: Sara, Maider, Jaione, Leiretxu, Jon M. y Mikel A. También a Jon Z. por mantener mi ordenador siempre a punto y a Edurne S. que me ayudo con los números.

Al Camino y mis personas del camino que me ayudaron a poner los pies en la tierra para afrontar la última etapa del largo camino: Pradi, David, Graeme, Françoise, Manuel, Redi, Ernesto, Jurgén, Marie-France, Robert, Holgram, Monica, Francesco, Steffen y al resto de peregrinos. Muchas gracias, eskerrik asko, moltes gràcies, thank you, merci beaucoup, danke, grazie mille, shum faleminderit, köszönöm szépen.

A todas las personas que de una manera u otra han contribuido a realizar este trabajo y estuvieron listas para brindarme todo su apoyo.

Eskerrik asko guztioi, bihotz-bihotzez.

# ÍNDICE DE LA TESIS

<b>RESÚMENES</b> .....	<b>11</b>
Resumen.....	<b>11</b>
Laburpena.....	<b>13</b>
Abstract.....	<b>15</b>
<b>ÍNDICE DE TABLAS</b> .....	<b>17</b>
<b>ÍNDICE DE FIGURAS</b> .....	<b>19</b>
<b>ÍNDICE DE ANEXOS</b> .....	<b>23</b>
<b>TABLA DE ABREVIATURAS</b> .....	<b>25</b>
<b>1-INTRODUCCIÓN</b> .....	<b>27</b>
1.1-Beneficios de la actividad física para las personas y las lesiones derivadas de su práctica en mujeres.....	<b>29</b>
1.2-Establecimiento de la magnitud del problema.....	<b>31</b>
1.3-Establecimiento de la etiología del problema.....	<b>33</b>
1.4-Las pruebas científicas y las revisiones sistemáticas.....	<b>36</b>
1.5-Métodos de evaluación de los factores de riesgo.....	<b>38</b>
1.6-Fiabilidad de las mediciones.....	<b>54</b>
<b>2-OBJETIVOS</b> .....	<b>59</b>
<b>3-PLANTEAMIENTO DE LA INVESTIGACIÓN</b> .....	<b>63</b>
3.1-Planteamiento secuencial de la tesis.....	<b>65</b>
3.2-Consideraciones éticas.....	<b>67</b>
<b>4-MATERIAL Y MÉTODOS</b> .....	<b>69</b>
4.1-Revisión sistemática sobre los factores de riesgo de las lesiones de tobillo.....	<b>71</b>
4.2-Revisión sistemática sobre la fiabilidad del <i>Star Excursion Balance Test</i> .....	<b>78</b>
4.3-Revisión sistemática sobre la fiabilidad de las pruebas de dorsiflexión en apoyo .....	<b>85</b>
4.4-Estudio de fiabilidad del <i>Star Excursion Balance Test</i> .....	<b>92</b>
4.5-Estudio de fiabilidad de las pruebas de dorsiflexión en apoyo.....	<b>101</b>
<b>5-RESULTADOS</b> .....	<b>113</b>
5.1-Revisión sistemática sobre los factores de riesgo de las lesiones de tobillo.....	<b>115</b>
5.2-Revisión sistemática sobre la fiabilidad del <i>Star Excursion Balance Test</i> .....	<b>130</b>
5.3-Revisión sistemática sobre la fiabilidad de las pruebas de dorsiflexión en apoyo .....	<b>145</b>
5.4-Estudio de fiabilidad del <i>Star Excursion Balance Test</i> .....	<b>166</b>
5.5-Estudio de fiabilidad de las pruebas de dorsiflexión en apoyo .....	<b>174</b>

<b>6-DISCUSIÓN</b> .....	<b>185</b>
6.1-Revisión sistemática sobre los factores de riesgo de las lesiones de tobillo .....	<b>188</b>
6.2-Star Excursion Balance Test .....	<b>193</b>
6.3- Pruebas de dorsiflexión en apoyo .....	<b>197</b>
6.5-Fortalezas de la tesis.....	<b>205</b>
6.6-Debilidades de la tesis.....	<b>206</b>
6.7-Implicaciones para la práctica y la investigación.....	<b>206</b>
<b>7-CONCLUSIONES</b> .....	<b>209</b>
<b>8-REFERENCIAS</b> .....	<b>213</b>

## RESUMEN

**Introducción:** La actividad física produce muchos beneficios a la persona que lo practica, pero también conlleva riesgos, entre ellos, la posibilidad de lesionarse. El 50% de las lesiones se producen en el miembro inferior, siendo las lesiones más frecuentes las lesiones agudas de tobillo. En la actualidad existe mucha controversia a la hora de identificar los factores de riesgo (FR) de esta lesión, pero sí se ha demostrado que los FR son diferentes en mujeres y hombres.

**Objetivo:** Analizar cuáles son los factores de riesgo intrínsecos para sufrir lesiones agudas de tobillo en mujeres deportistas y evaluar la fiabilidad de dos pruebas comúnmente empleadas para determinar el estado anatómico-funcional del tobillo (como predisponente de las lesiones de tobillo).

**Métodos:** Realizamos tres revisiones sistemáticas: 1) sobre los FR de las lesiones de tobillo, 2) sobre la fiabilidad del *Star Excursion Balance Test (SEBT)* y 3) sobre la fiabilidad de las pruebas de dorsiflexión en apoyo (PDFA). Analizamos las características y la calidad de los estudios incluidos y realizamos meta-análisis (MA) de sus resultados mediante Stata<sup>®</sup> siempre que fue posible. También realizamos dos estudios de campo con un diseño transversal de medidas repetidas (estudios de fiabilidad): 1) sobre el *SEBT* y 2) sobre las PDFA. Estudiamos la fiabilidad mediante el método de Bland y Altman, coeficiente de correlación intraclase (CCI), la *r* de Pearson, el error estándar de las medidas, la mínima diferencia detectable y el punto de corte empleando Medcalc<sup>®</sup> y SPSS<sup>®</sup>.

**Participantes:** Mujeres adultas practicantes de actividad física. Para los estudios de campo tomamos como muestra a jugadoras de voleibol: 9 (15,11 ± 3,56 años) para el *SEBT* y 46 para el estudio de fiabilidad del PDFA (16,26 ± 2,81 años).

**Resultados:** Se estudiaron 103 FR diferentes siendo factible realizar MA únicamente en 12 debido a la falta de réplica de los FR. No fue posible confirmar ni descartar ningún FR. El *SEBT* tiene un CCI adecuado pero un error aleatorio de 3,86 cm a 11,06 cm. Las PDFA tiene un CCI adecuado, y el error aleatorio es reducido en todos los métodos de evaluación.

**Conclusiones:** No se ha demostrado ningún FR para las lesiones de tobillo en mujeres deportistas. El *SEBT* no es una prueba fiable ya que tiene un error aleatorio elevado. Las PDFA son pruebas fiables para valorar la dorsiflexión.

**Palabras clave:** Mujeres, deportes, lesiones de tobillo, factores de riesgo, revisión sistemática, repetibilidad de los resultados.



## LABURPENA

**Sarrera:** Jarduera fisikoa oso onuragarria da pertsonarentzat, baina lesionatzeko arriskua ere badago hau praktikatzean. Lesioen %50 behe atalean gertatzen dira, eta hauen artean errepikatuena orkatileko lesio akutua dira. Gaur egun, eztabaida handia dago orkatileko lesio hauen barne arrisku faktoreak (AR) zehazteko orduan, hala ere faktore hauek emakume eta gizonen artea ezberdinak direla frogatua dago.

**Helburua:** Emakume kirolarientzat orkatileko lesio akutu bat izateko barne arrisku faktoreak aztertzea. Orkatilaren egoera anatomiko-funtzioanala zehazteko arrunki erabiltzen diren bi frogen fidagarritasuna ebaluatu orkatileko lesioen aurretiko baldintza bezala.

**Metodoak:** Hiru errebisio sistematiko (ERS) egin genituen: 1) orkatilen AR-ei buruz, 2) *Star Excursion Balance Test (SEBT)* –aren fidagarritasunari buruz y 3) bermean egindako dorsiflexio frogen (BEDF) fidagarritasunari buruz. ERS sartutako ikerlanen ezaugarriak eta kalitatea aztertu genuen eta emaitzen meta-analisia (MA) egin genuen Stata<sup>®</sup> bidez posible izan zenean. Honez gain, bi landa ikerlan egin genituen zeharkako neurketa errepikatu diseinua erabiliz (fidagarritasun ikerketak): 1) *SEBT*-ei buruz y 2) *BEDF*-ei buruz. Fidagarritasuna Bland eta Altman metodoa, *intraclass correlation coefficient (ICC)*, Pearsonen *r*-a, *standard error of the measurement* and *cut-off point* kalkulatuaz aztertu genuen Medcalc<sup>®</sup> y SPSS<sup>®</sup> erabiliz.

**Partehartzaileak:** Jarduera fisikoa egiten duten emakume helduak. Landa ikerketetarako boleibol jokalaria lagin bat hartu genuen: 9 (15,11 ± 3,56 urte) *SEBT*-rako eta 46 PDFA fidagarritasun frogarako (16,26 ± 2,81 urte).

**Emaitzak:** 103 AR desberdin ikertu dira, baina MA 12-rekin bakarrik posible izan da AR errepikakortasun faltagatik. Ezinezkoa izan da ikertutako AR-etatik bat ere berretsi ezta baztertu ere. *SEBT* ICC egokia dauka, baina 3,86 cm eta 11,06 cm bitarteko zorizko errorea dauka. PDFA-ak ICC egokia daukate, eta hauen zorizko errorea txikia da.

**Ondorioak:** Ez da ziurtatu emakume kirolarien orkatileko lesioentzako arrisku faktorerik. *SEBT* ez da froga fidagarria zorizko errore altua duelako. PDFA-k dorsiflexioa neurtzeko froga fidagarriak dira.

**Gako-hitzak:** Emakumeak, kirolak, orkatileko lesioak, arrisku faktoreak, errebisio sistematikoa, emaitzen errepikortasuna.



## **ABSTRACT**

**Introduction:** Physical activity has a number of benefits for the person who practises it, but it also carries the risk of injury. As much as 50% of injuries involve the lower limb, the most frequent being acute ankle injuries. Today there is a great deal of controversy in the identification of the different risk factors (RF) which result in an ankle injury, but it has been shown that risk factors are different for women and men.

**Objective:** To examine the intrinsic risk factors resulting in acute ankle injuries in sportswomen. To assess the reliability of two tests that are commonly used to determine the functional state of the ankle as a condition for ankle injuries.

**Method:** Three systematic reviews were conducted: 1) on the risk factors for ankle injuries, 2) on the reliability of the Star Excursion Balance Test (SEBT) and 3) on the reliability of the weight bearing dorsiflexion tests (WBDT). The characteristics and quality of the included studies were analyzed and meta-analyses (MA) of the results using Stata<sup>®</sup> were performed whenever possible. Two field studies (reliability studies) with repeated measures design were also conducted on SEBT and on WBDT. The reliability of the tests was assessed using different methods: the Bland-Altman method; intraclass correlation coefficient (ICC); Pearson's  $r$ ; standard error of the measurement; and cut-off point, using Medcalc<sup>®</sup> and SPSS<sup>®</sup> programmes for all of them.

**Participants:** Physically active adult sportswomen. The tests carried out in the reliability studies were applied to volleyball players: 9 (aged  $15.11 \pm 3.56$  years) for the SEBT and 46 (aged  $16.26 \pm 2.81$  years) for the WBDT.

**Results:** Although 103 risk factors were studied, MA could only be carried out on 12 because of the lack of replications of the RF. It was not possible to confirm or discard any of the RF. The SEBT has an adequate ICC, but the random error was between 3.86 cm and 11.06 cm. The WBDT has a high ICC and a small random error in all the measurement methods used in this study.

**Conclusions:** None of the variables studied have proven to be a risk factor for acute ankle injuries in sportswomen. The SEBT is not a reliable test since it has a high random error. The WBDT have been proven reliable to assess ankle dorsiflexion.

**Key words:** Women, sports, ankle injuries, risk factors, systematic review, reproducibility of results.



# ÍNDICE DE TABLAS

## 1-INTRODUCCIÓN

Tabla 1.6.1-Valoración de la precisión según los valores del Coeficiente de Correlación Intraclase	57
--	----

## 4-MATERIAL Y MÉTODOS

### 4.5-Estudio de fiabilidad de las pruebas de dorsiflexión en apoyo

Tabla 4.5.1-Tabla de métodos de evaluación empleado en el PDFAf para valorar la máxima dorsiflexión	107
Tabla 4.5.2-Tabla de métodos de evaluación empleado en el PDFAe para valorar la máxima dorsiflexión	109

## 5-RESULTADOS

### 5.1-Resultados revisión sistemática sobre factores de riesgo

Tabla 5.1.1-Calidad de los estudios incluidos en la RS FR	117
Tabla 5.1.2-Resumen de las características estudios RS FR	120
Tabla 5.1.3-Características cuantitativas estudios RS FR	121
Tabla 5.1.4-Características cualitativas estudios RS FR	123
Tabla 5.1.5-Resultados de los meta-análisis de la RS FR	127

### 5.2- Revisión sistemática sobre la fiabilidad del Star Excursion Balance Test

Tabla 5.2.1-Calidad de los estudios incluidos en la RS SEBT	132
Tabla 5.2.2-Resumen de las características estudios RS SEBT	134
Tabla 5.2.3-Características cuantitativas de los 14 documento incluidos en la RS SEBT	135

### 5.3- Revisión sistemática sobre la fiabilidad de las pruebas de dorsiflexión en apoyo

Tabla 5.3.1-Calidad de los estudios incluidos en la RS PDFA	148
Tabla 5.3.2-Resumen de las características estudios RS PDFA	149
Tabla 5.3.3-Características cuantitativas de los estudios RS PDFA	151

### 5.4-Estudio de fiabilidad del *Star Excursion Balance Test*

Tabla 5.4.1-Características de las participantes en estudio de fiabilidad del SEBT	166
Tabla 5.4.2-Resultados del método de MBA utilizando los valores absolutos	168
Tabla 5.4.3-Resultados del MBA utilizando los valores normalizados	169
Tabla 5.4.4-Resultados del CCI, r de Pearson, EEM, MDD y PDC valores absolutos	169
Tabla 5.4.5-Resultados del CCI, r de Pearson, EEM, MDD y PDC con valores normalizados	171

### **5.5-Estudio de fiabilidad de las pruebas de dorsiflexión en apoyo**

Tabla 5.5.1 Características de las participantes del estudio fiabilidad PDFA	<b>177</b>
Tabla 5.5.2-Resultados análisis de fiabilidad de la PDFAf con el MBA	<b>178</b>
Tabla 5.5.3-Resultados análisis de fiabilidad de la PDFAf con CCI, r de Pearson, EEM, PDC y MDD	<b>180</b>
Tabla 5.5.4-Resultados análisis de fiabilidad de la PDFAe con el MBA	<b>180</b>
Tabla 5.5.5- Resultados análisis de la fiabilidad de la PDFAe con CCI, r de Pearson, EEM, PDC y MDD	<b>182</b>

# ÍNDICE DE FIGURAS

## 1-INTRODUCCIÓN

Figura 1.1.1-La actividad física es beneficiosa para la salud	29
<b>1.1-Beneficios de la actividad física para las personas y las lesiones derivadas de su práctica</b>	
Figura 1.1.2-Modelo secuencial para la prevención de lesiones propuesto por Van Mechelen	30
<b>1.2-Establecimiento de la magnitud del problema</b>	
Figura 1.2.1-Distribución de las lesiones deportivas según ubicación	31
Figura 1.2.2-Las lesiones de tobillo son las más frecuentes entre mujeres practicantes de actividad física	32
<b>1.3-Establecimiento de la etiología del problema</b>	
Figura 1.3.1-Modelo completo sobre las causa de lesión	34
<b>1.4-Las pruebas científicas y las revisiones sistemáticas</b>	
Figura 1.4.1-Clasificación jerárquica de las pruebas científicas	37
<b>1.5-Métodos de evaluación para los factores de riesgo</b>	
Figura 1.5.1-Participante tocando la línea lateral (L) en el SEBT	39
Figura 1.5.2-Participante tocando la línea posterolateral (PL) empleando un aparato diseñado para realiza el SEBT	40
Figura 1.5.3-Movimiento de dorsiflexión y flexión plantar de tobillo	45
Figura 1.5.4-Evaluación de la DF sin apoyo mediante goniómetro	46
Figura 1.5.5-Evaluación de la DF sin apoyo con el participante decúbite supino empleando un inclinómetro	46
Figura 1.5.6-Evaluación de la DF en apoyo	48
Figura 1.5.7-Participante realizando la prueba de dorsiflexión en apoyo con la rodilla flexionada (PDFAf)	49
Figura 1.5.8-Participante realizando la prueba de dorsiflexión en apoyo con la rodilla en extensión (PDFAe)	50
Figura 1.5.9-Evaluación del DF mediante cristal marcando los ángulos	51
Figura 1.5.10-Aparato utilizado en Jones et al. (2005) para evaluar la prueba la PDFAf	52
Figura 1.5.11-Evaluación de la PDFAe midiendo la DH y la DV propuesto Montgomery et al (1989)	53
<b>1.6-Fiabilidad de las mediciones</b>	
Figura 1.6.1-Diferencia entre precisión y exactitud. Adaptado de Hulley et al. (2007)	54

## 4-MATERIAL Y MÉTODOS

### 4.4-Estudio de fiabilidad del *Star Excursion Balance Test*

Figura 4.4.1-Colocación de las líneas A, PM y PL en el SEBT	93
Figura 4.4.2-Colocación del pie en el SEBT con el centro geométrico del pie en el centro de la estrella	94
Figura 4.4.3-Colocación del pie en el SEBT con el primer dedo del pie en el centro de la estrella	95
Figura 4.4.4-Posición de inicio de las participantes en el SEBT	96
Figura 4.4.5-Participante tocando la línea anterior (A) durante el SEBT	97
Figura 4.4.6-Participante tocando la línea posterolateral (PL) durante el SEBT	98
Figura 4.4.7-Participante tocando la línea posteromedial (PM) durante el SEBT	99

### 4.5-Estudio de fiabilidad de las pruebas de dorsiflexión en apoyo

Figura 4.5.1-Colocación de las líneas para la PDFAf y PDF Ae	102
Figura 4.5.2-Diagrama temporal del estudio de fiabilidad de PDF A	103
Figura 4.5.3-Marcas previas a la realización de las pruebas	104
Figura 4.5.4-Colocación del pie en las PDF A	105
Figura 4.5.5-Participante realizando la PDF Af	106
Figura 4.5.6-Métodos de evaluación de la PDF Af	107
Figura 4.5.7-Métodos de evaluación de la PDF Ae	109
Figura 4.5.8-Método de evaluación la DVe en la PDF Ae	110

## 5-RESULTADOS

### 5.1-Revisión sistemática sobre los factores de riesgo de las lesiones de tobillo

Figura 5.1.1-Diagrama de flujo de la RS FR	116
Figura 5.1.2-Gráfico del meta-análisis FR historia de lesión previa	128
Figura 5.1.3-Gráfico del meta-análisis FR fuerza concéntrica en DF a 30°/s	128

### 5.2-Revisión sistemática sobre la fiabilidad del *Star Excursion Balance Test*

Figura 5.2.1-Diagrama de flujo de la RS SEBT	131
Figura 5.2.2-Gráfico del meta-análisis SEBT línea anterior (A)	139
Figura 5.2.3-Gráfico del meta-análisis SEBT línea posteromedial (PM)	140
Figura 5.2.4-Gráfico del meta-análisis SEBT línea posterolateral (PL)	141
Figura 5.2.5-Gráfico mínima diferencia detectable (MDD) SEBT línea anterior (A)	142
Figura 5.2.6-Gráfico mínima diferencia detectable (MDD) SEBT línea posteromedial (PM)	142
Figura 5.2.7-Gráfico mínima diferencia detectable (MDD) SEBT línea posterolateral (PL)	143
Figura 5.2.8-Gráfico mínima diferencia detectable (MDD) y punto de corte del SEBT con valores normalizados	143

### **5.3-Revisión sistemática sobre la fiabilidad de las pruebas de dorsiflexión en apoyo**

Figura 5.3.1-Diagrama de flujo de la RS PDFA	<b>146</b>
Figura 5.3.2-Gráfico del meta-análisis PDFAf FIA	<b>155</b>
Figura 5.3.3-Gráfico del meta-análisis PDFAf FIE	<b>156</b>
Figura 5.3.4-Gráfico del meta-análisis PDFAe FIA	<b>157</b>
Figura 5.3.5-Gráfico del meta-análisis PDFAe FIE	<b>158</b>
Figura 5.3.6-Gráfico resultados dimensionales PDFAf FIA métodos evaluación en grados	<b>160</b>
Figura 5.3.7-Gráfico resultados dimensionales PDFAf FIA métodos evaluación en cm	<b>160</b>
Figura 5.3.8--Gráfico resultados dimensionales PDFAf FIE métodos evaluación en grados	<b>161</b>
Figura 5.3.9-Gráfico resultados dimensionales PDFAf FIE métodos evaluación en cm	<b>161</b>
Figura 5.3.10-Gráfico resultados dimensionales PDFAe FIA	<b>162</b>
Figura 5.3.11-Gráfico resultados dimensionales PDFAe FIE	<b>163</b>

### **5.4-Estudio de fiabilidad del *Star Excursion Balance Test***

Figura 5.4.1-Diagrama de flujo participantes estudio fiabilidad SEBT	<b>167</b>
Figura 5.4.2-Gráfico Bland-Altman línea anterior empleando la media de los valores absolutos	<b>170</b>
Figura 5.4.3-Gráfico Bland-Altman línea posteromedial empleando la media de los valores normalizados	<b>171</b>

### **5.5-Estudio de fiabilidad de las pruebas de dorsiflexión en apoyo**

Figura 5.5.1-Diagrama de flujo de las participantes en el estudio PDFA1	<b>175</b>
Figura 5.5.2-Diagrama de flujo de las participantes en el estudio PDFA2	<b>176</b>
Figura 5.5.3-Gráfico Bland-Altman del inclinómetro ubicado en aquiles en la PDFAf	<b>179</b>
Figura 5.5.4-Gráfico Bland-Altman del ángulo derivado en la PDFAf	<b>179</b>
Figura 5.5.5-Gráfico Bland-Altman inclinómetro tibia en la PDFAf	<b>181</b>
Figura 5.5.6-Gráfico Bland-Altman del ángulo derivado en la PDFAe	<b>181</b>



# INDICE DE ANEXOS

## 1-INTRODUCCIÓN

1.4.1-Niveles de evidencia según el centro de medicina basada en la evidencia de Oxford

## 3-PLANTEAMIENTO DE LA INVESTIGACIÓN

3.2.1-Informe favorable de la tesis del CEISH UPV/EHU

3.2.2-Informe favorable de los estudios del CEISH UPV/EHU

## 4-MATERIAL Y MÉTODOS

### 4.1-Material y métodos de la revisión sistemática sobre factores de riesgo

4.1.1-Estrategias de búsqueda RS FR

4.1.2-Preguntas primera selección RS FR

4.1.3-Preguntas segunda selección RS FR

4.1.4-Preguntas calidad RS FR

### 4.2-Material y métodos de la revisión sistemática sobre el Star Excursion Balance Test

4.2.1-Primeras estrategias de búsqueda RS SEBT

4.2.2-Estrategias de búsqueda RS SEBT

4.2.3-Preguntas primera selección RS SEBT

4.2.4-Preguntas de la segunda selección RS SEBT

4.2.5-Preguntas calidad RS SEBT

### 4.3-Material y métodos revisión sistemática sobre las pruebas de dorsiflexión en apoyo

4.3.1-Estrategias de búsqueda RS PDFA

4.3.2-Preguntas primera selección RS PDFA

4.3.3-Preguntas segunda selección RS PDFA

4.3.4-Preguntas calidad RS PDFA

### 4.4-Material y métodos del estudio de fiabilidad del Star Excursion Balance Test

4.4.1-Consentimiento informado participantes SEBT

4.4.2-Autorizaciones para realizar el estudio SEBT

4.4.3-Cuestionario características de las participantes SEBT

4.4.4-Cuestionario previo sesiones de evaluación SEBT

### 4.4-Material y métodos del estudio de fiabilidad de las pruebas de dorsiflexión en apoyo

4.5.1-Consentimiento informado participantes PDFA

4.5.2-Autorizaciones para realizar el estudio PDFA

4.5.3-Cuestionario características de las participantes PDFA

4.5.4-Cuestionario previo sesiones de evaluación PDFA

## **5-RESULTADOS**

### **5.1-Resultados revisión sistemática sobre factores de riesgo**

5.1.1-Resultados acuerdo observadores RS FR

5.1.2-Razones calidad RS FR

5.1.3-Resumen estudios RS FR

5.1.4-Datos extraídos RS FR

5.1.5-Resultados MA RS FR

### **5.2-Resultados revisión sistemática sobre el Star Excursion Balance Test**

5.2.1-Resultados acuerdo observadores RS SEBT

5.2.2-Razones calidad RS SEBT

5.2.3-Resumen estudios RS SEBT

5.2.4-Datos subestudios RS SEBT

5.2.5-Resultados características subestudios RS SEBT

5.2.6-Resultados estudios primarios y MA RS SEBT

### **5.3-Resultados revisión sistemática sobre de las las pruebas de dorsiflexión en apoyo**

5.3.1-Resultados acuerdo observadores RS PDFA

5.3.2-Razones calidad RS PDFA

5.3.3-Resumen estudios RS PDFA

5.3.4-Datos subestudios RS PDFA

5.3.5-Resultados características subestudios RS PDFA

5.3.6-Resultados estudios primarios y MA RS PDFA

### **5.4-Resultados estudio de fiabilidad del Star Excursion Balance Test**

5.4.1-Datos estudio fiabilidad SEBT

5.4.2-Gráficos del MBA estudio fiabilidad SEBT

### **5.5-Resultados estudios de fiabilidad las pruebas de dorsiflexión en apoyo**

5.5.1-Datos estudio fiabilidad PDFA

5.5.2-Características participantes PDFA1 y PDFA2

5.5.3-Gráficos del MBA estudio fiabilidad PDFA

5.5.4-Análisis fiabilidad MBA PDFA1 y PDFA2

5.5.5-Análisis MBA acuerdo métodos evaluación PDFA

## TABLA DE ABREVIATURAS

ABREVIATURA	DEFINICIÓN
<b>A</b>	Línea anterior en el SEBT
<b>AD/ADe/ADf</b>	Método de evaluación para las PDFA. Se realiza calculando el ángulo de DF por razones trigonométricas empleando la DH y DV. ADe en PDFAe y ADf en PDFAf. Tablas 4.5.1 y 4.5.2.
<b>AF</b>	Actividad física
<b>An</b>	Línea anterior normalizada en relación de la longitud de la pierna en el SEBT
<b>BD</b>	Base de datos
<b>CCI</b>	Coefficiente de correlación intraclase
<b>CR</b>	Método de evaluación para las PDFAe. Figura 1.5.8.
<b>DD/DDe/DDf</b>	Método de evaluación para las PDFA. Distancia Dedo. Distancia entre el primer dedo del pie y la pared. DDe en PDFAe y DDf en PDFAf. Tablas 4.5.1 y 4.5.2.
<b>DE</b>	Desviación estándar
<b>DEM</b>	Diferencia estandarizada de medias
<b>DF</b>	Dorsiflexión
<b>DH/DHf/DHe</b>	Método de evaluación para las PDFA. Distancia Horizontal. Distancia entre el talón y vertical del borde supero externo de la rótula. DHe en PDFAe y DHf en PDFAf. Tablas 4.5.1 y 4.5.2
<b>DL</b>	DerSimonian-Laird
<b>DRe</b>	Método de evaluación para las PDFAe. Distancia Rodilla. Distancia entre la vertical proyectada sobre el suelo el borde supero externo de la rótula y la pared en PDFAe (tabla 4.5.2)
<b>DTe</b>	Método de evaluación para las PDFAe. Distancia Talón. Distancia entre el talón y la pared. Tabla 4.5.2
<b>DV/DVf/DVe</b>	Método de evaluación para las PDFA. Distancia Vertical. Distancia entre el suelo y el borde supero externo de la rótula. DVe en PDFAe y DVf en PDFAf. Tablas 4.5.1 y 4.5.2
<b>DX</b>	Método de evaluación para las PDFAf. Figura 1.5.10
<b>EEM</b>	Error estándar de la medida
<b>ES</b>	Error sistemático
<b>EV</b>	Eversión
<b>FIA</b>	Fiabilidad intra-observadores
<b>FIE</b>	Fiabilidad inter-observadores
<b>FP</b>	Flexión plantar
<b>FR</b>	Factor de riesgo
<b>GO</b>	Goniómetro
<b>IA/IAe/IAf</b>	Método de evaluación para las PDFA. Se realiza colocanco el inclinómetro en el tendón de aquiles. IAe en PDFAe y IAf en PDFAf. Tablas 4.5.1 y 4.5.2.
<b>IC 95%</b>	Intervalo de confianza al 95%
<b>IL</b>	Método de evaluación para las PDFA. Inclinómetro ubicado en la parte lateral de la pantorrilla
<b>IMC</b>	Índice de masa corporal
<b>IN</b>	Inversión
<b>IT/ITe/ITf</b>	Método de evaluación para las PDFA. Se realiza colocanco el inclinómetro en la tibia. ITe en PDFAe y ITf en PDFAf. Tablas 4.5.1 y 4.5.2.
<b>i<sup>2</sup></b>	I cuadrada
<b>LDA</b>	Límite de acuerdo
<b>MA</b>	Meta-análisis
<b>MBA</b>	Método de Bland y Altman
<b>ME</b>	Método de evaluación
<b>MDD</b>	Mínima diferencia detectable
<b>MI</b>	Miembro inferior
<b>Norm</b>	Normalizado
<b>PDC</b>	Punto de corte
<b>PDFA</b>	Pruebas de dorsiflexión en apoyo
<b>PDFA1</b>	Estudio de fiabilidad de las pruebas de dorsiflexión en apoyo realizado a principio de temporada la 2010-2011
<b>PDFA2</b>	Estudio de fiabilidad de las pruebas de dorsiflexión en apoyo realizado a final de temporada la 2010-2011
<b>PDFAe</b>	Prueba de dorsiflexión en apoyo con la rodilla en extensión
<b>PDFAf</b>	Prueba de dorsiflexión en apoyo con la rodilla flexionada

ABREVIATURA	DEFINICIÓN
<b>PL</b>	Línea posterolateral del SEBT
<b>PLn</b>	Línea PL normalizada en relación a la pierna de alcance
<b>PM</b>	Línea posteromedial del SEBT
<b>PMn</b>	Línea PM normalizada en relación a la pierna de alcance
<b>RDM</b>	Rango de movimiento
<b>RS</b>	Revisión sistemática
<b>SEBT</b>	<i>Star Excursion Balance Test</i>
<b>X</b>	Media
<b>X<sup>2</sup></b>	Ji cuadrado

# 1-INTRODUCCIÓN



## 1.1-BENEFICIOS DE LA ACTIVIDAD FÍSICA PARA LAS PERSONAS Y LAS LESIONES DERIVADAS DE SU PRÁCTICA

La actividad física produce una serie de beneficios a la persona que lo práctica entre los que cabe destacar la mejora y el mantenimiento de la salud tanto física como psíquica, además de suponer una apuesta divertida de ocio y una forma de socialización. Se ha observado que la práctica de actividad física en la población general sana reduce la mortalidad, independientemente de la causa<sup>1-4</sup> (figura 1.1.1). Además la calidad de vida de la población físicamente activa, es en general mejor en comparación de aquellas que no lo son<sup>5</sup>. Pero lo más importante es que la inactividad física es un factor de riesgo modificable para las enfermedades cardiovasculares<sup>2</sup> y para otro tipo de enfermedades como la diabetes, cáncer, obesidad, hipertensión, osteoporosis y osteoartritis y depresión<sup>6</sup>. Asimismo, la actividad física mejora la calidad de vida de las personas que padecen enfermedades crónicas<sup>7-16</sup>.



Figura 1.1.1. La actividad física es beneficiosa para la salud. Tomada el 21-11-2013 <http://www.biketours.com/family-friendly-bike-tours>.

No obstante, cuando realizamos algún tipo de actividad física es relativamente corriente que ocurra algún tipo de lesión, tanto para los practicantes aficionados como para los profesionales<sup>17</sup>. Una lesión, según la definición de la Unión Europea debe de cumplir al menos una de las siguientes características<sup>18</sup>: producir una

reducción de las marcas o del nivel competitivo, requerir consulta o tratamiento médico o tener consecuencias económicas y sociales. En general se ha observado que la incidencia de las lesiones depende mucho del tipo de actividad física y de la manera en que se realiza ésta. En un estudio que englobó 15 deportes diferentes y población de ambos sexos se encontró que la incidencia de lesiones variaba en los entrenamientos de 1,9 a 9,5 por cada 1000 exposiciones a la AF<sup>19</sup>. Powell y Barber-Foss (1999)<sup>20</sup> observaron que la incidencia en entrenamientos en baloncesto, voleibol, lucha, fútbol americano, fútbol, beisbol y softbol era 1,7 a 8,1 lesiones por 1000 exposiciones. Concretamente en mujeres, encontramos que la incidencia lesional por 1000 exposiciones también varía en función de la práctica deportiva de 2,7 a 6,1 lesiones<sup>19</sup> y de 1,7 a 5,3<sup>20</sup>. En fútbol femenino, por ejemplo, las lesiones producidas en las piernas ascienden a 5,49 lesiones por 1000 horas de actividad<sup>21</sup>. Independientemente del sexo de los participantes y la actividad estudiada las lesiones producidas en las competiciones son sustancialmente mayores en comparación con los entrenamientos<sup>19;21-24</sup>.

No obstante, este riesgo es inherente a la práctica deportiva, por lo que será preciso implementar planes de prevención adecuados al reducir estos significativamente el número de lesiones<sup>17;25-27</sup>. De esta manera las personas podrán seguir beneficiándose de la actividad física con un menor riesgo de lesión. Sin embargo, estos planes son muy generales y tratan varios posibles factores de riesgo (agilidad, fuerza, equilibrio...) para una determinada lesión y no se conoce cuáles son los componentes clave para un determinada lesión<sup>17</sup>. En consecuencia, hoy en día uno de los mayores retos de los programas de prevención es que sean lo más eficientes y rentable posibles<sup>28</sup>.

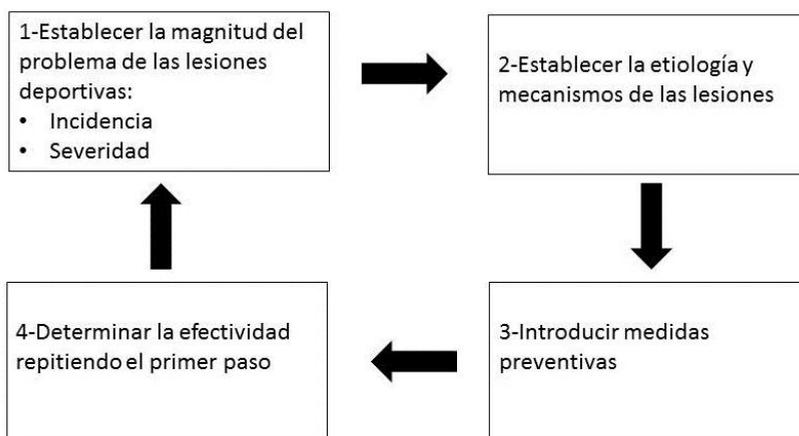


Figura 1.1.2. Modelo secuencial para la prevención de lesiones propuesto por Van Mechelen (1992)<sup>29</sup>.

En pro de la eficiencia Van Mechelen propuso en el año 1992<sup>29</sup> un modelo secuencial para desarrollar estrategias de prevención de lesiones (figura 1.1.2) que en la actualidad se sigue empleando y desarrollando<sup>17;30;31</sup>. Este autor planteó la necesidad de conocer la magnitud del problema, la etiología y los mecanismos para el tipo de lesión estudiado y posteriormente poder aplicar programas de prevención adecuados<sup>32</sup>. Conociendo los factores de riesgo (FR) de una lesión se podrán identificar a aquellos deportistas con mayor riesgo lesivo, y aplicarles, de manera individualizada los programas de prevención y de esa forma aumentar la eficiencia de los mismos.

## 1.2-ESTABLECIMIENTO DE LA MAGNITUD DEL PROBLEMA

La mayoría de las lesiones en deportes como el fútbol, baloncesto, voleibol, softbol, futbol americano o la lucha se producen en el miembro inferior<sup>19;22;23;33</sup> (figura 1.2.1). En baloncesto por ejemplo, en jugadoras aficionadas más de un 60% de las lesiones corresponden a esta zona<sup>22</sup> y en profesionales este porcentaje se eleva al 66%<sup>34</sup>. Las lesiones que se producen en las piernas son en general las más severas y las que más tiempo requieren para su recuperación<sup>22</sup>, ya que afectan directamente a la funcionalidad de la persona.

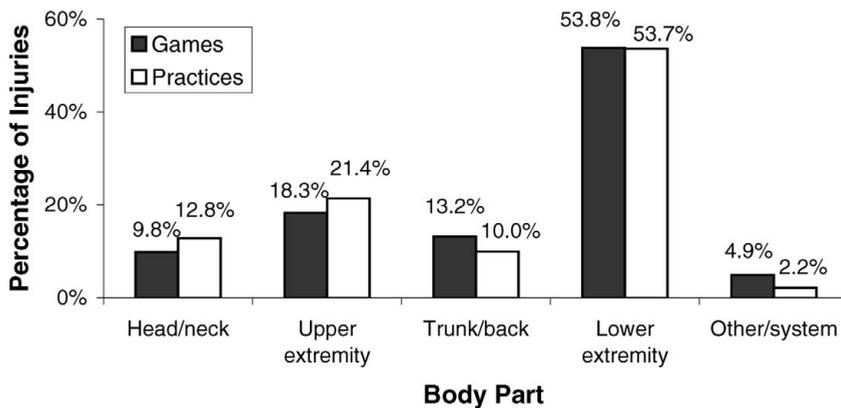


Figura 1.2.1. Distribución de las lesiones deportivas según ubicación<sup>19</sup>.

Hoy en día la atención de los investigadores se centra principalmente en la lesión de ligamento cruzado de rodilla (LCA) y en las lesiones de tobillo<sup>17</sup> (figura 1.2.2). La primera por su gravedad y la segunda por su alta frecuencia. Las lesiones de ligamento cruzado anterior suponen solo entorno a un 3% del total de lesiones, pero el 88% de ellas resultan con una baja de más 10 días<sup>19</sup>. Estas por lo tanto requieren mucho tiempo de recuperación y acarrear costes económicos muy

elevados<sup>28</sup>. En el caso de las mujeres la incidencia de esta lesión es significativamente mayor que la de los hombres, por lo que muchos investigadores se han centrado en la lesión ligamento cruzado anterior<sup>35-39</sup>.



**Figura 1.2.2.** Las lesiones de tobillo son las más frecuentes entre mujeres practicantes de actividad física. Tomada de [http://www.rpp.com.pe/2011-09-29-maria-sharapova-se-retira-de-torneo-de-tokio-por-lesion-en-el-tobillo-noticia\\_408506.html](http://www.rpp.com.pe/2011-09-29-maria-sharapova-se-retira-de-torneo-de-tokio-por-lesion-en-el-tobillo-noticia_408506.html) el 23-11-2013.

La lesión aguda de tobillo<sup>22;33;40-42</sup> es la que más se repite entre los deportistas, siendo los esguinces de tobillo<sup>40;43</sup> y específicamente los esguinces laterales de tobillo la lesión más habitual<sup>44</sup>. El porcentaje de estas lesiones varía según la actividad física que se realiza y el estudio consultado. Hootman et al. (2007) encontraron que el 14,8%<sup>19</sup> del total de lesiones eran lesiones de tobillo en un trabajo que agrupó 15 actividades diferentes. Por otro lado, Beynnon et al. (2001) hallaron que el 20% de su cohorte de jugadoras de baloncesto sufría esta lesión. Willems et al. (2005)<sup>54</sup> reportan una incidencia parecida (19%) en mujeres estudiantes de educación física que practicaban varios deportes. A diferencia de la lesión de ligamento cruzado anterior, no se han encontrado diferencias entre mujeres y hombres<sup>46;47</sup>.

Por lo general, las lesiones de tobillo no revisten mucha gravedad, pero la recurrencia de la lesión es tremendamente elevada<sup>48</sup>. Un deportista que ha sufrido un esguince severo vuelve a sufrir una nueva lesión en el 70% de las ocasiones<sup>48</sup>. Además es frecuente que se produzca una inestabilidad crónica de tobillo tras una lesión de tobillo dándose insuficiencias tanto mecánicas como funcionales<sup>49</sup>. Las

insuficiencias mecánicas incluyen laxitud patológica, problemas de movilidad articular y cambios sinoviales y degenerativos. Los cambios funcionales incluyen a su vez problemas propioceptivos, alteraciones en el control neuromuscular, déficits de fuerza y disminución del control postural<sup>49</sup>.

En la actualidad encontramos que los factores de riesgo para lesiones de ligamento cruzado están siendo ampliamente investigados, existiendo revisiones y consensos a nivel internacional<sup>35;36;38;39;50</sup> para reducir el número de esta lesiones. Sin embargo, y como veremos en el siguiente apartado, la causas para las lesiones de tobillo en mujeres son desconocidas a pesar de la alta incidencia de esta lesión y las consecuencias que ella acarrea, por lo que en esta tesis centramos nuestra atención en las lesiones de tobillo en mujeres.

### **1.3-ESTABLECIMIENTO DE LA ETIOLOGÍA DEL PROBLEMA**

Las lesiones deportivas tienen un origen multifactorial, es decir que se tienen que dar una serie de condiciones o factores para que esta se produzca. En el ámbito de las lesiones deportivas<sup>17;30;31;39</sup>, se emplea el modelo Meeuwisse (1994)<sup>51</sup> que fue adaptado por Bahr y Krosshaug (2005)<sup>17</sup> para poder explicar y comprender sus causas (figura 1.3.1). Cualquier lesión se produce cuando se dan una serie de factores de riesgo internos o condiciones, que son propios de la persona que realiza la actividad física y la predisponen en mayor o menor medida a sufrir una lesión. Durante la realización de la práctica deportiva la persona se enfrenta a una serie de factores externos como el tipo de actividad física, el calzado o el clima, haciéndola en un mayor o menor grado susceptible de sufrir una lesión. Pero, la combinación de los factores de riesgo intrínsecos y extrínsecos no es suficiente para producir la lesión, simplemente "prepara" a la persona para lesionarse<sup>51</sup>. Para que la lesión se produzca es necesario que ocurra el evento desencadenante (*Inciting event*) que se describe mediante una concreta situación de juego, la actuación del deportista (y de sus oponentes) y de una concreta descripción de la biomecánica general y de la zona a lesionarse. A pesar de que la lesión es una combinación de los factores de riesgo junto a un evento concreto, los factores de riesgo intrínsecos son los más determinantes<sup>17</sup>, y a su vez lo más desconocidos y difíciles de evaluar e identificar<sup>17</sup>.

Cuando hablamos de los factores de riesgo intrínsecos para las lesiones de tobillo, que están relacionados con las características individuales de cada persona, no hay unanimidad entre los investigadores. Esto es debido a la diferencia de métodos empleados, que muchas veces acarrea cierta contradicción en los resultados encontrados<sup>47;52</sup>. Además el número de investigaciones realizadas al respecto es muy reducida si tenemos en cuenta la incidencia de la lesión. De Noronha et al.

(2006)<sup>53</sup> realizaron un revisión sistemática en torno a esta cuestión y observaron que únicamente 21 artículos<sup>53</sup> habían estudiado los factores de riesgo, incluyendo población de ambos sexos. Estos autores<sup>53</sup> encontraron que la dorsiflexión (DF) puede predecir muy bien el riesgo de padecer una lesión, y que el balanceo postural, el rango de movimiento y la propiocepción también podrían ser factores de riesgo. Pero se desconocen los factores de riesgo para mujeres, ya que no proporcionan los resultados de manera segregada.

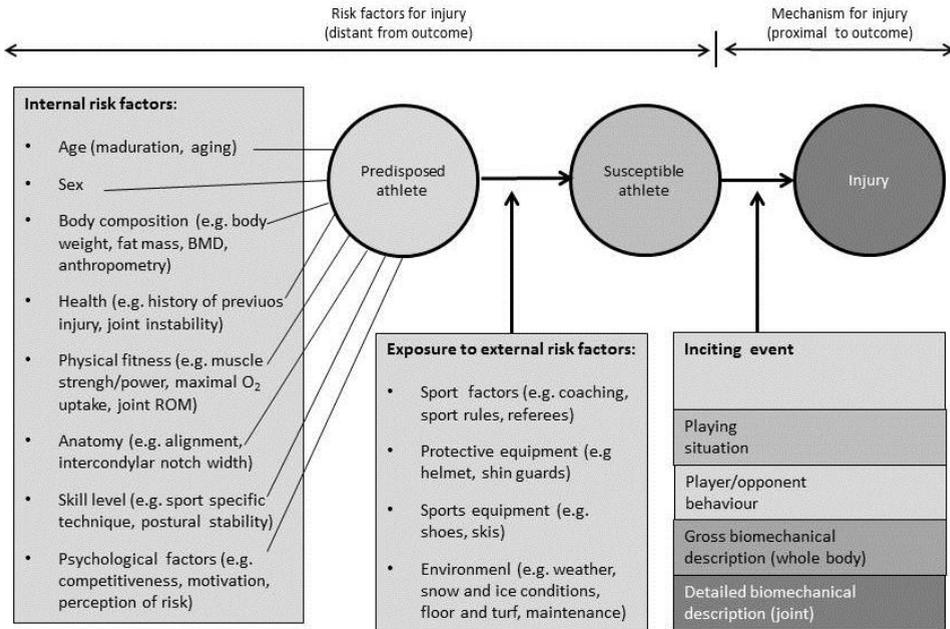


Figura 1.3.1. Modelo completo sobre las causas de lesión. Tomada de Barh y Krosshaugh (2005)<sup>17</sup>.

Cuando analizamos estudios primarios, encontramos que los factores de riesgo son muy diversos y abarcan muchas áreas como por ejemplo la fuerza o el equilibrio. Asimismo dentro de cada área se exploran diferentes variables o factores de riesgo.

Dentro del area de fuerza, que es la más investigada, encontramos que algunos FR evaluados en esta área están relacionados con un mayor riesgo de lesión y otros no. Willems et al. (2005)<sup>54</sup> encontraron que algunas variables de fuerza medidas de manera isocinética eran capaces de discriminar a las deportistas con mayor riesgo de lesión, en cambio otras no lo eran. Del mismo modo otros estudios<sup>45;52;54;55</sup> con población de ambos sexos han encontrado que ciertas

variables dentro de esta área podían discriminar a los deportistas con riesgo de lesión.

En cuanto al área de equilibrio se ha encontrado que algunos FR relacionados con el equilibrio estático, con datos registrados a principio de temporada parecen ser predictores de los esguinces de tobillo<sup>56</sup>. Cuando el equilibrio se mide de una forma más dinámica, tal y como se da en la práctica deportiva, se ha observado que es un factor que identifica a personas con inestabilidad crónica de tobillo<sup>57</sup>. También parece que es capaz de discriminar a mujeres con mayor riesgo de lesión en el miembro inferior<sup>58</sup>, por lo que quizás pudiera ser un FR para las lesiones de tobillo. Respecto a la propiocepción, algunas investigaciones han demostrado que un déficit en esta cualidad podría estar relacionada con un mayor riesgo de lesión. En una de las pocas investigaciones prospectivas sobre factores de riesgo de tobillo en mujeres realizada por Willems et al. (2005)<sup>54</sup>, encontraron que uno de los factores de riesgo del área de propiocepción discriminaba a las mujeres con mayor riesgo lesivo. Sin embargo otras de las variables dentro de esta área no lo hacían.

La técnica también podría ser otro factor de riesgo, por lo que se han propuesto estrategias para mejorar la técnica de diferentes gestos deportivos y disminuir así el número de lesiones<sup>48</sup>. Así, por ejemplo se ha observado una disminución de las lesiones en jugadores de voleibol mediante la mejora de la técnica de caída<sup>25</sup>.

No obstante, algunos factores de riesgo para las lesiones de tobillo normalmente incluidas en las investigaciones como, el peso, la laxitud del tobillo, la edad, la experiencia deportiva, la pierna dominante, el estado de forma física o las características antropométricas no parecen tener poder predictivo<sup>40;47;48;54</sup>. Otros factores de riesgo estudiados para las lesiones de tobillo, pero sin datos suficientes que los confirmen son: la disminución en la extensión de la primera metatarsofalángica, el mayor varo en la tibia y en consecuencia el incremento de la eversion del calcáneo, así como diferentes características del apoyo plantar<sup>54;55;59</sup>.

En conclusión, el número de estudios respecto a los factores de riesgo de las lesiones de tobillo en mujeres parece que es reducido. Asimismo los rasgos que llevan a las deportistas a lesionarse del tobillo son hasta el momento desconocidos, ya que existe mucha controversia en la literatura revisada a pesar del elevado número de lesiones de tobillo que se producen. Es por ello que consideramos que es pertinente con el estado actual de la ciencia en torno a los factores de riesgo intrínsecos para las lesiones agudas de tobillo en mujeres abordar esta cuestión de una manera rigurosa, mediante una revisión sistemática (RS).

## 1.4-LAS PRUEBAS CIENTÍFICAS Y LAS REVISIONES SISTEMÁTICAS

Las pruebas científicas o evidencias (*evidence*) se pueden clasificar en función del rigor científico que determinadas características del diseño científico tienen en comparación a otras posibles alternativas metodológicas<sup>60</sup>. De esta manera se crea una escala que de forma jerárquica diferencia los distintos niveles de la evidencia en función del diseño del estudio (figura 1.4.1). El esquema de gradación se construye a partir del reconocimiento de la capacidad diferencial de los diversos tipos de diseños para determinar la evidencia científica<sup>60</sup>. Las diferencias son inherentes al tipo de diseño y determinan a su vez la magnitud y el rigor en torno al hallazgo científico del diseño que se pretenda realizar.

Desde que en 1979 *Canadian Task Force on the Periodic Health Examination* realizara su escala jerárquica, diferentes organizaciones como U.S. *Preventive Services Task Force*, *National Health Service* (Reino Unido) y otras organizaciones han creado diferentes clasificaciones para las pruebas científicas en función del diseño. Aunque existen ligeras discrepancias, sobre todo en entorno a la ubicación de los ensayos clínicos, en el caso de los estudios pronósticos, es decir, aquellos que investigan los factores de riesgo, y los estudios diagnósticos (donde se incluyen los estudios de fiabilidad) estas organizaciones coinciden en que las revisiones sistemáticas (RS) son el diseño que mayor fuerza probatoria tiene. Concretamente el Centro de Medicina Basada en la evidencia de Oxford<sup>61</sup>, (anexo 1.4.1) ubica las RS en la posición más elevada en la jerarquía para este tipo de estudios. Este sistema también lo adopta la Colaboración Cochrane, impulsor principal de las revisiones y grupo de referencia para este tipo de diseño de estudios a nivel mundial<sup>62</sup>.

Las revisiones sistemáticas (RS) son revisiones de la literatura sobre una pregunta de investigación concreta que trata de identificar, seleccionar, evaluar y sintetizar todas las pruebas científicas pertinentes con la cuestión planteada. Este tipo de investigación de síntesis resulta tremendamente valiosa, ya que los investigadores y profesionales nos enfrentamos a cantidades inmanejables de información relativa a la evidencia. En el año 2005 se calculó que en la investigación médica para estar al día con los avances era preciso que los investigadores leyesen 1.500 artículos originales y 55 ensayos clínicos aleatorizados cada día<sup>63</sup>. Es poco probable que dispongamos del tiempo, las habilidades y los recursos para identificar, evaluar e interpretar esta evidencia e incorporarla a nuestra investigación y a la práctica. Por ello es que las RS resultan tremendamente útiles para ayudar a los profesionales a resumir de manera crítica la evidencia sobre un determinado tema.

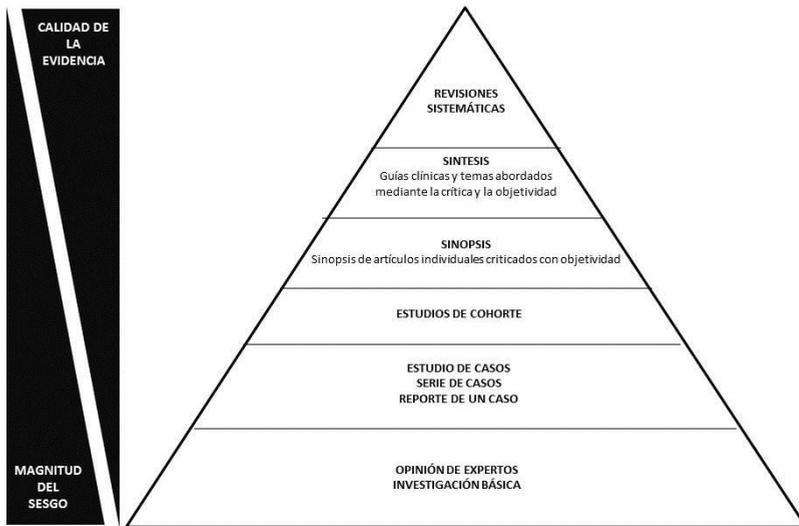


Figura 1.4.1. Clasificación jerárquica de las pruebas científicas.

Las RS utilizan unos métodos sistemáticos y explícitos que se eligen con el fin de minimizar sesgos, aportando así resultados más fiables a partir de los cuales se puedan extraer conclusiones y tomar decisiones. Los elementos fundamentales de una revisión sistemática son:

- Una metodología explícita y reproducible.
- Objetivos claramente descritos en los cuales se basan los criterios de elegibilidad de los estudios.
- Una búsqueda de la literatura que identifique todos los estudios que puedan cumplir los criterios de elegibilidad.
- Evaluación de la calidad de los trabajos incluidos.
- Una presentación sistemática y síntesis de las características y resultados de los documentos incluidos.
- Algunas revisiones sintetizan también los resultados de los estudios incluidos mediante la técnica de meta-análisis, pero esto no siempre es posible ni recomendable<sup>64</sup>.

Las RS permiten acumular y sintetizar conocimientos sobre una pregunta de investigación que se ha tratado de resolver previamente mediante estudios primarios. Mediante las RS los investigadores podemos aprender de las fortalezas y

de las debilidades de los estudios primarios y mejorar así el diseño y la investigación de campo de una cuestión científica por resolver o completar.

## **1.5-MÉTODOS DE EVALUACIÓN DE LOS FACTORES DE RIESGO**

La aparente escasez de estudios que hemos detectado en la revisión narrativa sobre los factores de riesgo (apartado 1.3) puede ser debida a la complejidad de los métodos de evaluación para los factores de riesgo, y a la poca importancia e inversión económica que se realiza en el deporte realizado por mujeres. Parece imprescindible, dado el estado actual de la ciencia respecto a las lesiones de tobillo en mujeres y la importancia que el deporte practicado por mujeres tiene en esta sociedad, encontrar métodos de evaluación accesibles y sencillos para este grupo de deportistas. Estos métodos de evaluación sencillos permitirían replicar los factores de riesgo estudiados en diferentes trabajos. De esta manera se aumentará el número de trabajos haciendo las pruebas científicas suficientes como para determinar o no la existencia de un FR. Dos de estos métodos de evaluación accesibles pueden ser el *Star Excursion Balance Test (SEBT)*, propuesto<sup>65;66</sup> para evaluar el equilibrio y la propiocepción, y las Pruebas de Dorsiflexión en apoyo (PDFA), que evalúan la dorsiflexión (DF) en apoyo.

### **1.5.1-STAR EXCURSION BALANCE TEST (SEBT)**

Existen diferentes métodos para evaluar el equilibrio, pero muchos de ellos requieren material sofisticado y poco accesible para la mayoría de la población<sup>67</sup>. Es por ello que han surgido numerosos métodos sencillos para valorar esta cualidad, como son *Single Leg Balance (SLB)*, el *Balance Error Scoring System (BESS)* y el *Star Excursion Balance Test (SEBT)*<sup>68</sup>. El *SLB* y *BESS* son pruebas que se emplean para valorar el equilibrio estático, ya que el objetivo de estas pruebas es el de mantener una determinada posición moviéndose lo menos posible. Sin embargo, estas pruebas no reflejan las exigencias de la actividad física, que son más dinámicas y exigen al deportista mover su centro de gravedad constantemente. El *SEBT* en cambio, es una prueba dinámica, donde el control postural del participante se ve retado hasta el máximo, y por lo tanto imita mejor las demandas de la actividad física que las pruebas estáticas. Además el *SEBT* ha obtenido una fiabilidad mejor que el *SLB* y el *BESS*<sup>68</sup>.

El *SEBT* se realiza marcando una serie de líneas en el suelo a modo de estrella (figura 1.5.1). El número máximo de líneas es de ocho, con una separación de 45° entre ellas, y con un mínimo de una. Cada una de las líneas ofrece un reto diferente para el participante, ya que el reclutamiento muscular es diferente<sup>69</sup> al requerir movimientos en los planos transversal, sagital y frontal diferentes

dependiendo de la dirección. El nombre de las líneas se denomina en función de la pierna evaluada como anterior (A), anteriomedial, (AM), medial (M), posteromedial (PM), posterior (P), posterolateral (PL), lateral (L) y anterolateral (AL). El número de líneas y que líneas utilizar varía según el estudio y el objetivo del mismo. Para establecer las líneas se adhiere una serie cintas métricas en el suelo, con lo que es una herramienta muy accesible, aunque se ha propuesto el uso de un aparato específico (figura 1.5.2) para la realización del *SEBT*<sup>70;71</sup>.



Figura 1.5.1. Participante tocando la línea lateral (L) en el *SEBT*. Tomada de Munro et al. (2010)<sup>72</sup>.

Para llevar a cabo la prueba, el participante coloca el pie a probar en la intersección de las líneas y la pierna libre a su lado adoptando una posición bípeda estable. Desde esta posición, el participante pasa a mantener el equilibrio sobre el pie que queda en el centro de la estrella e intenta tocar una de las líneas con el pie libre lo más lejos posible. Una vez tocada la línea, el participante tiene que retornar a la posición de inicio sin perder el equilibrio. Esta operación se repite a lo largo de cada una de las líneas en el número de intentos y en el orden que el procedimiento de la investigación establezca. La prueba no se considera válida si el

participante mueve el pie de apoyo, se apoya en el suelo con la pierna libre o pierde el equilibrio en algún momento de la prueba. El objetivo es que el participante consiga la máxima distancia en cada una de las líneas, lo cual refleja el nivel de equilibrio de la persona evaluada, cuanto mayor sea la distancia, mejor será éste. Generalmente, para evaluar el *SEBT* se mide la distancia desde el centro hasta el punto alcanzado por el participante en cada una de las líneas. Investigaciones más actuales han comenzado a medir el rango de movimiento también obtenido por diferentes articulaciones durante la prueba<sup>73;74</sup>, o empleando un acelerómetro<sup>75</sup>. Aunque estas fórmulas son menos habituales por lo costosas.



Figura 1.5.2. Participante tocando la línea posterolateral (PL) empleando un aparato específicamente diseñado para realiza el *SEBT*. Tomada de [www.philplisky.com](http://www.philplisky.com) el 1-08-2013.

El *SEBT* es una prueba que originalmente utilizaban algunos terapeutas para evaluar el equilibrio dinámico como alternativa a otros aparatos menos accesibles<sup>65;67</sup>. Esta prueba se planteó por primera vez en un libro del año 1995 donde se describen varias pruebas funcionales del miembro inferior. De un modo más científico, Kinzey y Armstrong estudiaron su fiabilidad por primera vez en año 1998<sup>67</sup>, momento desde el cual empezó a extenderse como herramienta para

valorar el equilibrio dinámico o el control postural. Estos autores comprobaron la fiabilidad de la prueba empleando dos de las ocho líneas planteadas en origen, la AM y la PM. Cada participante realizó 5 intentos por cada línea en cada uno de los dos días evaluados. Para evaluar la fiabilidad intra-observadores (FIA) de la prueba utilizaron el Coeficiente de Correlación Intraclass (CCI) y hallaron unos valores de 0,67 a 0,87 dependiendo de la línea y de la pierna. Este trabajo concluyó que el *SEBT* no era una prueba lo suficientemente fiable. Además, también estimaron mediante *Spearman-Brown prophecy* que el número de intentos necesarios para conseguir que la prueba fuera repetible era hacer el *SEBT* en seis ocasiones, realizando cinco intentos en cada de ellas, y utilizar la media de las tres mejores marcas para el análisis.

En el año 2000 Hertel et al.<sup>65</sup> realizaron un estudio para comprobar la FIA y fiabilidad inter-observadores (FIE) del *SEBT*, con las ocho líneas. Administraron el *SEBT* en dos ocasiones con separación de una semana a población activa de ambos sexos. Cada uno de los días realizaron dos bloques de tres intentos con cada línea y pierna, cada uno de ellos con un observador diferente. En este trabajo la FIA obtuvo un CCI de 0,78 a 0,96, mientras que la FIE fue de CCI de 0,35 a 0,84 el primer día y de 0,81 a 0,93 el segundo. Los autores achacaron la falta FIE en el primer día a un posible efecto de aprendizaje, ya que encontraron que la distancia obtenida por los participantes aumentaba hasta que entre los intentos 7-9 (la primera ronda de tres intentos del segundo día) se estabilizaban. Hertel et al. (2000)<sup>65</sup> concluyeron que el *SEBT* era fiable y recomendaron realizar seis intentos para minimizar los efectos de aprendizaje previo al registro de las marcas.

En una investigación más actual, Munro y Herrington (2010)<sup>72</sup> hallaron una muy buena FIA de la prueba empleando las líneas A, PM y PL. En este trabajo se aplicó la prueba en dos ocasiones con separación de siete días, siempre a la misma hora a hombres y mujeres practicantes de actividad física. Estos autores realizaron siete intentos seguidos por línea y pierna, en cada uno de los días, y emplearon la media de los intentos 5,6 y 7 para el análisis, obteniendo un CCI de 0,84 a 0,92.

Al obtener la prueba una buena fiabilidad en el trabajo de Hertel et al. (2000)<sup>65</sup>, y al ser una herramienta muy accesible, el *SEBT* se ha empleado en numerosas investigaciones como herramienta para evaluar equilibrio dinámico. A pesar de que la esencia de la prueba se mantiene como era en su origen, todavía podemos encontrar diferentes procedimientos para realizar el *SEBT*, tanto en los estudios sobre fiabilidad como en el resto de trabajos. Las dudas se refieren a qué número de líneas emplear, qué líneas emplear, el calentamiento previo a la prueba, el uso o no de calzado, la ayuda de las manos durante la prueba, la ubicación del pie en la estrella, qué número de intentos es necesario realizar en cada línea o qué estimador es el más adecuado utilizar.

Respecto al número de líneas, originalmente se tenían en cuenta las ocho líneas y por ello la mayoría de los estudios así han realizado<sup>65;69;74;76-82</sup> el *SEBT*. También es común usar tres líneas, aunque dependiendo del trabajo varía qué tres líneas emplear. En algunos casos los investigadores se decantan por la A, PM y PL<sup>57;58;83;84</sup> y en otros A, M y P<sup>85</sup>. Otros estudios han optado por emplear únicamente dos, como Kinzey y Armstrong (1998)<sup>67</sup> y Lanning et al. (2006)<sup>86</sup>, que hicieron el *SEBT* con la AM y la PM. Una de las principales razones de utilizar un menor número de direcciones, es que el tiempo de realización de la prueba con ocho líneas es muy elevado<sup>79;87</sup>, ya que realizan generalmente tres intentos por línea, y estos además tiene que ser realizados correctamente. Otra importante razón es que se ha encontrado redundancia entre líneas<sup>79</sup>, con lo que el reducir el número de líneas de una forma racional parece esencial para la evolución de prueba. Pero hoy en día no parece muy claro cuáles son las líneas más representativas para valorar el equilibrio dinámico, y con las cuales realizar el *SEBT*. Hertel et al. (2006)<sup>79</sup> encontraron que la línea PM era la que más asociada estaba a la inestabilidad crónica de tobillo, y recomendaban emplear las líneas AM, M y PM para detectar los déficits provocados por esta inestabilidad. Hubbard et al. (2007)<sup>49</sup> también identificaron que una asimetría entre piernas en las líneas A y PM identificaba a personas con esta dolencia. En cuanto la lesiones de rodilla, Herrington et al. (2009)<sup>88</sup> encontraron que las líneas A, M, L y PM diferenciaban a participantes con rotura del ligamento cruzado anterior de rodilla de los sanos. También se ha empleado el *SEBT* a la hora predecir lesiones del miembro inferior. Plisky et al. (2006)<sup>89</sup> encontraron que la suma de las líneas A, PM y PL al igual que la asimetría en la distancia de la línea A eran valores predictivos para las lesiones del miembro inferior en jugadores de baloncesto. Aun así, la activación muscular producida por cada una de las líneas es diferente<sup>69</sup>, por lo que quizás dependiendo del objetivo del estudio sea más conveniente utilizar unas líneas u otras.

Respecto al calentamiento, también se han realizado modificaciones. En origen, Kinzey y Armstrong (1998)<sup>67</sup> no mencionaron ningún tipo de calentamiento en su procedimiento, pero posteriormente se encontró que la prueba tenía efecto de aprendizaje<sup>65</sup>. Hertel et al.(2000)<sup>65</sup> descubrieron que las marcas se estabilizaban después del sexto intento, y por ello determinaron un calentamiento de 6 intentos por línea previo a realizar la prueba. Más adelante Robinson y Gribble (2008)<sup>87</sup> observaron que en la mayoría de las líneas, la distancia máxima y los desplazamientos cinemáticos asociados en la pierna de apoyo (la pierna evaluada) se estabilizaba en el cuarto intento, al igual que determinaron Munro y Herrington (2010)<sup>72</sup>. La diferencia entre los estudios respecto al calentamiento puede ser debida a diferencia de procedimiento empleado en las investigaciones. En el estudio de Hertel et al. (2000)<sup>65</sup> administraron los 12 intentos de la prueba en bloques de tres intentos, realizando dos bloques en cada día. Sin embargo,

Robinson y Gribble (2008)<sup>87</sup> y Munro y Herrington (2010)<sup>72</sup> realizaron 9 y 7 intentos seguidos en cada línea respectivamente, y observaron la estabilización al cuarto intento. En la actualidad parece más adecuado realizar la prueba con un calentamiento previo de cuatro intentos por línea, lo cual además reduce el tiempo de realización de la prueba. Aun así, la mayoría de los trabajos publicados hasta el momento emplean un calentamiento de 6 intentos por línea, porque la mayoría de las investigaciones se basan en el trabajo de Hertel et al.(2000)<sup>65</sup>. Sin embargo investigaciones más actuales<sup>90</sup> han empezado a introducir el calentamiento más reducido de 4 intentos por línea<sup>72</sup>.

Otra de las diferencias en el calentamiento es que algunos trabajos<sup>73;74;77;79</sup> introducen un descanso de cinco minutos entre el calentamiento específico del *SEBT* y la prueba, o incluso quienes realizan el calentamiento el día antes del registro de las marcas<sup>91</sup>. También hay algunos estudios que realizan otro tipo de calentamiento además del específico del *SEBT*, como el hacer cinco minutos de bicicleta y/o realizar estiramientos<sup>76;92</sup> previo al *SEBT*.

Atendiendo a la posición y condiciones de los participantes para realizar el *SEBT*, también persisten diferencias importantes entre estudios. Algunos autores han realizado la prueba con calzado<sup>65;69;70</sup>, otros sin calzado<sup>76;79</sup> y otros ni siquiera han controlado tal condición<sup>67;93</sup>. Respecto a la ubicación de los pies durante la prueba también hay diferencias entre trabajos. Algunos se decantan por colocar el centro del pie en la estrella<sup>65;79;87;94;95</sup>, otros colocan el primer dedo del pie<sup>58;96;97</sup>, otros ubican el pie en cuadrado en el centro de la estrella<sup>67;70;71;86</sup> o cambiando en función de la dirección línea a tocar<sup>69</sup>. En cuanto a la colocación de los brazos, ha habido algunos estudios en los que se ha permitido su uso libre<sup>65;69;79</sup>. Otros estudios sin embargo determinaron que los brazos debían de colocarse en las caderas durante la realización de la prueba<sup>82;93</sup>. En este caso, si las manos se soltaban de las caderas, el intento no se consideraba válido.

Respecto al número de intentos a realizar y qué estimador utilizar para el análisis, también existen discrepancias entre trabajos. Kinzey y Amstrong (1998)<sup>67</sup> realizaron cinco intentos por línea y utilizaron la media de los tres mejores intentos para su posterior análisis. Earl y Hertel (2001)<sup>69</sup> también realizaron cinco intentos aunque solo registraron los tres últimos. Aunque la mayoría de los trabajos realizan tres intentos por línea, existen diferencias a la hora de establecer el estimador. Algunos de los estudios utilizan la media los tres intentos<sup>72;73;76;79;81;88;98;99</sup>, otros en cambio emplean la máxima de los tres intentos<sup>58;71;86</sup>.

Además, cuando se quieren comparar participantes parece conveniente normalizar los resultados con la longitud de la pierna<sup>77</sup>, y aquí también existen discrepancias. Gribble y Hertel (2003) encontraron que la altura y la longitud de la pierna

explicaban las diferencias entre hombres y mujeres y propusieron normalizar los resultados respecto a la longitud de la pierna. De esta manera no se encontraron diferencias entre sexos. En cuanto a la forma de medir la pierna también existen diferencias, si bien es cierto que la mayoría de los trabajos emplean la distancia entre la cresta iliaca anterosuperior hasta el maleolo medial<sup>68;71-73;90;100</sup>, algunos emplean el maleolo lateral<sup>58;96</sup>.

En resumen, son pocos los estudios que han comprobado la fiabilidad del *SEBT*<sup>65;67;72</sup>, y todos con procedimientos diferentes, al igual que el resto de trabajos que han empleado la prueba. Por lo que hoy en día no hay un procedimiento estandarizado para realizar el *SEBT*, lo que hace muy difícil la comparación entre estudios. Además Kinzey y Amstrong (1998)<sup>67</sup> determinaron que la prueba no era fiable, con lo que persisten dudas acerca de la fiabilidad de la prueba. Por lo tanto es esencial establecer un procedimiento y comprobar su fiabilidad para poder estandarizar la prueba.

### 1.5.2-PRUEBAS DE DORSIFLEXIÓN EN APOYO (PDFA)

La dorsiflexión (DF) es un movimiento que se produce en la articulación del tobillo. Este movimiento hace que disminuya el ángulo entre el pie y la pierna de tal manera que la distancia entre los dedos del pie y complejo tibio-peroneo se ve reducida (figura 1.5.3). La DF también se puede denominar flexión dorsal, siendo el movimiento contrario la flexión plantar.

Este movimiento, además de la importancia que pudiera tener en la predicción de las lesiones<sup>53;101</sup>, es un componente que se evalúa de manera habitual en diferentes áreas de conocimiento. En las actividades físicas una adecuada DF es especialmente importante en acciones donde se produce apoyo en el suelo, y requieren que la rodilla se flexione más adelante que la vertical con el pie. En concreto, se ha podido observar que la DF es necesaria para saltar<sup>53;102</sup> y poder amortiguar adecuadamente la caída<sup>103</sup>. Fuera del ámbito exclusivamente deportivo, una adecuada DF es necesaria para realizar actividades de la vida diaria tales como caminar, subir escaleras, levantarse de una silla o incluso agacharse<sup>104;105</sup>. Es más, después de una lesión de tobillo es común que existan ciertas limitaciones en el movimiento de DF<sup>106;107</sup>. Por todo ello, la DF se evalúa de manera habitual para determinar la funcionalidad de tobillo y miembro inferior<sup>108;109</sup> así como para la monitorización en el tratamiento de las lesiones<sup>110</sup>.

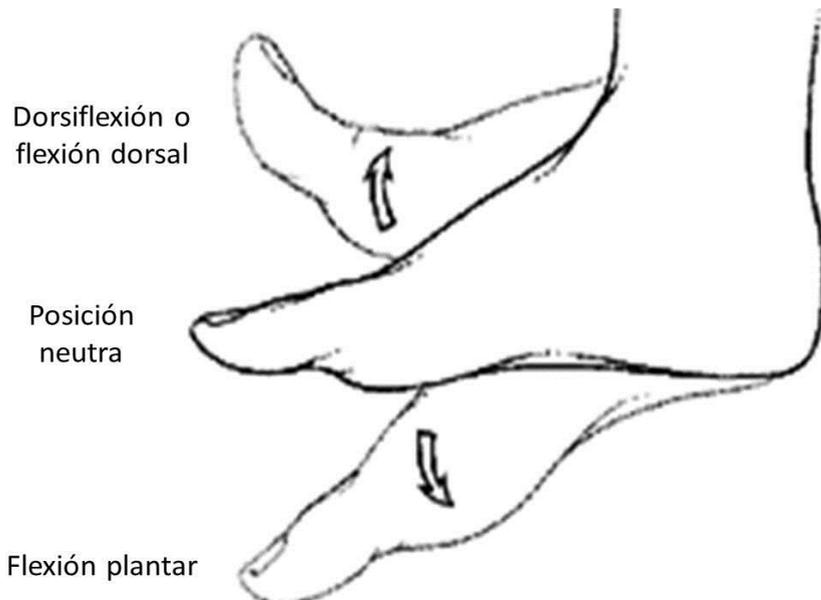


Figura 1.5.3. Movimiento de dorsiflexión y flexión plantar de tobillo. Tomada de [http://www.footdoc.ca/www.FootDoc.ca/Website%20Definitions%20\(Basic%20Terms\).htm](http://www.footdoc.ca/www.FootDoc.ca/Website%20Definitions%20(Basic%20Terms).htm) el 1-8-2013 y modificado por la autora de esta tesis.

A pesar de ser una variable comúnmente evaluada no existe consenso a la hora de determinar cuál es método más adecuado para su medición, y existiendo diferentes métodos para su evaluación<sup>111;112</sup>. Una reciente revisión realizada por Gratt & Chockalingam (2011)<sup>111</sup> sobre los métodos de evaluación de la DF concluyó que actualmente no existe un *gold standard* para su evaluación. Las diferencias entre métodos de evaluación propuestos en la literatura radican principalmente en si la prueba se realizaba en apoyo o sin apoyo, en el grado de flexión de la rodilla requerido durante la DF, y en el tipo de herramienta que se empleaba para la evaluación así como el modo de empleo de la misma.

Respecto a la herramienta de evaluación, el goniómetro (figura 1.5.4) e inclinómetro (figura 1.5.5) son las habituales para evaluar la DF<sup>103;108;111;113;114</sup>. Estas herramientas son también utilizadas en la medición rango de movimiento de otras articulaciones<sup>109;115-117</sup>. Pero la evaluación que se realiza empleando estos aparatos requiere cierta formación además del inconveniente de disponer de dicha herramienta, por lo que su utilidad queda limitada. A pesar de que dispongamos del material y personal necesario para su utilización, la fiabilidad estas herramientas es cuestionable. Se ha encontrado que su fiabilidad varía sustancialmente dependiendo del estudio consultado independientemente de la experiencia del observador<sup>112</sup>. Elveru et al. (1988)<sup>109</sup> encontraron una estimación puntual del CCI en la FIA evaluando varias características del pie (inversión, eversión, posición neutral del calcáneo respecto al astrágalo y DF) de 0,53 a 0,95

dependiendo de la variable evaluada y el tipo de participantes. Respecto a la FIE, esta varió de un CCI de 0,00 a 0,77.



Figura 1.5.4. Evaluación de la DF sin apoyo mediante goniómetro. Tomada de Johanson<sup>118</sup>.



Figura 1.5.5. Evaluación de la DF sin apoyo con el participante decúbiteo supino empleando un inclinómetro. Tomada de Calvo-Guisado<sup>119</sup>.

Si observamos la revisión sobre la fiabilidad de diferentes métodos para evaluar la DF de tobillo mediante goniómetro realizada por Martin y McPoil (2005)<sup>114</sup> encontramos que los CCI son diferentes dependiendo del estudio. La estimación puntual del CCI en el análisis FIA fue de 0,64 a 0,99, y de 0,29 a 0,81 en la FIE. Si bien es cierto que algunos estudios han presentado CCI adecuados para considerar la evaluación del DF fiable cuando se evalúa por un mismo observador, la FIE no lo ha sido, lo que limita mucho su utilización en la práctica<sup>109</sup>. A pesar de que la fiabilidad del goniómetro e inclinómetro es cuestionable, este sigue siendo la herramienta de evaluación preferida para evaluar la DF en el tobillo<sup>111</sup>.

No obstante, algunos investigadores, conscientes de la falta de fiabilidad del goniómetro e inclinómetro han tratado de mejorar la fiabilidad cuando se evalúa la DF, estandarizando el método de evaluación. Así Ekstrand et al. (1982)<sup>116</sup>, definieron mejor las marcas anatómicas para la ubicación de la herramienta de evaluación y la fijaron mejor consiguiendo mejorar la fiabilidad de todas las variables del miembro inferior evaluadas respecto a un estudio previo. En una investigación más actual Wilken et al. (2011)<sup>120</sup> evaluaron la DF mediante inclinómetro tratando de estandarizar al máximo la posición los participantes así como la fuerza aplicada en la DF, mediante un aparato diseñado para tal fin. La fiabilidad obtenida fue mejor que en los trabajos previos que habían estudiado la FIA y la FIE (el CCI para FIA fue de 0,94 (IC 95%: 0,78-0,98) a 0,99 (IC 95%: 0,96-0,99). La FIE también mejoró, con valores de 0,90 (IC 95%: 0,86-0,92) a 0,95 (IC 95%: 0,85-0,99).

Otra forma para mejorar la fiabilidad a la hora de evaluar la DF es realizar la prueba en apoyo (figura 1.5.6) en lugar de hacerla en cadena cinética abierta, tal y como propusieron Bennell et al. (1998)<sup>121</sup>. De esta manera la fuerza aplicada en el momento DF depende del participante y no de la fuerza aplicada por la persona que realiza la evaluación. Si observamos los CCI proporcionados por los estudios que han realizado la prueba con y sin apoyo, se encontró que en general la fiabilidad de la DF cuando se realiza en apoyo es mejor que cuando se realiza sin apoyo<sup>122;123</sup>.

En los estudios que han realizado la evaluación de la DF de las dos formas se estima que la fiabilidad también es mejor cuando esta se hace en apoyo. Denegar et al. (2002)<sup>124</sup> evaluaron la DF con los participantes tumbados o sentados obteniendo unos CCI de 0,96 a 0,97 y un error estándar de la medida (EEM) de 0,96° a 1,34°. Cuando la prueba se realizó en apoyo la FIA fue mejor, con CCI de 0,98 a 0,99 y unos EEM que variaban de 0,54° a 0,96°. Del mismo modo, Dinh et al. (2011)<sup>125</sup> encontraron una FIA ligeramente mejor en apoyo 0,97-0,99 que sin apoyo 0,97. En otro trabajo, Krause et al. (2011)<sup>126</sup> compararon diferentes métodos de evaluación de la DF sin apoyo con la prueba de DF en apoyo

empleando un inclinómetro (figura 1.5.6). Estos autores encontraron CCI más elevados realizando la prueba en apoyo tanto en FIA como en FIE. Concretamente en FIA la prueba DF sin apoyo obtuvo una mínima diferencia detectable (MDD) de  $5,70^\circ$  a  $7,00^\circ$  y CCI de 0,68 a 0,88. Sin embargo, la PDFAe logró un rango de MDD de  $6,1^\circ$  a  $6,4^\circ$ , y un rango de CCI de 0,88 a 0,89 dependiendo del observador y el método de evaluación. La FIE también fue mejor con unos CCI de 0,55 a 0,79 frente a un 0,82 de la PDFAe y una EEM de  $2,58^\circ$  a  $3,70^\circ$  en las pruebas sin apoyo, y de  $2,82^\circ$  con apoyo.



Figura 1.5.6. Evaluación de la DF en apoyo.

Además, evaluar la DF sin apoyo es una forma de evaluación parece poco práctica a la hora de valorar la funcionalidad del tobillo ya que no refleja los movimientos requeridos para actividades dinámicas tales como el caminar<sup>127</sup>, o la caída después de un salto<sup>103</sup>. Es más, se ha demostrado que la evaluación de la DF sin apoyo no refleja la DF en apoyo<sup>128;129</sup>. Por lo que la evaluación de la DF en cadena cinética abierta sólo parece adecuada cuando no es posible realizarla con el participante en apoyo<sup>120</sup>.

En resumen, resulta más fiable y práctico medir la DF en personas que realizan actividad física realizando la prueba de DF en apoyo (PDFA). Esta prueba se puede realizar tanto con la rodilla flexionada (PDFAf) como con la rodilla en extensión (PDFAe). En la PDFAf los participantes realizan una flexión máxima de tobillo pudiendo flexionar también la rodilla (figura 1.5.6). En la PDFAe, el participante tiene que encontrar la máxima DF sin flexionar la rodilla (figura 1.5.7). Generalmente en la pierna que se evalúa la vertical de la rodilla cae sobre el pie, que está en posición neutra respecto al plano frontal (sin eversión e inversión) y el participante mueve progresivamente el pie a evaluar hacia atrás para encontrar la

máxima DF. El método más extendido para evaluar la DF en apoyo es realizando la prueba con la rodilla flexionada. Pero es recomendable realizar ambas PDFA, porque al realizar únicamente la PDFAf, no se puede conocer si la DF está limitada por la acción del sóleo o del gemelo<sup>130</sup>.



Figura 1.5.7. Participante realizando la prueba de dorsiflexión en apoyo con la rodilla flexionada (PDFAf).

Respecto a las pruebas de dorsiflexión en apoyo (PDFA) existe cierta controversia a la hora de definir la colocación del pie durante las PDFA. En general los trabajos coinciden en ubicar el centro del talón en el centro de la línea<sup>121</sup>, no así en la posición de la parte anterior del pie. Algunos prefieren ubicar el pie con el primer dedo del pie en la línea, mientras que otros se decantan por colocar el segundo dedo en línea con el talón. Estos últimos trabajos defienden que la pronación y supinación del pie es menor colocando el segundo dedo del pie en línea con el talón, y por lo tanto el movimiento de DF es más puro. Algunos trabajos han colocado una rampa debajo del pie para que el pie este en posición neutra de pronación-supinación<sup>128;131</sup> y el movimiento sea de DF exclusivamente.

También existen dudas acerca del soporte que utilizan los participantes durante la prueba. Algunos estudios realizan la prueba próxima a la pared y permiten apoyarse a los participantes en ella<sup>104;106;121;128;130-133</sup> y otros trabajos proporcionan una silla para los participantes mantengan el equilibrio durante la prueba<sup>134</sup>. Otros

sin embargo no ayudan<sup>116;126;135-138</sup> a la persona evaluada a mantener el equilibrio durante la prueba.



Figura 1.5.8. Participante realizando la prueba de dorsiflexión en apoyo con la rodilla en extensión (PDFAe).

Pero la diferencia más importante en el método para realizar las PDFA se da a la hora de realizar la evaluación, ya que varía dependiendo del trabajo consultado. El método de evaluación más utilizado es el inclinómetro aunque su ubicación entre estudios varía, lo que hace difícil la comparación entre trabajos. Algunos colocan el inclinómetro en el tendón de aquiles (IA)<sup>131-133</sup> mientras que otros lo prefieren ubicar en la tibia (IT)<sup>121;123;137</sup> o en el lateral de la pierna evaluada (IL)<sup>124;126</sup>. Además de la ubicación del propia inclinómetro, también existen variaciones de la propia herramienta en todos los estudios nombrados, algunos emplean un inclinómetro digital<sup>133</sup>, otros líquido<sup>124</sup> y otros analógico<sup>121;131;132</sup>. Además, la fiabilidad de esta herramienta, aunque en general es mejor que la del goniómetro<sup>119;139</sup>, sigue siendo cuestionable y no soluciona el problema de la necesidad de la formación del personal encargado de la su empleo. Es por ello que algunos investigadores han empezado a emplear otro tipo de métodos de evaluación de la DF<sup>121;130;135</sup>, en busca de una mejor fiabilidad y sencillez de las PDFA.

El método de evaluación más extendido como alternativa a las herramientas que evalúan el ángulo de DF, es medir la distancia desde el primer dedo del pie hasta la pared (DD). Este método solo se ha aplicado en la PDFAf y fue propuesto por Bennell et al. (1998)<sup>121</sup> obteniendo CCI similares tanto en FIA como en FIE al

inclinómetro, con la ventaja de no requerir ninguna herramienta ni formación por parte de los observadores. Además, al encontrar en este estudio una alta correlación entre IT y DD, algunos autores pensaron que los métodos se podían intercambiar. Pero este método no es comparable entre participantes<sup>121</sup>, ya que a una misma DF con un pie mayor la distancia a la pared sería menor. Aun así, la DD se empezó a emplear de forma generalizada para la evaluar la DF en la PDFAf<sup>104;106;140-142</sup> como alternativa al inclinómetro y goniómetro. Asimismo otras investigaciones consultadas<sup>122;123</sup> indican que evaluar la distancia es más fiable que emplear el inclinómetro o el goniómetro, tanto cuando la evaluación se realiza por un mismo observador, como cuando se realiza por diferentes observadores.

Más allá de las diferencias entre métodos de evaluación de la DF, existen más dudas acerca de la realización de las PDFA. Cuando se emplea herramientas que evalúan la DF en ángulos, no parece claro que ángulo es que hay que medir. Algunos estudios definen claramente que evalúan el ángulo entre la vertical y la tibia<sup>101;130</sup>, pero hay algunos que ni siquiera definen que ángulo miden<sup>125;126;143</sup>.

Además de medir la distancia desde el primer dedo del pie hasta la pared, existen más alternativas para evaluar la DF. Munteanu et al (2009)<sup>130</sup> y Menz et al (2003)<sup>144</sup> emplearon un cristal con los ángulos marcados<sup>130;144</sup>, de tal manera que la DF se puede obtener transportando la referencias anatómicas previamente marcadas y obtener el ángulo de DF (figura 1.5.9).

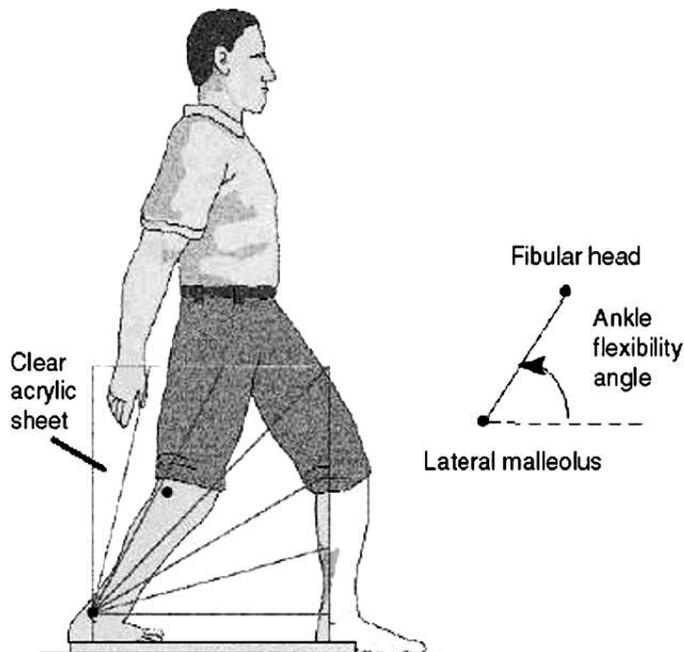


Figura 1.5.9. Evaluación del DF mediante cristal marcando los ángulos. Tomada de Gratt & Chockalingam (2011)<sup>111</sup>.

Otra alternativa posible es emplear electrogoniómetros tal y como propusieron Soper et al. (2004)<sup>138</sup>, obteniendo una buena fiabilidad en la PDFAf. Por otro lado, se podría emplear el aparato de Jones et al. (2005)<sup>135</sup>. Mediante esta herramienta (figura 1.5.10) se puede medir la distancia que la vertical de la rodilla sobrepasaba la línea de los dedos del pie (DX). No obstante el problema de estas alternativas, es que son poco prácticas ya que requieren material específico y/o formación para su empleo, y su fiabilidad también resulta cuestionable. Por lo tanto, parece que evaluar la distancia entre el primer dedo del pie y la pared, es el método sencillo y a la vez fiable para evaluar la DF.

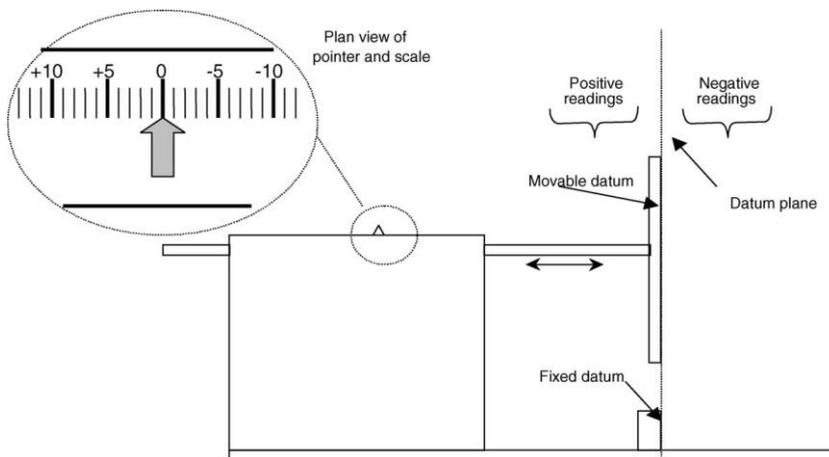


Figura 1.5.10. Aparato utilizado en Jones et al. (2005)<sup>135</sup> para evaluar la prueba de la PDFAf. Tomada de Jones et al. (2005)<sup>135</sup>.

Sin embargo el método de evaluación de DD tiene un problema, que la comparación entre participantes de diferentes características antropométricas resulta imposible<sup>121</sup>. Esto es debido a que a una misma DF, con un pie mayor la distancia entre pared y dedo sería menor. Por lo tanto, no es una herramienta apropiada si queremos comparar deportistas y establecer valores de lesionabilidad, puesto que nos encontraremos con diferentes parámetros antropométricos. Algunos estudios han tratado de resolver esta limitación relacionando la distancia con el ángulo, que es una medida que permite comparar participantes. En primer lugar, Bennell et al. (1998)<sup>121</sup> correlacionaron la distancia obtenida desde el primer dedo del pie hasta la pared y el ángulo entre la vertical y la tibia mediante un inclinómetro. Estos autores encontraron una alta correlación ( $r=0,93-0,96$ ,  $p<0,001$ ). En otro trabajo donde se correlacionaron ambas medidas, Dickson et al. (2008)<sup>123</sup> encontraron que no existía tal correlación. La correlación obtenida por Bennell et al. (1998)<sup>121</sup> puede ser debida al reducido tamaño de participantes y a

su posible homogeneidad. Además no parece lógico, ya que la distancia puede variar también en función de la longitud del pie, no solo de la DF. Por lo que es necesario buscar otras alternativas que mantengan la fiabilidad de la prueba, y permitan comparar a participantes con diferentes características antropométricas y que a su vez sean sencillas de realizar.

Una alternativa interesante podría ser el propuesto por Montgomery et al. (1989)<sup>136</sup> en un estudio donde se empleó la PDFAe (figura 1.5.11). Estos autores midieron la distancia desde el talón hasta la vertical la tuberosidad tibial (distancia horizontal, DH), y la distancia entre el suelo y la tuberosidad tibial (distancia vertical, DV). Mediante estas dos distancias, se derivó el rango de DF con una simple razón trigonométrica. Ésta técnica es comparable entre participantes puesto que se toma en cuenta la longitud de la tibia, y por lo tanto es relativa a la antropometría de cada participante. Este método también se ha aplicado en la PDFAf, en el estudio de Pope et al. (1998)<sup>101</sup>. Sin embargo, ninguno de los trabajos mencionados se comprobó la fiabilidad del método de evaluación y la descripción del procedimiento y del método de evaluación resulta parca.

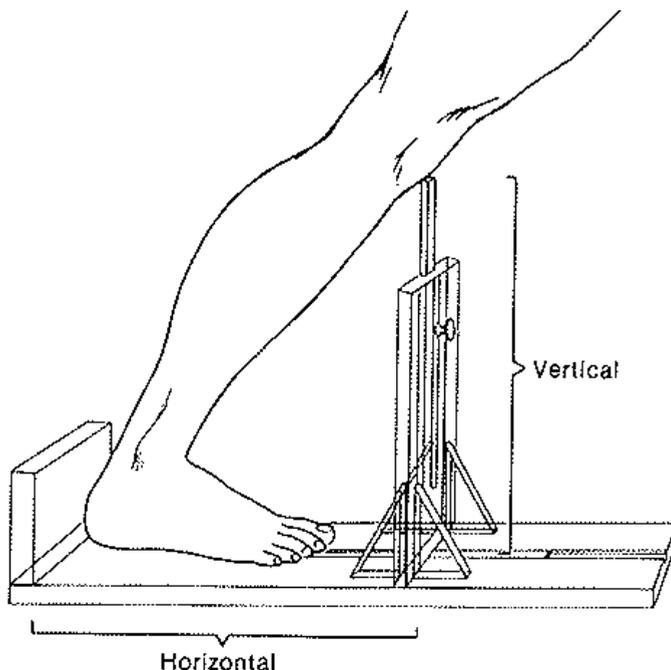


Figura 1.5.11. Evaluación de la PDFAe midiendo la DH y la DV propuesto Montgomery et al. (1989)<sup>136</sup>. Tomada de Montgomery et al. (1989)<sup>136</sup>.

Por lo tanto, parece que hoy en día no existe acuerdo en la comunidad científica acerca de cómo hay que evaluar la DF<sup>111</sup>. Tampoco parece claro el procedimiento para realizar la PDFAf y la PDFAe ya que éste no está estandarizado. Asimismo los métodos de evaluación propuestos hasta la actualidad por la literatura obtienen una fiabilidad cuanto menos cuestionable, además de ser complejos y necesitar material específico y personal formado para su realización. Por lo tanto, parece necesario estandarizar las PDFA de una manera práctica y sencilla que permita poder comparar los resultados de las diferentes investigaciones. Del mismo modo es necesario saber que método de evaluación es el más fiable y a su vez comparable entre participantes para evaluar la DF en las PDFA. Para ello es preciso evaluar la DF con las diferentes herramientas de manera simultánea, dentro de un mismo intento, ya que está manera estaremos midiendo el mismo fenómeno.

## 1.6-FIABILIDAD DE LAS MEDICIONES

Las características de las participantes que pretender discriminar a mujeres con riesgo de lesión (es decir los FR) van a ser evaluadas, medidas, empleando un determinado método de evaluación o instrumento. Estas mediciones que realizamos describen un determinado fenómeno como son la dorsiflexión o la distancia obtenida en el *SEBT* de una manera que puede ser analizada de manera estadística<sup>145</sup>. Por lo cual, la medición resulta un proceso esencial y primario en toda investigación y deberá garantizar su validez, puesto que las conclusiones que de la medición se deriven tendrán efectos en la propia validez de los estudios<sup>145</sup>.

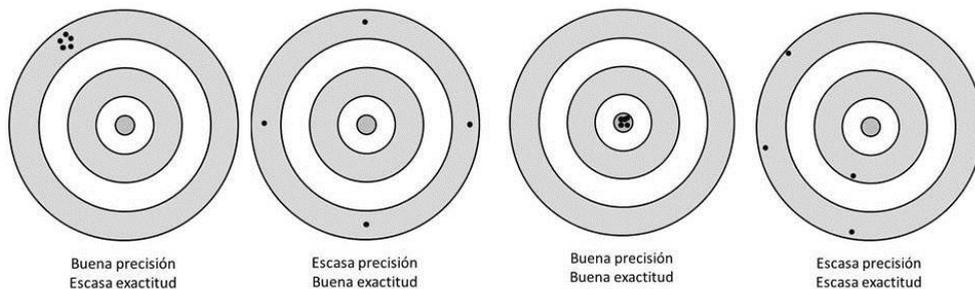


Figura 1.6.1. Diferencia entre precisión y exactitud. Adaptado de Hulley et al. (2007)<sup>145</sup>.

La validez de una medición va a depender tanto de su fiabilidad, precisión o repetibilidad como de su exactitud<sup>2</sup> (figura 1.6.1). La fiabilidad (en inglés *reliability* o *reproducibility of the results*) es el grado en el cual una variable es reproducible cada vez que se evalúa. En otras palabras, es el grado en que un instrumento proporciona resultados similares cuando se aplica varias veces a una misma

población en idénticas condiciones<sup>2</sup>. Cuanto más precisa sea una medición menor error aleatorio (variabilidad debida al azar) tendrá<sup>145</sup>. La exactitud (en inglés, *accuracy*) en cambio, se refiere es el grado en la que una medición representa lo que quiere representar<sup>145</sup>, es decir el grado de aproximación a la "verdad" de sus resultados<sup>145</sup>. Una medición exacta está exenta de error sistemático. La mejor manera de establecer la exactitud de una medición es disponer de un *gold standard*, es decir una medición que se considera exacta o verdadera, lo cual muy pocas veces es posible.

Algunas variables son relativamente fáciles de medir como el peso, la altura, el IMC y la historia de lesiones, sin embargo otras tienen un grado de subjetividad que hacen su medición sea más complicada, como la intensidad del esfuerzo, las medidas realizadas con el goniómetro o inclinómetro<sup>1</sup> y la mayoría de los métodos empleados para evaluar los factores riesgo para las lesiones de tobillo. En cualquier caso, el proceso de medición conlleva siempre algún grado de error<sup>2</sup>, independientemente de su simplicidad o complejidad, lo que va a determinar la validez de las mismas. El grado de error de una determinada medición va a depender de la variabilidad de 3 factores:

**1-Variabilidad del observador:** Se refiere a la variabilidad producida por la persona que realiza la medición como pueden ser; su habilidad en el uso de la herramienta de medición, su conocimiento del procedimiento de la prueba...

**2-Variabilidad del instrumento:** Se refiere a la variabilidad del método de evaluación, prueba o instrumento que se emplea para medir la variable en cuestión. Se puede deber a los cambios de temperatura, deterioro de los instrumentos, etc.

**3-Variabilidad del participante:** Se refiere a la variabilidad biológica intrínseca propia de los participantes en el estudio, como pueden ser cambios de humor, aumento de la musculatura debida al entrenamiento, etc.

La variabilidad que afecta al error de la medición en ocasiones está fuera del alcance del control del observador, como algunas de las variaciones propias del participante<sup>146</sup>. Sin embargo otras variables son fácilmente controlables como la posición de la persona evaluada, la técnica que emplea el observador así como la temperatura del lugar donde se realiza las mediciones<sup>146</sup>. Es importante, cuando realizamos cualquier medición que estas variables que podemos controlar estén correctamente definidas, para que así podamos atribuir las diferencias entre mediciones a una causa diferente a las discrepancias en el procedimiento de medición.

Para estudiar la fiabilidad o precisión de una medición se emplean un diseño transversal de medidas repetidas. Es decir, se realizan varias mediciones y se

comparan entre ellas para conocer el grado de error de la medición. Dentro de los estudios medidas repetidas se dan 3 tipos de estudios para estudiar la fiabilidad:

**1-Fiabilidad intra-observador (FIA):** Es el grado de consistencia de la medida de un mismo observador con un mismo instrumento en momentos diferentes sobre un mismo participante en idénticas condiciones.

**2-Fiabilidad inter-observadores (FIE):** Representa la consistencia entre dos observadores diferentes cuando evalúan con un mismo instrumento a un mismo grupo de participantes en idénticas condiciones.

**3-Concordancia entre dos instrumentos diferentes:** Representa la consistencia entre dos instrumentos diferentes. Esto sucede cuando se quieren intercambiar la medición de un instrumento por otro porque uno de ellos es más sencillo, más barato o portátil.

La fiabilidad o precisión de una medición debe ser entendida como el producto de las interacciones entre los instrumentos, los participantes y los observadores así como la interacción o contexto donde se realizan las mediciones. Es decir, la precisión no es propiedad exclusiva de las pruebas ya que está afectada por la variabilidad de circunstancias de evaluación<sup>146</sup>, así como del análisis estadístico empleado y su interpretación. Por lo tanto la fiabilidad tiene que ser considerada como nivel de consistencia de las medidas cuando es aplicada a un grupo específico en unas determinadas condiciones<sup>114</sup>. Es por ello que una medición debe ser interpretada de acuerdo circunstancias de evaluación, y deberá ser comprobada en la misma situación en la cual se quiere emplear. Quizá una prueba sea precisa para ser empleada dentro de un contexto, pero no en otro<sup>147</sup>, ya que dependiendo de las circunstancias pueden requerir un mayor grado de precisión para discriminar un determinado rasgo en los participantes.

Generalmente para evaluar la fiabilidad de una medición, se utiliza el coeficiente de correlación intraclass (CCI)<sup>148-150</sup> a pesar de que presenta gran complejidad en su empleo. La dificultad de la aplicación del CCI como análisis estadístico radica en que existen gran variedad de modelos de Fisher<sup>151</sup> y que en la elección del mismo es problemática<sup>152</sup>. Cada modelo es apropiado para una situación específica que viene determinada por el diseño experimental y el objetivo del estudio<sup>149;151</sup>. Pero resulta difícil elegir qué modelo es el más adecuado para cada estudio, y muchos investigadores no son conscientes de la diferencia entre modelos<sup>149;151;152</sup>. De hecho, la mayoría de libros de texto y artículos describen solo unos pocos modelos de todos los posibles, conteniendo muchos de ellos imprecisiones<sup>151</sup> y haciendo difícil la elección del modelo para los no expertos en estadística.

Elegir adecuadamente el modelo es esencial, ya que los modelos proporcionan resultados sustancialmente diferentes cuando se aplican a un mismo espectro de

datos<sup>149;151</sup> (tabla 1.6.2). Ejemplo de ello es un trabajo realizado por Müller y Büttner (1994) que aplicaron diferentes modelos de CCI existentes utilizando los mismos datos, y encontraron CCI de 0,15 (lo cual demuestra una escasa fiabilidad) hasta 0,92 (lo que demuestra una fiabilidad muy buena). Esto puede llevar a determinar que un método de evaluación es muy preciso o tremendamente impreciso dependiendo del modelo que se elija. Si la elección no es la adecuada podemos llegar a determinar que una medición no es fiable cuando sí lo es, o viceversa, lo que no va a reflejar la realidad de la medición. En la actualidad, la variedad de modelos ha aumentado y puede que muchos investigadores se decanten por el modelo que mejor fiabilidad reporte al ser la elección del modelo complicado e influenciados por el sesgo de la publicación.

A pesar de que se emplee el modelo adecuado, el CCI es una medida de fiabilidad relativa, adimensional y no proporciona un valor interpretable en la práctica<sup>153</sup>. Es decir, mediante el CCI u otras medidas adimensionales, como la *r* de Pearson, se determina si la prueba es precisa o no proporcionando un grado de precisión: fiabilidad muy buena, buena, moderada, regular o mala. Esta clasificación es engañosa, ya que sólo aporta información de manera estadística. El CCI no deja al especialista establecer de una manera práctica si un método de evaluación es fiable o no dependiendo del tipo de medición y el objetivo de la investigación. No es lo mismo la precisión que se necesite para detectar deficiencias entre participantes con el ligamento cruzado anterior de la rodilla roto de los que no, o para detectar pequeñas alteraciones funcionales en el tobillo que se puedan dar tras un esguince. La determinación de precisión de una medición es una decisión práctica, que la deben tomar los investigadores que son los especialistas sobre lo que se va a estudiar, empleando los métodos estadísticos adecuados y no debe dejarse exclusivamente en manos de la estadística.

Tabla 1.6.1. Valoración de la fiabilidad según los valores del Coeficiente de Correlación Intraclase (CCI)<sup>148</sup>.

Valor del CCI	Fuerza de la concordancia
>0,90	Muy buena
0,71-0,90	Buena
0,51-0,70	Moderada
0,31-0,50	Mediocre
<0,30	Mala o nula

Por todo ello, en la actualidad algunos investigadores emplean análisis estadísticos que proporcionan medidas dimensionales y reportan el error de la medición en las mismas unidades que se emplean en la evaluación. Algunos de estos análisis son el

error estándar de la medida (EEM), la mínima diferencia detectable (MDD), el punto de corte (PCD) y el método de Bland y Altman (MBA). Este tipo de análisis estadístico permite a los investigadores interpretar la precisión de sus mediciones de una manera práctica, ya que estos índices de fiabilidad, proporcionan el grado de error de una manera interpretable. Estos índices proporcionan el error de la medición en las unidades que realizamos la evaluación, con un intervalo de confianza (IC) al 95%, a excepción del EEM que es del 68%. Empleando este tipo de análisis estadístico los investigadores pueden decidir de manera práctica si un método de evaluación es suficientemente preciso dependiendo del objeto de estudio. Por el contrario con el CCI esto no es posible realizar, al decidir el mismo análisis estadístico si la fiabilidad es muy buena o nula con parámetros completamente teóricos.

Pero a pesar de que el EEM, MDD y PCD proporcionan un error de medición de manera interpretable, estos índices se basan en el CCI o en la  $r$  de Pearson para calcularse, lo que resulta complejo. Además estos índices solo proporcionan el error aleatorio de la medición. El método de Bland y Altman (1983)<sup>150;154;155</sup>, sin embargo es un método sencillo, y además proporciona más información que los otros índices. Esto es debido a que, además del error aleatorio con un intervalo confianza al 95% que son los límites de acuerdo (LDA), también proporciona el error sistemático (ES). Esta es la tendencia a que una de las mediciones muestre mayor o menor respecto a la otra. Si el error sistemático es 0, ó próximo a 0, se demostrará que no existe esa tendencia. Si se aleja será preciso decidir si estas diferencias son importantes desde el punto de vista práctico, al igual que con los LDA. Además MBA<sup>11</sup> se complementa con un análisis de la regresión, para comprobar que la relación entre la diferencia y la media de valores no muestra tendencia a disminuir o aumentar en función de la magnitud de la medida. Este se puede representar también de forma gráfica, lo cual permite visualizar dichos resultados de una manera visual lo que facilita la interpretación de los resultados derivados del análisis.

# 2-OBJETIVOS



## **El objetivo principal de esta tesis doctoral es:**

**Analizar cuáles son los factores de riesgo intrínsecos para sufrir lesiones agudas de tobillo en mujeres deportistas y evaluar la fiabilidad de las dos pruebas más comúnmente empleadas para determinar el estado anatómico-funcional del tobillo (como predisponente de las lesiones de tobillo).**

Este objetivo principal se desglosa en los siguientes objetivos específicos:

1. Determinar si existe algún factor de riesgo intrínseco (FR) asociado con lesiones agudas del tobillo en mujeres deportistas.
2. Determinar la fiabilidad del *Star Excursion Balance Test (SEBT)* en mujeres practicantes de actividad física, estableciendo un procedimiento sencillo que pueda ser estandarizado fácilmente.
  - 2.1. Determinar la fiabilidad del *Star Excursion Balance Test (SEBT)* a partir de una revisión sistemática (RS) con meta-análisis (MA) de los estudios que analizan la fiabilidad de esta prueba en mujeres.

Determinaremos tanto la fiabilidad intra-observador (FIA) como inter-observador (FIE).
  - 2.2. Proponer un procedimiento sencillo que pueda ser estandarizado fácilmente para el *Star Excursion Balance Test (SEBT)*.
  - 2.3. Analizar la fiabilidad intra-observador (FIA) del *Star Excursion Balance Test (SEBT)* empleando el procedimiento previamente propuesto en una muestra de mujeres deportistas.

3. Determinar la fiabilidad la prueba de dorsiflexión en apoyo con la rodilla flexionada (PDFAf) y extendida (PDFAe) en mujeres practicantes de actividad física, estableciendo un procedimiento y un método de evaluación sencillo que puedan ser estandarizados fácilmente y sean análogos para ambas pruebas:

3.1. Determinar la fiabilidad de la prueba de dorsiflexión en apoyo con la rodilla flexionada (PDFAf) y extendida (PDFAe) a partir de una revisión sistemática con meta-análisis de que analizan la fiabilidad de esta prueba en mujeres.

Determinaremos tanto la fiabilidad intra-observador (FIA) como inter-observador (FIE) de los diferentes métodos de evaluación propuestos por la literatura.

3.2. Proponer un procedimiento y un método de evaluación sencillo que pueda ser estandarizado fácilmente y que pueda ser realizado de manera similar en ambas pruebas de dorsiflexión en apoyo (PDFA).

3.3. Analizar la fiabilidad intra-observador (FIA) de la prueba de dorsiflexión en apoyo con la rodilla en flexión (PDFAf) y con la rodilla en extensión (PDFAe) empleando el procedimiento y método de evaluación que sea análogo en ambas pruebas sobre una muestra de mujeres deportistas.

3.4. Comparar los diferentes métodos de evaluación utilizados en las pruebas de dorsiflexión en apoyo (PDFA) respecto al grado de fiabilidad mostrado por cada uno.

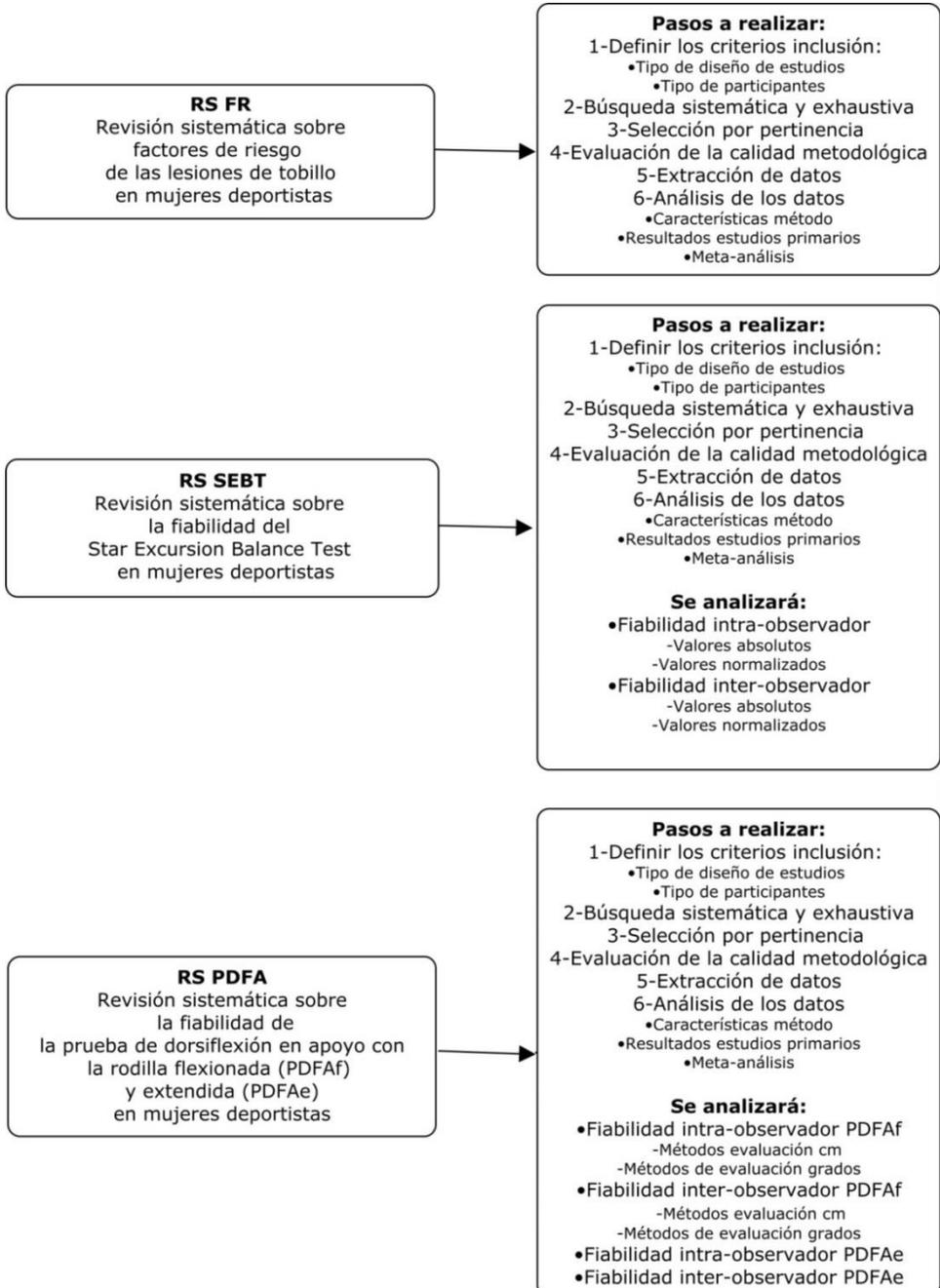
3.5. Determinar si existe alguna relación entre los resultados obtenidos mediante los diferentes métodos de evaluación en la prueba de dorsiflexión en apoyo con la rodilla flexionada (PDFAf) y extendida (PDFAe).

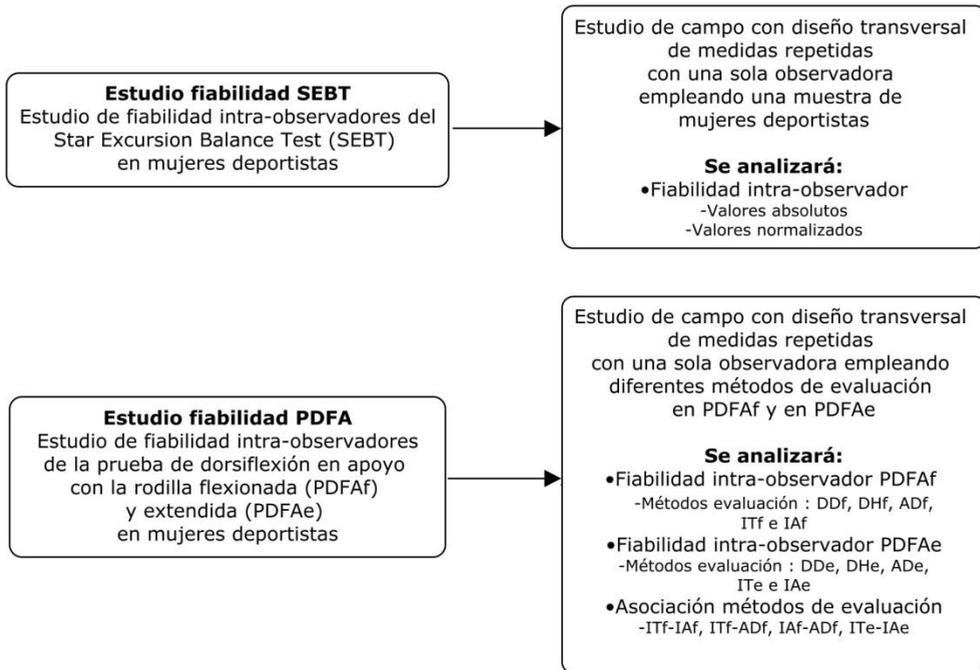
# **3-PLANTEAMIENTO DE LA INVESTIGACIÓN**



### 3.1-PLANTEAMIENTO SECUENCIAL DE LA TESIS

Para responder a los objetivos planteados en el capítulo anterior realizamos tres revisiones sistemáticas y dos estudios de fiabilidad siendo su planteamiento secuencial el siguiente:





### **3.2-CONSIDERACIONES ÉTICAS**

Tanto los estudios individuales como que se incluyen en esta tesis, así como la tesis en su conjunto fueron sometidos a la evaluación del Comité de Ética para las Investigaciones con Seres Humanos, sus muestras y sus datos (CEISH) de la UPV/EHU, habiendo sido aprobados tal y como se documenta en los informes emitidos por el CEISH que se adjunta en los anexos 3.2.1 y 3.2.2.



# 4-MATERIAL Y MÉTODOS



## **4.1-REVISIÓN SISTEMÁTICA DE LOS FACTORES DE RIESGO PARA LAS LESIONES DE TOBILLO EN MUJERES DEPORTISTAS**

### **4.1.1-CRITERIOS PARA CONSIDERAR ESTUDIOS EN LA RS**

#### **4.1.1.1-Tipo de estudios**

En la presente RS consideramos cualquier estudio que analizara los factores de riesgo intrínsecos sobre las lesiones agudas o traumáticas de tobillo en mujeres deportistas en edad fértil. Los trabajos que incluimos debían presentar los datos de manera separada en nuestro grupo de interés y en relación al tipo de lesión y a los factores de riesgo considerados.

Los estudios que incluimos fueron todos trabajos originales, por lo que presentaban por primera vez hallazgos científicos u ofrecían por primera vez una interpretación teórica de los mismos, incrementando o reestructurando los conocimientos. Sobre el tipo de trabajos estos podían ser artículos, comunicaciones de congresos, posters, tesis, tesinas o proyectos de fin de máster. No consideramos para la RS sobre FR, los trabajos de fin de grado ni los artículos en revistas no científicas, ni ningún otro trabajo que no hubiera sido revisado por pares. El tipo de diseño de los estudios fueron estudios de cohorte, de casos y controles y revisiones sistemáticas.

Consideramos factores de riesgo intrínsecos, las propias características del individuo como la edad, la composición corporal (peso, altura, IMC, porcentaje graso, medidas antropométricas...), estado de salud (historia de lesiones previas, estabilidad de la articulación...), condición física (fuerza o potencia muscular, VO<sub>2</sub> máxima, rango de movimiento articulaciones...), factores anatómicos (alineación miembros...), y nivel de destreza (técnica específica del deporte, estabilidad postural, equilibrio...)<sup>54</sup>. El tipo de lesiones registradas debían ser lesiones traumáticas o agudas, es decir una lesión producida en un momento específico e identificable<sup>156</sup>.

#### **4.1.1.2-Tipo de participantes**

Respecto a los participantes que consideramos, éstas fueron mujeres adultas que debían estar en edad fértil, es decir, que hubieran tenido la menarquía, y no hubiesen entrado en la menopausia, puesto que la gravedad y el tipo de lesión deportiva está influenciada por el grado de maduración biológica<sup>157</sup>. En el caso de no estar definido en los trabajos, debían ser mujeres en edad fértil según definición de la Organización Mundial de la Salud (OMS), es decir de 15 a 44

años<sup>158</sup>. Todas ellas debían de ser practicantes de actividad física, y dicho criterio estar definido como tal en los estudios. Del mismo modo, las participantes tenían que estar sanas en el momento de realizar las pruebas y por lo tanto, no padecer ningún tipo de lesión ni enfermedad.

## 4.1.2-MÉTODO DE BÚSQUEDA PARA IDENTIFICACIÓN DE LOS ESTUDIOS

### 4.1.2.1-Búsqueda electrónica

Para asegurar la inclusión de toda la literatura científica disponible minimizando el riesgo de sesgo de los estudios incluidos, la búsqueda se ha realizado en las bases de datos (BD) primarias y fuentes de información secundarias: Medline (desde 1946 a la 1ª semana de septiembre de 2012), Embase (desde 1974 a la 36ª semana de 2012), Cochrane Database of Systematic Reviews (desde el 2005 hasta agosto de 2012), Database of Abstracts of Reviews of Effects (tercer cuarto de 2012), Cochrane Central Register of Controlled Trials (septiembre de 2012), Sportdiscus (realizada el 12 de septiembre de 2012), Cinahl (realizada el realizada el 21 de septiembre de 2012), *Dissertations and Theses* (realizada el 2 de octubre de 2012) y Teseo (realizada entre el 7 y el 9 de octubre de 2012).

Para una búsqueda eficiente de la literatura científica, en el diseño de las estrategias se han incluido términos de lenguaje libre y del lenguaje documental o controlado (descriptores de los distintos tesauros de las bases de datos interrogadas), combinándolos convenientemente con operadores booleanos, truncamientos y otros elementos posicionales. Los términos empleados representaban los siguientes conceptos: "tobillo/*ankle*", "lesión/*injury*", "factores de riesgo/*risk factors*", "mujer/*woman*" y "deporte/*sport*" y sus sinónimos, adaptando cada búsqueda a la sintaxis y a las funcionalidades de cada base de datos, no limitando la búsqueda a ningún idioma. Cada estrategia ha sido revisada por una documentalista (MUC) a fin de garantizar el éxito de la consulta y logrando el objetivo de una búsqueda exhaustiva, sistemática y explícita. Todas las estrategias de búsqueda se pueden observar en el anexo 4.1.1.

### 4.1.2.2-Otros métodos de identificación de estudios

Además de la búsqueda electrónica, consultamos con expertos en el área y realizamos un seguimiento de referencias que los revisores identificaron como supuestamente pertinentes durante toda la revisión. Todos los trabajos identificados se incluyeron en la segunda fase de selección de documentos.

#### 4.1.2.3-Gestión de las referencias identificadas

Para la gestión de las referencias utilizamos en programa Reference Manager® versión 12 (Thomson Reuters), al cual importamos todas las referencias que encontramos mediante la búsqueda electrónica. En primer lugar, eliminamos los documentos que estaban duplicados mediante una revisora (ALR).

Consideramos como repetidos todos aquellos trabajos con el mismo título, autores, año de publicación, revista y número de revista. En los casos, en los que una misma referencia estaba referenciada en formato en papel y en formato *on-line*, también la consideramos duplicado, a pesar de que no coincidiesen el año y el número de la revista. Del mismo modo, también fueron clasificados como duplicados, cualquier referencia con mismo título y autores siendo uno de ellos *abstract* (debido a que es un poster o comunicación de un congreso o por cualquier otra razón). En estos casos, elegimos el artículo considerando el *abstract* un duplicado, siendo indiferente que hubieran sido publicados en la misma revista o en una diferente.

En segundo lugar la revisora comprobó que todas las referencias proporcionaran un *abstract*. En los casos, que no había resumen buscamos en otra BD o en el documento completo. En caso de no existir *abstract* adquirimos el documento completo para poder seleccionar o descartar el documento. En esta fase también eliminamos todas aquellas referencias incorrectamente referenciadas.

#### 4.1.3-SELECCIÓN DE ESTUDIOS

Para identificar los estudios que cumplían los criterios de inclusión de la RS (apartado 4.1.1) llevamos a cabo un proceso de selección de documentos que dividimos en dos partes. El fin de la primera selección, era descartar el mayor número de documentos posibles primando el título y resumen. En la segunda selección empleamos, el texto completo para comprobar, los estudios respondían a la pregunta de investigación de una manera más detallada.

Esta tarea fue realizada por dos revisores (ALR y ILIU) que fueron entrenados para realizar la selección de los documentos. Además diseñamos una herramienta de selección específicamente para esta tarea y realizamos una serie de pruebas con los revisores para su puesta a punto.

##### 4.1.3.1-Primera selección

Una vez confeccionada la herramienta, formamos a los dos revisores para seleccionar independientemente los documentos. A cada revisor le entregamos todos los documentos necesarios la primera selección:

1. **Documento de las referencias:** Referencia completa y resumen de cada uno de los trabajos identificados. Cada una de ellas, disponía de un número de identificación.
2. **Carpeta con los documentos completos:** De cada una de las referencias (tesis, artículos...) enumeradas de acuerdo con el documento de las referencias.
3. **Preguntas de selección (anexo 4.1.2):** Realizamos dos preguntas para comprobar si los documentos eran trabajos originales que analizaban los factores de riesgo intrínsecos para las lesiones de tobillo agudas en mujeres deportistas en edad fértil.
4. **Archivo de Excel:** Las preguntas de selección debían de contestarse en una hoja de Excel atendiendo a unos códigos predeterminados.

En base al título y resumen de cada referencia, los revisores contestaron cada una de las preguntas de selección mediante un "No" (0), "Sí" (1) o "No se puede saber" (2) en archivo de Excel.

Una vez finalizado el trabajo de selección individual, los revisores pasaron a comprobar los resultados para consensuar qué documentos pasaban a la segunda selección. En aquellos casos en los que no hubo acuerdo, recurrieron a un tercer revisor (JIE), cuyo juicio fue el que prevaleció.

#### 4.1.3.2-Segunda selección

Todos los documentos que fueron seleccionados en la primera parte de la selección y todos aquellos sin juicio definitivo debido a que la información proporcionada por el *abstract* era insuficiente, pasaron a la segunda selección. En esta fase también se incluyeron los trabajos aportados por los expertos y los identificados gracias al seguimiento de referencias. Así, entregamos a los revisores los siguientes documentos para escoger los documentos:

1. **Carpeta con los documentos completos:** De las referencias incluidas en la segunda selección.
2. **Las preguntas de la segunda selección (anexo 4.1.3):** Se diseñaron 8 preguntas para comprobar si los trabajos cumplían finalmente con todos los criterios de inclusión de la RS. Es decir, si en el estudio se analizaban los factores de riesgo intrínsecos para las lesiones de tobillo agudos en mujeres en edad fértil y los datos se presentaban de manera separada y adecuada para posterior análisis.
3. **Archivo de Excel:** Las preguntas tenían que contestarse en una hoja de Excel atendiendo a unos códigos predeterminados.

Los revisores debían contestar a las preguntas de selección utilizando el documento completo. Las respuestas podían ser “No” (0), “Sí” (1) o “No se puede saber” (2). Una vez finalizado el trabajo individual, los revisores consensuaron todas las respuestas. Los desacuerdos fueron resueltos por un tercer revisor (JIE), al igual que en la primera selección. Todos los trabajos que no cumplieron los criterios de inclusión de la RS, y aquellos en los que no se pudo determinar si los cumplían o no, quedaron excluidos de la revisión.

#### 4.1.4-RECOGIDA DE DATOS Y ANÁLISIS

Valoramos la calidad de aquellos trabajos que mediante los métodos de identificación y selección, los revisores habían determinado que habían estudiado algún factor de riesgo intrínseco para las lesiones de tobillo en mujeres adultas que practicaban actividad física. Posteriormente, extrajimos tanto sus características como sus resultados y realizamos varios análisis con los datos obtenidos de dichos estudios originales.

##### 4.1.4.1-Valoración de la calidad

Para la valoración de la calidad de los estudios utilizamos las cuestiones propuestas por Hayden et al (2006)<sup>159</sup> para estudios pronósticos. Esta valoración la realizaron de manera independiente dos revisores (ALR Y JIE), a los cuales se les entregó la siguiente documentación:

1. **Carpeta con los documentos completos:** De cada una de las referencias incluidas.
2. **Documento de referencias:** Referencia completa de cada uno los documentos incluidos para la valoración de la calidad con su correspondiente número de identificación.
3. **Plantilla con las preguntas sobre calidad (anexo 4.1.4).**
4. **Archivo de Excel:** Las preguntas debían de contestarse en una hoja de Excel atendiendo a unos códigos predeterminados en la hoja de instrucciones.

Ambos revisores respondieron las preguntas sobre la calidad de los trabajos valiéndose del documento completo mediante “No” (0), “Sí” (1) o “No se puede saber” (2) o “parcialmente (3)” en archivo de Excel. Cuando finalizaron el trabajo consensuaron todas las respuestas, y cuando no fue posible el acuerdo, nuestro protocolo estableció la consulta a un tercer revisor (ILIU), cuya decisión sería definitiva.

#### 4.1.4.2-Extracción de datos

Una revisora (ALR) extrajo las principales características de los estudios incluidos en la RS que se agruparon en variables identificativas de los trabajos, variables sobre participantes, variables sobre el método y sobre las conclusiones, en una plantilla diseñada para este fin. Todas las características extraídas fueron aquellas en relación al estudio de los FR de tobillo en mujeres practicantes de actividad física. Asimismo, extrajimos todos los FR que se habían analizado para el grupo de interés, clasificando estos en diferentes grupos: Características de las participantes, historia de lesiones previas, cualidades físicas, equilibrio, propiocepción, laxitud, fuerza específica del tobillo o MI, alineación articulación/rango de movimiento/flexibilidad y una última categoría con los FR no clasificables en ninguna de las categoría anteriores. Esta clasificación fue realizada respetando los criterios y las definiciones de los estudios originales sobre los FR que evaluaron.

#### 4.1.4.3-Análisis estadístico

En primer lugar valoramos el acuerdo existente entre revisores previo a realizar el consenso en ambas fases de selección al igual que la valoración del riesgo de sesgo. Para ello calculamos la kappa de Cohen (estimación puntual e intervalo de confianza al 95%) mediante el programa Medcalc® para windows, version 12.2.1.0 (Medcalc software, Mariakerke, Belgium). Así mismo, determinamos el acuerdo observado y su intervalo de confianza por aproximación a la normal.

Posteriormente elaboramos un análisis descriptivo de las variables extraídas de los trabajos incluidos en la RS. Para las variables cualitativas numeramos la cantidad de estudios y su correspondiente número de participantes de los cuales conocíamos dicha categoría. En lo que respecta a las variables cuantitativas estimamos la media, la DE, la máxima y la mínima de cada una de ellas. Cuando las variables hacían referencias a las características de las participantes, es decir edad, altura, peso, IMC y el tiempo de actividad física determinamos la media ponderando cada valor al número de participantes de los cuales conocíamos dicha cualidad.

En cuanto al tiempo de seguimiento durante los estudios, empleamos el número de meses como unidad para los análisis. Cuando los estudios primarios proporcionaron esta variable en otra medida diferente, se aplicaron los siguientes criterios para estimar el número de meses: cuatro semanas suponían un mes, y cada temporada se componía de 10 meses.

En lo referente a los resultados, realizamos un meta-análisis de los mismos mediante el modulo metan del paquete estadístico Stata® versión 12 (StataCorp,

College Station, Texas) siempre que esto fue posible. Es decir, cuando hubo más de un trabajo que estudió el mismo factor de riesgo siendo el *outcome* cualquier lesión de tobillo, y los resultados fueron presentados de manera analizable. Para que los resultados fueran susceptibles de incluirse en el MA, debían presentar en las variables continuas la media del grupo de lesionados y de no lesionados, siendo indiferente que proporcionaran la DE o no. Cuando los estudios no proporcionaron la DE para cada grupo, estimamos la DE como un tercio de la media observada. En cuanto a las variables dicotómicas, era necesario disponer del número de lesionados y no lesionados teniendo clasificados ambos grupos en expuestos al FR y no expuestos al FR.

Las variables continuas se meta-analizaron mediante el estadístico diferencia estandarizada de medias (DEM), mientras que los resultados dicotómicos se meta-analizaron empleando el estimador combinado riesgo relativo (RR). Los MA se realizaron con el modelo de efectos aleatorios empleando el modelo analítico de DerSimonian-Laird (DL), y de manera análoga en efecto fijos empleando la inversa de la varianza (IV) y Mantel-Haenszel (MH) según proceda (DEM o RR). Para estudiar la heterogeneidad de los estudios incluidos en MA empleamos la I cuadrada ( $I^2$ ), y seguimos los criterios de la Colaboración Cochrane<sup>160</sup> para su interpretación.

## **4.2-REVISIÓN SISTEMÁTICA DEL *STAR EXCURSION BALANCE TEST***

### **4.2.1-CRITERIOS PARA CONSIDERAR ESTUDIOS EN LA RS**

#### **4.2.1.1-Tipo de estudios**

En la presente RS consideramos cualquier documento que proporcionara resultados sobre la fiabilidad del *SEBT*, no siendo necesario que el objetivo sobre repetibilidad de la prueba estuviera definido. Con relación a las características del método científico, el diseño de los estudios debía ser descriptivo y de medidas repetidas sobre una misma muestra, evaluando el *SEBT* en más de una ocasión. Las evaluaciones podían ser realizadas por un mismo observador en diferentes ocasiones (FIA), o por diferentes observadores (FIE).

El método para evaluar la prueba debía ser la distancia desde el centro de la estrella, hasta el punto más lejano alcanzado por la participante en cada una de las líneas. El número de líneas podía variar entre estudios, la única condición para su inclusión en este trabajo, era que la fiabilidad de cada una de ellas hubiera sido estudiada de manera independiente, y por lo tanto, se proporcionara datos sobre la precisión de cada una de manera separada.

Todos los trabajos que incluimos en esta revisión fueron trabajos originales, donde se presentaban por primera vez hallazgos científicos u ofrecían una interpretación teórica de los mismos, incrementando o reestructurando los conocimientos. Estos trabajos fueron artículos de revistas científicas, comunicaciones de congresos y tesis.

#### **4.2.1.2-Tipo de participantes**

Los participantes en los trabajos que consideramos en esta revisión fueron de sexo femenino, o en su defecto, los datos exclusivos de este grupo debían poder extraerse para su análisis de manera independiente de los datos masculinos. En cuanto a las características del grupo de mujeres, debían estar en edad fértil, es decir, que hubieran tenido la menarquia, y no haber entrado en la menopausia. En el caso de no estar definido en los trabajos, debían ser mujeres en edad fértil según definición de la Organización Mundial de la Salud (OMS), es decir de 15 a 44 años. Todas ellas debían de ser practicantes de actividad física, y dicho criterio estar definido como tal en los estudios. Del mismo modo, las participantes tenían que estar sanas en el momento de realizar las pruebas y por lo tanto, no padecer ningún tipo de lesión ni enfermedad.

Al no encontrar ningún estudio que cumpliera con los criterios de inclusión referentes al tipo de participantes previamente establecido, decidimos modificar dichos criterios. De tal manera que finalmente incluimos en la revisión todos aquellos estudios que habían comprobado la fiabilidad del *SEBT* en una muestra que hubiera incluido mujeres.

## 4.2.2-MÉTODOS PARA LA IDENTIFICACIÓN DE LOS ESTUDIOS

### 4.2.2.1-Búsqueda electrónica

Realizamos diferentes búsquedas en BD primarias y secundarias para asegurar la inclusión de toda la literatura científica disponible minimizando el riesgo de sesgo en la inclusión de los trabajos incluidos. Con intención de realizar una búsqueda eficiente de la literatura científica existente, en el diseño de las estrategias incluimos términos de lenguaje libre y del lenguaje documental o controlado (descriptores de los distintos tesauros de las bases de datos interrogadas), combinándolos convenientemente con operadores booleanos, truncamientos y otros elementos posicionales.

Realizamos una primera búsqueda empleando términos que representaban los siguientes conceptos: "*Star Excursion Balance Test*", "fiabilidad/*reliability*", "deporte/*sport*" y "mujeres/*women*" y sus sinónimos, adaptando cada búsqueda a la sintaxis y a las funcionalidades de cada base de datos, no limitando la búsqueda a ningún idioma. Las bases de datos interrogadas en esta búsqueda fueron: Medline (desde 1948 a la 3ª semana de noviembre de 2011), Embase (desde 1980 a la semana nº 52 del 2011) y Sportdiscus (realizada el 26 de diciembre de 2011). Pero el número de documentos identificados mediante esta estrategia de búsqueda fue insuficiente (anexo 4.2.1), y a pesar de que ésta búsqueda respondía mejor al objetivo de la investigación, decidimos modificarla con intención de ser más eficientes en la búsqueda de la literatura.

Por lo que la finalmente empleamos un único concepto para realizar la búsqueda que fue "*Star Excursion Balance Test*". El objetivo de esta búsqueda fue el de abarcar todo la literatura disponible, conscientes de que el objetivo de fiabilidad sería en muchos trabajos secundario y que la selección de estos no podía ser realizada de otra manera que no fuera manual. Las bases de datos interrogadas para esta búsqueda final fueron: Medline (desde 1946 a la 4ª semana de diciembre de 2011), Embase (desde 1974 a la semana nº 52 del 2011), Pascal (desde 1984 a la semana nº 51 de 2011), Cochrane Database of Systematic Reviews (desde 2005 a Diciembre de 2011), Database of Abstracts of Reviews of Effects (cuarto cuatrimestre de 2011), Sportdiscus (realizada el 26 de diciembre de 2011), Cinahl (realizada el 26 de diciembre de 2011), PEDro (realizada el 2 enero

de 2012), *Dissertations and Theses* (realizada el 2 de enero de 2012) y Teseo (realizada el 2 enero de 2012). Todas las estrategias de búsqueda se pueden observar en el anexo 4.2.2.

Todas las estrategias de búsqueda empleadas en esta revisión sistemática han sido revisadas por una documentalista (MUC) a fin de garantizar el éxito de la consulta y logrando el objetivo de una búsqueda exhaustiva, sistemática y explícita.

#### **4.2.2.1-Otros métodos de identificación de estudios**

Para completar la búsqueda electrónica nos pusimos en contacto con expertos en el área y realizamos un seguimiento de referencias que se prolongó durante toda la duración de la revisión. El protocolo de esta revisión estableció la inclusión de estos documentos en la primera fase selección.

#### **4.2.2.2-Gestión de las referencias identificadas**

Una vez realizada la búsqueda electrónica, se importaron todas las referencias al gestor Reference Manager® versión 12 (Thomsom Reuters) descartándose todos los documentos repetidos con el mismo título, autores, revista, número y año. Aquellos trabajos con el mismo título y autores, siendo uno de ellos *abstract*, elegimos el artículo considerando el *abstract* un duplicado, siendo indiferente que hubieran sido publicados en la misma revista o en una diferente. Siempre que el resumen hiciera referencia a un documento más completo (tesis, artículo...) escogimos el documento más completo considerándose el resumen un duplicado.

Si el resumen suministrado por la BD, no fue informativo se buscó otro que facilitara más información en otra BD. Si aun así, la información seguía siendo escasa, se empleó el texto completo por defecto en lugar del resumen, siempre y cuando éste fuera informativo. De no ser así, la referencia quedo descartada.

Cuando eliminamos todos los documentos repetidos iniciamos la fase de selección de estudios para su análisis en la RS.

#### **4.2.3-SELECCIÓN DE ESTUDIOS**

Con los trabajos encontrados gracias a la búsqueda electrónica y a la consulta con expertos, realizamos una selección que dividimos en dos partes. El objetivo era el extraer aquellos documentos que cumplieran con los criterios de inclusión de la RS previamente definidos (apartado 4.2.1). Esta tarea se realizó de manera independiente por dos revisores (ALR y JCP). El fin de la primera selección era descartar el mayor número de documentos posibles primando el título y resumen, empleando el texto completo cuando la información no fue suficiente. En la segunda selección empleamos el texto completo para comprobar si efectivamente

los estudios respondían a la pregunta de investigación de una manera más detallada.

Diseñamos la herramienta de selección específicamente para esta tarea y la probamos con los revisores encargados de la selección. Una vez refinada, se inició el proceso de selección.

#### 4.2.3.1-Primera selección

Para realizar esta selección a cada revisor se le entregó la siguiente documentación:

1. **Documento de las referencias:** Referencia completa y resumen de cada uno de los trabajos identificados. Cada una de ellas disponía de un número de identificación.
2. **Carpeta con los documentos completos:** De cada una de las referencias (tesis, artículos...) enumeradas de acuerdo con el documento de las referencias.
3. **Preguntas de selección (anexo 4.2.3):** Formulamos una serie de preguntas para comprobar si el documento trataba sobre algún estudio que había realizado el *SEBT*, participaban mujeres, se estudiaba la fiabilidad de la prueba, y decidir si finalmente el documento se seleccionaba o no.
4. **Archivo de Excel:** Las preguntas de selección, debían de contestarse en una hoja de Excel atendiendo a unos códigos predeterminados.

Los revisores respondieron a todas las preguntas de selección de manera individual mediante un "No" (0) o un "Sí" (1). Una vez finalizado el trabajo de selección individual, los revisores pasaron a comprobar los resultados para consensuar que documentos pasaban a la segunda selección. En aquellos casos en los que no hubo acuerdo, el protocolo establecía recurrir a un tercer revisor (JIE), cuyo juicio fue el que prevalecería.

#### 4.2.3.2-Segunda selección

Todos los artículos seleccionados por consenso en la primera selección, pasaron a la segunda, que también fue realizada por dos revisores independientes a los cuales entregamos una serie de documentos:

1. **Carpeta con los documentos completos:** De cada una de las referencias que habían pasado la primera selección.
2. **Las preguntas de la segunda selección (anexo 4.2.4):** Se diseñaron dos preguntas para comprobar si los trabajos cumplían finalmente con todos los criterios de inclusión de la RS. Es decir si el estudio de fiabilidad

se realizaba sobre nuestro grupo de interés (mujeres deportistas en edad fértil) y de cada una cada línea de manera separada.

3. **Archivo de Excel:** Las preguntas tenían que contestarse en una hoja de Excel atendiendo a unos códigos predeterminados.

Los encargados de realizar la selección contestaron a todas las cuestiones mediante un "No" (0) o un "Sí" (1) en la hoja de Excel correspondiente. Una vez finalizado el trabajo individual, los revisores consensuaron todas las respuestas. Según el protocolo, los desacuerdos debían resolverse por un tercer revisor (JIE).

#### 4.2.4-RECOGIDA DE DATOS Y ANÁLISIS

En esta fase de la revisión incluimos aquellos artículos que habían estudiado la fiabilidad del *SEBT* y habían incluido mujeres en el mismo. En primer lugar realizamos la valoración de la calidad, y después procedimos a extraer y analizar las características y resultados de los estudios.

##### 4.2.4.1-Valoración de la calidad

Al igual que la fase de selección, la valoración de la calidad de los estudios se realizó mediante dos revisores que realizaron la tarea de manera independiente (ALR y JCP). Para este trabajo, adaptamos las preguntas de la plantilla de diagnóstico de CASPE<sup>161</sup> confeccionando siete preguntas sobre calidad del *SEBT*. Probamos las preguntas con los revisores encargados de la valoración de la calidad, y una vez refinada la herramienta se entregó a los revisores la siguiente documentación:

1. **Carpeta con los documentos completos:** De cada una de las referencias incluidas.
2. **Documento de referencias:** Referencia completa de cada uno los documentos incluidos para la valoración de la calidad con su correspondiente número de identificación.
3. **Plantilla con las preguntas sobre calidad (anexo 4.2.5).**
4. **Archivo de Excel:** Las preguntas debían de contestarse en una hoja de Excel atendiendo a unos códigos predeterminados en la hoja de instrucciones.

Los revisores contestaron cada una de las preguntas de calidad mediante un "No" (0) o un "Sí" (1) en archivo de Excel. Una vez realizada la valoración de la calidad de manera independiente se consensuaron todas las respuestas, así como las causas para determinar dicha respuesta. Al igual que la fase de selección, el protocolo estableció un tercer revisor en aquellos casos donde el consenso no fuera posible (JIE).

#### 4.2.4.2-Extracción de datos

Mediante una revisora (ALR) extrajimos las características de los estudios incluidos en la revisión, así como sus principales resultados. Agrupamos las características de los trabajos en variables identificativas de los estudios, sobre participantes, sobre los métodos empleados, sobre los resultados y sobre las conclusiones de los autores. Todas las características extraídas fueron aquellas en relación al estudio de fiabilidad del *SEBT*. En primer lugar extrajimos los datos para cada trabajo incluido en la revisión, y posteriormente los separamos en dos subestudios en función de la fiabilidad que habían comprobado; FIA *SEBT* y FIE *SEBT*.

#### 4.2.4.3-Análisis estadístico

En primer lugar, evaluamos el acuerdo entre ambos revisores antes de consensuar las respuestas en la primera y segunda selección y en la valoración de la calidad de los estudios. Para ello calculamos la kappa de Cohen (estimación puntual y con un intervalo de confianza al 95%), que llevamos a cabo mediante el programa Medcalc® para windows, version 12.2.1.0 (Medcalc software, Mariakerke, Belgium). También calculamos el acuerdo observado y su intervalo de confianza al 95% por aproximación a la normal. En la primera selección analizamos el acuerdo alcanzado en la cuarta pregunta (si se selecciona el estudio o no). En la segunda selección y la valoración de la calidad analizamos todas las preguntas.

Respecto a los estudios incluidos en la RS sobre *SEBT* realizamos un análisis descriptivo de las variables extraídas de los subestudios. En lo referente a las variables cuantitativas computamos el número de subestudios y su correspondiente número de participantes para cada una de las categorías. En cuanto a las variables cuantitativas, calculamos la media, la desviación estándar (DE), la máxima y la mínima de los trabajos que proporcionaban los datos en cada una de las categorías analizadas. Respecto a las variables sobre los participantes (edad, altura y peso), calculamos la media ponderando su valor al número de participantes de los cuales conocíamos dicha cualidad.

En cuanto a los resultados obtenidos en los estudios de fiabilidad analizados en esta RS, representamos gráficamente los resultados de las líneas A, PM, y PL empleando *forest plots*. Para ello los trabajos debían presentar sus resultados de manera analizable, es decir debían proporcionar datos para las líneas A, PM y/o PL de manera separada, la muestra analizada debía ser conocida y debían proporcionar resultados para cada tipo de fiabilidad de manera independiente. En las investigaciones donde se analizó la fiabilidad de más de un observador, presentamos sus datos de manera segregada. De esta manera, presentamos los CCI de cada uno de los estudios y su intervalo de confianza por un lado, y los índices dimensionales por otro (MDD y PDC). Para estos valores ubicamos su eje la

línea 0 extendiendo el valor proporcionado para cada estudio a positivo y negativo con intención de figurar el IC al 95%. Cuando el estudio primario proporcionó el EEM, calculamos la MDD de la siguiente manera:

$$MDD = 1,96 \times \sqrt{2} \times EEM$$

Realizamos un meta-análisis (MA) de los resultados siempre que esto fue posible. Para ello empleamos el modulo metan del paquete estadístico Stata<sup>®</sup> versión 12 (StataCorp, College Station, Texas). Debido al tipo de distribución del estadístico CCI que impide su MA, transformamos los CCI de los estudios individuales en tamaño del efecto mediante la transformación Z de Fisher. Calculamos la inconsistencia por los métodos habituales y realizamos MA con el modelo de efectos aleatorios utilizando el modelo analítico de DerSimonian-Laird (DL). Realizamos un MA de CCI independiente para las líneas A, PM y PL. En cada línea realizamos tres MA, uno para cada grupo de valores normalizado y absolutos, y otro para el conjunto de ambos. Se presentan los resultados del MA tras la transformación inversa (es decir en valores de CCI), para mejorar la interpretabilidad.

## **4.3-REVISIÓN SISTEMÁTICA DE PRUEBAS DE DORSIFLEXIÓN EN APOYO**

### **4.3.1-CRITERIOS PARA CONSIDERAR ESTUDIOS EN LA RS**

#### **4.3.1.1-Tipo de estudios**

En esta revisión incluimos cualquier trabajo que comprobara la fiabilidad de PDFAf o PDFaE, siendo indiferente que el objetivo de fiabilidad estuviera definido en el trabajo o no. La razón por la cual consideramos también trabajos que no habían definido como objetivo del trabajo principal comprobar la fiabilidad alguna PDFa, fue que en muchas ocasiones los trabajos de fiabilidad se consideran secundarios<sup>147</sup>, e incluyen el estudio de fiabilidad dentro del método del estudio principal.

Respeto al diseño de los estudios, incluimos cualquier trabajo descriptivo de medidas repetidas en que se realizó alguna PDFa sobre una misma muestra en más de una ocasión. Las repeticiones de las evaluaciones fueron realizadas por un mismo observador (FIA) o por diferentes observadores (FIE). El método para evaluar el desplazamiento entre el pie y el complejo tibioperoneo en las PDFa, fue la distancia angular (en grados) o lineal (en cm), sin tener en cuenta la herramienta que se empleó para dicha evaluación. Los resultados de la fiabilidad de los estudios incluidos presentaron los valores de fiabilidad de manera separada para cada una de las pruebas, métodos de evaluación y tipos de fiabilidad (FIA o FIE).

El periodo entre sesiones de evaluación no fue mayor de 15 días a no ser que el estudio pudiera garantizar que las características físicas de las participantes en los diferentes momentos de evaluación fueron las mismas.

Los trabajos que incluidos en esta revisión fueron todos originales, tales como artículos científicos, comunicaciones de congresos y tesis.

#### **4.3.1.2-Tipo de participantes**

Respecto al tipo de participantes, el protocolo de la RS estableció que éstas debían ser mujeres y, por lo tanto los resultados de la fiabilidad de este grupo debían poder analizarse de manera separada a la de los hombres. Respecto a la edad de las participantes, éstas debían estar en edad fértil, es decir, que hubieran tenido la menarquía, y no haber entrado en la menopausia. En el caso de no estar definido en los trabajos, la mujeres debían tener entre 15 y 44 años, que es la edad que la Organización Mundial de la Salud (OMS) considera edad fértil. Del mismo modo,

las participantes también debían ser practicantes de actividad física y estar sanas, es decir que no padecieran ningún tipo de lesión ni enfermedad en el momento de realizar el estudio.

Al no encontrar ningún estudio que cumpliera con los criterios de inclusión referentes al tipo de participantes previamente establecido, decidimos modificar dichos criterios. De tal manera que finalmente incluimos en la revisión todos aquellos estudios que habían comprobado la fiabilidad de alguna PDFa en una muestra mixta de mujeres y hombres.

### **4.3.2-MÉTODOS PARA LA IDENTIFICACIÓN DE LOS ESTUDIOS**

#### **4.3.2.1-Búsqueda electrónica**

Para esta revisión sistemática a fin de garantizar el éxito de la consulta y logrando el objetivo de una búsqueda exhaustiva, sistemática y explícita, diseñamos estrategias de búsqueda que fueron revisadas por una documentalista (MUC). Estas búsquedas fueron realizadas en bases de datos primarias y fuentes de información secundarias, a saber: Medline (desde 1946 a la 3ª semana de enero de 2012), Embase (desde 1974 a la semana 4ª de 2012), Cochrane Database of Systematic Reviews (desde el 2005 hasta enero de 2012), Database of Abstracts of Reviews of Effects (primer cuarto de 2012, realizada el 12 de febrero de 2012), Sportdiscus (realizada el 31 de enero de 2012), Cinahl (realizada el 9 de febrero de 2012), PEDro (realizada el 1 de febrero de 2012), Dissertations and Theses (realizada el 1 de febrero de 2012) y Teseo (realizada el 1 de febrero de 2012).

Para una búsqueda eficiente de la literatura científica, en el diseño de las estrategias incluimos términos de lenguaje libre y del lenguaje documental o controlado (descriptores de los distintos tesauros de las bases de datos interrogadas), combinándolos convenientemente con operadores booleanos, truncamientos y otros elementos posicionales. Los términos empleados representaban los siguientes conceptos: "dorsiflexión/*dorsiflexion*", "fiabilidad/*reliability*", "deporte/*sport*" y "mujeres/*women*" adaptando cada búsqueda, al tesauro, a la sintaxis y a las funcionalidades de cada BD, y sin limitar las búsqueda a ninguna fecha ni idioma. Todas las estrategias de búsqueda se pueden observar el anexo 4.3.1.

#### **4.3.2.2-Otros métodos de identificación de estudios**

Además de la búsqueda electrónica, realizamos un seguimiento de aquellas referencias que consideramos que podían responder al objetivo de la RS encontradas en los documentos recuperados que habían comprobado la fiabilidad de alguna PDFa, en los documentos no originales y todos aquellos que no se

incluyeron en la revisión. Además, también consultamos con expertos en este área de conocimiento.

#### 4.3.2.3-Gestión de las referencias identificadas

Después de realizar la búsqueda electrónica, y la consulta con expertos, importamos todas las referencias al gestor Reference Manager® versión 12 (Thomson Reuters) y descartamos todos aquellos documentos que estaban duplicados o no proporcionaban suficiente información para realizar la selección, mediante un revisor.

Consideramos duplicados todos aquellos documentos con mismo título, autores y revista. También consideramos duplicados los trabajos que tenían el mismo título y autores, siendo uno de ellos *abstract* (debido a que es un poster o comunicación de un congreso o por cualquier otra razón), y el otro artículo o tesis. En todos los casos elegimos el documento más completo, siendo indiferente que hubieran sido publicados en la misma revista o en una diferente.

Respecto a la información proporcionada por los resúmenes, cuando ésta fue escasa buscamos otro *abstract* que fuera más informativo en otra BD. Cuando no pudimos hallar ninguno, optamos por el proporcionado por el texto completo y en su defecto pasamos a seleccionar el documento basándonos en texto completo. Todos aquellos documentos que no aportaran la información suficiente para realizar la selección quedaron directamente descartados.

Una vez eliminados los duplicados, se inició el proceso de selección de los trabajos para su inclusión en la RS.

#### 4.3.3-SELECCIÓN DE ESTUDIOS

El objetivo de esta fase de la RS fue extraer aquellos documentos que cumplieran con los criterios de inclusión que previamente habíamos definido (apartado 4.3.1). Esta tarea se realizó de manera independiente por dos revisores (ALR y OUA) que realizaron la selección en dos fases. En la primera selección, descartaron o eligieron las referencias en base al título y resumen. En la segunda selección, emplearon el documento completo para seleccionar finalmente las referencias que cumplieran con los criterios de inclusión de la RS. Durante ambas selecciones, uno de los revisores (ALR) realizó el seguimiento de las referencias, para incorporar en la segunda selección aquellos documentos que no habían sido identificados mediante la búsqueda electrónica y la consulta con expertos.

Para la realización de ambas selecciones confeccionamos específicamente unas herramientas que los revisores probaron de manera independiente con una muestra de referencias. Posteriormente, los revisores se reunieron para mejorar y

refinar dicha herramienta por consenso, e iniciaron la correspondiente fase de selección.

#### 4.3.3.1-Primera selección

En esta selección a cada revisor se le entregaron los siguientes documentos:

1. **Instrucciones revisores:** Documento donde se definían los términos a emplear en la revisión y el procedimiento para realizar la selección.
2. **Documento de las referencias:** Referencia completa y resumen de cada uno de los trabajos identificados. Cada una de ellas, disponía de un número de identificación.
3. **Preguntas de la primera selección (anexo 4.3.2):** Realizamos una serie de preguntas para identificar si en los trabajos se realizaba alguna de las pruebas de interés, si se incluían participantes mujeres en el estudio y si se realizaba algún estudio de fiabilidad.
4. **Archivo de Excel:** Las preguntas de selección, debían de contestarse en una hoja de Excel atendiendo a unos códigos predeterminados que se describieron en la hoja de instrucciones.

En base al título y resumen de cada referencia, los revisores contestaron cada una de las preguntas de selección mediante un "No" (0), "Sí" (1) o "No se puede saber" (2) en archivo de Excel. En caso de responder "No" a alguna de las preguntas los revisores pasaron directamente a la última de las preguntas, no seleccionando el documento. Todos los documentos que fueron seleccionados en esta fase y todos aquellos sin juicio definitivo de selección, debido a que la información proporcionada por el *abstract* era insuficiente, pasaron a la segunda selección.

Cuando cada revisor de manera independiente finalizó el trabajo de selección, pasaron a comparar los resultados entre ellos. Se evaluó el acuerdo entre ambos revisores antes de consensuar las respuestas. Para valorar el acuerdo empleamos el porcentaje de acuerdo y la kappa de Cohen, calculando para ambos la estimación puntual y por intervalo de confianza al 95%), que llevamos a cabo mediante el programa Medcalc<sup>®</sup> para windows, version 12.2.1.0 (Medcalc software, Mariakerke, Belgium).

Cualquier desacuerdo entre revisores fue resuelto por procedimiento de consenso, en caso de no poder llegar a un acuerdo, el protocolo de la revisión establecía recurrir a un tercer revisor (JIE), cuyo juicio fue el que prevaleció.

#### 4.3.3.2-Segunda selección

Los revisores recibieron la siguiente documentación para realizar la segunda selección:

1. **Instrucciones revisores:** Documento donde se definían los términos a emplear en la revisión y el procedimiento para realizar la segunda selección.
2. **Carpeta con los documentos completos:** De cada una de las referencias incluidas en esta selección, cada una con un número de identificación.
3. **Documento de referencias:** Referencia completa de cada uno los documentos incluidos en esta selección con su correspondiente número de identificación.
4. **Preguntas de la segunda selección (anexo 4.3.3):** Se diseñaron tres preguntas para cada uno de las dos pruebas (PDFAf y PDF Ae) para comprobar si los trabajos cumplían finalmente con todos los criterios de inclusión de la RS. Estas preguntas se referían a sí se comprobaba al fiabilidad de la prueba, y sí se presentaban los resultados de manera exclusiva del grupo de interés (mujeres sanas practicantes de actividad física en edad fértil).
5. **Archivo de Excel:** Las preguntas de selección debían de contestarse en una hoja de Excel atendiendo a unos códigos predeterminados en la hoja de instrucciones.

Las preguntas de ésta selección debían contestarse con el documento completo mediante un "No" (0) o un "Sí" (1). Cuando los revisores respondieron que no a alguna de las cuestiones, pasaron directamente a la última de las cuestiones de cada una de las pruebas (PDFAf o PDF Ae). Al igual que en la primera selección, cuando los revisores finalizaron la selección individual, pasaron a consensuar todas las respuestas. En caso de desacuerdo, el protocolo estableció la consulta a un tercer revisor cuyo juicio fue el que prevaleció (JIE).

#### 4.3.4-RECOGIDA DE DATOS Y ANÁLISIS

En esta parte la revisión incluimos aquellos trabajos que habían estudiado la fiabilidad alguna PDFa y valoramos de manera separada la calidad de los estudios sobre PDFAf y PDF Ae.

##### 4.3.4.1-Valoración de la calidad

Para este trabajo dos revisores (ALR y OUA) respondieron de manera independiente y valiéndose de los documentos completos, una serie de preguntas para valorar la calidad de los estudios. Creamos esta plantilla específicamente para la tarea, adaptando las preguntas de plantilla de diagnóstico de CASPe<sup>161</sup> a las peculiaridades de los estudios de fiabilidad y de las características de las pruebas (anexo 4.3.4). Las preguntas fueron probadas por los revisores, y una vez

mejoradas y aclaradas todas las dudas, se inició la valoración de la calidad de los estudios, para lo cual se entregó a cada revisor los siguientes documentos:

1. **Carpeta con los documentos completos:** De cada una de las referencias incluidas.
2. **Documento de referencias:** Referencia completa de cada uno los documentos incluidos para la valoración de la calidad con su correspondiente número de identificación.
3. **Plantilla con las preguntas sobre calidad (anexo 4.3.4).**
4. **Archivo de Excel:** Las preguntas debían de contestarse en una hoja de Excel atendiendo a unos códigos predeterminados en la hoja de instrucciones.

Los revisores contestaron cada una de las preguntas de calidad mediante un "No" (0) o un "Sí" (1) en archivo de Excel. Una vez realizada la valoración de la calidad de manera independiente, se consensuaron todas las respuestas, así como las causas para determinar dicha respuesta. Al igual que la fase de selección, el protocolo estableció un tercer revisor (JIE) en aquellos casos donde el consenso no fuera posible.

#### 4.3.4.2-Extracción de datos

Una revisora (ALR) pasó a extraer las principales características de los trabajos y sus resultados en relación al estudio de fiabilidad del PDF/A. Las características de los estudios se agruparon en variables identificativas de los estudios, del diseño y objetivo, sobre participantes, sobre los métodos empleados, de resultados y de conclusiones de los autores. Estos datos se extrajeron de manera general para cada trabajo, para posteriormente separarlos en función de la prueba de DF realizada y del tipo de fiabilidad estudiada, creándose 4 subgrupos, a saber; estudios de FIA PDF/Af, FIA PDF/Ae, FIE PDF/Af y FIE PDF/Ae. En los casos donde se emplearon dos métodos de evaluación o más, consideramos cada uno de ellos como subestudios diferentes dentro de cada su grupo.

#### 4.3.4.3-Análisis estadístico

En primer lugar, estudiamos el acuerdo obtenido por los revisores previo a consensuar las respuestas en la fase de selección así como en la valoración de la calidad. Para valorar el acuerdo utilizamos la kappa de Cohen (estimación puntual y por intervalo de confianza al 95%), que llevamos a cabo mediante el programa Medcalc® para windows, version 12.2.1.0 (Medcalc software, Mariakerke, Belgium). Así mismo, calculamos el acuerdo observado y su intervalo de confianza por aproximación a la normal. Para este análisis consideramos las respuesta "No" y "No se puede saber" como iguales.

Posteriormente, realizamos un análisis descriptivo de las variables tanto cualitativas como cuantitativas de cada uno de los subestudios identificados. Respecto a las variables cualitativas calculamos el número de estudios así como el número de participantes para cada una de las categorías. Para el total, calculamos la proporción de estudios y participantes respecto al total de trabajos. En cuanto a las variables cuantitativas calculamos la media, la DE, la máxima y la mínima de los trabajos que proporcionaban los datos en cada una de las categorías analizadas.

En cuanto a los resultados de fiabilidad de los estudios incluidos en la RS presentamos estos gráficamente y de manera separada para FIA PDFAf, FIA PDFAe, FIE PDFAf y FIE PDFAe mediante *forest plot* los diferentes índices de fiabilidad empleados. En las investigaciones donde se analizó la fiabilidad de más de un observador, presentamos sus datos de manera segregada. Así, representamos en gráficos por un lado la estimación puntual del CCI y su intervalo de confianza al 95% siempre que fue posible, para los métodos de evaluación empleando grados y cm como unidades de medición. Por otro, representamos los límites de acuerdo (LDA) y error sistemático (ES) del método Bland y Altman y la mínima diferencia detectable (MDD). Para el MDD ubicamos el eje de la línea en 0, extendiendo las líneas a valores positivos y negativos para representar el IC al 95%. Para estas figuras dividimos los resultados de los métodos de evaluación que utilizaron como unidad de medición los cm, y los grados. En los casos que los autores no proporcionaron el MDD, calculamos mediante el EEM el MDD de la siguiente manera:

$$MDD = 1,96 \times \sqrt{2} \times EEM$$

En lo referente a los resultados, realizamos un meta-análisis de los mismos mediante el módulo metan del paquete estadístico Stata<sup>®</sup> versión 12 (StataCorp, College Station, Texas) siempre que esto fue posible. Dada el tipo de distribución del estadístico CCI que impide su MA, transformamos los CCI de los estudios individuales en tamaño del efecto mediante la transformación Z de Fisher. Realizamos los MA calculando la inconsistencia por los métodos habituales, y utilizamos el modelo de efectos aleatorios y el modelo analítico de DerSimonian-Laird (DL). Realizamos un meta-análisis de CCI para cada subgrupo analizado; FIA PDFAf, FIA PDFAe, FIE PDFAf y FIE PDFA. Primero analizamos los métodos de evaluación en grados y en cm de manera separada y posteriormente analizados ambos de manera conjunta. Se presentan los resultados del MA tras la transformación inversa, para mejorar la interpretabilidad, es decir en CCI.

## **4.4-ESTUDIO DE FIABILIDAD DEL STAR EXCURSION BALANCE TEST**

### **4.4.1-TIPO DE ESTUDIO**

La fiabilidad test-retest se evaluó mediante un diseño transversal de medias repetidas con una sola observadora que aplicó dos veces el *Star Excursion Balance Test (SEBT)*, con una separación máxima de veintidós días sobre una muestra de jugadoras de voleibol.

### **4.4.2-PARTICIPANTES**

Esta investigación se llevó a cabo con jugadoras de voleibol pertenecientes a *Getxoko Aixerrota Boleibol Kirol Taldea* de los equipos senior de primera división de liga vasca (LV) y los tres equipos infantiles de liga escolar (LEI1, LEI2 y LEI3), que eran los equipos de mayor edad y mayor volumen de entrenamiento dentro del club la temporada 2007-2008. Todas jugadoras que estaban entrenando y jugando partidos en cada uno de los equipos previamente mencionados, fueron susceptibles de ser incluidas en el estudio. De este grupo de jugadoras, únicamente incluimos a aquellas que habían tenido la primera regla al estar la ubicación, la gravedad y el tipo de lesión deportiva influenciadas por el grado de maduración biológica<sup>157</sup>. Como criterios de exclusión establecimos los siguientes: haber sufrido un esguince de tobillo en las seis semanas previas al estudio, haber sufrido una contusión cerebral tres meses antes de la realización de la prueba, padecer desórdenes vestibulares o enfermedades crónicas de oído y sufrir cualquier tipo de enfermedad o lesión que no les permitiese entrenar con normalidad en el momento de realizar la prueba.

Todas las participantes y en caso de ser menores, sus tutores, fueron informados sobre la naturaleza del proyecto, y firmaron un consentimiento escrito (anexo 4.4.1). El club al que pertenecían las jugadoras también fue informado y firmó la correspondiente autorización, al igual que el centro donde se realizó el estudio (anexo 4.4.2). Los entrenadores del club también fueron informados, y el desarrollo de los entrenamientos no se modificó debido al estudio.

Toda la información obtenida de las participantes fue tratada de acuerdo con la Declaración de Helsinki y la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal. Este estudio fue aprobado por el Comité de Ética para las Investigaciones relacionadas con Seres Humanos (CEISH) de la EHU/UPV (anexo 3.2.2).

#### 4.4.3-INSTRUMENTOS

En la revisión previa a la realización del trabajo de campo, observamos la existencia de diferentes formas de realizar el *SEBT*. Al ser una prueba que todavía no estaba estandarizada, decidimos emplear el procedimiento más sencillo y que a su vez pudiera proporcionar una buena fiabilidad tratando de minimizar las posibles fuentes de variabilidad.

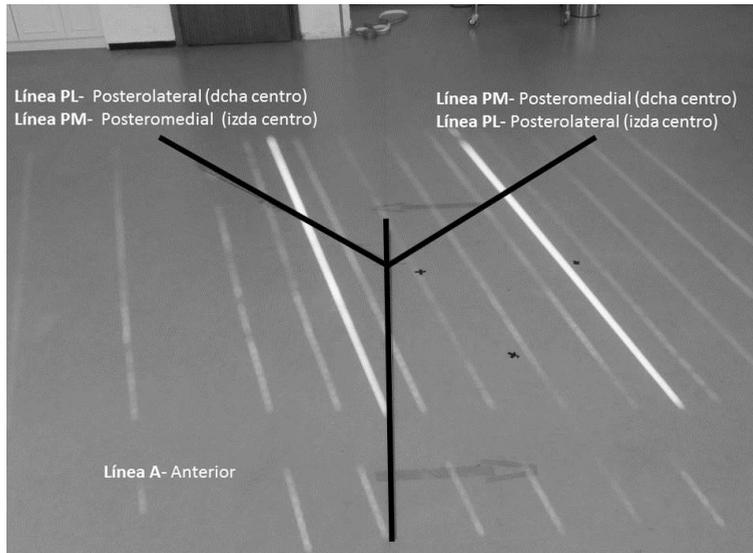


Figura 4.4.1. Colocación de las líneas A, PM y PL en el *SEBT*

Para la realización del *SEBT* escogimos tres líneas (A, PM y PL) del total de ocho posibles, porque en los trabajos sobre predicción del riesgo de lesión<sup>58</sup> y en otros sobre lesiones del miembro inferior y tobillo se había encontrado más adecuado utilizar estas<sup>57;79</sup>. Para la confección de la prueba, en primer lugar, pegamos en el suelo una cinta de 160 cm de longitud (línea A) y seguidamente, a 135°, tanto a la izquierda como a la derecha, pegamos otras dos líneas de la misma longitud. Estas dos líneas fueron la PL y la PM, la referencia de medial o lateral se consideró siempre en relación al pie probado (figura 4.4.1). Se decidió que a diferencia de otros trabajos<sup>67;68;72</sup> las líneas tuvieran los centímetros y milímetros marcados, para eliminar el posible error al observar las marcas y medirlas (desde el centro al lugar donde alcanza la participante). De esta manera también se reducía el tiempo de realización del *SEBT*, ya que con las líneas lisas, el observador debería medir con una cinta métrica las marcas desde el centro de la estrella hasta el punto alcanzado por la participante, en al menos dos veces. En el centro longitudinal de cada una de las líneas se dibujó una línea recta con rotulador imborrable. El lugar

donde las líneas se juntaban se marcó con un punto negro, éste se consideró el centro de la estrella. Prolongamos la línea A 30 cm hacia atrás, donde marcamos también una línea longitudinal central.

#### 4.4.4-PROCEDIMIENTOS

Realizamos el *SEBT* en dos momentos diferentes con una separación media de 6,6 (DE  $\pm 6,5$ ) días. La recogida de datos comenzó el 8 de mayo de 2008 y finalizó el 16 de mayo del 2008. El mes previo al comienzo del estudio, es decir en abril, las participantes rellenaron un cuestionario para comprobar si eran susceptibles de tomar parte en la investigación. Para ello rellenaron un cuestionario (anexo 4.4.3), y ese mismo día se recogieron las características antropométricas de las participantes. Medimos la altura con las participantes de pie valiéndonos de una cinta métrica adherida a la pared. También cuantificamos la longitud de las piernas, estando las participantes decúbito supino. La medición la realizamos con una cinta métrica milimetrada en material de plástico inextensible (*Siber Hegner Maschinen*, Zürich, Suiza), desde la espina ílica anterosuperior hasta el borde más distal del maléolo medial. Determinamos la pierna dominante preguntando a las participantes con que pierna chutarían lo más fuerte posible un balón.

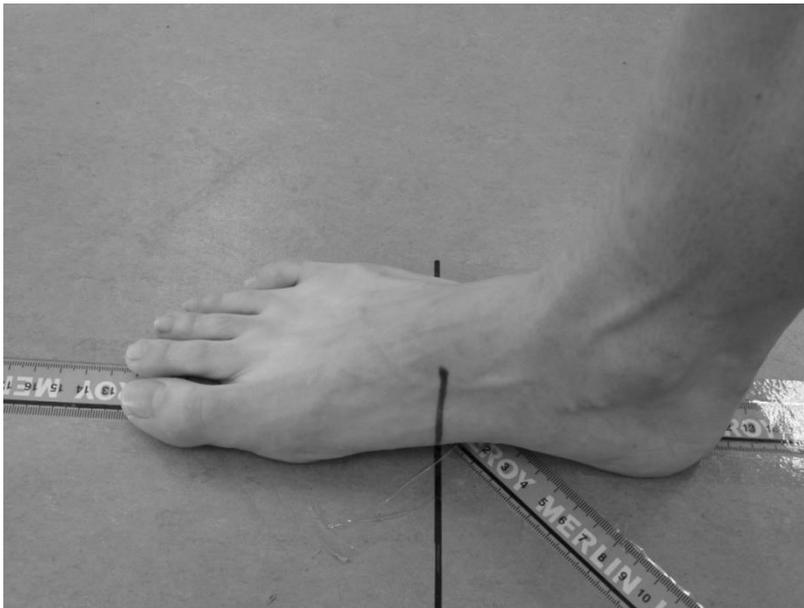


Figura 4.4.2. Colocación del pie en el *SEBT* con el centro geométrico del pie en el centro de la estrella.

Para la realización de la prueba la participante se descalzó para facilitar la colocación del pie, y eliminar la variabilidad debida al uso del calzado. Para iniciar el *SEBT*, se colocó el pie de la pierna a probar con el centro geométrico del pie o con el primer dedo del pie en el centro de la estrella. Si la prueba se realizaba con centro del pie, se medía la longitud del pie y se colocaba la mitad del pie en el punto central, quedando el talón y el primer dedo del pie alineados en la línea A (figura 4.4.2). Si la posición a medir era el primer dedo del pie, el primer dedo del pie se colocaba en el centro de la estrella, estado también el talón y el primer dedo del pie alineados en la línea A (figura 4.4.3). En cualquiera de los dos casos el pie libre de la participante se colocó en paralelo al testado en una posición cómoda para la participante, y las manos se agarraron a la cintura, siendo considerada esta la posición de inicio (figura 4.4.4). Decidimos colocar las manos en la cintura para realizar el *SEBT* para facilitar el control por parte del observador, así se si la participante perdía el equilibrio era fácilmente identificable porque soltaba las manos de la cintura.

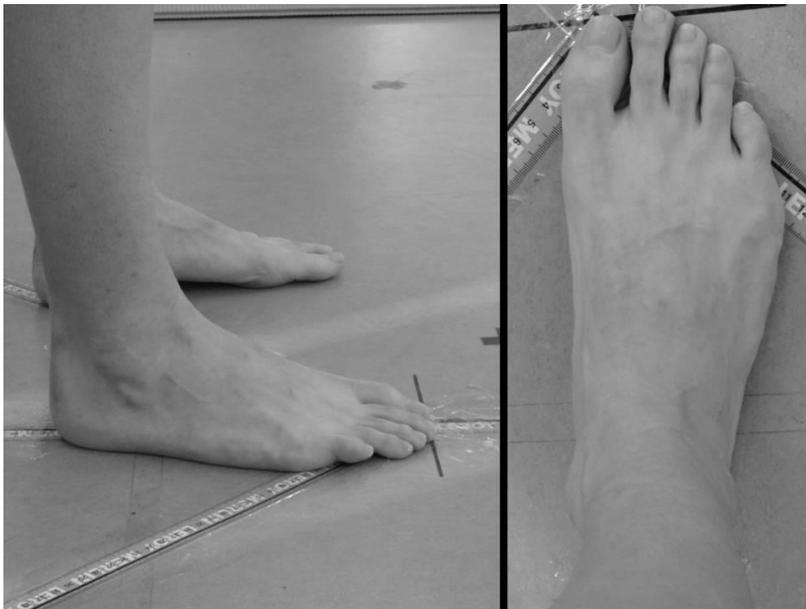


Figura 4.4.3. Colocación del pie en el *SEBT* con el primer de dedo del pie en el centro de la estrella.

Partiendo de la posición de inicio (figura 4.4.4) se le pidió a la participante que tocara suavemente la línea correspondiente, con la parte más distal del pie, y que retornara a la posición de inicio. El intento no se consideró válido y se repitió si la participante: soltaba las manos de la cadera, movía o levantaba el pie de apoyo en

algún momento de la prueba, apoyaba el pie libre en el suelo, no tocaba la línea, perdía el equilibrio desde que abandonaba la posición de inicio hasta que la volvía a retomar, no era capaz de mantener la posición de inicio al menos un segundo después de retornar a la misma. En caso de errar en el intento, se recolocaba a la participante en la posición de inicio y se volvía a repetir el intento. La investigadora marcó hasta dónde había llegado la participante con un rotulador, y si el intento era bueno, cuando la participante volvía al centro de la estrella, la investigadora leía la marca (a la altura de la línea longitudinal central) y anotaba el resultado.



Figura 4.4.4. Posición de inicio de las participantes en el SEBT.

El primer día de prueba, y en caso de ser requerido por las participantes, también el segundo, explicamos la prueba a cada una de ellas. Se explicaron las normas verbalmente, realizamos una demostración y aclaramos todas las dudas. Una vez entendido el funcionamiento del SEBT, la participante realizó un calentamiento de seis intentos seguidos por línea<sup>65</sup>, en cada una de las tres direcciones con cada pierna. Para que el intento del calentamiento fuera válido, debía de ser realizado correctamente. El objetivo del calentamiento era minimizar los efectos del aprendizaje<sup>65</sup>. Antes de iniciar el SEBT la participante descansó cinco minutos<sup>79</sup>.

En la prueba cada participante hizo tres intentos seguidos a cada línea<sup>65</sup> y entre cada intento se esperaron diez segundos aproximadamente, en los cuales la participante salía de la estrella para que la investigadora observase la marca

obtenida y la registrase. Una vez anotados los datos, la participante se colocaba en la posición de inicio con la ayuda de la investigadora, y realizaba el siguiente intento.

El orden para determinar con qué pie realizar el *SEBT* en primer lugar, si la prueba se hacía con el centro geométrico del pie o con el primer dedo del pie en el centro de la estrella, y el orden en el que se debían tocar las diferentes líneas en el test *SEBT*, se estableció aleatoriamente. Para ello utilizamos una tabla de permutación de seis elementos tomados de seis en seis, y eligiéndose los números de forma consecutiva para cada participante. Así, el primer número determinó con qué pie se comenzaba, el segundo la parte del pie que se colocaba en la estrella, y el tercero y el cuarto el orden de las líneas en *SEBT*. La última línea a tocar se establecía por eliminación. En caso de que los números repitieran la misma línea cogimos el siguiente.



Figura 4.4.5. Participante tocando la línea anterior (A) durante el *SEBT*.

En caso de que el primer número fuera impar se probó en primer lugar el miembro derecho, y cuando fue par, el izquierdo. Para determinar qué parte del pie colocar en el centro de la estrella, cuando fue impar se colocó el primer dedo, y cuando fue par el centro geométrico del pie. El orden por el cual debían de tocar las líneas en el *SEBT* cuando el número fue 1 ó 2 la línea a tocar debía ser la A (figura 4.4.5), si el número fue el 3 ó el 4 la PL (figura 4.4.6), y si fue 5 ó 6 la PM (figura 4.4.7).

El segundo día de test también se asignó aleatoriamente con qué pie comenzar el test y el orden de las líneas a tocar, pero no qué parte del pie colocar en el centro del *SEBT*. Cada participante colocó el pie tal y como asignado el primer día. El orden de las líneas y de las piernas en el calentamiento fue el mismo en el que se debía realizar la prueba cada día.

Tanto el primer día como el segundo día de prueba se pasó a las jugadoras un cuestionario previo (anexo 4.4.4), haciendo referencia a las posibles variaciones intrínsecas de las participantes de un momento a otro, con intención de garantizar la similitud de los momentos evaluados. Anotamos los resultados del *SEBT* de cada uno de los días de medición en hojas separadas, de tal manera que la investigadora no conocía las marcas obtenidas por las participantes el primer día de prueba.



Figura 4.4.6. Participante tocando la línea posterolateral (PL) durante el *SEBT*.

#### 4.4.5-ANÁLISIS ESTADÍSTICO

Para el análisis de los datos, empleamos por un lado los datos absolutos obtenidos en cada línea, y por otro lado, los datos normalizamos en relación a la longitud de la pierna de alcance<sup>77;90;100</sup>. En este valor normalizado se expresa la distancia como porcentaje de la longitud de la pierna. Para ello dividimos las distancia obtenida en cada una de las líneas, entre la longitud de la pierna de alcance y la multiplicamos

por cien<sup>77;90</sup>. En todos los análisis utilizamos como estimadores la media y la máxima de los 3 intentos realizados en cada línea y día.

Estudiamos la fiabilidad del *SEBT* mediante el método propuesto por Bland y Altman (1995)<sup>162</sup> que realizamos con el programa Medcalc<sup>®</sup> para windows, version 12.2.1.0 (Medcalc software, Mariakerke, Belgium), calculando el error sistemático (ES), límite de acuerdo (LDA), límite de acuerdo inferior, límite de acuerdo superior e intervalos de confianza para el ES y el LDA inferior y superior. Además, realizamos una regresión lineal, para completar el análisis del MBA y con intención de comprobar que la relación entre la diferencia y la media de valores no muestra tendencia a disminuir o aumentar en función de la magnitud de la medida. La regresión la calculamos mediante el programa SPSS versión 19.0.01 (IBM<sup>®</sup>) empleando como la variable predictora la media de cada par de mediciones (día 1 y día 2) de cada una de las participantes y la variable dependiente la diferencia entre cada par de mediciones de cada una de los participantes. El umbral para la significación estadística en la regresión se determinó a priori en  $p < 0,05$ .



Figura 4.4.7. Participante tocando la línea posteromedial (PM) durante el *SEBT*.

Además, calculamos de manera análoga a otros trabajos sobre la fiabilidad del *SEBT* y a pesar de las consideraciones metodológicas del presente trabajo, los siguientes índices de fiabilidad: el Coeficiente de Correlación Intraclase (CCI), el

Coefficiente de Correlación de Pearson ( $r$  de Pearson), el Error Estándar de la Medida (EEM), la mínima diferencia detectable (MDD) y el punto de corte (PDC) descrito por Jacobson y Truax (1991)<sup>163</sup> y utilizado por Sabin (2011)<sup>90</sup>.

Para calcular el CCI y la  $r$  de Pearson utilizamos el programa SPSS. En el caso del CCI aplicamos el modelo alfa, tipo de CCI mixto, y el acuerdo absoluto, proporcionando los valores individuales.

El EEM utilizado por Munro et al. (2010)<sup>72</sup> (EEM) se define de la siguiente manera:

$$EEM = DE \times \sqrt{1 - CCI}$$

La mínima diferencia detectable (MDD) se define de la siguiente manera<sup>72</sup>:

$$MDD = 1,96 \times \sqrt{2} \times EEM$$

El punto de corte (PDC)<sup>90;163</sup> se define:

$$PDC = 1,96 \times \sqrt{2} \times EEM + (ES)$$

Para calcular el EEM en el caso del punto de corte (PDC) empleamos la  $r$  de Pearson en lugar del CCI al igual que Sabin et al. (2010)<sup>90</sup> y así poder comparar los resultados.

## **4.5-ESTUDIO DE FIABILIDAD DE LA PDFAf Y LA PDF Ae EMPLEANDO DIFERENTES MÉTODOS DE EVALUACIÓN**

### **4.5.1-TIPO DE ESTUDIO**

La fiabilidad test-retest se evaluó mediante un diseño transversal de medidas repetidas con una sola observadora que aplicó la prueba de dorsiflexión en apoyo tanto con la rodilla flexionada (PDFAf) como con la rodilla en extensión (PDF Ae) con una separación máxima de siete días sobre una muestra de jugadoras de voleibol, en dos ocasiones, una al principio de temporada y otra al final de la misma.

### **4.5.2-PARTICIPANTES**

El presente estudio se llevó a cabo con jugadoras pertenecientes al club de voleibol *Getxoko Aixerrota Boleibol Kirol Taldea* de los equipos de categoría senior de primera (LV1) y segunda división de liga vasca (LV2), y el equipo de liga escolar en la categoría cadete (LEC). Escogimos estos equipos dentro del puesto que eran los tres con jugadoras de mayor edad, y mayor carga de entrenamiento dentro club.

Al inicio del estudio, todas jugadoras que estaban entrenando y jugando partidos en cada de uno de los equipos, fueron susceptibles de entrar en el estudio. De este grupo de jugadoras, sólo incluimos a aquellas que habían tenido la primera regla, ya que la ubicación, la gravedad y el tipo de lesión deportiva está influenciada por el grado de maduración biológica<sup>157</sup>. También excluimos a las jugadoras que habían tenido un esguince de tobillo en las seis semanas previas al estudio, o que habían sufrido una contusión cerebral tres meses antes de la realización de la prueba, así como las que padecían desórdenes vestibulares ó enfermedades crónicas de oído. Tampoco incluimos a aquellas jugadoras que padecieran algún tipo de enfermedad que no les permitiese entrenar con normalidad. Ninguna de las participantes había sido operada del miembro inferior.

Todas las participantes y en caso de ser menores, sus tutores, fueron informados sobre la naturaleza del proyecto, y firmaron un consentimiento escrito (anexo 4.5.1). El club al que pertenecían las jugadoras también fue informado y firmó la correspondiente autorización al igual que el centro donde se realizó el estudio (anexo 4.5.2) Los entrenadores conocían la realización del estudio y la dinámica de los entrenamientos no fue modificada debida al mismo.

Toda la información obtenida de las participantes fue tratada de acuerdo con la Declaración de Helsinki y la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal. Este estudio fue aprobado por el Comité

de Ética para las Investigaciones con Seres Humanos (CEISH) de la UPV/EHU (anexo 3.2.2).

### 4.5.3-INSTRUMENTOS

Para la realización de las pruebas utilizamos una pared lisa que en relación con el suelo mantenía un ángulo de  $90^\circ$ . Pegamos en el suelo 2 m de cinta milimetrada partiendo desde la pared y en perpendicular a esta. En la pared, a continuación de la línea del suelo, pegamos otra línea de 100 cm que partía desde el suelo (figura 4.5.1). En el medio de la cinta del suelo se dibujó una línea longitudinal negra con un rotulador imborrable, para fijar el centro de la línea. Todo ello se tapó con un celo transparente más ancho, para evitar que la cinta se levantase e impedir que la línea central se borrara. También utilizamos para evaluar la DF máxima en realizar la PDFAf y la PDF Ae un inclinómetro Unilevel (ISOMED, Inc., Portland, OR).



Figura 4.5.1. Colocación de las líneas para la PDFAf y PDF Ae.

### 4.5.3-PROCEDIMIENTOS

Realizamos la PDFAf y la PDF Ae en dos días diferentes con una separación máxima de siete días, en dos momentos diferentes, uno al principio de la temporada 2010-2011 (PDFA1), y el segundo al final de la misma (PDFA2) (figura 4.5.2).



Figura 4.5.2. Diagrama temporal del estudio de fiabilidad de PDFA.

El primer día de prueba se pasó a las participantes un cuestionario para determinar si eran susceptibles de entrar en el estudio con preguntas referentes a los criterios de inclusión y exclusión (anexo 4.5.3). Además en ese cuestionario, también se formularon preguntas sobre las características de las participantes. En caso de ser susceptibles a ser incluidas en el estudio tomamos las características antropométricas de las participantes. Medimos la altura con las participantes de pie valiéndonos de una cinta métrica adherida a la pared y pesamos a todas las participantes. También cuantificamos la longitud de las piernas, estando ellas decúbito supino. La medición la realizamos con una cinta métrica milimetrada en material de plástico inextensible (*Siber Hegner Maschinen, Zürich, Suiza*), desde el espina iliaca anterosuperior hasta el borde más distal del maléolo lateral.

El orden para determinar qué prueba se hacía en primer lugar y con qué pie comenzaban las pruebas se estableció aleatoriamente. Para ello utilizamos una tabla de permutación de seis elementos tomados de seis en seis, eligiéndose los números de forma consecutiva para cada participante. Así, el primer número determinó qué prueba se realizaba en primer lugar y el segundo con qué pie comenzar. Si el primer número fue par realizamos la PDFAf primero, y si fue impar la PDFAe. En el caso del segundo número, si fue par empezábamos ambas pruebas con la pierna derecha, y en caso de ser impar con la izquierda. Tanto el primer como el segundo día el orden de las pruebas y de las piernas fue el mismo para cada participante.

Al principio de cada sesión de prueba se marcaron una serie de puntos anatómicos para la correcta ubicación de los segmentos corporales y del inclinómetro durante

la prueba. Con la participante en pie, marcó el centro del talón y el punto a 7,5 cm de la parte más distal del talón a lo largo del tendón de Aquiles (figura 4.5.3). Estando la participante decúbiteo supino la investigadora realizó una marca 15 cm por debajo de la tuberosidad tibial sobre el borde anterior de la tibia, y otra en el borde supero externo de la rótula (figura 4.5.3).

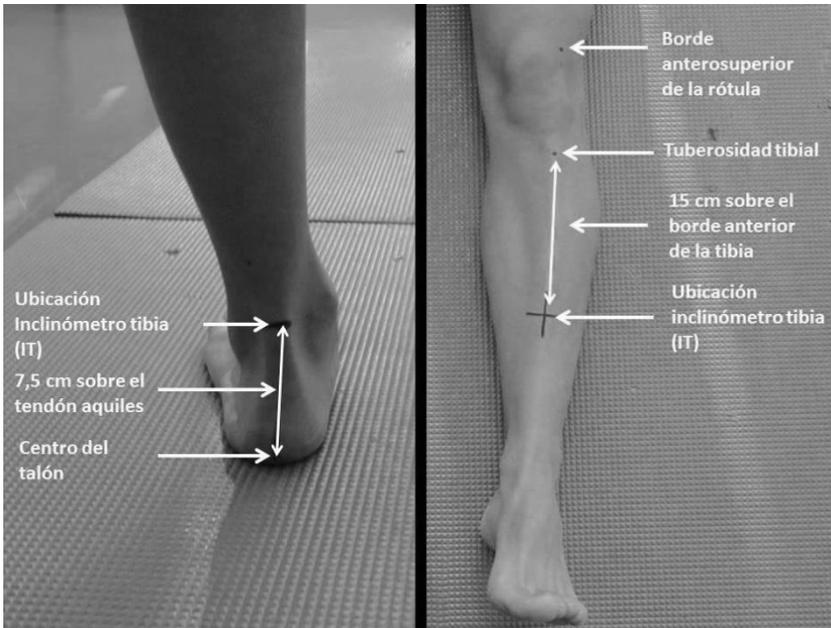


Figura 4.5.3. Marcas previas a la realización de las pruebas.

Antes de realizar la PDFAf y la PDF Ae la investigadora explicó el procedimiento de cada prueba, y realizó una demostración de cada una. En caso de que el participante lo requiriese se aclararon todas las dudas que pudiera haber. No se realizó calentamiento previo puesto que los estudios sobre fiabilidad de las pruebas no habían definido si se realizaba o no, y los que lo habían hecho habían realizado calentamientos diferentes (tabla 5.2.3).

Realizamos todas las mediciones siempre en el mismo lugar, con las mismas herramientas y con una temperatura estable<sup>164</sup>. Además, medimos a cada participante a la misma hora en su par de mediciones. Con intención de garantizar la similitud de las participantes en las mediciones, antes de realizar la segunda medición pasamos a las jugadoras otro cuestionario. En él incluimos cuestiones para comprobar que seguían cumpliendo los criterios de inclusión, y que nada extraordinario que pudiera afectar a la fiabilidad de prueba había ocurrido (anexo 4.5.4). La investigadora anotó los resultados de las pruebas de cada uno de los

días de medición en hojas separadas, de tal manera que no conocía las marcas obtenidas por las participantes el primer día de prueba durante la recogida de las marcas del segundo día.

#### 4.5.4.1-Prueba de dorsiflexión en apoyo con la rodilla flexionada (PDFAf)

Para realizar la PDFAf la participante colocó el pie del tobillo a medir con el centro del talón que previamente la investigadora había marcado, y el primer dedo del pie alineados con la línea negra marcada en la cinta (figura 4.5.4).



Figura 4.5.4. Colocación del pie en las PDFAf.

Con el pie sobre la línea, la jugadora realizó una flexión de rodilla, tobillo y cadera, con el objetivo de tocar con la rodilla la línea marcada en la pared pero sin levantar el talón del pie evaluado (figura 4.5.5). Se permitió a la participante apoyarse en la pared para mantener el equilibrio durante la prueba, y colocar la pierna no testada en una posición cómoda. De esta manera, la persona realizó todos los intentos que consideró oportunos para encontrar la máxima distancia posible desde la pared hasta el pie, manteniendo el talón en contacto con el suelo y teniendo la rodilla tocando la línea marcada en la pared. En cada intento la participante tenía que despegar la rodilla de la pared para la que marca realizada en la piel no se adhiriera a la misma. Cuando la participante consideró que había alcanzado la máxima distancia, se lo indicó a la investigadora. Una vez comprobado que el

intento era válido, realizó las mediciones correspondientes al protocolo establecido (figura 4.5.6 y tabla 4.5.1) desde la parte lateral de la pierna evaluada mientras la participantes mantenía la DF máxima.



Figura 4.5.5. Participante realizando la PDFAf.

En primer lugar la investigadora realizó las marcas de las líneas tal y como se recoge en la tabla 4.5.1: distancia entre el talón y la pared (DHf), distancia entre el primer dedo del pie y la pared (DDf) y la distancia entre suelo y el borde superior externo de la rótula (DVf). Tras marcar las líneas, en la PDFA1 la observadora evaluó la DF mediante el inclinómetro en dos ubicaciones: Inclinómetro ubicado en el talón de Aquiles (IAf) e inclinómetro ubicado en la tibia (ITf). En cada intento se realizaron dos mediciones utilizando el inclinómetro. Para todas las evaluaciones la observadora se valió de las marcas anatómicas previamente marcadas (figura 4.5.3). Una vez tomadas estas mediciones con el inclinómetro y/o marcados todos los puntos anatómicos sobre las líneas, la participante abandonó la posición de máxima DF y la investigadora recogió las marcas de las líneas: DHf, DDf y DVf. Cada participante realizó tres intentos seguidos con cada pierna. Además de las variables medidas durante la prueba, calculamos por razones trigonométricas el ángulo entre la vertical y la tibia (ADf) valiéndonos de DVf y de la DHf en cada uno de los intentos (tabla 4.5.1). En PDFA1 evaluamos la máxima DF en esta prueba mediante IAf, ITf, DHf, DDf y DVf, y en el PDFA2, DHf y la DVf.

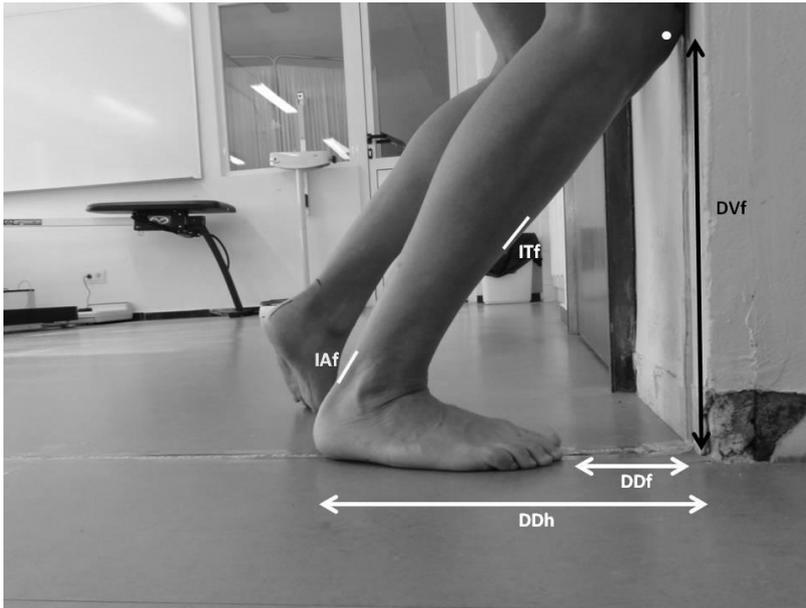


Figura 4.5.6. Métodos de evaluación de la PDFAf.

Tabla 4.5.1. Tabla de métodos de evaluación empleado en el PDFAf .

Método	Definición	Procedimiento	Unidades	Estudio
<b>Distancia Horizontal (DHf)</b>	Distancia entre el talón y vertical del borde supero externo de la rótula	Marcar en el centro de la línea del suelo la parte más distal del talón	Centímetros	PDFA1 y PDFA2
<b>Distancia Dedo (DDf)</b>	Distancia entre el primer dedo del pie y la pared	Marcar en el centro de la línea del suelo la parte más distal del primer dedo del pie	Centímetros	PDFA1
<b>Distancia Vertical (DVf)</b>	Distancia entre el suelo y el borde supero externo de la rótula	Marcar en la línea de la pared la altura que alcanza el borde supero externo de la rotula	Centímetros	PDFA1 y PDFA 2
<b>Inclinómetro Aquiles (IAf)</b>	Ángulo entre la vertical y la tibia medido con el inclinómetro ubicado en el tendón de aquiles	Colocar el inclinómetro sobre el tendón de aquiles a 7,5 cm del calcáneo	Grados	PDFA1
<b>Inclinómetro Tibia (ITf)</b>	Ángulo entre la vertical y la tibia medido con el inclinómetro ubicado en la tibia	Colocar el inclinómetro 15 cm por debajo de la tuberosidad tibial, sobre el borde anterior de la tibia	Grados	PDFA1
<b>Ángulo Derivado (ADf)</b>	Ángulo entre la vertical y la tibia derivado de DHf y DVf por razones trigonométricas	1-Tan $\alpha$ = DVf/DHf. 2-AD=90-Arctan $\alpha$	Grados	PDFA1 y PDFA2

Arctan  $\alpha$  - Arcontangente de alfa; Tan  $\alpha$  - Tangente de alfa.

#### 4.5.4.2-Prueba de dorsiflexión en apoyo con la rodilla en extensión (PDFAe)

En la PDFAe la participante colocó el pie de la pierna a evaluar del mismo modo que en el PDFAf (figura 4.5.4). Manteniendo el pie sobre la línea se pidió a la participante que llevara ese pie lo más atrás posible, sin elevar el talón, manteniendo la rodilla extendida, y la rodilla y el primer dedo del pie la pierna libre en contacto con la pared (figura 4.5.7). La pierna no evaluada debía estar con el pie y la rodilla en contacto con la pared, en el lugar que la participante considerara más oportuno. Cuando la participante logró la máxima distancia desde la pared hasta el pie evaluado, se lo indicó a la investigadora y ésta realizó las mediciones si el intento era correcto (tabla 4.5.2 y figura 4.5.7).

Al igual que en PDFAf, la observadora marco las líneas del suelo en primer lugar según el protocolo (tabla 4.5.2): distancia entre el talón y la pared (DTe), distancia entre el primer dedo del pie y la pared (DDe), distancia desde el borde supero externo de la rótula hasta la pared (DRe) y distancia entre suelo y el borde supero externo de la rótula (DVe) (figura 4.5.8). Cuando se midió la DVe (figura 4.5.8), la plomada debía caer sobre la línea del suelo para que el intento se considerase bueno y poder evaluar DRe. Tras marcar las líneas, la observadora evaluó la DF mediante el inclinómetro en dos ubicaciones: inclinómetro ubicado en el talón de Aquiles (IAe) e inclinómetro ubicado en la tibia (ITe). En la PDFA1 empleamos como métodos de evaluación IAe, ITf, DTe y DDe, y en la PDFA2 DTe, DRe, y DVe. Una vez tomadas las mediciones en la máxima DF de la misma manera que en PDFAf, la participante abandono la posición y la investigadora recogió las marcas de las líneas.

A partir de las variables DTe y DRe calculamos la distancia entre el talón y el borde supero externo de la rótula (DHe) como medición análoga al DHf de la PDFAf. Posteriormente estimamos el ángulo derivado de las distancias horizontal y vertical (ADe) de mismo modo que PDAFe; por razones trigonométricas (tabla 4.5.2).

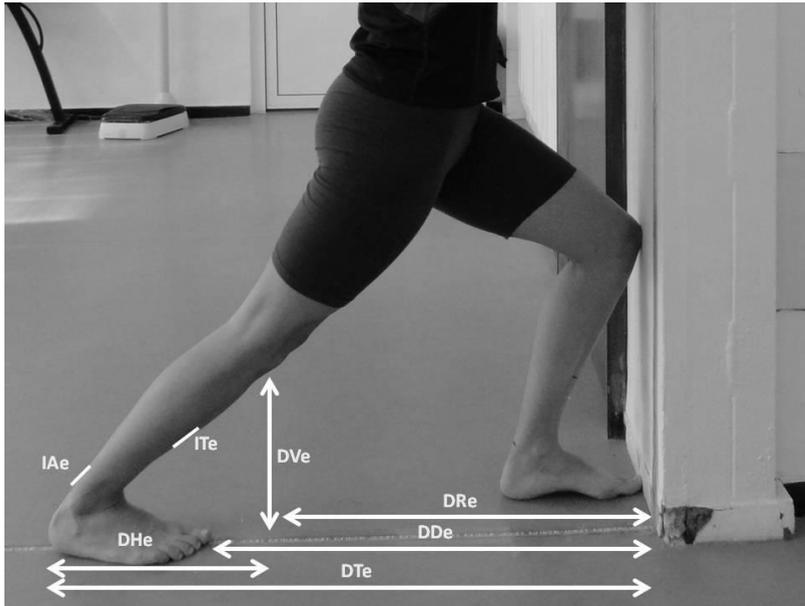


Figura 4.5.7. Métodos de evaluación de la PDFAe.

Tabla 4.5.2. Tabla de métodos de evaluación empleado en el PDFAe.

Método	Definición	Procedimiento	Unidades	Estudio
<b>Distancia talón (DTe)</b>	Distancia entre el talón y la pared	Marcar en el centro de la línea del suelo la parte más distal del talón	Centímetros	PDFA1 y PDFA2
<b>Distancia Rodilla (DRe)</b>	Distancia entre el borde supero externo de la rótula y la pared	Marcar el punto donde cae la plomada ubicada en el borde supero externo de la rótula	Centímetros	PDFA2
<b>Distancia horizontal (DHe)</b>	Distancia entre el talón y vertical del borde supero externo de la rótula	DHe= DTe-DRe	Centímetros	PDFA2
<b>Distancia dedo (DDe)</b>	Distancia entre el primer dedo del pie y la pared	Marcar en el centro de la línea del suelo la parte más distal del primer dedo del pie	Centímetros	PDFA1
<b>Distancia vertical (DVe)</b>	Distancia entre el suelo y el borde supero externo de la rótula	Poner el vertical la plomada desde el borde supero externo de la rótula de la pierna evaluada, y medir su longitud	Centímetros	PDFA2
<b>Inclinómetro Aquiles (IAe)</b>	Ángulo entre la vertical y la tibia medido con el inclinómetro ubicado en el tendón de aquiles	Colocar el inclinómetro sobre el tendón de aquiles a 7,5 cm del calcáneo	Grados	PDFA1
<b>Inclinómetro tibia (ITe)</b>	Ángulo entre la vertical y la tibia medido con el inclinómetro ubicado en la tibia	Colocar el inclinómetro 15 cm por debajo de la tuberosidad tibial, sobre el borde anterior de la tibia	Grados	PDFA1
<b>Ángulo derivado (ADe)</b>	Ángulo entre la vertical y la tibia derivado de DHe y DVe por razones trigonométricas	1-Tan $\alpha$ = DVe/DHe. 2-ADe=90-Arctan $\alpha$	Grados	PDFA2

Arctan  $\alpha$  - Arcontangente de alfa; Tan  $\alpha$  - Tangente de alfa.

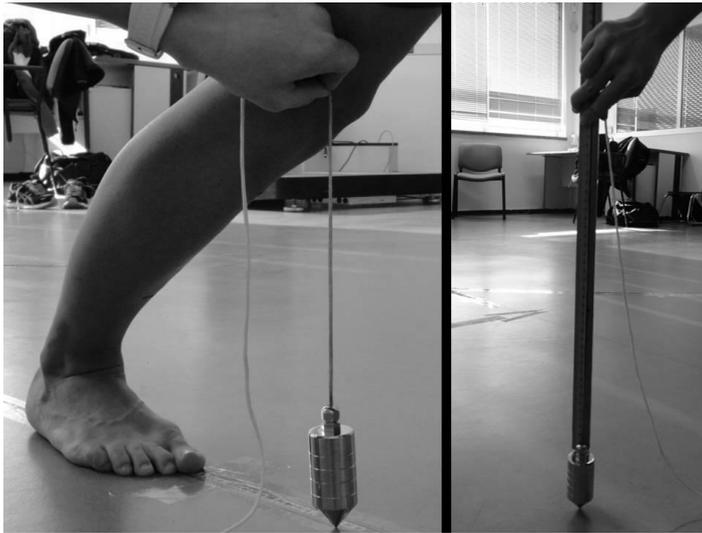


Figura 4.5.8. Método de evaluación la DVe en la PDF Ae.

#### 4.5.5-ANÁLISIS ESTADÍSTICO

Para todos los análisis realizados en este estudio, utilizamos la media de los tres intentos. En las medidas realizadas con el inclinómetro (IT e IA), realizamos dos mediciones por intento, y consideramos la media de ambas como estimador de cada intento.

En este trabajo comprobamos la fiabilidad de los diferentes métodos de evaluación empleados en las dos pruebas de dosiflexión, que fueron: IAf, ITf, DHf, DDf, DVf, ADf, IAe, ITE, DTe, DRe, DVe, DHe y ADe. Para determinar su fiabilidad utilizamos el método propuesto por Bland y Altman (1995)<sup>165</sup> que realizamos con el programa Medcalc<sup>®</sup> para windows, version 12.2.1.0 (Medcalc software, Mariakerke, Belgium), calculando el error sistemático (ES), límite de acuerdo (LDA), límite de acuerdo inferior, límite de acuerdo superior e intervalos de confianza para el ES y el LDA inferior y superior. Además, para complementar este método realizamos una regresión lineal, con intención de comprobar que la relación entre la diferencia y la media de valores no muestra tendencia a disminuir o aumentar en función de la magnitud de la medida. Para ello, utilizamos el programa SPSS versión 19.0.01 (IBM<sup>®</sup>) realizando una regresión lineal siendo la variable predictora la media de cada par de mediciones (día 1 y día 2) de cada una de las participantes y la variable dependiente la diferencia entre cada par de mediciones de cada una de los participantes. El umbral para la significación estadística en la regresión se determinó a priori en  $p < 0,05$ .

Además, calculamos de manera análoga a otros trabajos sobre la fiabilidad y al estudio de la fiabilidad del SEBT, y a pesar de las consideraciones metodológicas de la tesis, los siguientes índices de fiabilidad: el Coeficiente de Correlación Intraclass (CCI), el Coeficiente de Correlación de Pearson ( $r$ ), el Error Estándar de la Medida (EEM), la mínima diferencia detectable (MDD) y el punto de corte (PDC) descrito por Jacobson y Truax (1991)<sup>163</sup>.

Para determinar el CCI y la  $r$  de Pearson utilizamos el programa SPSS® versión 19.0.01 (IBM®). En el caso de la CCI aplicamos el modelo alfa, tipo de CCI mixto, y el acuerdo absoluto, proporcionando los valores individuales.

El EEM utilizado por Munro et al. (2010)<sup>72</sup> (EEM) se define de la siguiente manera:

$$EEM = DE \times \sqrt{1 - CCI}$$

La mínima diferencia detectable (MDD) se define de la siguiente manera<sup>72</sup>:

$$MDD = 1,96 \times \sqrt{2} \times EEM$$

El punto de corte (PDC) se define:

$$PDC = 1,96 \times \sqrt{2} \times EEM + (ES)$$

Siendo el EEM para el PDC calculado mediante la  $r$  de Pearson en lugar del CCI, y el ES la diferencia media las diferencias entre los resultados del segundo día respecto al primero, es decir el error sistemático de la prueba.

También estudiamos la fiabilidad de las variables medidas tanto en PDFa1 como en PDFa2 (DHf, DVf, ADf y DTe) de manera separada para observar si existían diferencias en la fiabilidad a principio y al final del estudio. Este análisis lo realizamos contrastando las evaluaciones realizadas el primer día con el segundo,

tanto en PDFa1 como en PDFa2. Para su análisis utilizamos únicamente el método Bland-Altman.

Por último, analizamos el grado de acuerdo de las siguientes variables, es decir la relación entre las medidas de resultado: IAf-ITf, ITf-ADf, IAf-ADf, ITE-IAe y ADf-ADe. Para lo que utilizamos el método propuesto por Bland y Altman<sup>155,165</sup> que llevamos a cabo mediante el programa Medcalc<sup>®</sup>.

# 5-RESULTADOS



## **5.1-REVISIÓN SISTEMÁTICA DE LOS FACTORES DE RIESGO PARA LAS LESIONES DE TOBILLO EN MUJERES DEPORTISTAS**

### **5.1.1-SELECCIÓN DE ESTUDIOS**

Tras la búsqueda electrónica identificamos 1.091 referencias (figura 5.1.1) susceptibles a ser incluidas en la RS, que una revisora analizó (ALR). Ésta eliminó todos los duplicados (n=434) y comprobó que estas proporcionaran un *abstract* y títulos informativos. Por esta última razón 12 documentos fueron descartados, ya que la BD *Sportdiscus* nos proporcionó referencias que correspondían a las actas completas de diferentes congresos, y en el apartado de *abstract* aparecía el título de todas las presentaciones realizadas en las mismas. Todos estos títulos fueron examinados.

La primera selección se inició con 645 referencias, de las cuales 567 (87,91%) fueron descartadas en base a su título y resumen (figura 5.1.1). 532 (80,97%) referencias fueron suprimidas por no estudiar los factores de riesgo para las lesiones de tobillo y 35 (6,17%) por no ser trabajos originales o tener un diseño apropiado para estudiar los FR. En total 78 referencias que identificamos mediante la búsqueda en bases de datos pasaron a la segunda selección junto con las 21 referencias halladas gracias al seguimiento de referencias y a la consulta con expertos.

Así, los dos revisores analizaron 99 documentos (8,90 % de todos los documentos encontrados) a texto completo para determinar si finalmente cumplían los criterios de inclusión de la revisión. Finalmente, en la segunda selección 89 trabajos fueron descartados por diversas razones. 64 de los documentos no estudiaron los FR para las lesiones de tobillo, un trabajo no proporcionó datos sobre los FR y 19 de los documentos los eliminamos por no poder extraerse los resultados del grupo de mujeres practicantes de actividad física.

En esta fase además identificamos otros cinco documentos que suprimimos de la revisión por otras causas. En dos de los documentos existía un duplicado de datos con otros dos documentos también incluidos, por lo que incluimos aquellos que más información nos proporcionaban. Un documento resultó ser una RS sobre FR de lesiones de tobillo incluyendo participantes de ambos sexos en la misma. En el análisis sobre FR femeninos encontramos que los estudios primarios que se estudiados ya estaban incluidos en nuestra revisión, por lo que esta referencia fue descartada. Otro trabajo lo descartamos al no acceso a él, tras ponernos en contacto con autor. El último de los documentos lo descartamos tras ponernos en contacto con el autor y declarar éste que no había incluido mujeres.

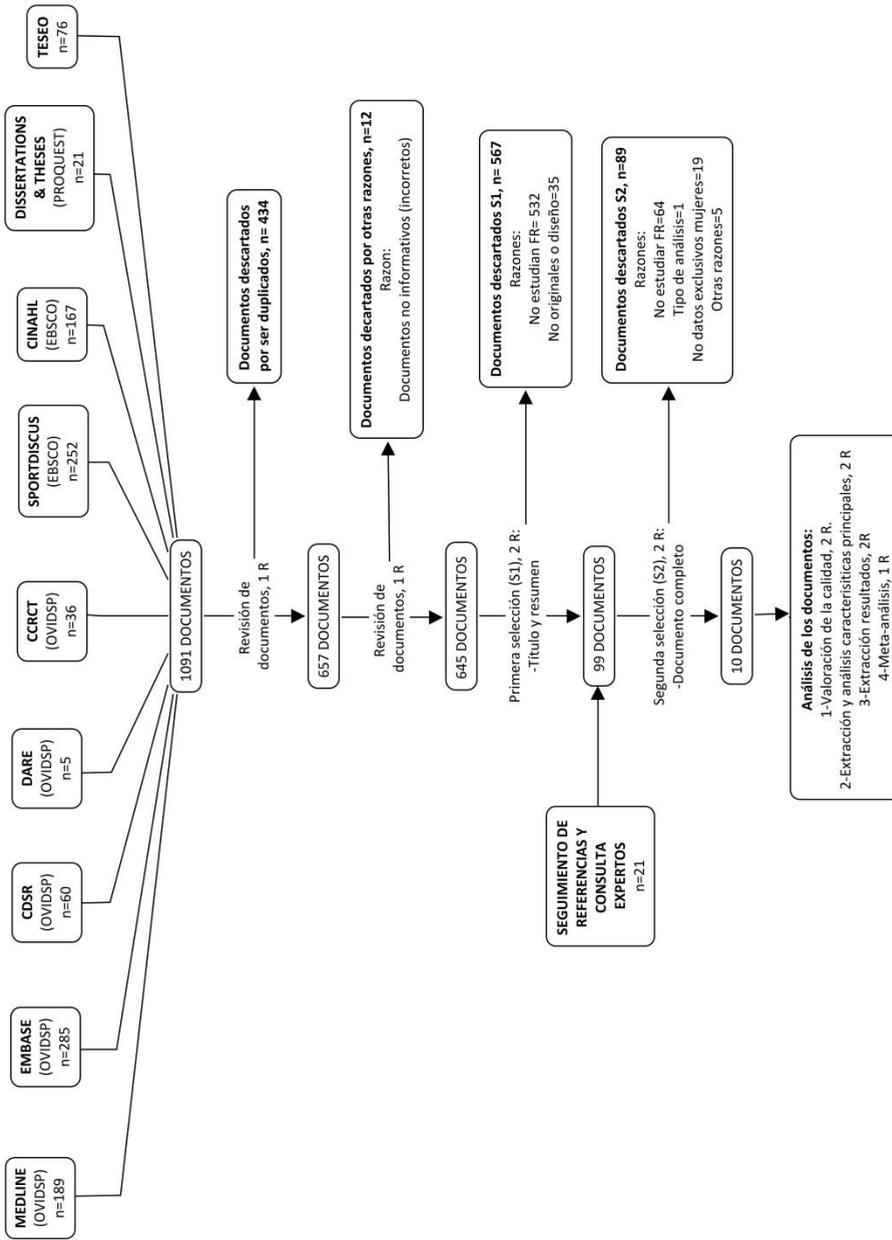


Figura 5.1.1. Diagrama de flujo de la RS FR. n - número de documentos; R - Revisores; S1 - Primera selección; S2 - Segunda selección.

En ambas selecciones analizamos el grado de acuerdo obtenido entre nuestros revisores que estimamos mediante la kappa de Cohen y el acuerdo observado, calculamos para ambos la estimación puntual y el intervalo de confianza (anexo 5.1.1). Finalmente incluimos en la RS sobre FR de lesiones de tobillo mujeres adultas practicantes de actividad física 10 trabajos<sup>18;40;44;52;55;166-170</sup>, del total de 1.112 referencias identificadas (1.091 mediante búsqueda electrónica y 21 mediante la consulta con expertos y seguimiento de referencias).

### 5.1.2-CALIDAD DE LOS ESTUDIOS

Con intención de valorar el riesgo de sesgo de los resultados que proporcionaban los trabajos incluidos, evaluamos su calidad por dos investigadores que trabajaron de manera independiente. Una vez realizada la valoración, ambos evaluadores consensuaron todas las respuestas (tabla 5.1.1) así como las razones para dar tal valoración (anexo 5.1.2). El acuerdo entre ambos observadores se valoró de la misma manera que en el proceso de selección (anexo 5.1.1).

Tabla 5.1.1. Calidad de los estudios incluidos en la RS FR.

Estudio	P1	P2	P3	P4	P5	P6
DuRant et al. (1992) <sup>166</sup>	Sí	No	Sí	Sí	No	Parcial
Krivickas & Feinberg (1996) <sup>167</sup>	?	Sí	Sí	?	No	Sí
Payne et al. (1997) <sup>52</sup>	No	?	Sí	Sí	No	Parcial
Beynnon et al. (2001) <sup>55</sup>	No	?	Sí	Sí	Parcial	Sí
Mei-Dan et al. (2004) <sup>169</sup>	Sí	No	Sí	Sí	No	Sí
Willems (2004) <sup>18</sup>	No	?	Sí	Sí	Parcial	Sí
Faude et al.(2006) <sup>168</sup>	Sí	?	Sí	Sí	No	Parcial
Kofotolis & Kellis (2007) <sup>40</sup>	Sí	?	Sí	Sí	Parcial	Sí
Steffen et al. (2008) <sup>170</sup>	Sí	No	Parcial	Sí	Parcial	Sí
Waterman et al. (2010) <sup>44</sup>	Sí	?	Sí	Sí	Parcial	?
<b>TOTAL SÍ (n/%)</b>	<b>6/ 60%</b>	<b>1/10%</b>	<b>9/90%</b>	<b>9/90%</b>	<b>0/0%</b>	<b>6/60%</b>

P1 – Participación en el estudio; P2 - Pérdida diferencial de participantes; P3 – Evaluación del factor pronóstico; P4 – Evaluación del *outcome*; P5 – Evaluación de los factores de confusión; P6 – Análisis. Para más detalles sobre las cuestiones ver anexo 4.1.4. Parcial – Parcialmente; ? – No se puede saber.

Atendiendo a la calidad de los trabajos en cuanto a los criterios participación (P1), cabe decir que el 60% de los trabajos<sup>40;44;166;168-170</sup> los cumplen, un 30%<sup>18;52;55</sup> tiene un claro riesgo de sesgo al respecto y un trabajo<sup>167</sup> presenta dudas acerca de esta cuestión. En lo que referente a la pérdida diferencial de participantes durante el tiempo en el que se desarrollan los estudios (P2), un único estudio<sup>167</sup> no presenta riesgo de sesgo en este aspecto. En seis de los trabajos, a saber; Beynnon et al. (2001)<sup>55</sup>, Payne et al. (1997)<sup>52</sup>, Willems (2004)<sup>18</sup>, Faude et al. (2006)<sup>168</sup>, Kofotolis & Kellis (2007)<sup>40</sup>, y Waterman et al. (2010)<sup>44</sup>, se describen únicamente la muestra analizada, por lo que no podemos saber si se dan pérdidas de participantes o no para un periodo de seguimiento que varía de dos<sup>52</sup> a 24

meses<sup>44</sup>. Kofotolis & Kellis (2007)<sup>40</sup>, que describe las características de las participantes para un tamaño muestral de 204 jugadoras de baloncesto, presenta inexplicablemente dos participantes analizadas menos cuando observamos los resultados encontrados. Por otro lado, tres de los estudios pierden participantes de manera importante desde el inicio del estudio hasta el final. DuRant et al. (1992)<sup>44;166</sup> pierden el 54% de participantes en el que se evalúa los FR. A pesar de que estos autores detallan el número de participantes de las que obtienen el *outcome* y las que no, consideramos que la pérdida de participantes durante el seguimiento es excesiva para poder garantizar la veracidad de los resultados. Meidan et al. (2004)<sup>169</sup> pierde un 21,7% de la muestra con la que inician el estudio en periodo de seguimiento de 4 meses. Por último, Steffen et al. (2008)<sup>170</sup> proporcionan una variación importante en el tamaño de la muestra cuando describen los datos sobre las características de las participantes así como cuando reporta los resultados. Más concretamente, el estudio describe en texto una muestra de 1.430, sin embargo en la tabla se describen las características de 1.425 participantes. Asimismo cuando reportan los resultados para las lesiones de tobillo, proporcionan datos de 2.758 piernas lo que supondría 1.379 participantes.

En cuanto a los FR evaluados, los estudios incluidos los describieron de manera adecuada según los criterios previamente establecidos. La mayoría empleaban FR sencillos de evaluar como la edad, el IMC y la altura, y cuando empleaban FR más sofisticados citaba la referencia de origen, o los describían correctamente, por lo que los revisores determinaron adecuado el modo de medir los FR. El trabajo de Steffen et al. (2008)<sup>170</sup> obtiene una calificación de parcialmente debido que no sabemos cuál es la proporción de participantes a los que se le mide el factor pronóstico.

Respecto al *outcome*, está RS determinó que el 90% de estudios analizados no presentaban riesgo en esta cuestión ya que consideramos que la lesión estudiada no es difícil de diagnosticar. Únicamente el estudio de Krivickas & Feinberg (1996)<sup>167</sup> plantea dudas puesto que no incluyen los esguinces que hayan permitido continuar la práctica deportiva, lo cual es frecuente en este tipo de lesiones, pudiéndose perder algunas lesiones durante el seguimiento.

Respecto al criterio sobre los FR de confusión (P5), cabe decir que ninguno de los estudios los tiene en cuenta de manera adecuada a la hora de analizar los FR. De los 10 trabajos incluidos, el 50%<sup>52;166-169</sup> no incluye ni menciona nada sobre los ellos ni los tiene en cuenta a la hora de analizar los FR. El resto de trabajos tienen en cuenta la exposición a la actividad física para evaluar la susceptibilidad de lesión respecto al FR, unos en número de exposiciones<sup>44;55</sup> y otros en tiempo de exposición<sup>18;40;170</sup>. Beynnon et al. (2001)<sup>55</sup>, Payne et al. (1997)<sup>52</sup> y Willems (2004)<sup>18</sup>

establecieron como criterio de exclusión el haber sufrido alguna lesión previa en tobillo o miembro inferior.

Respecto al análisis estadístico realizado (P6), la mayoría lo presenta de manera adecuada a los criterios de calidad establecidos al menos parcialmente. DuRant et al (2000)<sup>166</sup> Beynnon et al. (2001)<sup>55</sup>, Payne et al. (1997)<sup>52</sup>, Willems (2004)<sup>18</sup>, presentan un análisis adecuado a su objeto de estudio, aunque cuando estudian los FR de tobillo, solo proporcionan los datos del grupo masculino. Respecto a Payne et al. (1997)<sup>166</sup> proponen un análisis adecuado pero no reportan todos los resultados derivados del mismo. Además estos resultados no se proporcionan de manera analizable para realizar un MA. Por otro lado, Faude et al. (2006)<sup>168</sup> realiza, en opinión de los revisores, un análisis reducido respecto al objeto del estudio y el diseño del mismo. Respecto a Waterman et al. (2010)<sup>44</sup>, éste trabajo no proporciona los resultados del análisis de regresión, con lo que crea dudas respecto a la adecuación del análisis y su riesgo de sesgo.

### 5.1.3-ANÁLISIS DE LAS CARACTERÍSTICAS DE LOS ESTUDIOS INCLUIDOS Y RESULTADOS

Tras la valoración de la calidad de los trabajos incluidos en la revisión, pasamos a extraer sus características en las categorías previamente establecidas (anexo 5.1.3 y 5.1.4), recogiendo un resumen de las mismas en la tabla 5.1.2. Una vez obtenidos los datos, procedimos a analizar las características de los trabajos incluidos así como sus resultados (tablas 5.1.3 y 5.3.4).

Los 10 documentos estudiados en esta RS presentaban diferentes nacionalidades, pero la mayoría (50 %) habían sido realizados en EEUU<sup>44;52;55;166;167</sup>. El origen del resto fue variado; Bélgica<sup>18</sup>, Grecia<sup>40</sup>, Alemania<sup>168</sup>, Israel<sup>169</sup> y Noruega<sup>170</sup>. En cuanto al objetivo descrito por los autores, cuatro estudios<sup>18;44;52;55</sup> examinaron los FR de las lesiones de tobillo, proporcionando éstos los resultados de mujeres y hombres de manera diferenciada, mientras que dos trabajos<sup>40;169</sup> estudiaron exclusivamente a una cohorte femenina. Los restantes cuatro documentos<sup>166-168;170</sup>, tenían un objetivo diferente al de explorar los FR de las lesiones del tobillo, pero presentaron algún resultado para el grupo y la lesión de interés de esta RS, o pudimos extraer los datos para el grupo de interés.

Los estudios analizados presentaron un mínimo de 31 participantes y un máximo de 1.430, siendo el total de de 3.287. La media de mujeres por estudio fue de 328,70 ( $\pm$  465,16). Una vez eliminados los dos estudios<sup>44;170</sup> con la muestra sustancialmente más elevada al resto, la media fue de 111,50 ( $\pm$  61,18) con un tamaño muestral máximo de 202 mujeres.

Tabla 5.1.2. Resumen de las características estudios RS FR.

Estudio	País	Obj	n analizadas (caract)	Población/AF	T seguimiento	n FR	FR evaluados	outcome/n	Análisis estadístico	MA
DuRant et al. (1992) <sup>166</sup>	EEUU	No	202 (de 13 a 20 años)	E Instituto/ ≠ AF	1 temporada	1	Tobillo normal/anormal	Lesión tobillo/15	2x2	NA
Krivickas & Feinberg (1996) <sup>167</sup>	EEUU	No	70 (19,6 años, 164 cm y 63,4 kg)	E Universitarias/ ≠ AF	1 temporada	2	Laxitud y rigidez muscular	Lesión tobillo/7	X <sup>2</sup>	NA
Payne et al. (1997) <sup>162</sup>	EEUU	Sí	31 (20 años, 176,9 cm y 70,4 kg)	E Universitarias/ baloncesto	9 semanas	13	Flexibilidad, propiocepción, alineación	Lesión tobillo/7	X <sup>2</sup> , R stepwise	NA
Beynonn et al. (2001) <sup>165</sup>	EEUU	Sí	68 (De 8 a 23 años, 165,7 cm y 59,4 kg)	E Universitarias/ ≠ AF	2 temporadas	36	Características, equilibrio, fuerza, laxitud, alineación	Esguince tobillo por inversión/13	T-S, X <sup>2</sup> -R Cox	Sí
Mei-Dan et al. (2004) <sup>169</sup>	Israel	Sí EM	65 (18,5 años, 160,7 cm+ y 57 kg)	Militares/AF ND	4 meses	2	Historia, arco longitudinal de pie	Esguince tobillo/27	RR	NA
Willems (2004) <sup>118</sup>	Bélgica	Sí	159 (18,3 años, 167,3 cm y 60,16 kg)	E Universitarias EF/ ≠ AF	3 cursos académicos	58*	Características, condición física, alineación, propiocepción, fuerza, equilibrio	Esguince tobillo por inversión/32	R Cox	Sí
Faude et al. (2006) <sup>168</sup>	Alemania	No	143 (22,4 años, 169 cm y 61 kg)	Liga nacional fútbol	10 meses y medio	1	Historia LP	Esguince tobillo/33	OR, 2x2	Sí
Kofotolis & Kellis (2007) <sup>170</sup>	Grecia	Sí EM	204† (24,8 años, 175 cm, 20,3 kg/m2)	Liga profesional baloncesto	2 años	4	Características, historia	Esguince tobillo/32	X <sup>2</sup> , R stepwise, OR, 2x2	Sí (LP)
Steffen et al. (2008) <sup>170</sup>	Noruega	No	1.425 <sup>‡</sup> (15,4 años, 166,0, 56 y 20,00 ± 2 kg/m2)	Liga regional de fútbol sub 17	8 meses	1	Historia LP	Lesión tobillo/111	RR, 2x2	Sí
Waterman et al. (2010) <sup>144</sup>	EEUU	Sí	917 (165,5 cm, 60,60 kg 23,17kg/m2)	Militares/ ≠ AF	2 años	6	Características, condición física	Esguince tobillo/133	T-S, X <sup>2</sup> , R Poisson	Sí

AF - Actividad física practicada por las participantes; Caract - Características; E - Estudiantes; EM - Exclusivamente con mujeres; FR- Factor de riesgo; LP - lesión previa; MA - Meta-análisis; NA - No aplicable; Obj - ¿Es objetivo del estudio? Los FR riesgo en de las lesiones de tobillo y se proporcionan resultados exclusivo del grupo de mujeres?; OR - Odds ratio; R - Regresión; RR - Riesgo relativo; T - Tiempo; T-S - T-Student; X<sup>2</sup> - Ji Cuadrado; ≠ - Diferentes. \* = no conocemos el número de FR evaluados en la fuerza de reacción ni en el European Fitness Test. Estas variables se toman como una, aunque sospechamos que miden varias variables en cada grupo; † = la n analizada para FR tobillo 202; ‡ = datos características participantes. N estudiada 1,430 y datos análisis FR tobillo 2.645 piernas; +=características n de inicio (83), no la analizada. Las características de este estudio no fueron tomadas en cuenta en el análisis. Para más información consultar el anexo 5.1.3.

En cuanto a las características referentes a las participantes (edad, altura, peso, IMC y tiempo de actividad física semanal), ponderamos el valor extraído del estudio a la muestra que analizaron los estudios primarios para obtener la media de las participantes. De este análisis, descartamos el trabajo de Mei-Dan et al. (2004)<sup>169</sup> ya que el trabajo reportaba las características de las participantes que iniciaron el estudio, 83 participantes, y en el análisis de los FR se realizó con 65 mujeres. Consideramos esta pérdida excesiva como para incluirla en este análisis.

Tabla 5.1.3. Características cuantitativas estudios RS FR.

Variable	n est	% est	n part	% part	Media	DE	Mínima	Máxima
n mujeres	10	100,00	3.287	100,00	328,70	465,16	31,00	1.430,00
Edad (años)	6	60,00	2.035	61,91	17,27*	-	15,40	24,80
Altura (cm)	8	80,00	3.020	91,88	166,73*	-	164,00	176,90
Peso (kg)	7	70,00	2.818	85,73	58,42*	-	56,00	70,40
IMC (kg/m <sup>2</sup> )	4	40,00	2.708	82,39	21,18*	-	20,00	23,16
T AF (horas/semana)	3	30,00	1.791	54,49	7,59*	-	2,73	15,33
T seg (meses)	10	100,00	3.287	100,00	13,22	8,02	2,20	24,00
n FR	10	100,00	3.287	100,00	12,40	19,31	1,00	58,00
n outcomes	10	100,00	3.287	100,00	41,00	44,16	7,00	133,00

DE – Desviación estándar; FR – Factores de riesgo; IMC – Índice de masa corporal; n est – Número de estudios en los que la característica está descrita; n part – Número de participantes en los que la característica está descrita; T AF – Tiempo de actividad física; T seg – Tiempo de seguimiento; % est – Porcentaje de estudios en los que la característica está descrita; % part – Porcentaje de participantes en los que la característica está descrita; \* - Variable ponderada al número de participantes.

Las mujeres analizadas estas tenían una edad media de 17,27 años con una edad media mínima de 15,40 y máxima de 24,80, en los seis estudios<sup>18;40;52;167;168;170</sup> de los cuales hemos podido conocer este rasgo. Respecto a la altura, la media de las participantes estudiadas fue de 166,73 cm, variando esta de 164,00 cm a 176,90 cm en los ocho trabajos<sup>18;40;44;52;55;167;168;170</sup> (91,88% de participantes) que describieron esta característica. Las participantes más altas corresponden a los trabajos que examinan los FR en jugadoras de baloncesto, como son los trabajos de Payne et al. (1997)<sup>52</sup> y Kofotolis y Kellis (2008)<sup>40</sup>. Si descartamos estos trabajos del análisis la altura varía poco, de 166 cm a 169 cm. Respecto al peso, estas tenían un peso medio de 58,52 kg (en siete estudios y el 85,73% de participantes), con una mínima de 56 y máxima de 70,40 kg. Del mismo modo que con la altura, las participantes más pesadas fueron las jugadoras de baloncesto en el trabajo de Payne et al. (1997)<sup>52</sup>, no describiendo Kofotolis y Kellis (2008)<sup>40</sup> esta característica. Si descartamos el trabajo de Payne et. al (1997)<sup>52</sup>, el peso medio de las participantes varía de 56,00 kg a 63,4 kg. En cuanto al IMC, solo fue posible extraer los datos del IMC en 4 trabajos<sup>18;40;44;170</sup>, variando está de 20,00 kg/m<sup>2</sup> a 23,16 kg/m<sup>2</sup>. El IMC medio en los cuatro estudios (n participantes 82,39%) que describieron esta fue de 21,18 kg/m<sup>2</sup>.

Respecto al tipo de población estudiada en los trabajos, la mayoría de ellos (50%) utilizó a estudiantes, bien universitarias<sup>18;52;55;167</sup> o procedentes de un instituto<sup>166</sup> representado un 16,12% de la población estudiada. Dos trabajos<sup>44;169</sup> examinaron los FR en mujeres militares, lo que supuso un 29,88% de las participantes incluidas en la RS y tres trabajos<sup>40;168;170</sup> eligieron a una cohorte de alguna liga deportiva (54% de la participantes). En lo que respecta al tipo de actividad física que practicaban las cohortes, en general la mayoría (50%)<sup>18;44;55;166;167</sup> de los estudios examinaron a participantes de diferentes actividades. Cuatro trabajos estudiaron a una población procedente a un deporte específico, dos a jugadoras de baloncesto<sup>40;52</sup> y otros dos estudios<sup>168;170</sup> a jugadoras de fútbol. Uno de los trabajos<sup>169</sup> no definió el tipo de actividad física práctica por la cohorte analizada. Sin embargo, cuando observamos el número de participantes, la mayoría de ellas fueron jugadoras de fútbol (47,86%). Esto es debido a que el estudio de Steffen et al. (2008)<sup>170</sup>, empleo a jugadoras de fútbol, y su muestra supuso el 43,50% del total de la muestra de esta RS. Si extraemos este estudio del análisis, observamos que la mayor parte de las participantes (75,25%) observadas practican varias actividades.

En cuanto al tiempo dedicado por estas a la actividad física, sólo fue descrito por 3 estudios<sup>18;40;170</sup>, siendo el tiempo mínimo de actividad física de 2,73 horas por semana y el máximo de 15,33, lo que fue una media de 7,59 horas por semana. No creemos que estos datos reflejen la realidad de la población estudiada. En primer lugar porque solo se describe esta variables en tres de los 10 estudios. Es más, uno de los trabajos, el de Willems (2004)<sup>18</sup> examina los FR en estudiantes de educación física, incluyendo la práctica deportiva puramente académica y la extra-académica. Esta población por lo tanto realizara más actividad física que cualquier otra muestra de población. Los otros dos trabajos presentan unas medias muy diferentes, y describe que las jugadoras de baloncesto profesionales 2,73 horas por semana, mientras que las jugadoras de fútbol practican 7,42 horas por semana.

Respecto al tiempo de seguimiento de las participantes, la media fue de 13,22 meses ( $\pm 8,02$  meses) con un periodo de seguimiento mínimo de dos meses y un máximo de 24. Tuvimos que estimar el tiempo de seguimiento en meses ya que los trabajos proporcionaban el tiempo en unidades diferentes (número de temporada, curso académico, número de semanas...).

En cuanto a los FR estudiados la media de los estudios fue 12,40 ( $\pm 19,31$ ), siendo el número mínimo de uno y el máximo de 58. El 70% de los trabajos<sup>40;44;166-170</sup> evaluaron de uno a seis FR, uno<sup>55</sup> 13 y los dos<sup>18;55</sup> restantes empleando de 36 y 58. Respecto a esta variable es necesario precisar que la tesis de Willems (2004) presenta dudas acerca del número exacto de variables susceptibles a ser

predictoras de las lesiones de tobillo estudiadas. Esto es debido a que en material y métodos describen parcamente que va a realizar el *European Test of Physical Fitness* y la fuerza de reacción, con lo que no sabemos qué van a medir. Asimismo, en resultados no reporta los resultados de los FR correspondientes a la batería de pruebas, con lo que resulta imposible determinar en número exacto de FR estudiados. Es por ello que en este trabajo se contabilizaron ambos grupos como un solo FR.

Tabla 5.1.4. Características cualitativas estudios RS FR.

Variables	Categorías	RS FR					
		n est	n part	% est	% part	% est t	% part t
Objetivo	Sí	6	1.442	60,00	43,87	60,00	43,87
	No	4	1.845	40,00	56,13	40,00	56,13
Población	Estudiantes	5	530	50,00	16,12	50,00	16,12
	Liga deportiva	3	1.775	30,00	54,00	30,00	54,00
	Militares	2	982	20,00	29,88	20,00	29,88
	ND	0	0	0,00	0,00	0,00	0,00
Tipo de AF	Varias AF	5	1.416	50,00	43,08	50,00	43,08
	Fútbol	2	1.573	20,00	47,86	20,00	47,86
	Baloncesto	2	233	20,00	7,09	20,00	7,09
	ND	1	65	10,00	1,98	10,00	1,98
Outcome	Lesión tobillo	4	1.733	40,00	52,72	40,00	52,72
	Esguince	4	1.327	40,00	40,37	40,00	40,37
	Esguince inversión	2	227	20,00	6,91	20,00	6,91
	ND	0	0	0,00	0,00	0,00	0,00
Unidad de análisis	Una pierna	2	227	20,00	6,91	20,00	6,91
	Dos piernas	2	1.461	20,00	44,45	20,00	44,45
	Participante	1	917	10,00	27,90	10,00	27,90
	ND	5	682	50,00	20,75	50,00	20,75
Tiempo de exposición	Horas exposición	3	1.791	30,00	54,49	30,00	54,49
	n exposiciones	2	985	20,00	29,97	20,00	29,97
	ND	5	511	50,00	15,55	50,00	15,55

AF – Actividad física; DE – Desviación estándar; IMC – Índice de masa corporal; ND – No definido; n est – Número de estudios en los que la característica está descrita; n part – Número de participantes en los que la característica está descrita; T AF – Tiempo de actividad física; T seg – Tiempo de seguimiento; % est – Porcentaje de estudios en los que la característica está descrita; % part – Porcentaje de participantes en los que la característica está descrita; t – Número o porcentaje respecto al total de estudios o participantes incluidos en la RS.

Respecto a la relación entre el objetivo del estudio y los FR, seis han definido el objetivo de estudiar los FR empleando población mixta y proporcionando los resultados de manera separada o solo han estudiado mujeres<sup>18;40;44;52;55;169</sup>. El trabajo de Mei-Dan et al. (2004)<sup>169</sup> es el único trabajo que tiene como objetivo comprobar los FR de tobillo en mujeres y estudia dos FR. Kofotolis y Kellis (2008)<sup>40</sup> y Waterman et al. (2010)<sup>44</sup> estudian la incidencia y su relación FR, estudiando cuatro y seis FR respectivamente, siendo todas las variables de FR de las áreas de

las características de las participantes y cualidades físicas. Los restantes tres estudios, los de Payne et al. (1997), Beynnon et al. (2001)<sup>55</sup> y Willems et al. (2004)<sup>18</sup>, sólo estudian los FR para las lesiones de tobillo examinando de 13 a 58 FR. Los restantes cuatro<sup>166-168;170</sup> estudios podemos extraer o realizan algún análisis de manera separada para la lesión y el grupo de interés de esta RS, analizando estos uno<sup>166;168;170</sup> o dos<sup>167</sup> FR diferentes.

En los estudios incluidos se han analizado 103 FR diferentes sobre lesiones de tobillo en mujeres (anexo 5.1.4), de la cuales el 85% (88 variables) se ha estudiaron exclusivamente en un trabajo. En lo que respecta a las restantes 15 variables, éstas fueron estudiadas en dos o más estudios. La altura<sup>18;40;44;55</sup> y la historia de lesiones<sup>40;168-170</sup> previas se estudiaron en cuatro trabajos diferentes, y el peso<sup>18;40;44</sup> y el IMC<sup>18;40;44</sup> en 3. Las once variables restantes: edad, propiocepción del tobillo activa de tobillo a 15º de inversión, la fuerza concéntrica del tobillo en dorsflexión, flexión plantar, eversión (EV) e inversión (IN) y la excéntrica en EV e IN a 30º/s, la laxitud general evaluado por el método propuesto por Beighton, el rango de movimiento (RDM) en IN del calcáneo sin apoyo y el RDM EV calcáneo sin apoyo, se estudiaron en dos trabajos.

En lo que respecta al número de FR analizados por áreas, las características de las participantes fueron evaluadas en cuatro estudios<sup>18;40;44;55</sup> diferentes. Las variables medidas en este grupo de FR fueron cinco; la edad, la altura, el peso, el IMC y la dominancia en el miembro de las participantes. A excepción de la pierna dominante, todas las variables de este grupo fueron analizadas en dos o más trabajos. Ningún estudio detectó que dichos FR fueran susceptibles de discriminar población con riesgo de lesión de las que no.

En cuanto a la historia de lesiones previas en el tobillo, cuatro estudios analizaron estas variables como posible FR, encontrando tres<sup>40;168;169</sup> que no estaba asociado a un mayor riesgo de lesión, y uno<sup>170</sup> que sí.

Respecto a las cualidades físicas como FR, estas han sido examinadas en dos estudios, Waterman et al. (2010)<sup>44</sup> y Willems (2004)<sup>18</sup>, lo que supone un total de cuatro FR diferentes, sin que haya habido replica de ninguno. Cabe decir que Willems (2004)<sup>18</sup>, realiza el *European Test of Physical Fitness* que se compone de varias pruebas, pero lo hemos contabilizado como uno porque la autora no determina qué variables emplea para el análisis, describiéndose en resultados que las cualidades físicas evaluados con esta prueba no resultaron FR en su estudio. Ningún estudio primario detectó que las características de las participantes y cualidades físicas fueran predictoras de lesión de tobillo.

Atendiendo al área de propiocepción del tobillo, en la RS detectamos 15 variables diferentes entre los dos trabajos<sup>18;52</sup> que midieron los FR de este grupo. Una única

variable fue evaluada por ambos estudios, que fue la propiocepción del tobillo a 15° en inversión cuando se proporcionan resultados absolutos, encontrando únicamente Payne et al. (1997)<sup>52</sup> una asociación entre la lesión de tobillo y esta variable. Respecto a este grupo, de manera individual Payne et al. (1997)<sup>52</sup> encontraron que la propiocepción activa a 10° de DF en valores absolutos<sup>52</sup> discriminaba bien mujeres susceptibles de lesión de las que no lo eran. Por otro lado, Willems (2004)<sup>18</sup> detectó tres FR más en este grupo, a saber; inversión pasiva a 15° en valores absolutos, inversión pasiva a 15° en valores exactos y inversión pasiva máxima menos 5° en valores exactos.

Alguna variable de equilibrio se evaluó como posible FR en dos trabajos<sup>18;55</sup> haciendo un total de 19 variables diferentes, no replicándose ninguna. De este grupo, en la tesis de Willems (2004)<sup>18</sup> se identificaron dos variables como FR; el límite de estabilidad y límite de estabilidad en distancia máxima.

En lo referente a los FR del área de laxitud, esta se evaluó en dos estudios<sup>55;167</sup> haciendo un total de 3 variables diferentes; *talar tilt test*, la prueba del cajón anterior y el método Beighton para evaluar la laxitud general. Solo esta última, fue evaluada en ambos estudios. Ninguna de las variables evaluadas en este grupo resultó ser FR para las lesiones de tobillo, aunque la *p* en Krivickas et al. (1997)<sup>167</sup> en el análisis del método Beighton fue de 0,09 lo que sugiere cierta tendencia a a las atletas a lesionarse en opinión de los autores.

En cuanto al área de FR de alineación, rango de movimiento y flexibilidad, tres estudios<sup>18;52;55</sup> examinaron alguna variable perteneciente a este grupo. En total se estudiaron 20 variables diferentes, de las cuales dos (el RDM en IN del calcáneo sin apoyo y el RDM en EV del calcáneo sin apoyo) se midieron en dos trabajos diferentes, los de Willems (2004)<sup>18</sup> y Beynnon et al. (2001)<sup>55</sup>. El RDM en EV del calcáneo sin apoyo fue predictora de la lesiones de tobillo en el trabajo de Beynnon et al. (2001)<sup>55</sup> y pero no así para Willems (2004)<sup>18</sup>. De este grupo individualmente, se identificaron como FR un mayor RDM en extensión de la primera metatarsofalánge<sup>18</sup> y Varo/Valgo tibia apoyo<sup>55</sup>.

El grupo de factores de riesgo que más veces se evaluó en los estudios primarios fue la fuerza específica de tobillo. En este apartado, es preciso destacar que Willems (2004) evaluó como FR varias variables de la fuerza de reacción, pero no reportó cuales fueron, con lo que para el análisis se contabilizó como una sola. En consecuencia, 33 variables de este grupo fueron detectadas en tres estudios diferentes<sup>18;52;55</sup>, replicándose seis de ellos en dos estudios<sup>18;55</sup> (anexo 5.1.4). De este grupo, una mayor la fuerza isocinética concéntrica en DF a 120°/s fue detectada como FR en el estudio de Willems (2004)<sup>18</sup>. En cuanto al resto de

variables los estudios originales no encontraron que estuvieran asociadas a un mayor riesgo de lesión.

Fuera de las áreas establecidas, se han evaluado tres FR en los trabajos incluidos en la revisión: Normalidad o anormalidad del tobillo<sup>166</sup>, la rigidez muscular<sup>167</sup> y la longitud arco del pie<sup>169</sup>. Ninguna de estas variables tuvo réplica ni resultó ser predictora de las lesiones de tobillo.

Respecto a los *outcomes* evaluados, cuatro trabajos registraron la lesiones de tobillo<sup>52;166;167;170</sup>, otros cuatro<sup>40;44;168;169</sup> contabilizaron los esguinces de tobillo y cuatro<sup>18;55</sup> los esguinces de tobillo por inversión. En total se produjeron 410 *outcomes*, lo que supuso una media de 41,00 ( $\pm$  44,16) lesiones por estudio, con un mínimo de siete lesiones registradas y máximo de 133.

En cuanto a la unidad empleada en el análisis de los FR, cinco estudios no determinaron cuál tomaron en cuenta<sup>40;166-169</sup>. Willems (2004)<sup>18</sup> y Beynnon et al. (2001)<sup>55</sup> emplearon la pierna lesionada en los casos, y una pierna elegida aleatoriamente en los controles, es decir en las no lesionadas. Dos trabajos<sup>52;170</sup> incluyeron ambas piernas en el análisis. Por un lado, Payne et al. (1997)<sup>52</sup> estudio de manera diferenciada ambas piernas, mientras que Steffen et al. (2008)<sup>170</sup> incluyeron ambas piernas en un mismo análisis. El estudio de Waterman et al. (2010)<sup>44</sup>, emplearon las variables estudiadas como una unidad atribuible a la persona, independientemente de la pierna.

Respecto a la exposición a la actividad física durante el seguimiento, dicha variable se ha evaluado en horas de exposición en tres trabajos<sup>18;40;170</sup>, y en número de exposiciones en dos trabajos<sup>44;55</sup>, aunque en uno de estos trabajos, el de Kofotolis y Kellis (2008)<sup>40</sup>, no tomaron en cuenta esta variable para el análisis. El resto de estudios no mide esta variable.

### 5.3.4-META-ANÁLISIS

Realizamos MA con aquellos FR que se replicaron y además los estudios proporcionaban sus resultados de manera analizable, o podían ser extraídos. De las 15 variables estudiadas en más de un trabajo en tres casos no pudimos realizar MA. La edad no pudo meta-analizarse ya que Willems (2004)<sup>18</sup> proporcionó las medias del grupo de lesionados y no lesionados y Kofotolis y Kellis (2007) calcularon el *odd ratio*. La laxitud evaluada por Krivickas et al. (1997)<sup>167</sup> no proporcionó resultados de manera analizable. Ambas variables resultaron no ser FR para los estudios primarios. La tercera variable no incluida en el MA, fue una de las de propicepción (15º en inversión cuando se proporcionan resultados absolutos). Esta fue identificada como FR para el trabajo de Payne et al. (1997)<sup>52</sup>, pero no así

para Willems (2004)<sup>18</sup>, no pudimos meta-analizar esta variable al no proporcionar datos analizables los primeros.

Por lo que tras la extracción de datos identificamos 12 variables susceptibles (12% del total de FR estudiados) de ser incluidas en el MA (anexo 5.1.4). Realizamos dos MA por cada variable, uno con el modelo de efectos fijos y otro con el de efectos aleatorios, no encontrándose diferencias reseñables entre ambos modelos, por lo que aquí se presentan los resultados del MA de efectos aleatorios (tabla 5.1.5 y anexo 5.1.5).

Tabla 5.1.5. Resultados de los meta-análisis de la RS FR.

Factor de riesgo	n estudios	n	DEM/ RR	DEM/ RR IC LI	DEM/ RR IC LS	I <sup>2</sup>
RR combinado LP DL*	3	3106	1,549	1,159	2,069	0,00%
DEM combinada altura DL	3	1144	0,000	-0,161	0,160	0,00%
DEM combinada peso DL	3	1144	-0,005	-0,165	0,155	0,00%
DEM combinada IMC DL	2	1076	-0,018	-0,184	0,148	0,00%
DEM combinada RDM IN DL	2	227	0,154	-0,172	0,481	0,00%
DEM combinada RDM EV DL	2	227	-0,172	-0,582	0,238	29,40%
DEM combinada FC IN 30°/s DL	2	227	0,033	-0,294	0,359	0,00%
DEM combinada FE IN 30°/s DL	2	227	0,266	-0,062	0,593	0,00%
DEM combinada FC EV 30°/s DL	2	227	0,147	-0,180	0,474	0,00%
DEM combinada FE EV 30°/s DL	2	227	0,011	-0,316	0,337	0,00%
DEM combinada FC DF 30°/s DL*	2	227	0,479	0,150	0,809	0,00%
DEM combinada FC FP 30°/s DL	2	227	0,133	-0,193	0,460	0,00%

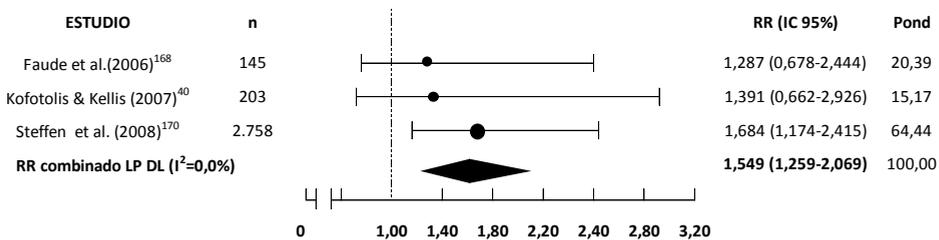
EV – Eversión; DEM – Diferencia Estandarizada medias; DF - Dorsiflexión; DL - DerSimonian-Laird; FC – Fuerza concéntrica (isocinética); FE - Fuerza excéntrica (isocinética); FP – Flexión plantar ; IC LI – Intervalo de confianza al 95% límite inferior; IC LS – Intervalo de confianza al 95% límite superior; IN - Inversión; IMC – Índice de masa corporal; I<sup>2</sup> - I cuadrada; LP - Historia de lesión previa; RDM – Rango de movimiento; RR - Riesgo relativo; \* - Identificado como FR.

En cuanto a la historia de lesión previa, los trabajos de Kofotolis y Kellis (2007)<sup>40</sup> Faude et al. (2006)<sup>168</sup> y Mei-Dan et al. (2004)<sup>169</sup> muestran que el haber sufrido una lesión previa no es un FR. Sin embargo, Steffen et al. (2008)<sup>170</sup>, sí que encuentran asociación entre esta variable y la lesión. El meta-análisis, descartando el trabajo de Mei-Dan et al. (2004)<sup>169</sup> que no proporcionó resultados de manera analizable, muestra que indiferentemente de ser lesión de tobillo<sup>40;170</sup> o esguince de tobillo<sup>168</sup>, este hecho incrementa el riesgo en 1,55 (IC 95%: 1,26-2,07) para una nueva lesión<sup>170</sup> o esguince de tobillo<sup>40;168</sup> (figura 5.1.2).

Respecto a las características de las participantes realizamos MA para la altura<sup>18;44;55</sup> y peso<sup>18;44;55</sup> con tres estudios, y el IMC con dos<sup>18;44</sup>. Cabe decir, que no incluimos en el MA de IMC y altura los resultados de Kofotolis y Kellis (2007)<sup>40</sup>,

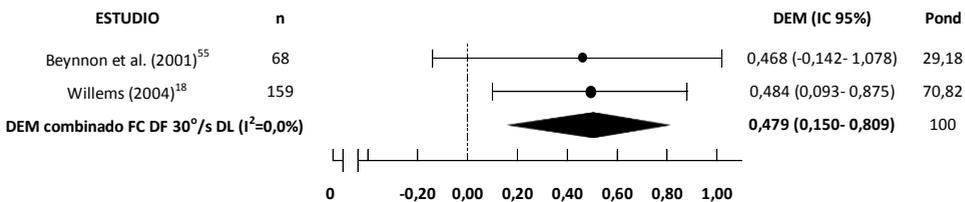
ya que no fue posible extraer los datos, al proporcionar solo estos el *odd ratio*. Para realizar el MA de las características tuvimos que estimar como el 30% de la media la DE, en dos casos en la variable altura<sup>44;55</sup>, uno en peso<sup>44</sup> y otro en IMC<sup>44</sup>. Ninguna de ellas mostró un DEM combinado que demostrara asociación entre variable y la lesión de tobillo (tabla 5.1.5), confirmando lo que los estudios primarios de manera individual habían determinado.

Respecto a las dos variables sobre rango de movimiento, el RDM en EV del calcáneo sin apoyo discriminó a la población susceptible de lesionarse en el trabajo de Beynnon et al. (2001)<sup>55</sup>, pero no en el de Willems et al. (2001)<sup>18</sup>. La DEM muestra que dicha variable empleando el método de evaluación de Elveru et al. (1988)<sup>109</sup> no es FR (tabla 5.1.5). El RDM en IN medido con el mismo método<sup>109</sup> no está asociado a un mayor riesgo de lesión, cuando observamos los estudios de manera individual ni tampoco cuando analizamos el DEM (tabla 5.1.5) .



**Figura 5.1.2.** Gráfico del meta-análisis FR historia de lesión previa. DL - DerSimonian-Laird; IC 95% - Intervalo de confianza al 95; I<sup>2</sup> - I cuadrada; LP - Historia de lesión previa; Pond - Ponderación; RR - Riesgo relativo. ; |—| - Intervalo de confianza medidas del RR; ● - Estimación puntual del RR; ◆ - RR combinado con IC 95%.

Respecto a las seis variables de fuerza de tobillo, no encontramos relación entre las lesiones y las variables evaluadas, a excepción de la fuerza concéntrica en DF a 30°/s (tabla 5.1.5). Esta variable fue evaluada en dos estudios, no encontrándose asociada al riesgo de lesión en ninguno de ellos cuando observamos sus resultados de manera individual (figura 5.1.3), sin embargo el DEM muestra una asociación pequeña.



**Figura 5.1.3.** Gráfico del meta-análisis FR fuerza concéntrica en DF a 30°/s. DEM - Diferencia estandarizada de medias; DL - DerSimonian-Laird; FC DF- Fuerza concéntrica (isocinética) en dorsiflexión; IC 95% - Intervalo de confianza al 95; I<sup>2</sup> - I cuadrada; Pond - Ponderación; |—| - Intervalo de confianza medidas de la DENM; ● - Estimación puntual de la DEM; ◆ - DEM combinado con IC 95%.

### REVISIÓN SISTEMÁTICA SOBRE LOS FACTORES DE RIESGO: SÍNTESIS DE LOS RESULTADOS

- **10 estudios** de **calidad adecuada**, pese a una pérdida diferencial de participantes y un control insuficiente de los factores de confusión, estudiaron un total **103 factores de riesgo** diferentes.
- Muchos de los factores de riesgo en áreas de **equilibrio, fuerza específica de tobillo, rango de movimiento y alineación del miembro inferior**, han sido estudiados **solo una vez** y se han asociado a un mayor riesgo de lesión.
- De los **factores de riesgo** estudiamos más de una vez y que **permiten realizar meta-análisis**:
  - La **historia de lesión previa** aumenta el riesgo un 55%.
  - Una reducción de la **fuerza concéntrica en dorsiflexión a 30°/s** aumenta ligeramente el riesgo.

## **5.2-RESULTADOS REVISIÓN SISTEMÁTICA DEL STAR EXCURSION BALANCE TEST**

### **5.2.1-SELECCIÓN DE ESTUDIOS**

Mediante la búsqueda electrónica y la consulta con expertos identificamos 321 estudios sobre el *SEBT*, de los cuales 149 fueron duplicados. Revisamos uno por uno los 172 documentos no repetidos y eliminamos once referencias; ocho por no ser los documentos informativos, dos por ser literatura gris (*no peer-review*) y una por estar el documento en japonés (figura 5.2.1).

Por lo tanto, en la primera selección incluimos 161 referencias. Los revisores leyeron el abstract, y cuando éste no fue suficiente consultaron el documento completo descartando de esta manera 147 documentos. El 27,9% (45 documentos) de las referencias fueron descartadas por no realizar la prueba de la estrella, el 19,3% (31 documentos) por no incluir participantes mujeres en el trabajo y el resto, un 44,1% (71 documentos), por no comprobar la fiabilidad de la prueba. En total 14 documentos pasaron a la segunda selección de los cuales ninguno proporcionó resultados exclusivos al grupo de mujeres practicantes de actividad física.

Tanto en la primera como en la segunda selección se evaluó el acuerdo alcanzado por los revisores antes de consensuar las respuestas mediante la Kappa de Cohen y el acuerdo observado, proporcionando la estimación puntual y los intervalos de confianza de cada uno de ellos (anexo 5.2.1).

Al comprobar que ningún estudio había analizado la fiabilidad de la prueba proporcionando datos exclusivos del grupo de mujeres deportistas decidimos analizar los documentos que habían estudiado la fiabilidad del *SEBT* incluyendo mujeres. Estos documentos fueron los 14 estudios<sup>58;65;67;68;71-73;86;90;92;96;97;99;100</sup> que habían superado la primera fase de la selección (figura 5.2.1).

### **5.2.2-CALIDAD DE LOS ESTUDIOS**

En primer lugar realizamos la valoración de la calidad de los estudios incluidos (tabla 5.2.1 y anexo 5.2.2). En cuanto a la separación entre sesiones de evaluación, cinco de los trabajos<sup>65;67;72;73;100</sup> las separaron entre uno y quince días. En el resto de los nueve trabajos<sup>58;68;71;86;90;92;96;97;99</sup> este hecho no se describe o bien no se hace. Respecto a sí se completó o no el estudio de fiabilidad ninguno de los trabajos lo menciona.

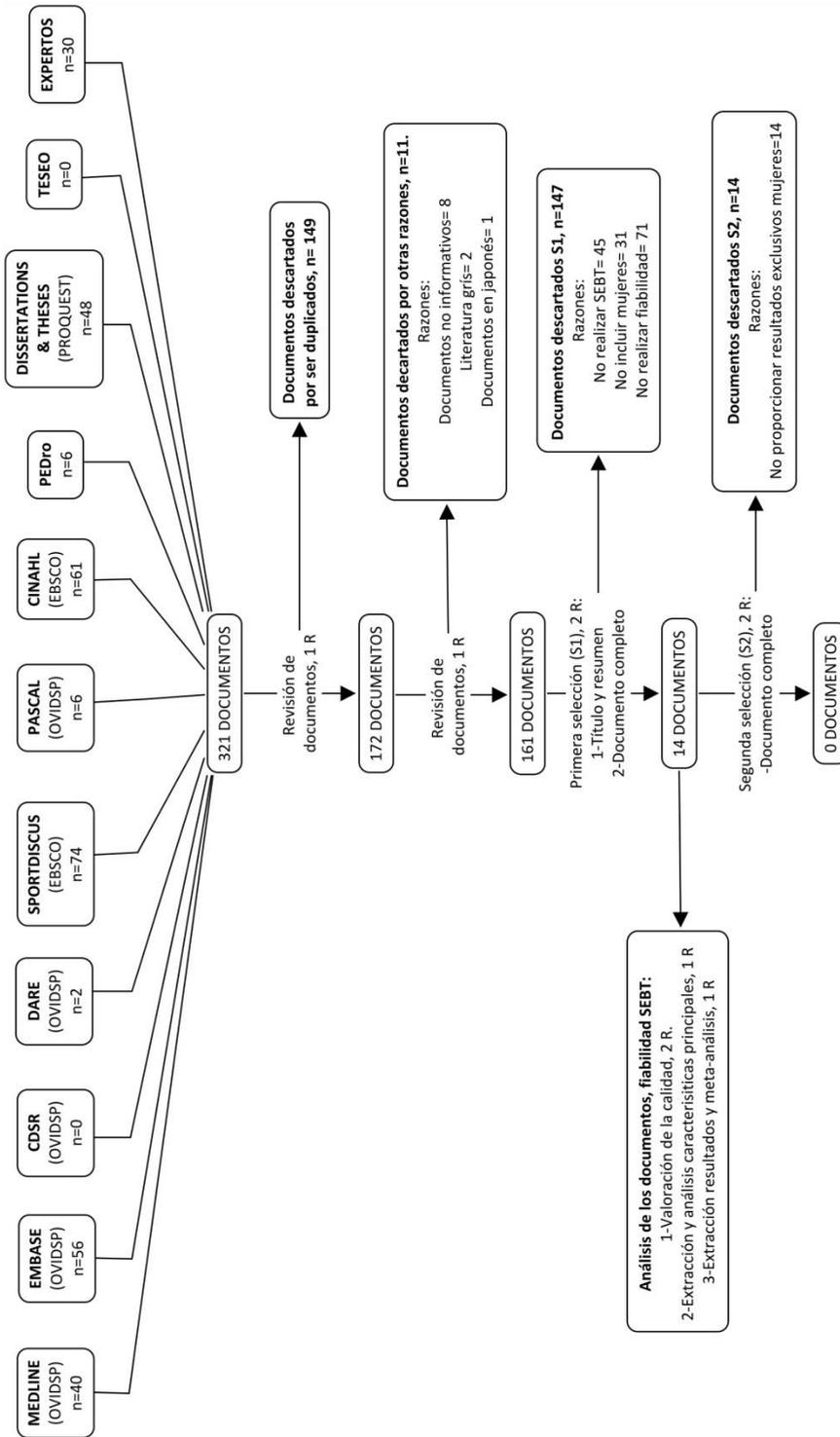


Figura 5.3.1. Diagrama de flujo de la RS SEBT.  
n - número de documentos; R - Revisores; S1 - Primera selección; S2 - Segunda selección.

En cuanto a las pérdidas de participantes, Sabin (2011)<sup>90</sup> perdió a 3,6% de los reclutados inicialmente para el estudio, y Clark et al. (2010)<sup>68</sup> no perdió ningún participante durante el estudio de fiabilidad. Los restantes 12 (85,71%) trabajos<sup>58;65;67;71-73;86;92;96;97;99;100</sup> no describen si perdieron participantes o no y en qué número lo hicieron.

Tabla 5.2.1. Calidad de los estudios incluidos en la RS SEBT.

Estudio	P1	P2	P3	P4	P5	P6	P7
Kinzey & Armstrong (1998) <sup>67</sup>	Sí	No	No	No	No	No	No
Hertel et al. (2000) <sup>65</sup>	Sí	No	No	No	No	Sí	No
Lanning et al. (2006) <sup>66</sup>	No	No	No	No	No	No	No
Plisky et al. (2006) <sup>58</sup>	No	No	No	No	No	Sí	No
English & Malone (2009) <sup>99</sup>	No	No	No	No	No	No	No
Gribble et al. (2009) <sup>73</sup>	Sí	No	No	No	Sí	Sí	No
Clark et al. (2010) <sup>68</sup>	No	No	0%	No	No	No	No
Filipa et al. (2010) <sup>96</sup>	No	No	No	No	No	No	No
Leavey et al. (2010) <sup>92</sup>	No	No	No	No	No	No	No
Munro et al. (2010) <sup>72</sup>	Sí	No	No	No	Sí	No	No
Nagano et al. (2010) <sup>71</sup>	No	No	No	No	No	No	No
Sabin et al. (2010) <sup>100</sup>	Sí	No	No	No	No	No	No
Dannelly et al. (2011) <sup>97</sup>	No	No	No	No	No	No	No
Sabin (2011) <sup>90</sup>	No	No	3,60%	No	Sí	No	No
<b>TOTAL Sí (n/%)</b>	<b>5/ 35,71%</b>	<b>0/ 0%</b>	<b>2/ 14,29%</b>	<b>0/ 0%</b>	<b>3/ 21,43%</b>	<b>3/ 21,43%</b>	<b>0/ 0%</b>

P1 - ¿La separación entre las mediciones fue igual o mayor a un día y menor de 15 días (en caso de ser mayor, fueron estas controladas)?; P2 - ¿Se completó el estudio de la fiabilidad?; P3 - ¿Cuál fue el porcentaje de pérdidas de mujeres (participantes)?; P4 - ¿Se realizó la prueba en las mismas circunstancias en todos los momentos registrados?; P5 - ¿Están adecuadamente descritos los participantes?; P6 - ¿Existe una adecuada descripción de la prueba?; P7 - ¿Hubo evaluación "ciega" de los resultados?

Ninguno de los estudios controló que las circunstancias en todos los momentos medidos fueran iguales para las participantes o no describieron este hecho, atendiendo a la cuarta pregunta sobre calidad. La descripción de las participantes fue completa atendiendo a los criterios de calidad de esta RS en tres de los trabajos<sup>72;73;90</sup>. El resto no describió cómo era los participantes empleados en el estudio de fiabilidad<sup>58;65;67;68;71;86;92;96;97;99;100</sup>.

Respecto a la pregunta número 6, tres estudios<sup>58;65;73</sup> describieron de manera adecuada el procedimiento empleado para realizar el SEBT, el resto de trabajos<sup>67;68;71;72;86;90;92;96;97;99;100</sup> no lo hicieron o no se puede saber. En lo que respecta a si se realizó una evaluación ciega, ninguno de los trabajos hace o describe tal condición. El acuerdo entre observadores se recoge en el anexo 5.2.1.

### 5.2.3-ANÁLISIS DE LAS CARACTERÍSTICAS DE LOS ESTUDIOS INCLUIDOS

Después de valorar la calidad de los estudios pasamos a extraer sus principales características y resultados mediante una revisora (ALR), para posteriormente analizarlos (tablas 5.2.2 y 5.2.3 y anexo 5.2.3).

De los 14 documentos analizados, 12 fueron artículos de revistas<sup>58;65;67;68;71-73;86;92;96;97;100</sup>, uno fue un *abstract* de las actas de un congreso<sup>99</sup> y el último de los documentos identificados una tesis<sup>90</sup>. De los 14 trabajos incluidos, el 85,71% (12) fueron de Estados Unidos<sup>58;65;67;68;73;86;90;92;96;97;99;100</sup>, uno de Japón<sup>71</sup> y otro de Reino Unido<sup>72</sup>. Respecto al año de publicación, uno fue del año 1998<sup>67</sup>, y otro del 2000<sup>65</sup>. En 2006 se realizaron dos estudios<sup>58;86</sup>, al igual que en 2009<sup>73;99</sup>. En el año 2010 se publicaron seis estudios de fiabilidad<sup>68;71;72;92;96;100</sup>, y en el 2011 dos<sup>90;97</sup>.

En cuanto al objetivo de comprobar la fiabilidad del *SEBT*, 5 trabajos<sup>65;67;68;72;90</sup> tuvieron como objetivo principal o secundario comprobar la fiabilidad de la prueba. Los restantes 9 trabajos<sup>58;71;73;86;92;96;97;99;100</sup> presentaron los resultados de fiabilidad sin definir la repetibilidad como objetivo. El tipo de fiabilidad estudiada, en 10 trabajos fue de intra-observadores del *SEBT*<sup>58;67;71-73;86;90;96;99;100</sup>, y en dos tanto la intra- como la inter-observadores<sup>65;68</sup>. En dos de los trabajos<sup>92;97</sup> no conocemos la clase de fiabilidad analizada. Para estudiar la fiabilidad, seis de los trabajos (42,86% del total) no describieron el número de sesiones empleadas en la investigación, siete realizaron 2 sesiones y un trabajo analizó 5 sesiones de evaluación diferentes. En cuanto al tiempo entre estas sesiones, la separación mínima fue de 0 días<sup>58</sup> y la máxima de 55 días<sup>90</sup>, siendo la media de  $14,25 \pm 18,48$  días.

En cuanto a los participantes, todos los trabajos a excepción de uno<sup>71</sup> definieron el número de participantes en el estudio de fiabilidad, habiéndose comprobado la fiabilidad del *SEBT* en un total de 290. Los estudios incluyeron una muestra media de  $22,31 \pm 12,25$  participantes. Conocemos que 10 trabajos incluyeron mujeres, siendo la media de inclusión de  $16,10 \pm 7,98$  mujeres. La edad media de los participantes en los ocho trabajos<sup>65;67;68;72;73;90;92;96</sup> que proporcionaron dicha información fue de 20,96 años. Respecto a la altura y el peso, siete documentos<sup>65;67;72;73;90;92;96</sup> proporcionaron dicha información siendo la altura media de los 178 participantes conocidos de 169,02 cm, con un peso de 67,68 kg. El tipo de población empleada, en el 50% de los estudios<sup>58;65;72;73;86;96;100</sup> (107 participantes) fueron practicantes de actividad física, en 21,42% de los trabajos<sup>90;92;97</sup> fueron estudiantes universitarios y 34 los participantes eran militares<sup>68</sup>. En los restantes tres documentos<sup>67;71;99</sup> no conocemos el tipo de

Tabla 5.2.2. Resumen de las características estudios RS SEBT.

Estudio	País	T íab	Obj	n ses/t (días)	n participantes (edad)	Tipo de población	Observadores : experiencia	Calentamiento (intentos por línea)	Líneas/ marcadas	Intentos/ estimador	Normalizado/ método	Análisis estadístico	MA
<b>Kinzy &amp; Armstrong (1998)</b> <sup>67</sup>	EEUU	FIA	Sí	2/7	11 M/9H (18-35 años)	ND	ND	ND	AM, PM/No	5/√ de 3 mejores	No	CCI, EEM, Spearman-Brown	Sí
<b>Hertel et al. (2000)</b> <sup>65</sup>	EEUU	FIA, FIE	Sí	2/7	8M/8H (21,3±1,3años)	Practicantes AF	A y B: ND	1	A, AM, M, PM, P, PL, L, AL/No	3/intento	No	CCI, EEM, ANOVA	Sí
<b>Lanning et al. (2006)</b> <sup>86</sup>	EEUU	FIA	No	ND/ND	5 P	Practicantes AF	ND	6	AM, PM/Sí	3/max	No	CCI, EEM	No, rango CCI para todas las líneas
<b>Pliisky et al. (2006)</b> <sup>58</sup>	EEUU	FIA	No	2/0	10 M y 4 H	Practicantes AF	ND	6	A, PM, PL/Sí	3/max	Sí/lateral	CCI (IC 95%), EM, CVEM	Sí
<b>English &amp; Malone (2009)</b> <sup>99</sup>	EEUU	FIA	No	ND/ND	29 P (30-60 años)	ND	A: Sí	6	AM, PM/ND	3/√	Sí/ND	CCI	No, rango CCI para todas las líneas y otra prueba
<b>Gribble et al. (2009)</b> <sup>73</sup>	EEUU	FIA	No	5/7	8 M y 8 H (22,5 años)	Practicantes AF	ND	6	A, M, P/Sí	3/√	Sí/medial	CCI	No, rango CCI para todas las líneas SEBT
<b>Clark et al. (2010)</b> <sup>68</sup>	EEUU	FIA, FIE	Sí	2/28	14 M y 20 H (20,71 años)	Militares	A y B: ND	1 a 6	A, PM, PL/No	3/√	Sí/ND	CCI	No, único valor CCI para todas las líneas
<b>Filipa et al. (2010)</b> <sup>96</sup>	EEUU	FIA	No	ND/ND	20 M (15,16 años)	Practicantes AF	ND	6	A, PM, PL/ND	1/intento	Sí/lateral	CCI	Sí
<b>Leavey et al. (2010)</b> <sup>92</sup>	EEUU	ND	No	ND/ND	23 M y 25 H (22,06 ± 1,58)	E universitarios	ND	6	A, AM, M, PM, P, PL, L, AL/ND	3/√	Sí/ND	CCI	No, rango CCI para todas las líneas
<b>Munro et al. (2010)</b> <sup>72</sup>	RU	FIA	Sí	2/7	11 M y 11 H (22,55 años)	Practicantes AF	A: ND	0	A, AM, M, PM, P, PL, L, AL/No	7/√ 3 últimos	Sí/medial	CCI, EEM, MDD	Sí, valores absolutos y
<b>Nagano et al. (2010)</b> <sup>71</sup>	Japón	FIA	No	ND/ND	14 piernas	ND	ND	3	A, PM, PL/Sí	3/max	Sí/medial	CCI	Sí
<b>Sabin et al. (2010)</b> <sup>100</sup>	EEUU	FIA	No	2/3	10 P	Practicantes AF	ND	3 (día previo)	A, P/No	5/max	Sí/medial	CCI	Sí
<b>Dannelly et al. (2011)</b> <sup>97</sup>	EEUU	ND	No	ND/ND	29 M	E universitarios	ND	ND	A, PM, PL/ND	3/√	Sí/ND	CCI	No, rango CCI para todas las líneas
<b>Sabin (2011)</b> <sup>90</sup>	EEUU	FIA	Sí	2/55,2 2	27 M (22,7±4,21 años)	E universitarios	ND	4	A, M, P/No	3/ intento, X y max	Sí/medial	CCI, PDC	Sí, estimador media y máxima

AF-Actividad física; Calentamiento - Realizan un estudio calentamiento previo a la prueba; Experiencia: Sí- Titulados universitarios y con experiencia en el área; No- Sin título universitario y/o sin experiencia en el área; CCI- Coeficiente de correlación intraclass; E- Estudiantes; EEM- Error estándar de la medida; FIA- Fabilidad intra-observadores; FIE- Fabilidad inter-observadores; H- hombres; IC 95%- Intervalo de confianza al 95%; Lateral - Cresta ilíaca anterosuperior maleo medial; M- mujeres; MA- ¿Se incluye el estudio en el meta-análisis?; Max - Máxima; MBA - Método de Bland y Altman; Medial - Cresta ilíaca anterosuperior maleo medial; Método - ¿Que método se emplea para medir la longitud de la pierna?; MA- No aplicable; ND - No definido; Normalizado - ¿Se normaliza la distancia en relación a la longitud de la pierna?; N ses - Número de sesiones para evaluar la fiabilidad; MDD - Mínima diferencia de detectable; Obj - ¿Es objetivo del estudio el comprobar la fiabilidad del SEBT?; Observadores - Número de observadores, cada uno representado con una letra; P - Participantes; t (días) - Tiempo entre sesiones medido en días; T íab - Tipo de fiabilidad estudiada en el trabajo; X- media; + - Proporción mismo valor para FIA y FIE.

personas incluidas. Respecto a la manera de seleccionar y a los factores de inclusión y exclusión, sólo un trabajo<sup>68</sup> explicó como lo realizó.

Atendiendo al número de observadores, el 21,42% de estudios<sup>65;72;99</sup> definió cuantos observadores emplearon al estudiar la fiabilidad. Respecto a la experiencia de los mismos, únicamente English y Malone (2009)<sup>99</sup> describieron la experiencia del observador, siendo éste experto.

Respecto a las características del *SEBT*, la media de líneas empleadas fue de 3,79  $\pm$  2,33, siendo la mínima de dos líneas y la máxima de ocho. Las tres líneas más empleadas en los 14 estudios incluidos en la RS fueron la PM (78,57%), la línea A (78,57%) y la PL (57,14%) de los trabajos. Las líneas L y AL fueron las menos utilizadas en los estudios (21,43%). El número de intentos medio por línea fue de 3,43  $\pm$  1,40, con una mínima de un intento y máxima de siete intentos. El estimador más empleado (50% de los estudios)<sup>67;68;72;73;92;97;99</sup> fue la media. La máxima de los intentos se empleó en cuatro trabajos<sup>58;71;86;100</sup> (28,57%), y dos trabajos<sup>65;96</sup> utilizaron cada intento como estimador. Sabin (2011)<sup>90</sup> utilizó la media, la máxima y los intentos para comprobar la fiabilidad del *SEBT*. Respecto al calentamiento previo a la prueba, dos estudios<sup>67;97</sup> no definieron si lo realizaron o no, los restantes doce trabajos<sup>58;65;68;71-73;86;90;92;96;99;100</sup> realizaron una media de 3,91 intentos por línea previo a la prueba. Dos trabajos<sup>92;100</sup> realizaron un calentamiento general además del específico del *SEBT*.

Tabla 5.2.3. Características cuantitativas de los 14 documento incluidos en la RS SEBT.

Variable	n est	% est	n part	% part	Media	DE	Mínima	Máxima
<b>n participantes</b>	13	92,86	290	100,00	22,31	12,25	5,00	48,00
<b>n mujeres</b>	10	71,43	161	55,52	16,10	7,98	8,00	29,00
<b>Edad (años)</b>	8	57,14	212	73,10	20,96*	-	15,16	22,70
<b>Altura (cm)</b>	7	50,00	178	61,38	169,02*	-	162,51	174,00
<b>Peso (kg)</b>	7	50,00	178	61,38	67,68*	-	54,49	75,72
<b>IMC (kg/m<sup>2</sup>)</b>	0	0,00	0	0,00	-	-	-	-
<b>n líneas</b>	14	100,00	290	100,00	3,79	2,33	2,00	8,00
<b>n intentos</b>	14	100,00	290	100,00	3,43	1,40	1,00	7,00
<b>n observadores</b>	3	21,43	67	23,10	1,33	0,58	1,00	2,00
<b>T ses (días)</b>	8	57,14	159	54,83	14,25	18,48	0,00	55,00

DE – Desviación estándar; IMC – Índice de masa corporal; n est – Número de estudios en los que la característica está descrita; n intentos – Número de intentos a realizar en el *SEBT*; n líneas – Número de intentos a realizar en cada línea; n mujeres – Número de mujeres que participan en los estudios; n observadores – Número de observadores que se utilizan en los estudios; n part – Número de participantes en los que la característica está descrita; T ses– Tiempo entre sesiones de evaluación; % est – Porcentaje de estudios en los que la característica está descrita; % part – Porcentaje de participantes en los que la característica está descrita; \* - Variable ponderada al número de participantes.

Si tenemos en cuenta el uso de calzado en la realización de la prueba, en cuatro de los trabajos<sup>65;86;92;96</sup> los participantes realizaron la prueba con calzado, en seis lo hicieron descalzos<sup>58;68;71-73;90</sup> y en cuatro<sup>67;97;99;100</sup> no se describió tal condición. Respecto al uso de los brazos, en dos estudios<sup>65;92</sup> se permitió al participante emplear los brazos para mantener el equilibrio, en seis<sup>58;71-73;96;100</sup> se les obligo a poner las manos en la cadera y en otros seis<sup>67;68;86;90;97;99</sup> no se describió la colocación de los brazos durante el *SEBT*.

En cuanto a la normalización de los resultados en función de longitud de la pierna, los tres primeros trabajos<sup>65;67;86</sup> no la realizaron. Los restantes once trabajos<sup>58;68;71-73;90;92;96;97;99;100</sup> normalizaron los resultados en función de la longitud de la pierna de cada participantes. El método para medir la longitud de la pierna fue desde la cresta íliaca antero superior hasta el maleolo medial en el 35,71% (n=5)<sup>71-73;90;100</sup> de los trabajos y hasta el lateral en 14,29% (n=2)<sup>58;96</sup>. Cuatro de los trabajos<sup>68;92;97;99</sup> no describieron el método para medir la pierna.

Respecto al método estadístico para evaluar la fiabilidad de la prueba, todos los estudios emplearon el CCI aportando su estimación puntual y 1 trabajo<sup>58</sup> proporcionó su intervalo de confianza al 95%. Cinco de los trabajos<sup>58;65;67;72;86</sup> calcularon también el EEM de la prueba, uno<sup>72</sup> el MDD y otro el PCD<sup>90</sup>. Todo los trabajos incluidos en esta RS, a excepción del de Kinzey y Armstrong (1998)<sup>67</sup> concluyeron que el *SEBT* era una prueba fiable.

### 5.3.4-ANÁLISIS DE LAS CARACTERÍSTICAS DE LOS SUBESTUDIOS

Una vez extraídos todos los datos, dividimos los estudios en subestudios empleando el criterio de tipo de fiabilidad; intra- o inter-observadores (anexo 5.2.4). Tuvimos que descartar dos estudios, el de Leavey et al. (2010)<sup>92</sup> y el de Dannelly et al. (2010)<sup>97</sup> de este análisis puesto que no conocíamos el tipo de fiabilidad que analizaron en sus respectivos estudios. Realizamos un análisis descriptivo para las características extraídas (anexo 5.2.5).

Identificamos 12 subestudios<sup>58;65;67;68;71-73;86;90;96;99;100</sup> que habían estudiado la fiabilidad intra-observadores del *SEBT*, de los cuales once<sup>58;65;67;68;72;73;86;90;96;99;100</sup> reportaron el número de participantes siendo un total de 213. Un trabajo<sup>71</sup> describió el número de piernas analizadas. El objetivo de comprobar la fiabilidad fue principal o secundario en cinco (41,67%)<sup>65;67;68;72;90</sup> de los estudios sobre FIA *SEBT* y en los dos estudios<sup>65;68</sup> que comprobaron la FIE de la prueba.

El número de mujeres estudiadas en los trabajos de FIA *SEBT* fue de 109 (51,17%), estando definido el número de mujeres en ocho trabajos<sup>58;65;67;68;72;73;90;96</sup>. La media de participación de los estudios fue de 19,36 ± 8,47, con 13,62 ± 6,67 mujeres. En cuanto a la FIE *SEBT*, ésta se estudió en dos

ocasiones<sup>65;68</sup> con un total de 50 participantes siendo 22 de ellos mujeres. La media de participantes por estudio fue de  $25,00 \pm 12,73$  y  $11,00 \pm 4,24$  mujeres. Respecto a las características de los participantes, en la FIA *SEBT* la edad media fue de 20,87 años en los seis estudios<sup>65;68;72;73;90;96</sup> que reportaron dicho dato, y de 21,58 en los dos estudios<sup>65;68</sup> sobre FIE. La altura media de los participantes fue de 168,41 cm y 66,10 kg en los 101 participantes de los cinco estudios<sup>65;72;73;90;96</sup> que describieron estas características. El 58,33%<sup>58;65;72;73;86;96;100</sup> de los estudios sobre FIA emplearon participantes practicantes de actividad física, en un trabajo militares<sup>68</sup> y en otro estudiantes universitarios<sup>90</sup>, siendo desconocida esta variables en tres trabajos<sup>67;71;99</sup> y un total de 49 participantes. En FIE *SEBT* los participantes fueron en un estudio militares (n=34)<sup>68</sup> y en el otro practicantes de actividad física (n=16)<sup>65</sup>.

Atendiendo a las características del procedimiento del *SEBT*, el número medio de líneas por estudio fue de  $3,50 \pm 2,15$  líneas, con un mínimo de dos líneas y un máximo de ocho en caso del FIA. Las líneas más empleadas fueron la PM y la A que fueron utilizadas en el 66,67% de los trabajos, y 41,67% respectivamente. Estas líneas también fueron la combinación de líneas más empleada en los trabajos, ya que se analizaron en un total de cuatro subestudios<sup>58;68;71;96</sup>. En FIE, la A, PM y PL también fueron las más comunes, empleando en el trabajo de Hertel et al. (2000)<sup>65</sup> las combinación de las 8 líneas y en el de Clark et al. (2010)<sup>68</sup> la A, PM y PL. Las líneas estuvieron en la mayoría de los casos sin los cm marcados, en seis subestudios<sup>65;67;68;72;90;100</sup> FIA y en todos<sup>65;68</sup> los de FIE.

En cuanto a la colocación del pie para la realización del *SEBT*, en FIA éste se colocó en diferentes posiciones en función del estudio. Tres trabajos<sup>67;71;86</sup> ubicaron en el pie en cuadrado dibujado en el centro de la estrella, otros tres subestudios<sup>65;68;73</sup> posicionaron el centro del pie en la intersección de la líneas, dos trabajos<sup>72;90</sup> colocaron el talón y otros dos<sup>58;96</sup> el primer dedo del pie en el centro de la estrella. En la mayoría (n=6, 50%)<sup>58;68;71-73;90</sup> de los trabajo se prefirió realizar la prueba sin calzado y con las manos agarradas en las caderas (n=6, 50%)<sup>58;71-73;96;100</sup>. Hubo dos trabajos<sup>99;100</sup> que no definieron la ubicación del pie, tres<sup>67;99;100</sup> el uso o no del calzado y cinco<sup>67;68;86;90;99</sup> el empleo de manos. En FIE, ambos subestudios<sup>65;68</sup> colocaron el centro del pie en la intersección de las líneas, realizando un estudio<sup>65</sup> la prueba con calzado y el otro<sup>68</sup> con los participantes descalzos. El uso de las manos fue libre en trabajo de Hertel et al. (2000)<sup>65</sup> y no definió en Clark et al. (2010)<sup>68</sup>.

Los trabajos que estudiaron la FIA realizaron en su mayoría (n=5, 41,67%)<sup>58;73;86;96;99</sup> un calentamiento previo de seis intentos por línea. El resto de trabajos varió de un intento<sup>65</sup> hasta cuatro<sup>90</sup>, dejando a la elección realizar de uno

a seis intentos por línea en el trabajo de Clark et al. (2010)<sup>68</sup>. En cuanto a la FIE ambos trabajos<sup>65;68</sup> realizaron diferentes calentamientos.

En cuanto a la normalización de los resultados obtenidos en cada línea en función de la longitud de la pierna, nueve de los trabajos lo hicieron para comprobar la FIA, cinco<sup>71-73;90;100</sup> valiéndose del maléolo medial y dos<sup>58;96</sup> del lateral y los restantes dos<sup>68;99</sup> no describieron el método. En el caso de FIE, un trabajo no definió<sup>68</sup> el método para medir la pierna y en otro caso<sup>65</sup> no se normalizó los resultados obtenidos en cada línea.

En cuanto al número de intentos a realizar para evaluar la FIA del *SEBT*, la media fue de  $3,50 \pm 1,50$  intentos por línea, siendo el número mínimo de intentos uno y el máximo de siete. Respecto al estimador empleado, la media se empleó en tres estudios<sup>68;73;99</sup> y la máxima en cuatro<sup>58;71;86;100</sup>. Dos de los trabajos<sup>65;96</sup> emplearon cada intento como estimador. En FIE, se realizó en ambos estudios tres intentos, empleándose en uno<sup>65</sup> cada intento como estimador y el otro<sup>68</sup> la media. En cuanto a los observadores, en FIA tres estudios<sup>65;72;99</sup> reportaron el número de los mismos y un único trabajo describió su experiencia<sup>99</sup>. En FIA, Hertel et al. (2000)<sup>65</sup> empleo dos observadores, no describiendo la experiencia de los mismos.

Los subestudios que analizaron la FIA se separaron las sesiones de evaluación una media de  $14,25 \pm 18,48$  días, siendo el número que reporta este dato de ocho estudios (66,67%). En FIE, un estudio<sup>68</sup> realizó las sesiones de evaluación el mismo día y el otro<sup>65</sup> separando las sesiones siete días. Respecto a que evaluaciones fueron las que compararon, en FIA compararon las evaluaciones entre las diferentes sesiones de evaluación en los nueve estudios<sup>58;67;68;71-73;86;90;100</sup>, otro comparó los intentos dentro de la misma sesión<sup>65</sup> y en dos no definieron como lo hicieron<sup>96;99</sup>. En FIE, un trabajo<sup>65</sup> comparó los intentos realizados y el otro<sup>68</sup> las evaluaciones realizadas dentro de la misma sesión de evaluación.

Atendiendo a los análisis estadísticos, todos los estudios emplearon el CCI para comprobar tanto la FIA como el FIE aportando su estimación puntual. Un trabajo<sup>58</sup> que estudio la FIA *SEBT* proporcionó además su intervalo de confianza al 95%. En cuanto al modelo de Fisher de CCI empleando, en FIA dos trabajos<sup>65;67</sup> utilizaron el 2,1 y otros dos<sup>58;72</sup> el 3,1. Sabin (2011)<sup>90</sup> utilizó el modelo 2,1 cuando empleó la máxima como estimador y el 2,k para la media, y Nagano et al. (2010)<sup>71</sup> utilizó el 1,1. El 50% de los subestudios de FIA no reportaron el modelo empleado. En FIE, Hertel et al. (2000)<sup>65</sup> emplearon el modelo 2,1 y Clark et al. (2010)<sup>68</sup> no describieron cual utilizaron. En cuanto a los métodos dimensionales, el EEM se calculó en cuatro subestudios<sup>65;67;72;86</sup> de FIA y en uno de FIE<sup>65</sup>. La MDD se obtuvo en un trabajo<sup>72</sup> de FIA *SEBT* y el PDC en otro<sup>90</sup> también sobre FIA. El método de Bland y Altman (MBA) no se empleó en ningún trabajo.

En lo que respecta a las conclusiones, todos los autores concluyeron que la prueba era fiable o consideraron que el *SEBT* era fiable empleando la prueba en el estudio principal, a excepción de Kinzey & Armstrong (1998)<sup>67</sup> que consideraron que la prueba no obtenía suficiente fiabilidad.

### 5.2.5-RESULTADOS DEL CCI DE LOS ESTUDIOS PRIMARIOS Y META-ANÁLISIS

Extrajimos todos los resultados de los estudios que presentaban los datos de manera analizable, los cuales se pueden observar en el anexo 5.2.6. Estos resultados los representamos en función de líneas mediante *forest plot*, (figuras 5.2.2-5.2.4). Los valores que ilustramos en los gráficos son los obtenidos tras la transformación Z de Fisher y su posterior reconversión en CCI.

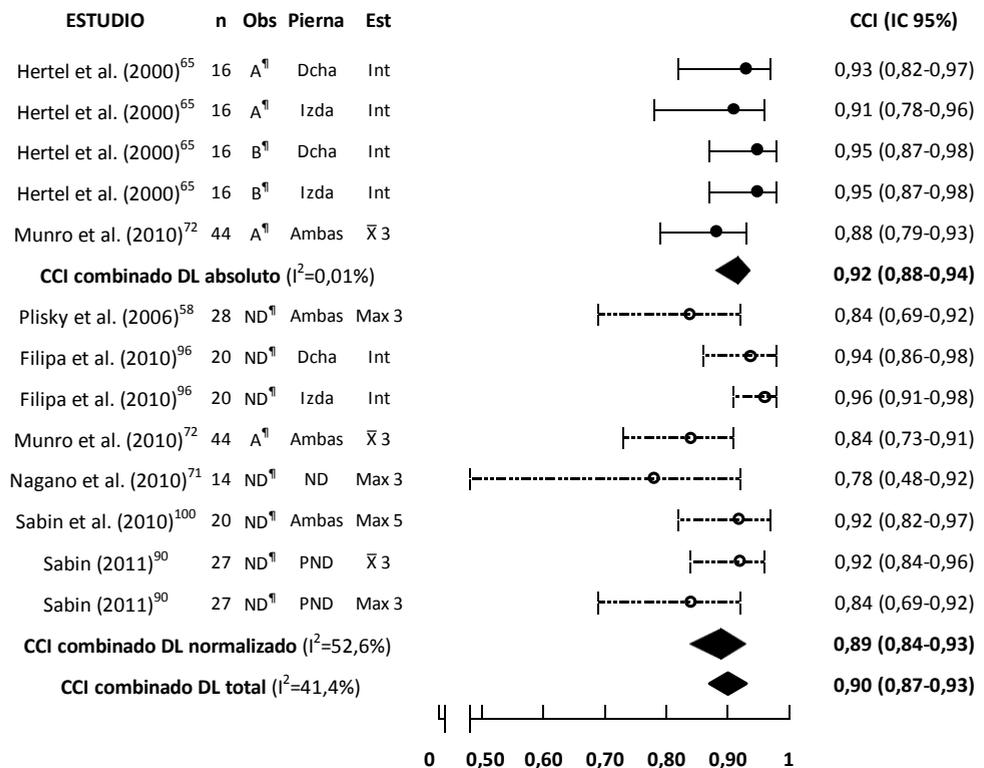


Figura 5.2.2. Gráfico del meta-análisis *SEBT* línea anterior (A).  
 DL - DerSimonian-Laird; Est – Estimador; IC 95% – Intervalo de confianza al 95; Int – Intento;  $I^2$  - I cuadrada;  
 Max – Máxima; ND – No definido; PND – Pierna no dominante;  $\bar{X}$  - media; ¶ - Experiencia observador no  
 definida; |—| - Intervalo de confianza en valores absolutos; |-| - Intervalo de confianza en valores  
 normalizados; ● - Estimación puntual en valores absolutos; ○ – Estimación puntual en valores normalizados; ◆  
 - CCI combinado con IC 95%..

En cuanto a la inclusión de los resultados de los trabajos en el meta-análisis (MA) siete de los trabajos<sup>58;65;67;71;90;96;100</sup> proporcionaron datos de manera analizable. Los restantes siete trabajos<sup>73;86;92;97;99</sup> aportaron un rango de CCI o un único valor de CCI<sup>68</sup> para todas las líneas empleadas. Más concretamente, en la FIA ocho<sup>58;65;67;71;72;90;96;100</sup> de los 12 (66,67%) trabajos analizados en esta fase proporcionaron los datos de una manera analizable. En el trabajo de Munro et al. (2010)<sup>72</sup> empleamos tanto los valores normalizados como los absolutos para el análisis. Del trabajo de Sabin et al. (2011)<sup>90</sup> empleamos los análisis que realizaron los autores utilizando la media y la máxima como estimadores. En el caso del FIE, no realizamos MA puesto que solo un trabajo<sup>65</sup> proporcionó datos de manera analizable.

Todos los MA realizados (figuras 5.2.2-5.2.4) muestran una consistencia aceptable con unas I<sup>2</sup> entorno al 50 % o menores. Los resultados analizados, a excepción del de Nagano et al. (2010)<sup>71</sup>, muestran unos CCI por encima de 0,80 con unos IC al 95% de muy similares, independientemente del estimador o de si se emplea valores absolutos o en relación a la longitud de la pierna.

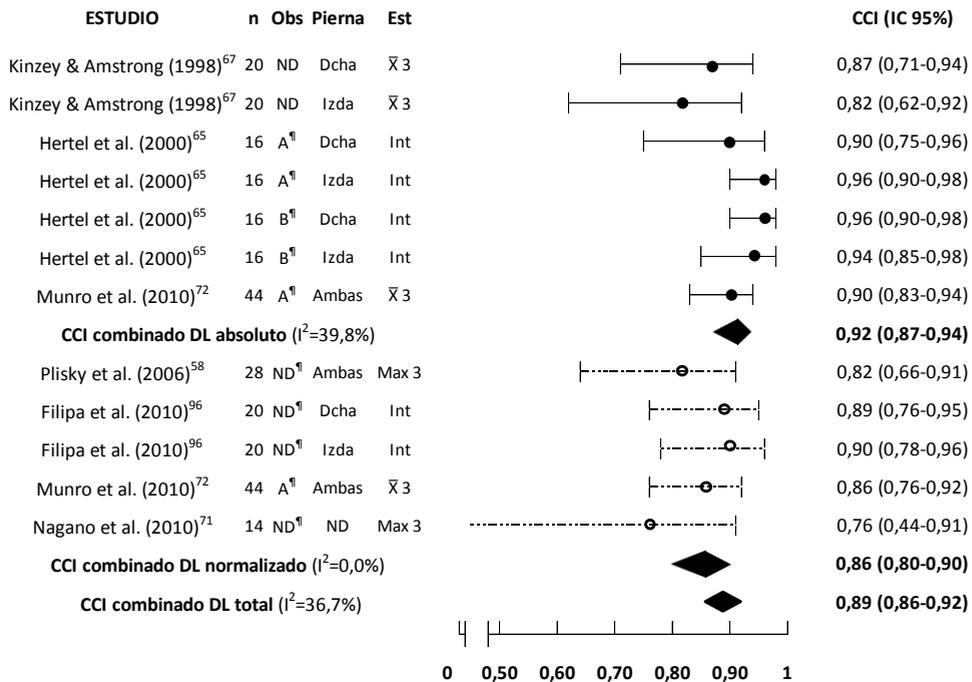


Figura 5.2.3. Gráfico del meta-análisis SEBT línea posteromedial (PM).  
 DL - DerSimonian-Laird; Est – Estimador; IC 95% – Intervalo de confianza al 95; Int – Intento; I<sup>2</sup> - I cuadrado; Max – Máxima; ND – No definido; X - media; ¶ - Experiencia observador no definida; |—| - Intervalo de confianza en valores absolutos; |-|-| - Intervalo de confianza en valores normalizados; ● - Estimación puntual en valores absolutos; ○ – Estimación puntual en valores normalizados; ◆ - CCI combinado con IC 95%..

Los CCI combinados muestran una fiabilidad muy buena del *SEBT* cuando se utilizan valores absolutos, no habiendo diferencias importantes entre la fiabilidad de las líneas. Los CCI en este caso son de 0,92 (IC 95%: 0,88-0,94) en la línea A, de 0,92 (IC 95%: 0,87-0,94) en la PM y de 0,93 (IC 95%: 0,89-0,95).

El CCI combinado de los valores normalizados es ligeramente menor, pero muestra que la fiabilidad también es buena y no habiendo diferencias entre líneas. Los valores en relación a la longitud de la pierna la línea A obtiene un CCI de 0,89 (IC 95%: 0,84-0,93), la PM de 0,86 (IC 95%: 0,80-0,90) y la PL de 0,86 (IC 95%: 0,78-0,91).

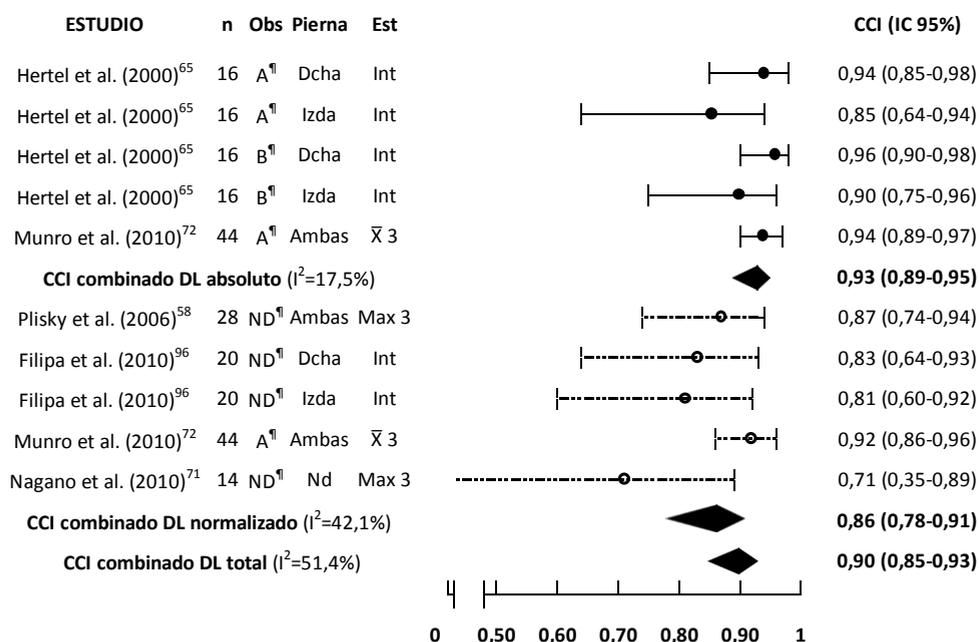


Figura 5.2.4. Gráfico del meta-análisis *SEBT* línea posterolateral (PL).

DL - DerSimonian-Laird; Est - Estimador; IC 95% - Intervalo de confianza al 95; Int - Intento;  $I^2$  - I cuadrado; Max - Máxima; ND - No definido;  $\bar{X}$  - media; ¶ - Experiencia observador no definida; |—| - Intervalo de confianza en valores absolutos; |—| - Intervalo de confianza en valores normalizados; ● - Estimación puntual en valores absolutos; ○ - Estimación puntual en valores normalizados; ◆ - CCI combinado con IC 95%.

### 5.2.6-RESULTADOS DIMENSIONALES DE LOS ESTUDIOS PRIMARIOS

De los 14 estudio incluidos en la revisión sistemática, cuatro proporcionaron datos dimensionales de manera analizable: Kinzey y Armstrong (1998)<sup>67</sup>, Hertel et al. (2000)<sup>65</sup>, Munro et al. (2011)<sup>72</sup> y Sabin (2011)<sup>90</sup>.

En estos resultados encontramos en valores absolutos errores de medición analizados mediante MDD que van de 4,71 cm a 11,06 cm (figuras 5.2.5-5.2.7). El error de medición más alto lo aporta el trabajo Kinzey y Armstrong (1998)<sup>67</sup>. Si obviamos este trabajo encontramos que el error aleatorio más alto es de 8,40 cm, y que la mayoría de los trabajo reportan un MDD de 5 o 6 cm.

Si comparamos las MDD entre líneas encontramos que la línea A es aparentemente la más precisa, y la PL la menos fiable. Aunque existen discrepancias en función del observador en el trabajo del Hertel et al. (2000)<sup>65</sup> y entre los trabajos consultados.

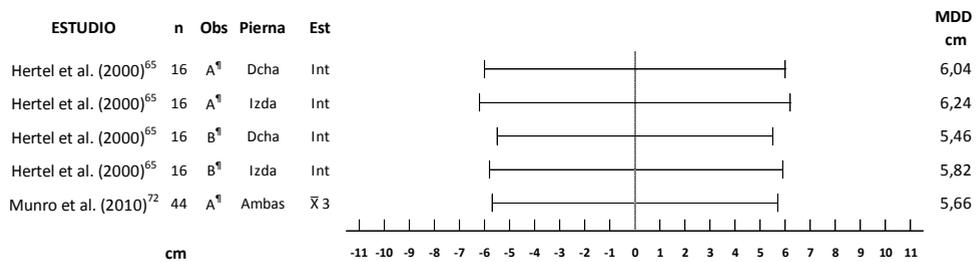


Figura 5.2.5. Gráfico mínima diferencia detectable (MDD) SEBT línea anterior (A).  
Est – Estimador; Int – Intento; X̄ – media; ¶ – Experiencia observador no definida; |—| – MDD en cm.

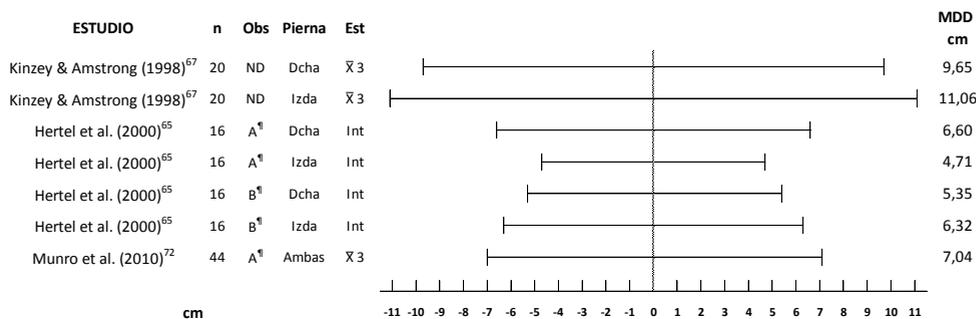


Figura 5.2.6. Gráfico mínima diferencia detectable (MDD) SEBT línea posteromedial (PM).  
Est – Estimador; Int – Intento; X̄ – media; ¶ – Experiencia observador no definida; |—| – MDD en cm.

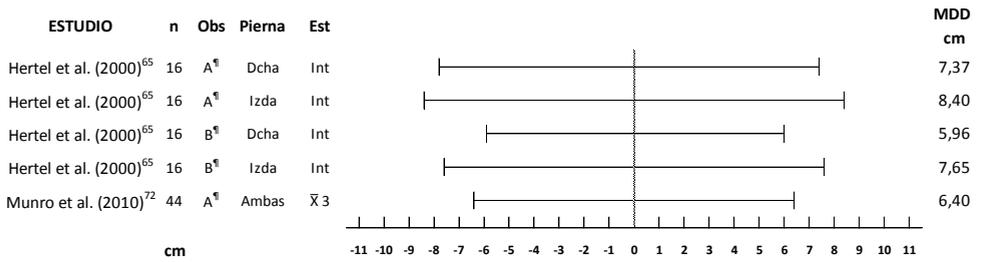


Figura 5.2.7. Gráfico mínima diferencia detectable (MDD) SEBT línea posterolateral (PL).  
Est – Estimador; Int – Intento; X̄ - media; ¶ - Experiencia observador no definida; |—| – MDD en cm.

En lo que respecta a los valores normalizados únicamente tenemos datos de dos trabajos, el de Munro et al. (2010)<sup>72</sup> y el de Sabin (2011)<sup>90</sup>, reportando este último trabajo el error del método en PCD. Presentamos los resultados en un único gráfico debido a la escasez de resultados (figura 5.2.8). El porcentaje de error va de 6,87% a 8,15% dependiendo de la línea. En cuanto al PDC, este es superior, de 13%.

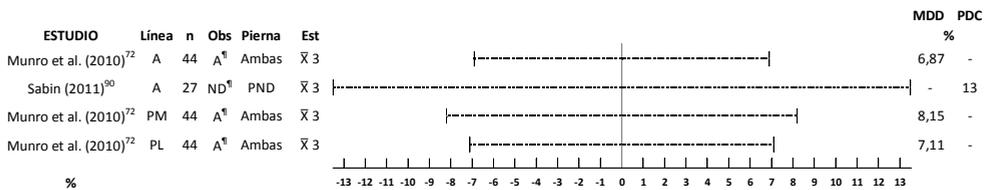


Figura 5.2.8. Gráfico mínima diferencia detectable (MDD) y punto de corte (PDC) del SEBT con valores normalizados.

Est – Estimador; Int – Intento; X̄ - media; ¶ - Experiencia observador no definida; |—| – MDD o PDC según proceda en valores noremalizados.

**REVISIÓN SISTEMÁTICA SOBRE EL *STAR EXCURSION BALANCE TEST*:  
SÍNTESIS DE LOS RESULTADOS**

- **14 documentos** han comprobado la **fiabilidad del *Star Excursion Balance Test*** con muestras que ha incluido **mujeres**:
  - La **fiabilidad intra-observadores** se ha comprobado en **12** ocasiones, y la **fiabilidad inter-observadores** en **2**.
- La **calidad** de los estudios es **mala** y los **métodos** empleados son **dispar** (heterogéneos).
- **Todos** los estudios han empleado el **coeficiente de correlación intraclase** y **algunos** proporcionan también **medidas dimensionales**.
- Los meta-análisis concluye que el **coeficiente de correlación intraclase combinado es bueno o muy bueno**.
- Las medidas **dimensionales** muestran un **error aleatorio muy alto**.

## **5.3-RESULTADOS DE LA REVISIÓN SISTEMÁTICA DE PRUEBAS QUE VALORAN LA DORSIFLEXIÓN EN APOYO**

### **5.3.1-SELECCIÓN DE ESTUDIOS**

A partir de la búsqueda electrónica y a la consulta con expertos identificamos 527 documentos supuestamente pertinentes con el objetivo del estudio. Una revisora examinó los documentos uno a uno descartando 128 por ser duplicados, basándose en los criterios previamente establecidos. Una vez descartadas todas las copias, la revisora verificó que los 399 documentos disponían de *abstract*, y que esté proporcionaba suficiente información para realizar la selección. Por esta razón siete documentos fueron descartados, ya que la base de datos *Sportdiscus* nos proporcionó siete referencias que correspondían a las actas completas de diferentes congresos, y en el apartado de *abstract* aparecía el título de todas las presentaciones realizadas en las mismas. Todos estos títulos fueron revisados para comprobar que ninguna referencia pertinente con el objeto del estudio quedara descartada (figura 5.3.1).

Iniciamos la primera selección con 392 referencias mediante la cual descartamos empleando dos revisores independientes 330 documentos, no siendo necesario recurrir a un tercer revisor en ninguno de los casos. El número de documentos eliminados por no realizar ninguna prueba de DF en apoyo fue de 319 (96,6%). Dos de los trabajos fueron descartados por no incluir mujeres (0,6%) en el mismo y 9 (2,73%) por no comprobar a fiabilidad de PDFa. De tal manera que 62 referencias, un 15,82% del total de documentos incluidos en la primera selección, pasaron a la segunda selección, todos ellos por ser imposible determinar si cumplían o no los criterios de inclusión mediante el título y el *abstract*. Respecto al seguimiento de referencias, no identificamos ningún documento que no se hubiera incluido en la RS previamente.

Finalmente, en la segunda selección en la cual los revisores utilizaron el documento completo, descartamos las 62 referencias incluidas en esta fase de selección, 52 de ellas por no comprobar la fiabilidad de ninguna PDFa, y 10, por no proporcionar datos exclusivos del grupo de mujeres.

Al comprobar que ningún documento cumplía con los criterios de inclusión de la RS, tuvimos que modificar dichos criterios. De tal manera que incluimos en la revisión todos aquellos estudios que habían estudiado la fiabilidad de alguna PDFa, y que habían incluido mujeres en el mismo, es decir, los 10 documentos en los que los revisores habían respondido "Sí" en la primera pregunta de la segunda selección (figura 5.3.1).

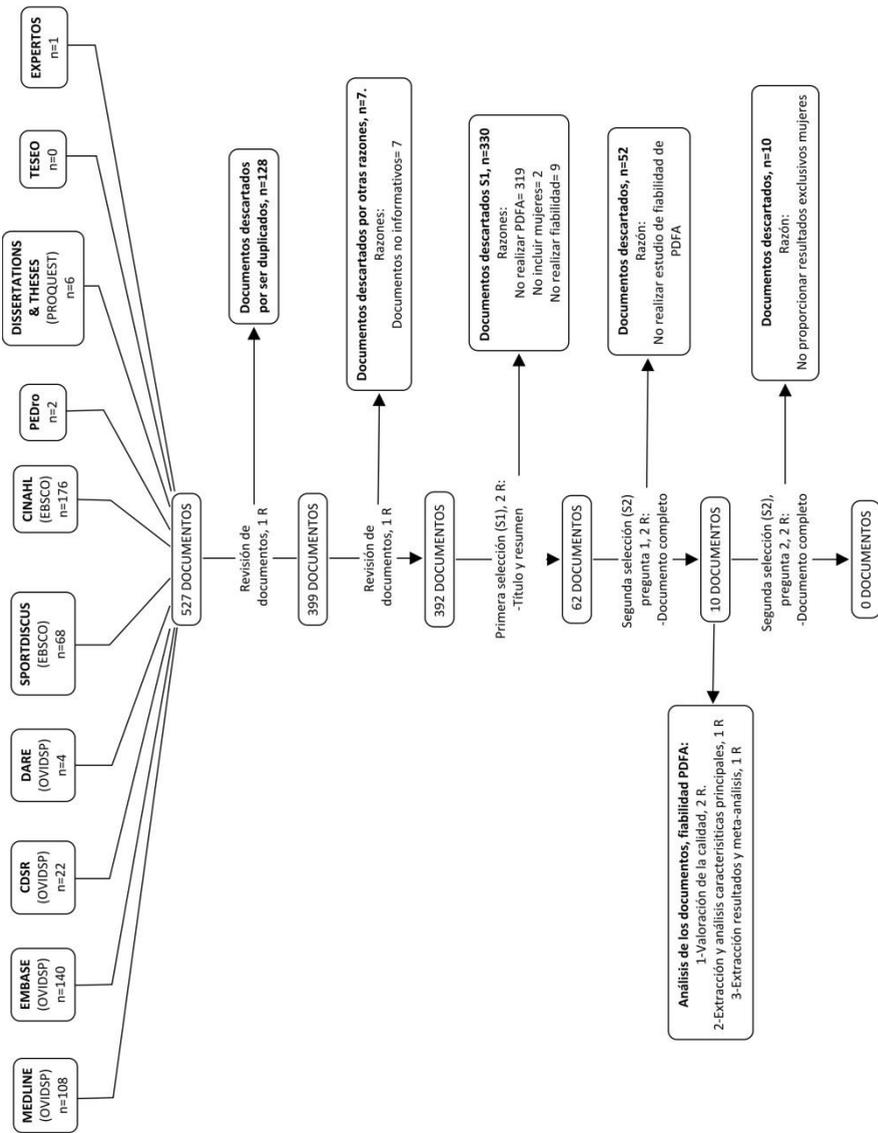


Figura 5.2.1. Diagrama de flujo de la RS PDFAs.  
n - número de documentos; R - Revisores; S1 - Primera selección; S2 - Segunda selección.

Tanto en la primera como en la segunda selección se evaluó el acuerdo alcanzado por los revisores antes de consensuar las respuestas mediante la Kappa de Cohen y el acuerdo observado, proporcionando la estimación puntual y los intervalos de confianza de cada uno de ellos (anexo 5.3.1).

### 5.3.2-CALIDAD DE LOS ESTUDIOS

Con los 10 documentos que habían comprobado la fiabilidad de alguna PDFA, realizamos la valoración del riesgo de sesgo o calidad mediante dos revisores independientes (tabla 5.3.1 y anexo 5.3.2). La valoración del riesgo de sesgo se realizó de manera separada para cada una de las pruebas, de manera que el trabajo de Barton et al. (2005)<sup>171</sup>, se valoró en dos ocasiones, una para cada prueba. Al ser las respuestas a las cuestiones sobre calidad idénticas para la PDFAf y la PDF Ae, se presenta en la tabla como un único trabajo, no así para valorar el acuerdo entre revisores (anexo 5.3.1). En esta fase tampoco fue necesaria la consulta a un tercer revisor, puesto que tras de la revisión independiente los revisores consensuaron todas las respuestas.

Respecto a la valoración del riesgo de sesgo, únicamente dos estudios (20% de los incluidos en la RS), los de Burne et al. (2004)<sup>143</sup> y Crossley et al. (2007)<sup>172</sup>, separaron las evaluaciones a comparar más de un día pero menos de quince (P1). Seis de los trabajos<sup>121;126;130;134;135;171</sup>, evaluaron ambos tipo de fiabilidad separando la FIA siguiendo los criterios de calidad de esta revisión, pero para la FIE compararon las evaluaciones correspondientes a una misma sesión, es decir día. Del mismo modo, Rabin y Kozol (2010)<sup>173</sup>, que evaluaron únicamente la FIE, realizaron una única sesión y Dinh et al. (2011)<sup>125</sup> no definieron el tiempo entre sesiones en el método (P1). Respecto a sí se completó el estudio de fiabilidad (P2) ninguno de los trabajos analizados lo definió, así como tampoco determinaron si hubo alguna pérdida de participantes (P3) desde el inicio hasta el final del estudio. Del mismo modo, ninguno de los trabajos pudo garantizar que la situación de los participantes era la misma en cada par de mediciones comparadas (P4), ya que o bien no lo describen o no lo hacen.

En cuanto a la descripción de la prueba (P5) y de los participantes (P6), ningún trabajo cumple los criterios de cuantificación del sesgo establecidos en esta RS, a excepción de Munteanu et al. (2009)<sup>130</sup> que describe correctamente la prueba. Atendiendo a si la evaluación de los resultados fue ciega (P7), 5 estudios describieron tal condición (50%), los de Jones et al. (2005)<sup>135</sup>, Munteanu et al. (2009)<sup>130</sup>, Rabin y Kozol (2010)<sup>173</sup>, Dinh et al. (2011)<sup>125</sup> y Krause et al. (2011)<sup>126</sup>, el resto de trabajos<sup>121;134;143;171;172</sup> no mencionaron nada al respecto o bien no lo hicieron.

Tabla 5.3.1. Calidad de los estudios incluidos en la RS PDF/A.

Estudio	Prueba	P1	P2	P3	P4	P5	P6	P7
Bennell et al. (1998) <sup>121</sup>	PDFAf	No	No	No	No	No	No	No
Burne et al. (2004) <sup>143</sup>	PDFAe	Sí	No	No	No	No	No	No
Nitz y Low Choy (2004) <sup>134</sup>	PDFAf	No	No	No	No	No	No	No
Jones et al. (2004) <sup>135</sup>	PDFAf	No	No	No	No	No	No	Sí
Crossley et al. (2007) <sup>172</sup>	PDFAf	Sí	No	No	No	No	No	No
Munteanu et al. (2009) <sup>130</sup>	PDFAe	No	No	No	No	No	Sí	Sí
Barton et al. (2010) <sup>171</sup>	PDFAf y PDFAe	No	No	No	No	No	No	No
Rabin y Kozol (2010) <sup>178</sup>	PDFAf	No	No	No	No	No	No	Sí
Dinh et al. (2011) <sup>173</sup>	PDFAe	No	No	No	No	No	No	Sí
Krause et al. (2011) <sup>126</sup>	PDFAe	No	No	No	No	No	No	Sí
<b>TOTAL (n/%)</b>	<b>10/100%</b>	<b>2/20%</b>	<b>0/0%</b>	<b>0/0%</b>	<b>0/0%</b>	<b>0/0%</b>	<b>1/10%</b>	<b>5/50%</b>

P1 - ¿La separación entre las mediciones fue igual o mayor a un día y menor de 15 días (en caso de ser mayor, fueron estas controladas)?; P2 - ¿Se completó el estudio de la fiabilidad?; P3 - ¿Cuál fue el porcentaje de pérdidas de mujeres(participantes)?; P4 - ¿Se realizó la prueba en las mismas circunstancias en todos los momentos registrados?; P5 - ¿Están adecuadamente descritos los participantes?; P6 - ¿ Existe una adecuada descripción de la prueba?; P7 - ¿Hubo evaluación "ciega" de los resultados? Ver anexo 5.3.2 para ver las causas de la valoración.

### 5.3.3-ANÁLISIS DE LAS CARACTERÍSTICAS DE LOS ESTUDIOS INCLUIDOS

Posteriormente pasamos a extraer y analizar las principales características de los trabajos que incluimos en la RS (anexos 5.3.3 y 5.3.4, tablas 5.3.2 y 5.3.3) mediante una revisora. La mayoría de los estudios se habían realizado en Australia (60%)<sup>121;130;134;143;171;172</sup>, habiéndose desarrollado dos en EEUU<sup>125;126</sup>, uno en Israel<sup>173</sup> y otro en Reino Unido<sup>135</sup>.

En cuanto la prueba de DF que estudiaron los trabajos, identificamos cuatro estudios que habían comprobado la fiabilidad del PDFAe<sup>125;126;130;143</sup>, y cinco la del PDFAf<sup>121;134;135;172;173</sup>. Uno de los documentos incluidos en la RS, el de Barton et al. (2010)<sup>171</sup>, analizaron la fiabilidad de ambas pruebas. Los trabajos estudiados en esta revisión analizaron tanto la FIA como la FIE en siete ocasiones<sup>121;125;126;130;134;135;171</sup>, comprobándose exclusivamente la FIE en un trabajo<sup>173</sup> y la FIA en dos<sup>143;172</sup>.

Respecto a si el objetivo del estudio era el de comprobar la fiabilidad, en cuatro (50%) estudios sí que lo fue<sup>121;126;130;135;171</sup>. En el resto de trabajos (50%) los resultados de fiabilidad se presentaron dentro del método o resultados sin que ni siquiera se planteara su objetivo<sup>125;134;143;172;173</sup>. El 90% de los trabajos estudiaron la fiabilidad realizando dos sesiones de evaluación. El único trabajo<sup>173</sup> que no realizó dos sesiones valoró la fiabilidad inter-observadores realizando una única sesión. Respecto a la separación media entre sesiones de evaluación, esta fue de 9,62 días, con una mínima de siete días y la máxima de 21.

Tabla 5.3.2. Resumen de las características estudios RS PDFA

Estudio	País	Prueba	T fiab	Obj	n ses/ t (días)	n participantes (edad)	Tipo de población	Observadores: experiencia	Calentamiento	Variable(s) evaluada	Método evaluación	Alineación segmentos	Intentos/ Estimador	Análisis estadístico	MA
Bennell et al. (1998) <sup>21</sup>	Australia	PDFAF	FIA, FIE	Sí	2/7	5M y 8H (18,8 ± 2 años)	E universitarios	A: No; B, C y D: Sí	ND	1-Ángulo vertical-tibia, 2- distancia dedo pared	Inclinómetro tibia, 2-cinta métrica	Talón-dedo-rodilla	3/√	CCI (IC 95%), EEM	Sí
Burne et al. (2004) <sup>21</sup>	Australia	PDFAe	FIA	No	2/7	14	Militares	A:ND	ND	ND	Goniómetro	ND	ND	CCI	Sí
Nitz y Low Choy (2004) <sup>24</sup>	Australia	PDFAF	FIA, FIE	No	2/14	15M	ND	A, B y C: ND	ND	Ángulo vertical-tibia	Inclinómetro tibia	ND	ND	CCI†	No, único valor CCI para FIA y FIE
Jones et al. (2005) <sup>25</sup>	Reino Unido	PDFAF	FIA, FIE	Sí	2/7	9M (40,00 ± 5,27) y 8H (39,9 ± 5,54)	ND	A, B Y C: Sí	Movilidad tobillo, estramientos	Distancia que la rodilla sobrepasa pie	Aparato con cinta métrica	Pie-pantorrilla	6/√	CCI (IC 95%), MBA	Sí
Crossley et al. (2007) <sup>172</sup>	Australia	PDFAF	FIA	No	2/7	10	ND	ND	ND	Ángulo vertical-tibia	Inclinómetro tibia	Talón-dedo-rodilla	3/√	CCI, EEM	Sí
Munteanu et al. (2009) <sup>180</sup>	Australia	PDFAe	FIA, FIE	Sí	2/7	20M y 10H (22,1 ± 5,6)	E universitarios	A, B Y C: Sí; D: No	Estramientos, familiarización prueba	1-Ángulo vertical-tibia, 2- ángulo maleolo-perone	1- Inclinómetro tibia, 2- cristal lateral	Talón-dedo	2/√	CCI (IC 95%), EEM, MBA	Sí
Barton et al. (2010) <sup>21</sup>	Australia	PDFAe, PDFAF	FIA, FIE	Sí	2/21	15	E universitarios	A, B Y C: Sí	Caminar 1 min, estramientos	ND	Inclinómetro tibia	ND	ND	CCI (IC 95%), MBA	Sí
Rabin y Kezool (2010) <sup>178</sup>	Israel	PDFAF	FIE	No	1/NA	29M (24,3 ± 3,2 años)/ND	E universitarios	A y B: Sí	ND	Ángulo vertical-tibia	Inclinómetro tibia	Talón-dedo-rodilla	3/√	CCI (IC 95%)	Sí
Dinh et al. (2011) <sup>22</sup>	EEUU	PDFAe	FIA, FIE	No	2/ND	9	ND	A, B Y C: ND	ND	ND	Goniómetro	Pie-pantorrilla	3/√	CCI, EEM†	Sí
Krause et al. (2011) <sup>24</sup>	EEUU	PDFAe	FIA, FIE	Sí	2/7	26M y 13 H (24,2 ± 2,7 años)	ND	A y B: ND	Caminar 5 min	ND	Inclinómetro lateral	Pie-pantorrilla	1	CCI (IC 95%), EEM, MDD	Sí

Calentamiento - ¿Realizan en el estudio calentamiento previo a la prueba?; E - Estudiantes; EEM - Error estándar de la medida; Experiencia: Sí - Titulados universitarios y/o con experiencia en el área. No - Sin título universitario y/o sin experiencia en el área; CCI - Coeficiente de correlación intraclass; FIA - Fiabilidad intra-observadores; FIE - Fiabilidad inter-observadores; H - hombres; IC 95% - Intervalo de confianza al 95%; M - mujeres; MA - ¿Se incluye el estudio en el meta-análisis?; MBA - Método de Bland y Altman; NA - No aplicable; ND - No definido; N ses - Número de sesiones para evaluar la fiabilidad; MBA - Método de Bland y Altman; MDD - Mínima diferencia de delecible; Obj - ¿Es objetivo del estudio el comprobar la fiabilidad de alguna PDFA?; Observadores - Número de observadores, cada uno representado con una letra; t (días) - Tiempo entre sesiones medido en días; † - Proporcionan mismo valor para FIA y FIE.

Respecto a los participantes, se emplearon un total de 191 participantes para comprobar la fiabilidad de alguna PDFA, de los cuales sabemos que 104 (54,45%) eran mujeres. Respecto a las características de los participantes referentes a la edad, altura y peso (o IMC) cabe decir que de los 10 trabajos incluidos en la RS, 5 (50%) describieron estas variables<sup>121;126;130;135;173</sup>, en un total de 128 participantes (67%). La edad media de los participantes fue de 25,28 años, siendo la media de los estudios de 18,8 y la máxima de 40,00 años. En lo referente a la altura la media fue de 169,19 cm, con una media mínima de 164,1 y una máxima de 171,8 cm. Respecto al peso de los participantes la media de los cuatro estudios<sup>121;126;135;173</sup> que reportaron dicha variable en un total de 98 participantes fue de 67,78 kg, siendo la mínima de 58 kg y la máxima de 74,4 kg. En el trabajo de Munteanu et al. (2009)<sup>130</sup> el IMC fue de  $22,8 \pm 3,3 \text{ kg/m}^3$ . En cuanto al tipo de población cuatro de los trabajos<sup>121;171;173;174</sup> emplearon estudiantes universitarios y uno utilizó militares<sup>143</sup>. El resto de estudios no describieron el tipo de población. En lo referente a la selección de participantes y sus criterios de inclusión y exclusión, ninguno describió como lo realizaron.

Atendiendo a los observadores, dos trabajos<sup>143;172</sup> no describieron el número de observadores que emplearon para el estudio. Los restantes ocho trabajos emplearon una media de 2,75 observadores, siendo la mínima de los estudios dos y la máxima de cuatro. En cuanto a la experiencia de los evaluadores, cinco de los estudios describieron la experiencia de los mismos (50% de los trabajos incluidos en la RS)<sup>125;126;134;143;172</sup>, siendo los observadores expertos en tres de los estudios<sup>135;171;173</sup> y combinando expertos con no expertos (mixto) en dos de los trabajos<sup>121;130</sup>.

En cuanto a los métodos de evaluación, en ocho trabajos (80%) únicamente comprobaron la fiabilidad de un método<sup>125;126;134;135;143;171-173</sup>, y en dos casos se emplearon dos métodos diferentes<sup>121;130</sup>. En siete de estos estudios se emplearon el inclinómetro como método de evaluación, siendo su ubicación preferente (60%) en la tibia<sup>121;130;134;171-173</sup> (IT), y un trabajo en la parte lateral de la pierna<sup>126</sup> (IL). Los estudios que emplearon el inclinómetro midieron en su mayoría el ángulo de la vertical con la tibia<sup>121;130;134;172;173</sup>, aunque dos de los trabajos no definieron la variable evaluada<sup>126;171</sup>. Bennell et al. (1998) además de IT midieron en cm la distancia entre el primer dedo del pie y la pared. Por otro lado, Munteanu et al. (2009)<sup>130</sup> midieron el ángulo existente entre la maléolo lateral y la cabeza del peroné transportando dichos puntos anatómicos a un cristal con los ángulos marcados (CR) (figura 1.3.1). Dos de los trabajos<sup>125;143</sup> emplearon el goniómetro (GO) como método de evaluación de la DF, pero sin definir que variable era la que evaluaban. Por último, Jones et al (2005)<sup>135</sup> emplearon un método de evaluación diferente al resto de estudios, diseñando un aparato específico para el estudio

que evaluaba en cm la PDFAf midiendo la distancia que la rodilla superaba la vertical de los dedos del pie (DX) (figura 1.3.2).

Los estudios incluidos en esta revisión emplearon diferentes estrategias para controlar la DF del tobillo alineado los segmentos de diversas maneras. Bennell et al. (1998)<sup>121</sup>, Crossley et al. (2007)<sup>172</sup> y Rabin y Kozol (2010)<sup>173</sup> alinearon el talón, uno de los dedos del pie y la rodilla empleando una línea continua marcada en el suelo y la pared. Munteanu et al. (2009)<sup>130</sup> alinearon el talón y el segundo dedo del pie sobre una línea en el suelo, dejando la rodilla libre. Finalmente, tres de los trabajos<sup>125;126;135</sup> alinearon el pie y la pantorrilla de manera visual para mantener la DF controlada. Los restantes 3 estudios<sup>134;143;171</sup> no definieron como controlaron la posición del pie y la pierna durante la prueba.

Respecto al tipo de análisis estadístico empleado, todos los trabajos emplearon el CCI para comprobar la fiabilidad de la prueba, aunque únicamente proporcionaron su intervalo de confianza al 95% en seis de los estudios<sup>121;126;130;135;171;173</sup>. Además, en cinco trabajos acompañando este índice con el EEM<sup>121;125;126;130;172</sup>. Respecto al método propuesto por Bland y Altman, tres de los mismos (30%) lo emplearon<sup>130;135;171</sup>, y uno de los trabajos empleó la mínima diferencia detectable<sup>126</sup>. Munteanu et al (2009)<sup>130</sup> fue el único trabajo que proporcionó dos índices de fiabilidad dimensionales, el EEM y el MBA.

Tabla 5.3.3. Características cuantitativas de los estudios RS PDFAf.

Variable	n est	% est	n part	% part	Media	DE	Mínima	Máxima
<b>n participantes</b>	10	100,00	191	100,00	19,10	10,00	9,00	39,00
<b>n mujeres</b>	6	60,00	104	54,45	17,33	9,43	5,00	29,00
<b>Edad (años)</b>	5	50,00	128	67,02	25,28*	-	18,80	40,00
<b>Altura (cm)</b>	5	50,00	128	67,02	169,19*	-	164,10	171,80
<b>Peso (kg)</b>	4	40,00	98	51,31	65,78*	-	58,00	74,40
<b>IMC (kg/m<sup>2</sup>)</b>	1	10,00	30	15,71	-	-	-	-
<b>n intentos</b>	7	70,00	147	76,96	3,00	1,53	1,00	6,00
<b>n observadores</b>	8	80,00	167	87,43	2,75	0,70	2,00	4,00
<b>T ses (días)</b>	8†	80,00†	153†	80,10†	9,62	5,21	7,00	21,00

DE – Desviación estándar; IMC – Índice de masa corporal; n est – Número de estudios en los que la característica está descrita; n intentos – Número de intentos a realizar en la PDFAf; n mujeres – Número de mujeres que participan en los estudios; n observadores – Número de observadores que se utilizan en los estudios; n part – Número de participantes en los que la característica está descrita; T ses– Tiempo entre sesiones de evaluación; % est - Porcentaje de estudios en los que la característica está descrita; % part - Porcentaje de participantes en los que la característica está descrita; † - Rabin y Kozol (2011) solo estudian PDFAf y FIE y realizan una sesión, por lo que esta variables no es aplicable, se contabiliza como ND para no interferir en la media, DE, mínima y máxima.

### 5.3.4-ANÁLISIS DE LAS CARACTERÍSTICAS DE LOS SUBESTUDIOS

En segundo lugar, pasamos a extraer los datos referentes a cada tipo de fiabilidad, tipo de prueba y método de evaluación empleado (anexo 5.3.4), dividiendo los trabajos en su subestudios (PDFAf FIA, PDFAf FIE, PDFAe FIA y PDFAe FIE). Una vez extraídos los datos analizamos las características de cada subestudio (anexo 5.3.5). Atendiendo a estos criterios la fiabilidad alguna PDFa se comprobó en 23 ocasiones con un total de 445 participantes, siendo 229 mujeres. La media de participación por subestudio fue de 19,35 personas ( $\pm 9,38$ ). En cuanto a la media de participación femenina, en los 17 trabajos que definen el número de mujeres participantes la media fue de  $15,27 \pm 8,49$  por trabajo.

Respecto a la fiabilidad intra-observadores (FIA) de la PDFAf está se comprobó en seis<sup>121;134;135;171;172</sup> ocasiones diferentes con un total de 83 participantes, siendo el 40,46% de ellos mujeres. En estos trabajos participaron una media de  $13,83 \pm 2,40$ , personas y  $8,50 \pm 4,73$  mujeres (en los cuatro subestudios en los que reportan dicha variable<sup>121;134;135</sup>). La FIE de la PDFAf se comprobó en otras seis<sup>121;134;135;171;173</sup> ocasiones con un total de 63 mujeres de 102 participantes. La media de participación media de mujeres en los cinco estudios en los que participaron<sup>121;134;135;173</sup> fue de  $12,60 \pm 10,04$ .

En cuanto a la PDFAe, la FIA se probó en seis trabajos<sup>125;126;130;143;171</sup> diferentes con un total de 137 participantes (48,18% de mujeres), con una media de  $22,83 \pm 11,79$  participantes y  $22,00 \pm 3,46$  mujeres en los tres subestudios donde definieron la n de mujeres<sup>126;130</sup>. La FIE de la PDFAe se estudió en cinco<sup>125;126;130;171</sup> trabajos empleando 123 sujetos con una media de  $24,60 \pm 12,25$ , siendo 66 mujeres (media de  $22,00 \pm 3,46$  por cada uno de los seis subestudios<sup>126;130</sup> que incluyeron mujeres).

En los subestudios donde se incluyeron datos sobre la FIA PDFAf se definió el objetivo de fiabilidad en cuatro<sup>121;135;171</sup> de los seis incluidos en la RS, al igual que en FIE PDFAf<sup>121;135;171</sup> y FIA PDFAe<sup>126;130;171</sup>. Respecto a la FIE PDFAe, un único subestudio<sup>125</sup> no definió el objetivo de fiabilidad como tal, de los cinco incluidos en esta revisión. En total, 69,57%<sup>121;126;130;135;171</sup> de los 23 subestudios definieron el objetivo de comprobar la fiabilidad de la PDFa como tal.

Respecto a la descripción de las características de los participantes en los estudios de fiabilidad en las diferentes PDFa, de los 23 subestudios 13 describieron la edad y altura (56,52%)<sup>121;126;130;135;173</sup>, nueve trabajos (39,13%)<sup>121;126;135;173</sup> proporcionaron el peso y cuatro<sup>130</sup> (17,39%) el IMC. En cuanto a la edad media de los participantes, la más alta fue de 27,18 años en la PDFAf FIA y la más baja en PDFAe en FIE con una edad media de 22,93 años. Respecto a la estatura, los participantes más altos se emplearon en la PDFAe FIE con 179,29 cm de altura y

los más bajos de 167,68 cm en PDFAf FIE. La mayoría de los participantes (58,65%) fueron estudiantes universitarios<sup>121;130;171;173</sup>, no conociendo el origen de los mismos en 38,20%<sup>125;126;134;135;172</sup> de los 445 del total.

En lo referente al número de intentos realizado en las pruebas, la media de los 16 subestudios<sup>121;125;126;130;135;172;173</sup> que describieron dicha variable fue de 2,88 intentos con una DE 1,40 intentos. La mínima de los intentos en los trabajos fue de un intento y la máxima de 6 intentos. El estimador empleado para el análisis en el 60,87% de los subestudios incluidos en la RS fue la<sup>121;125;130;135;172;173</sup> media, no describiéndose esta variable en 7 subestudios<sup>134;143;171</sup> o empleando el único intento realizado en 3<sup>126</sup>.

En general los subestudios emplearon 2,88 observadores, con una mínima de un observador y una máxima de 4. Estos observadores tenían un título universitario y/ experiencia en el área de conocimiento en el 30,43%<sup>135;171;173</sup> de los 23 subestudios. En 34,78% de los trabajos<sup>121;130</sup> combinaron observadores con y sin experiencia, y un 34,78% de los subestudios<sup>125;126;134;143;172</sup> no describió la experiencia de las personas que realizaron las evaluaciones.

Los estudios sobre FIA separaron las sesiones de evaluación 10,50 días en el PDFAf y 9,80 días en PDFAe. La mínima separación fue de siete días y la máxima en estos estudios fue de 21. Todos los trabajos que definieron que evaluaciones compararon para la FIA, un 83,33%<sup>121;126;130;135;143;171;172</sup>, compararon las evaluaciones de cada observador entre las diferentes sesiones de evaluación. Respecto al FIE, la separación entre sesiones de evaluación fue de 11,20 días en los 5 subestudios<sup>121;134;135;171</sup> de PDFAf que definieron el lapso de tiempo entre sesiones. En PDFAe la separación media entre sesiones fue de 10,4 días, definido por 4 de los subestudios<sup>126;130;171</sup>. De los 11 subestudios que comprobaron la FIE<sup>121;125;126;130;134;135;171;173</sup> un 72,73%<sup>121;130;135;171;173</sup> describieron que evaluaciones compararon, contrastando todos ellos las evaluaciones realizadas dentro de la misma sesión.

Para la evaluación el método más empleado fue el de colocar el inclinómetro la tibia (52,17% de los subestudios)<sup>121;130;134;171-173</sup>. Las distancias lineales, la DD y la DX, se utilizaron en 17,40% de las ocasiones<sup>121;135</sup>, y sólo en la PDFAf. Para la PDFAf, el inclinómetro en la tibia fue el método más empleado, puesto que 8<sup>121;134;171-173</sup> de los 12 subestudios lo emplearon. El resto de trabajos utilizaron o bien la distancia dedo (DD)<sup>121</sup> pared o la distancia que sobrepasaba la rodilla la vertical de los dedos del pie (DX)<sup>135</sup>. En cuanto la PDFAe, la IT se utilizó en 4 ocasiones<sup>130;171</sup>, el inclinómetro en el lateral en 2<sup>126</sup> y el goniómetro en 3 trabajos<sup>125;143</sup>. El goniómetro no se utilizó en ningún subestudio para cuantificar la DF durante la PDFAf.

La variable evaluada fue el ángulo existente entre la vertical y la tibia en 43,48% de los trabajos<sup>121;130;134;172;173</sup>. En el 39,13% de los subestudios<sup>125;126;171</sup> se desconoce que ángulo midieron bien con el inclinómetro o con el goniómetro. El resto subestudios evaluaron alguna distancia lineal u otro ángulo diferente. Respecto a la herramienta de evaluación, la favorita para evaluar la DF fue el inclinómetro (60,78% de los estudios)<sup>121;126;130;134;171-173</sup>, seguido del goniómetro con 13,04% de estudios que lo emplearon. También se utilizaron otras herramientas; cristal con los ángulos marcados (CR)<sup>130</sup>, cinta métrica (CM)<sup>121</sup> o aparato que mide la distancia entre la vertical del dedo del pie y la rodilla (DX)<sup>135</sup>.

Respecto a los análisis estadísticos, todos los estudios emplearon el CCI para determinar la fiabilidad de la prueba. De todos los subestudios, un 26,09%<sup>125;134;143;172</sup> no definió el modelo de Fisher de CCI que eligieron para el análisis, los mismos que no proporcionaron el IC 95% de CCI. Para comprobar a fiabilidad intra-observadores se utilizó el modelo 2,1 en tres<sup>135;171</sup> de los 12 trabajos que estudiaron esta fiabilidad, en dos subestudios<sup>121</sup> se utilizó el 3,3, en otros dos<sup>130</sup> el 2,2 y en uno<sup>126</sup> el 3,1. Respecto a la FIE, cuatro subestudios<sup>126;135;171</sup> emplearon el modelo 2,1, dos<sup>121</sup> el 2,3, otros dos<sup>130</sup> el 2,4 y uno<sup>173</sup> el 3,2.

Los estudios incluidos en la RS también emplearon otros análisis estadísticos como EEM, el MBA y la MDD. Para la FIA el EEM se utilizó en el 58,33% (n=7)<sup>121;125;126;130;172</sup> de los subestudios y en FIE en cinco trabajos (42,7%)<sup>121;125;126;130</sup>. Respecto al método propuesto por Bland y Altman (MBA) en total 10 subestudios<sup>130;135;171</sup> (43,48% del total) lo aplicaron, cinco<sup>130;135;171</sup> para la FIA y otros cinco<sup>130;135;171</sup> para la FIE. Respecto a la MDD está únicamente se estimó en un trabajo<sup>126</sup> que comprobó la FIA PDFAe.

Todos los trabajos concluyeron que los métodos de evaluación empleado eran fiables tanto para PDFAf como para PDFAe.

### 5.3.5-RESULTADOS DEL CCI DE LOS ESTUDIOS PRIMARIOS Y META-ANÁLISIS

Extrajimos todos los resultados de los estudios que presentaban los datos de manera analizable, los cuales se pueden observar en el anexo 5.3.6. Estos resultados los representamos en función de los subestudios previamente descritos (PDFAf IFA, PDFAf FIE, PDFAe FIA y PDFAe FIE) mediante *forest plot*, diferenciando el método de evaluación, el observador y la pierna evaluada, así como el día para los estudios de FIE. Los valores que ilustramos en los gráficos son los obtenidos tras la transformación Z de Fisher y su posterior reconversión en CCI. Las diferencias entre el CCI original y sus respectivos IC al 95% y la transformación realizada son prácticamente inapreciables tal y como se puede observar en el anexo 5.3.6. Todos los estudios a excepción del realizado por Nitz &

Low Choy (2004)<sup>134</sup> fueron incluidos en los meta-análisis. Este trabajo no pudo incluirse en dicho análisis puesto que proporcionó un mismo CCI para la fiabilidad intra- e inter-observadores.

Realizamos varios MA en cada subestudio, uno para cada tipo de unidad de evaluación, cm o grados, y uno que englobó ambas unidades intentando detectar las fuentes de heterogeneidad<sup>160</sup> mediante la  $I^2$  para poder conseguir un CCI combinando con una consistencia adecuada.

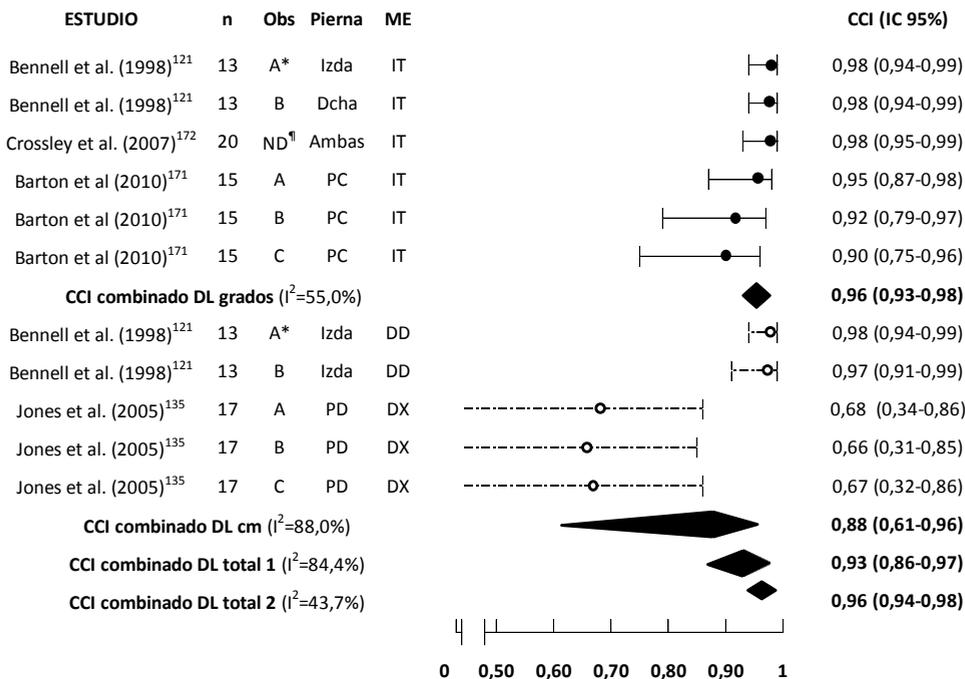


Figura 5.3.2. Gráfico del meta-análisis PDFAf FIA.

DD - Distancia entre el primer dedo del pie y la pared; DL - DerSimonian-Laird; DX - Distancia que la vertical de la rodilla sobrepasa los dedos del pie; IC 95% - Intervalo de confianza al 95%; IT - Inclinómetro tibia;  $I^2$  -  $I^2$  cuadrado; ME - Método de evaluación; Obs - Observador; PC - Pierna del par de los casos en grupo control; PD - Pierna dominante; Total 1 - Todos los estudios, Total 2 - Todos los estudios a excepción de Jones et al. (2005); ¶ - Experiencia observador no definida; \* - Observador sin experiencia; |—| - Intervalo de confianza en medidas en grados; |—| - Intervalo de confianza medidas en cm; ● - Estimación puntual medidas en grados; ○ - Estimación puntual medidas en cm; ◆ - CCI combinado con IC 95%.

Respecto a los resultados de los PDFAf FIA (figura 5.3.2) la estimación puntual de los estudios primarios CCI fue de 0,66 a 0,98. Cuando combinamos los resultados de los trabajos encontramos un CCI de 0,93 (IC 95%: 0,86-0,97), siendo la heterogeneidad alta ( $I^2=84,4\%$ ). Para los métodos que emplean los grados como unidad de evaluación, el valor de CCI mínimo fue de 0,90 y el máximo de 0,98 siendo. El CCI combinado fue de 0,96 (IC 95%: 0,93-0,98) con una  $I^2$  de 55%. Los trabajos que estudiaron la fiabilidad de la PDFAf evaluando la DF en cm obtuvieron

un valor de 0,66 a 0,98, con un CCI combinado de 0,88 (IC 95%: 0,61-0,96). La heterogeneidad fue alta ( $I^2=88,0\%$ ). Realizamos otro MA con todos los métodos de evaluación empleados (total 2) previa extracción de los resultados de Jones et al. (2005)<sup>135</sup> que empleó un método de evaluación exclusivo en su trabajo (DX) y además tenía un CCI sustancialmente diferente que el resto de trabajos. La consistencia del MA en este caso fue adecuada ( $I^2<50\%$ ) obteniendo un CCI combinando de 0,96 (IC 95%: 0,94-0,98).

En lo referente a la FIE de la PDFAf (figura 5.3.3) los CCI van de 0,83 a 0,97 en métodos de evaluación en grados, y en cm de 0,82 a 0,99. Cuando se combina el CCI, la heterogeneidad es elevada ( $I^2=75,7\%$ ) y cuando se extrae el estudio con método de evaluación diferente al resto (DX), la  $I^2$  mejora poco, 75,6%. El CCI combinado más consistente ( $I^2=56,2\%$ ) es el realizado con los métodos que emplean grados como herramienta de evaluación con un valor 0,93 (IC 95%: 0,86-0,96). La estimación puntual en todos los casos es similar, 0,93 ó 0,94, siendo el IC al 95% más amplio en el CCI combinado de los métodos de evaluación que utilizan cm.

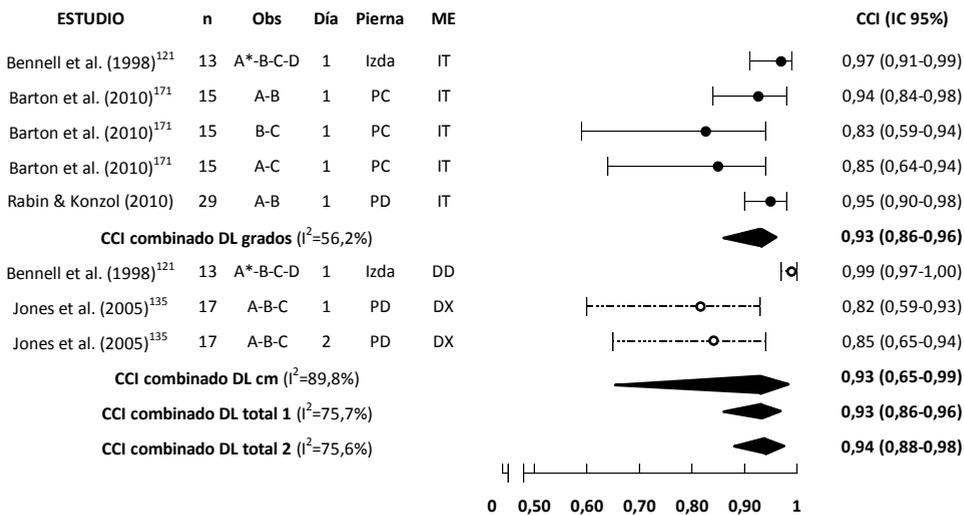


Figura 5.3.3. Gráfico del meta-análisis PDFAf FIE.

DD - Distancia entre el primer dedo del pie y la pared; DL - DerSimonian-Laird; DX - Distancia que la vertical de la rodilla sobrepasa los dedos del pie; IC 95% - Intervalo de confianza al 95%; IT - Inclínometro tibia;  $I^2$  -  $I^2$  cuadrado; ME - Método de evaluación; Obs - Observador; PC - Pierna del par de los casos en grupo control; PD - Pierna dominante; Total 1 - Todos los estudios; Total 2 - Todos los estudios a excepción de Jones et al. (2005); \* - Observador sin experiencia; |—| - Intervalo de confianza en medidas en grados; |—| - Intervalo de confianza medidas en cm; ● - Estimación puntual medidas en grados; ○ - Estimación puntual medidas en cm; ◆ - CCI combinado con IC 95%.

En cuanto a la PDFAe (figura 5.3.4), en la fiabilidad intra-observadores la estimación puntual del CCI varió de 0,67 a 0,99. Al combinar todos los CCI

obtuvimos un valor de 0,88 (IC 95%: 0,84-0,92) con una  $I^2$  de 65,9%. Exploramos las posibles causas de la inconsistencia del valor combinado y decidimos descartar el trabajo de Dinh et al. (2011)<sup>125</sup> por ser el único trabajo que presentó un CCI combinado para sus tres observadores de manera conjunta, además de presentar un CCI substancialmente diferentes al resto de los trabajos. En este caso el CCI combinado fue de 0,87 (IC 95%: 0,83-0,90) mostrando una heterogeneidad adecuada ( $I^2 = 48,0\%$ ).

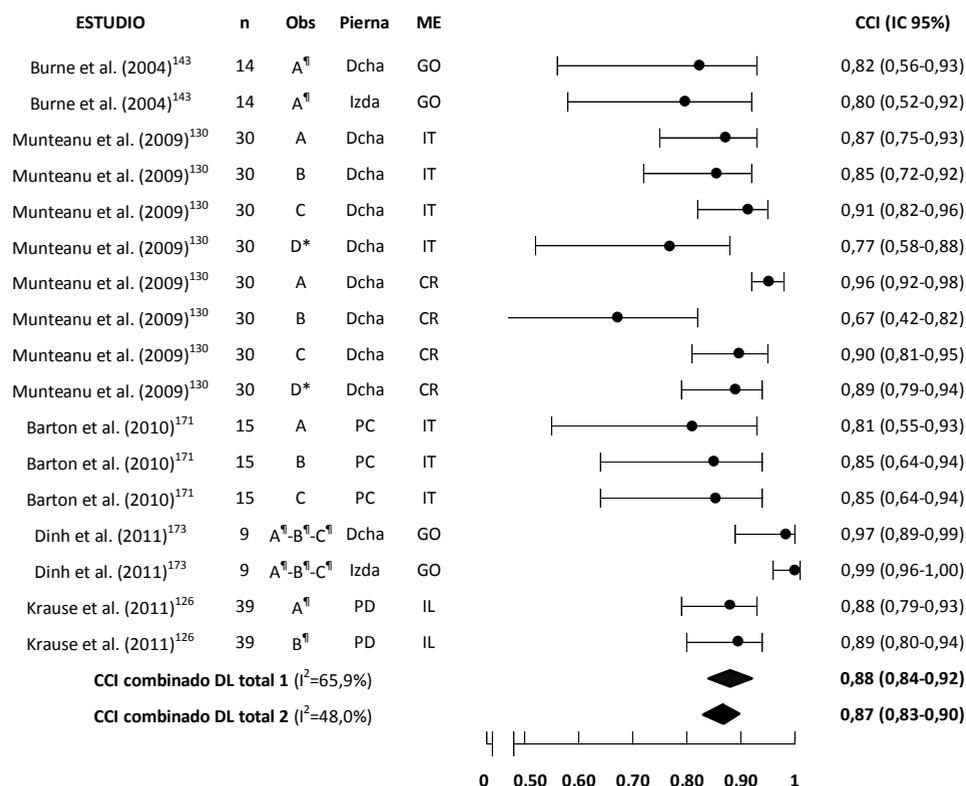


Figura 5.3.4. Gráfico del meta-análisis PDFAe FIA.

DL - DerSimonian-Laird; GO - Goniómetro; IC 95% - Intervalo de confianza al 95%; IL - Inclínometro lateral; IT - Inclínometro tibia;  $I^2$  - I cuadrado; ME - Método de evaluación; Obs - Observador; PC - Pierna del par de los casos en grupo control; PD - Pierna dominante; Total 1 - Todos los estudios; Total 2 - Todos los estudios a excepción de Dinh et al. (2011); ¶ - Experiencia observador no definida; \* - Observador sin experiencia; |—| - Intervalo de confianza en medidas en grados; ● - Estimación puntual medidas en grados; ◆ - CCI combinado con IC 95%.

Respecto a los CCI de la PDFAe FIE (figura 5.3.5), los índices van de 0,66 a 0,97 dependiendo del observador, el método empleado y la pierna evaluada. En cuando el CCI combinado obtenemos un valor de 0,90 con un IC al 95% de 0,84 a 0,94.

La heterogeneidad es de  $I^2=72,8\%$ . En este caso no pudimos detectar la fuente de variabilidad para lograr una mayor consistencia del MA combinado.

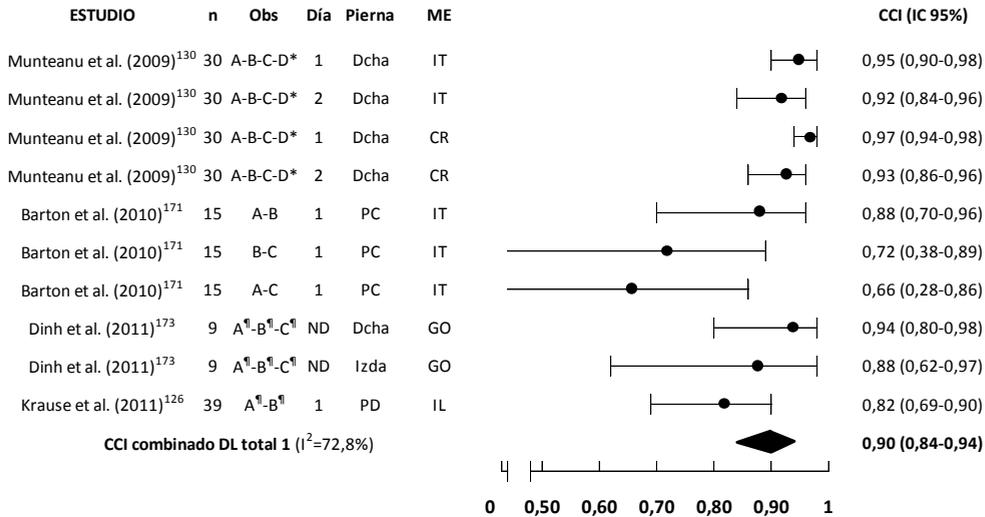


Figura 5.3.5. Gráfico del meta-análisis PDFAe FIE.

CR – Cristal con los ángulos marcados; DL – DerSimonian-Laird; GO - Goniómetro; IL - Inclínometro lateral; IT - Inclínometro tibia;  $I^2$  - I cuadrado; ME - Método de evaluación; ND – No definido; Obs - Observador; PC - Pierna del par de los casos en grupo control; PD - Pierna dominante; Total 1 – Todos los estudios; ¶ - Experiencia observador no definida; \* - Observador sin experiencia; |—| - Intervalo de confianza en medidas en grados; ● - Estimación puntual medidas en grados; ◆ - CCI combinado con IC 95%.

Cuando comparamos la FIA y FIE de la prueba de PDFAf (figuras 5.3.2 y 5.3.3) observamos que el CCI va de 0,82 a 0,99 en FIE, y de 0,66 a 0,98 en FIA. Los CCI combinados muestra pequeñas diferencias entre ambos tipos de fiabilidad. En el primer MA realizado (total 1), la estimación puntual en ambas fiabilidades es idéntico, de 0,93, aunque el IC al 95% es ligeramente mayor en FIE. Una vez eliminando el trabajo de Jones et al. (2005)<sup>135</sup> (total 2) el valor combinado es mejor en FIA que FIE; 0,96 (IC 95%: 0,94-0,98) frente a 0,94 (IC 95%: 0,88-0,98). Del mismo modo el MA con los métodos de evaluación en grados, también muestran un CCI combinado ligeramente mayor FIA que FIE. En cuanto al CCI combinado en cm, éste muestra una mejor FIE que FIA, con una estimación puntual en el valor combinado de 0,88 (IC 95%: 0,61-0,96) en FIA y de 0,93 (IC 95%: 0,65-0,99) en FIA.

Si analizamos los resultados de los estudios primarios (figuras 5.3.2 y 5.3.3) observamos que todos aquellos trabajos que han estudiado la FIA y la FIE dentro de un mismo estudio<sup>121;171</sup> muestran CCI ligeramente mayores en FIA que FIE, cuando se evalúa en grados. En la evaluación en cm, la FIE es mayor que la FIA tal y como muestran los trabajos de Jones et al. (2005)<sup>135</sup> y de Bennell et al (1998)<sup>121</sup> (cuando emplean como método de evaluación DD). Jones et al. (2005)<sup>135</sup> obtienen

unos CCI de 0,66 a 0,68 en FIA, y de 0,82 a 0,85 en FIE. Bennell et al (1998)<sup>121</sup> proporcionan unos CCI de 0,97 a 0,98 en FIA, y de 0,99 en FIE.

En cuanto a la comparación del FIA y el FIE en PDFAe (figuras 5.3.4 y 5.3.5), los MA realizados muestra un CCI combinado mayor en FIA que en FIE, con una estimación puntual de 0,87 frente a 0,90. Si estudiamos de manera individual los trabajos que han comprobado ambas fiabilidades, encontramos que Munteanu et al. (2009)<sup>130</sup> obtiene mayores CCI en FIE que en FIA. En este trabajo los CCI en FIA van de 0,77 a 0,96, y en FIE de 0,92 a 0,97. En el caso de Barton et al. (2010)<sup>171</sup>, cuando compara los observadores A y B, en FIA obtienen un CCI de 0,81 y 0,85 respectivamente, siendo la FIE entre ambos observadores de 0,88. El resto de trabajos que miden tanto FIA como FIE<sup>125;126</sup> presentan unos índices de fiabilidad mayores en FIA que FIE.

Respecto a la comparación de la FIA entre PDFAf y PDFAe (figura 5.3.2. y 5.3.4) encontramos que el CCI va de 0,67 a 0,98 en PDFAf y de 0,67 a 0,99 en PDFAe. Al realizar el MA, éste muestra que la PDFAf es más fiable que la PDFAe, ya que la PDFAf muestra un mayor CCI combinado que la PDFAe, de 0,96 (IC 95%: 0,94 a 0,98) y 0,87 (IC 95%: 0,83-0,90). Cuando combinamos los CCI de los métodos de evaluación que han empleado grados en PDFAf (la única unidad de evaluación que se ha empleado en PDFAe) el resultado índice también es mayor, de 0,96 (IC 95%: 0,93-0,98) con una I<sup>2</sup> de 55,0%. Barton et al. (2010)<sup>171</sup> que estudia la FIA tanto en PDFAf como en PDFAe, también encuentra mayores CCI en PDFAf.

Respecto a la FIE, ésta también es mejor en PDFAf que en PDFAe (figura 5.3.3. y 5.3.5). Los CCI de los estudios primarios presentan índices de 0,82 a 0,99 en PDFAf y de 0,66 a 0,97 en PDFAf. Cuando se realiza el MA, la PDFAe obtiene un valor combinado de 0,90 (IC 95%: 0,84-0,94), y la PDFAf de 0,93 (IC 95%: 0,86-0,96). En PDFAf, el MA combinado de los métodos de evaluación en grados 0,93 (IC 95%: 0,86-0,96), mejor que el MA de PDFAe. En cuanto a Barton et al. (2010)<sup>171</sup>, que es el único estudio que ha estudiado la FIE de ambas pruebas, encuentra un CCI de 0,83 a 0,94 en PDFAf y de 0,88 a 0,97 en PDFAe.

### 5.3.6-RESULTADOS DIMENSIONALES DE LOS ESTUDIOS PRIMARIOS

En cuanto a las medidas dimensionales de fiabilidad, como son MDD y los LDA y ES del MBA, separamos los *forest plot* para cada subestudios y observador en las unidades proporcionadas en grados por un lado, y las dadas en cm por otro (figuras 5.3.6-5.3.11). El EEM proporcionado por Dinh et al. (2011)<sup>125</sup> no pudo ser incluido en este análisis al proporcionar un único valor para ambos tipos de fiabilidad.

Respecto a los resultados interpretables desde el punto de vista práctico, la MDD en el caso de la PDFaf FIA en grados es de 2,77° a 3,05°, y en LDA de 4,45° a 6,35° (figura 5.3.6) con un erro sistemático (ES) de -0,20° a 0,95°.

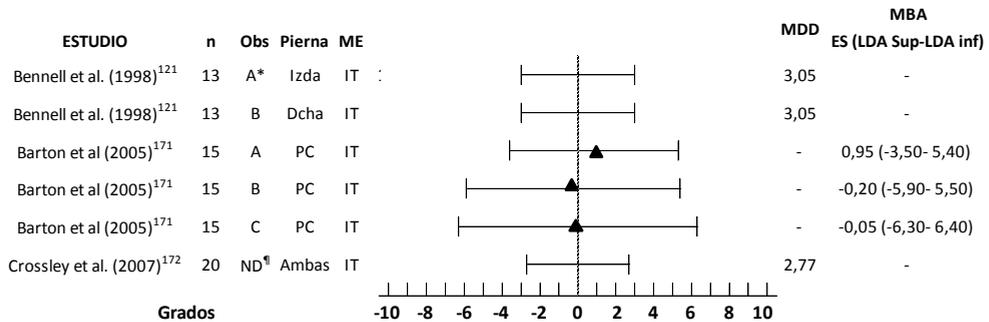


Figura 5.3.6. Gráfico resultados dimensionales PDFaf FIA métodos evaluación en grados.

ES- Error sistemático; IT - Inclinómetro tibia; LDA inf – Límite de acuerdo inferior; LDA sup – Límite de acuerdo superior; MBA – Método de Bland y Altman; MDD – Mínima diferencia detectable; ME - Método de evaluación; Obs - Observador; PC - Pierna del par de los casos en grupo control; ¶ - Experiencia observador no definida; \* - Observador sin experiencia; |—| - MDD o límite de acuerdo superior e inferior según proceda en medidas tomadas en grados; ▲ - Error sistemático (solo MBA).

En la PDFaf FIA cuando la evaluación de la DF se realiza en cm el error de método representado por MDD es de 1,39 cm a 1,66 cm y el LDA de 3,30 cm a 3,66 cm siendo el ES de -0,21 cm a 0,90 cm (figura 5.3.7).

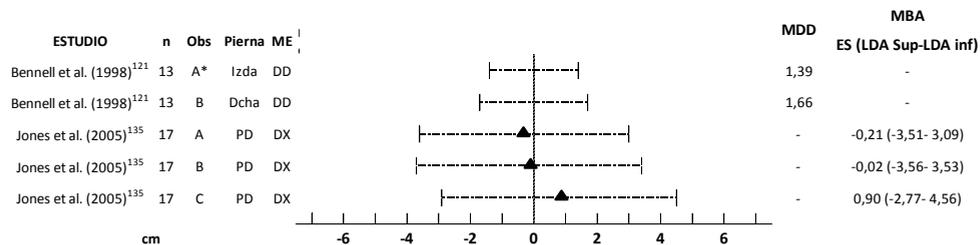


Figura 5.3.7. Gráfico resultados dimensionales PDFaf FIA métodos evaluación en cm.

DD - Distancia entre el primer dedo del pie y la pared; DX - Distancia que la vertical de la rodilla sobrepasa los dedos del pie; ES - Error sistemático; LDA inf – Límite de acuerdo inferior; LDA sup – Límite de acuerdo superior; MBA – Método de Bland y Altman; MDD – Mínima diferencia detectable; ME - Método de evaluación; Obs - Observador; PD - Pierna dominante; \* - Observador sin experiencia; |—| - MDD o límite de acuerdo superior e inferior según proceda en medidas tomadas en cm; ▲ - Error sistemático (solo MBA).

Respecto al FIE de la PDFaf en grados (figura 5.3.8), cuando observamos el MDD éste es de 3,88° en el único estudio que proporciona el EEM. Respecto al MBA, el ES va de -0,10° a -2,45° con unos LDA de 4,90° a 7,05° dependiendo de los observadores comparados. En cm (figura 5.3.9), el error de método es de 1,11 cm

en MDD y de 1,57 cm a 2,65 cm en LDA. El ES de los métodos en cm es -0,48 cm a 1,10 cm.

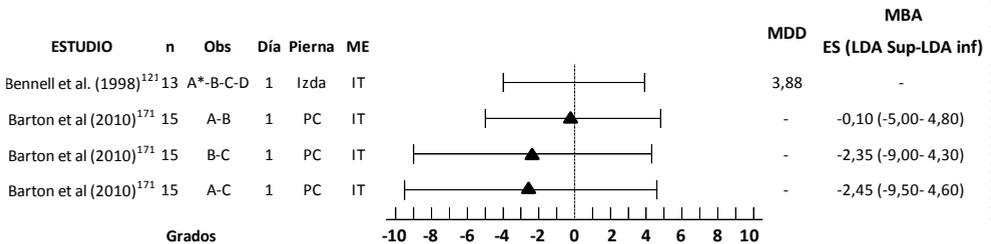


Figura 5.3.8. Gráfico resultados dimensionales PDFAf FIE métodos evaluación en grados.

ES- Error sistemático; IT - Inclinómetro tibia; LDA inf – Límite de acuerdo inferior; LDA sup – Límite de acuerdo superior; MBA – Método de Bland y Altman; MDD – Mínima diferencia detectable; ME - Método de evaluación; Obs - Observador; PC - Pierna del par de los casos en grupo control; ¶ - Experiencia observador no definida; \* - Observador sin experiencia; |—| - MDD o límite de acuerdo superior e inferior según proceda en medidas tomadas en cm; ▲ - Error sistemático (solo MBA).

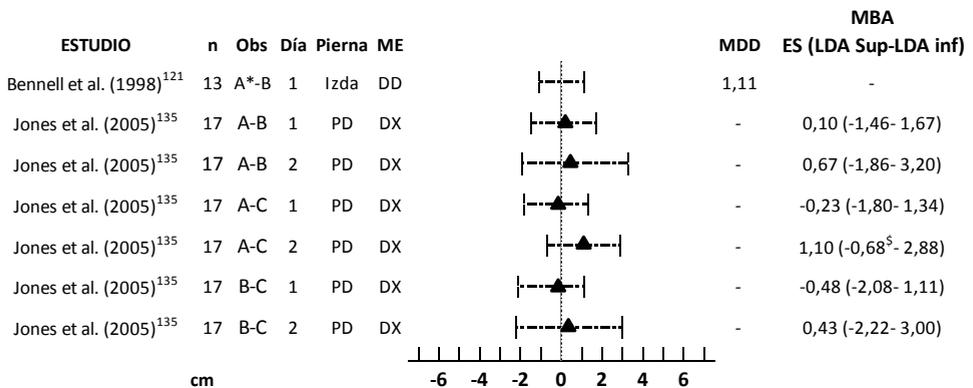


Figura 5.3.9. Gráfico resultados dimensionales PDFAf FIE métodos evaluación en cm.

DD - Distancia entre el primer dedo del pie y la pared; DX - Distancia que la vertical de la rodilla sobrepasa los dedos del pie; ES- Error sistemático; LDA inf – Límite de acuerdo inferior; LDA sup – Límite de acuerdo superior; MBA – Método de Bland y Altman; MDD – Mínima diferencia detectable; ME - Método de evaluación; Obs - Observador; PD - Pierna dominante; \* - Observador sin experiencia; \$- Valor corregido por este estudio; |—| - MDD o límite de acuerdo superior e inferior según proceda en medidas tomadas en cm; ▲ - Error sistemático (solo MBA).

En cuanto a la PDFAe FIA (figura 5.3.10), tres<sup>126;130;171</sup> trabajos proporcionaron resultados dimensionales de fiabilidad. Munteanu et al. (2009)<sup>130</sup> aportan tanto MBA como MDD y Barton et al. (2010) el EEM, mediante el cual calculamos MDD. Krause et al. (2011)<sup>126</sup> proporcionó directamente el MDD además del EEM. El MDD de la PDFAe FIA fue de 3,33° a 6,93° y el LDA de 3,80° a 8,35°. En cuanto al error sistemático, en ambos trabajos<sup>130;171</sup> se observa una tendencia a que los observadores midan más el primer día de observación que en el segundo, ya que el ES en se expresa en negativo en el 90% de los resultados proporcionados, con valores de -0,40° a -2,05°.

En lo referente a la FIE de la PDFAe los resultados se recogen la figura 5.3.11. En cuanto a la MDD, Munteanu et al. (2009)<sup>130</sup> muestra índices de 3,60° a 5,98°, y Krause et al. (2011)<sup>126</sup> de 7,82°. Respecto a los LDA, éstos van 4,70° a 6,30° en Munteanu et al. (2009)<sup>130</sup>, siendo los de Barton et al. (2010) los más elevados, de 5,40° a 13,00°. Respecto al ES, el trabajo de Munteanu et al. (2009)<sup>130</sup> no muestra error sistemático (0,00°). Sin embargo el ES Barton et al. (2010)<sup>171</sup> aunque es despreciable (0,30°) entre los observadores A y B, entre los B y C, y A y C, es de -5,00° y -5,30° respectivamente.

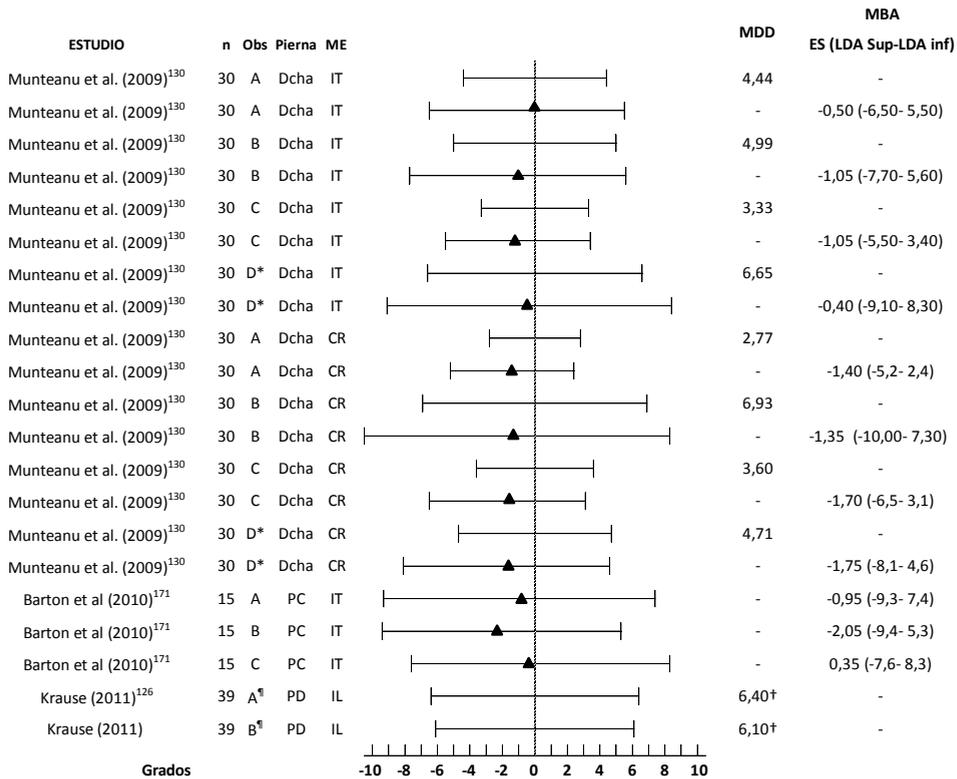


Figura 5.3.10. Gráfico resultados dimensionales PDFAe FIA.

CR – Cristal con los ángulos marcados; ES - Error sistemático; IL - Inclinómetro lateral; IT - Inclinómetro tibia; LDA inf – Límite de acuerdo inferior; LDA sup – Límite de acuerdo superior; MBA – Método de Bland y Altman; MDD – Mínima diferencia detectable; ME - Método de evaluación; Obs - Observador; PC - Pierna del par de los casos en grupo control; PD - Pierna dominante; ¶ - Experiencia observador no definida; \* - Observador sin experiencia; † - MDD proporcionado por los autores de los estudios primarios; I-I - MDD o límite de acuerdo superior e inferior según proceda en medidas tomadas en cm; ▲ - Error sistemático (solo MBA).

Cuando comparamos la FIA y la FIE de la PDFAf, al igual que cuando con el CCI, en grados la FIA es ligeramente mejor que la FIE (figuras 5.3.6 y 5.3.8). Bennell et al. (1998)<sup>121</sup> obtuvieron una MDD 3,05° en FIA, y de 3,88° en FIE. Respecto a

Barton et al. (2005)<sup>171</sup> los LDA van de 4,45° a 6,35° en FIA y de 4,90° a 7,05° en FIE. En los métodos de evaluación en cm la FIE el resulta mejor (figuras 5.3.7 y 5.3.9). Bennell et al. (1998)<sup>121</sup> muestran una MDD de 1,39 cm y 1,66 cm en FIA, y de 1,11 cm en FIE. Por otro lado, Jones et al. (2005)<sup>135</sup> obtienen unos LDA de 3,30 cm a 3,66 cm en FIA, y de 1,57 cm a 2,65 cm en FIE.

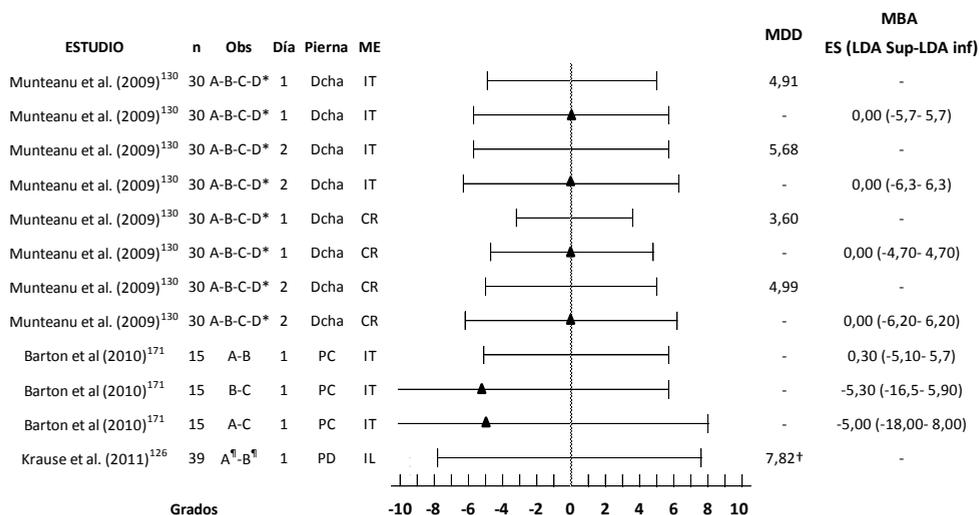


Figura 5.3.11. Gráfico resultados dimensionales PDFAe FIE.

CR – Cristal con los ángulos marcados; ES - Error sistemático; IL - Inclinómetro lateral; IT - Inclinómetro tibia; LDA inf – Límite de acuerdo inferior; LDA sup – Límite de acuerdo superior; MBA – Método de Bland y Altman; MDD – Mínima diferencia detectable; ME - Método de evaluación; Obs - Observador; PC - Pierna del par de los casos en grupo control; PD - Pierna dominante; † - Experiencia observador no definida; \* - Observador sin experiencia; † - MDD proporcionado por los autores de los estudios primarios; |—| - MDD o límite de acuerdo superior e inferior según proceda en medidas tomadas en cm; ▲ - Error sistemático (solo MBA).

En cuanto a la comparación de FIA y la FIE en PDFAe (figura 5.3.10 y 5.3.11), en general tanto la MDD como el LDA son ligeramente mayores en FIA que FIE, aunque las diferencias son muy pequeñas y en ocasiones la FIE resulta más precisa. En Munteanu et al. (2009)<sup>130</sup> los valores en FIE son de 3,60° a 5,68° (MDD), y de 4,70° a 6,30° (LDA), y en FIA 2,77° a 6,65° (MDD) y de 3,80° a 8,65° (LDA). En cuanto a Barton et al. (2005)<sup>171</sup>, los LDA son más pequeños en FIA, de 7,35° a 8,35°, cuando estos límites en FIE van de 5,40° a 13,00°. Sin embargo cabe destacar que en FIA el observador A obtiene un LDA de 8,35° y el B de 7,35°, siendo la LDA en FIE entre ambos de 5,40°. En cuanto al ES, en FIA casi siempre es en negativo (90%), pero en FIE se diferente dependiendo del trabajo.

En lo referente a la comparación de entre la PDFAf y PDFAe en FIA, cabe decir que en general en la PDFAf encontramos MDD y LDA menores que en PDFAe (figura 5.3.6 y 5.3.10). Las MDD en PDFAf van de 3,05° a 2,77°, y en PDFAe de 2,77° a 6,93°. En cuanto a los LDA estos van 4,45° a 6,35° en PDFAf, y de 3,80° a 8,65°

en la PDFAe. El único estudio que comprueba la FIA de ambas pruebas, el de Barton et al. (2010), también muestra que los LDA son menores en la prueba con la rodilla flexionada, siendo de 7,35° a 8,35° en PDFAe, y de 4,45° a 6,35° en PDFAf.

En la FIE, la PDFAf también muestra una mejor precisión que la PDFAe (figuras 5.3.8 y 5.3.11). Respecto a la MDD encontramos que Bennell et al. (1998)<sup>121</sup> obtiene unos valores en PDFAf de 3,88°, y Munteanu et al. (2009)<sup>130</sup> en PDFAe de 3,60 a 5,68°. En cuanto a los LDA, estos van de 4,70° a 13,00° en PDFAe, y de 4,90° a 7,05° en PDFAe. Barton et al. (2010)<sup>171</sup> que estudia ambas pruebas, encuentra también un menor LDA en PDFAf que en PDFAe, con valores de 4,90° a 7,05°, y de 5,40° a 13,00° respectivamente.

Respecto a la similitud de los índices fiabilidad dimensionales, Muntenanu et al. (2009)<sup>130</sup> que es el único que proporciona resultados del MBA y EEM (figuras 5.3.10 y 5.3.10), cabe decir que el LDA es 2,05° a 1,03° superior al MDD en FIA y 0,62° a 1,21° en FIE, siendo la diferencia media de 1,31°.

**REVISIÓN SISTEMÁTICA SOBRE LAS PRUEBAS DE DORSIFLEXIÓN EN  
APOYO:  
SINTESIS DE LOS RESULTADOS**

- **10 documentos** han comprobado la fiabilidad de **alguna prueba de dorsiflexión en apoyo** con muestras que han incluido **mujeres**:
  - La **fiabilidad intra-observadores y fiabilidad inter-observadores** de la prueba de dorsiflexión en apoyo con la **rodilla flexionada** se ha comprobado en **6 ocasiones**.
  - La **fiabilidad intra-observadores** de la prueba de dorsiflexión en apoyo con la **rodilla en extensión** se ha comprobado en **6 ocasiones** y su **fiabilidad inter-observadores** en **5**.
- **La calidad** de los estudios es **mala** y los **métodos** empleados son **dispares** (heterogéneos).
- **Todos** los estudios han empleado el **coeficiente de correlación intraclase** y **algunos** proporcionan también **medidas dimensionales**.
- Los meta-análisis concluye que el **coeficiente de correlación intraclase combinado es bueno o muy bueno**.
- Las medidas **dimensionales** muestran un **error aleatorio** de los métodos de evaluación **de aceptable a muy alto** dependiendo del estudio.

## 5.4-RESULTADOS DEL ESTUDIO DE FIABILIDAD DEL STAR EXCURSION BALANCE TEST

En el momento de comenzar el presente estudio, 47 jugadoras del club estaban entrenando y jugando partidos con su equipo, de las que 37 aceptaron tomar parte en el mismo. De este grupo no incluimos a 11 jugadoras por no haber tenido la menarquía, por lo que en total 26 jugadoras fueron susceptibles de comenzar el estudio. Iniciamos el estudio de fiabilidad del *SEBT* el 8-05-2008, y al completar el par de mediciones en 9 jugadoras el 16-06-2008, realizamos un análisis previo de la fiabilidad del *SEBT* empleando el método de Bland-Altman. En este análisis comprobamos que las diferencias entre días eran extremadamente elevadas para considerar la prueba repetible, y decidimos no continuar con la recogida de datos. Los datos que esta tesis se presentan corresponden a esas nueve jugadoras iniciales. En total, 17 participantes que aceptaron tomar parte en el estudio, cumplieron los criterios de inclusión y no incurrieron en los de exclusión, no fueron estudiadas (figura 5.4.1). Todos los datos recogidos en este trabajo se pueden observar en anexo 5.4.1. Respecto a las participantes cabe decir que en sus características (tabla 5.4.1) la media y mediana en todas las variables es muy próxima. En cuanto al lapso de tiempo entre sesiones, este fue de una media 6,56 días, siendo la mínima de un día y la máxima de 21 días.

Tabla 5.4.1. Características de las participantes en estudio de fiabilidad del *SEBT*.

Variable	n	Edad (años)	AVol (años)	Voleibol (min/sem)	Altura (cm)	LPD (cm)	LPND (cm)	Tiempo ses (días)
<b>Media</b>	9	15,1	3,6	320	163,7	85,0	85,2	6,6
<b>DE</b>	9	2,2	1,8	60	6,1	4,4	4,5	6,5
<b>Mínima</b>	9	13,0	1,0	180	158,0	81,9	82,0	1,0
<b>Máxima</b>	9	19,0	7,0	360	176,0	96,1	96,7	21,0
<b>Mediana</b>	9	16,0	4,0	360	161,5	83,9	84,1	6,0
<b>Q1</b>	9	13,0	2,0	300	160,5	82,8	82,8	1,0
<b>Q3</b>	9	16,0	4,0	360	164,0	84,2	84,2	6,0
<b>RIQ</b>	9	3,0	2,0	60	3,5	1,4	1,4	5,0

AVol - Número de años practicando voleibol; DE - Desviación Estándar; Tiempo ses - Separación entre cada par de sesiones de evaluación; LPD - Longitud pierna dominante; LPND - Longitud pierna no dominante; Q1 - Primer cuartil; Q3 - Tercer cuartil; RIQ - Rango intercuartil; Voleibol- Minutos de entrenamiento de voleibol semanales.

Para comprobar la fiabilidad de la prueba, empleamos el método de Bland y Altman<sup>155</sup> (MBA) utilizando como estimadores la media y la máxima de los tres intentos registrados en cada una de las líneas, tanto con valores absolutos (tabla 5.4.2 y anexo 5.4.2) como con valores normalizados (tabla 5.4.4 y anexo 5.4.2).

Para esta tarea analizamos ambas piernas de las participantes de manera conjunta, puesto que un trabajo previo reveló que no existirían diferencias respecto la fiabilidad entre la pierna dominante y no dominante. Para completar el análisis con el MBA, realizamos un análisis de regresión (tabla 5.4.2 y 5.4.4).

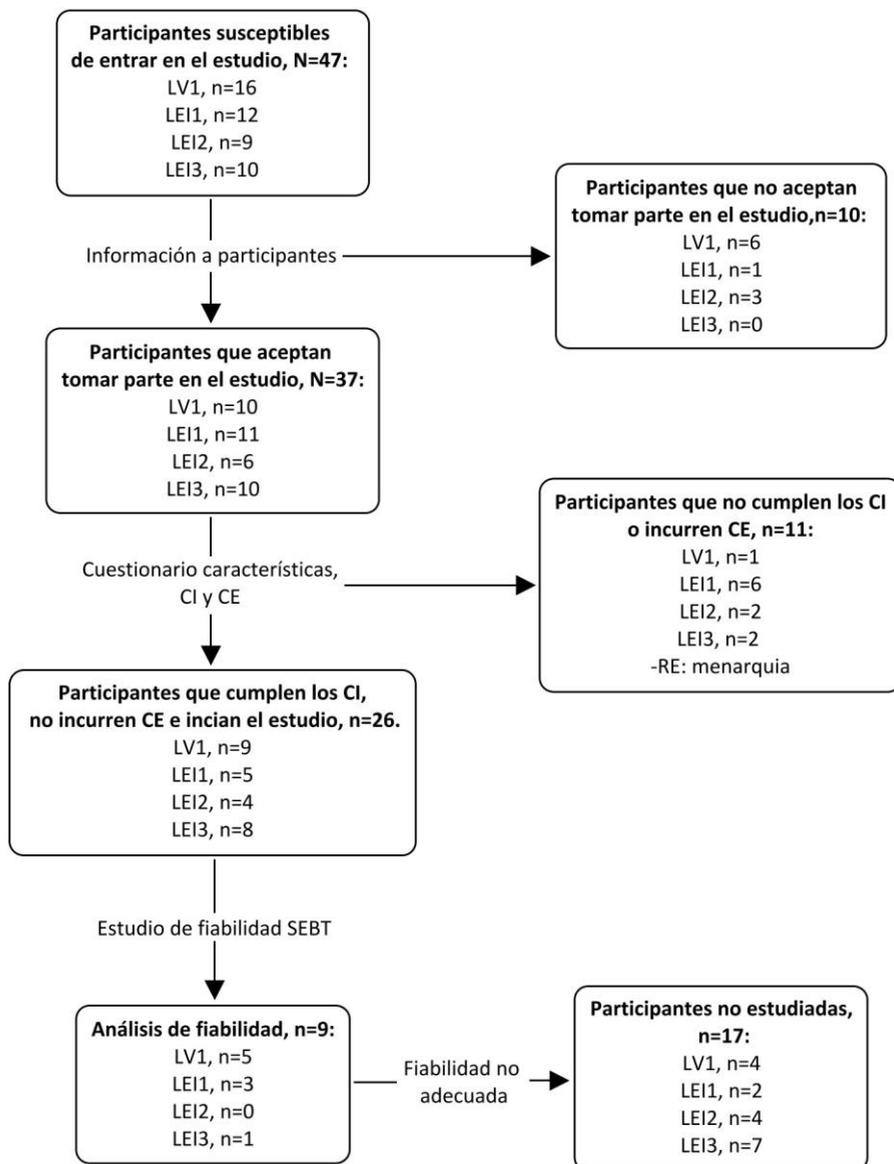


Figura 5.4.1. Diagrama de flujo participantes estudio fiabilidad SEBT.

LV1 - Liga vasca primera división; LEI1 - Equipo liga escolar infantil; LEI2 - Equipo liga escolar infantil 2; LEI3 - Equipo liga escolar infantil 3; CI - Criterios de inclusión; CE - Criterios de exclusión; RE - Razón de exclusión.

En cuanto a los valores absolutos obtenidos en cada una de las líneas, observamos que existe tendencia a lograr una mayor distancia el primer día de prueba con ambos estimadores. Este hecho se refleja la diferencia entre días, o error sistemático (ES), que en todos los casos se obtiene un valor positivo. Los gráficos derivados de este análisis (anexo 5.4.2) también evidencian esta tendencia mostrando la línea que representa el ES por encima del 0 en eje Y, como por ejemplo en la línea A empleando la media de los valores absolutos (figura 5.4.2). El ES es menor empleando la media como estimador en las líneas A y PL. En la línea PM, utilizando la máxima como estimador el ES es menor que utilizando la media, sin embargo la DE del ES es mayor.

Tabla 5.4.2. Resumen de los resultados del método de MBA utilizando los valores absolutos.

Variable	n	ES	ES IC 95%	LDA	LDA inf	LDA inf IC 95%	LDA sup	LDA sup IC 95%	r	p
<b>A X (cm)</b>	18	1,08	-0,22 a 2,38	5,11	-4,03	-6,29 a -1,77	6,19	3,93 a 8,45	0,12	NS
<b>PM X (cm)</b>	18	0,98	0,00 a 1,96	3,86	-2,87	-4,58 a -1,17	4,84	3,13 a 6,54	0,12	NS
<b>PL X (cm)</b>	18	0,92	-0,35 a 0,22	5,01	-4,09	-6,30 a -1,87	5,94	3,72 a 8,15	0,29	NS
<b>A max (cm)</b>	18	1,34	-0,04 a 2,72	5,44	-4,10	-6,50 a -1,70	6,78	4,38 a 9,18	0,26	NS
<b>PM max (cm)</b>	18	0,77	-0,53 a 2,06	5,11	-4,34	-6,60 a -2,08	5,88	3,62 a 8,13	0,10	NS
<b>PL max (cm)</b>	18	1,26	0,00 a 2,52	4,97	-3,71	-5,91 a -1,51	6,23	4,04 a 8,43	0,31	NS

ES - Error sistemático; IC 95% - Intervalo de confianza al 95%; max. - Máxima; LDA - Límite de acuerdo; LDA Inf - Límite de acuerdo inferior; LDA sup - Límite de acuerdo superior; NS - No significativo; r - Regresión; p - Nivel de significación; X - Media.

Respecto a los LDA, cabe decir que la línea A es la que valores más elevados obtiene, tanto con la media como con la máxima. En el caso de la media, el segundo LDA más elevado es el de la PL, y con la máxima como estimador la PM. Comparando los estimadores, los LDA son más pequeños empleando la media a excepción de en la línea PL, donde el LDA es menor con la máxima, pero esta diferencia es de 0,07 cm. Los LDA con media como estimador van de 3,86 cm en la línea PM hasta 5,11 cm en la línea A, y con la máxima de 4,97 cm a 5,44 cm en las líneas PL y A respectivamente. No observamos tendencia a que las diferencias tiendan a aumentar o disminuir en función de la magnitud de la medida, puesto la regresión no es significativa en todos los casos.

Atendiendo a los valores normalizados en relación a la longitud de la pierna, también observamos que se obtienen valores más elevados el primer día de prueba al ser ES siempre un valor positivo (tabla 5.4.4 y anexo 5.4.2). El ES del SEBT con los valores normalizados en las líneas A y PL, al igual que con valores absolutos, es menor empleando la media como estimador. Sin embargo, en la línea PM observamos un menor ES en la PM con la máxima distancia de los tres intentos, pero el LDA con la media como estimador resulta más pequeña.

Respecto a los LDA, éstos son más altos tanto con la máxima como estimador como con la media, en la línea A. Al igual que con valores absolutos, el LDA es más

pequeño es el de la línea PM empleando la media, y en la línea PL utilizando la máxima. Respecto a la diferencia entre estimadores los LDA son menores con la media como estimador, menos en la línea PL. Los LDA empleando la media los tres intentos como estimador van de un 6,04% en la línea A, a un 4,57% en la línea PM en relación a la longitud la pierna (figura 5.4.3). Con la máxima el LDA mayor es el de la línea A con un 6,42% y el menor en la línea PL con un 5,90%. En ninguna línea ni con ningún estimador observamos tendencia a que las diferencias aumenten o disminuyan en relación a la magnitud de las medidas ya que la regresión es no significativa en todos los casos.

Tabla 5.4.3. Resultados del MBA utilizando los valores normalizados.

Variable	n	ES	ES IC 95%	LDA	LDA inf	LDA inf IC 95%	LDA sup	LDA sup IC 95%	r	p
An X (%)	18	1,23	-0,30 a 2,76	6,04	-4,80	-7,47 a -2,14	7,27	4,60 a 9,94	0,25	NS
PMn X (%)	18	1,17	0,02 a 2,33	4,57	-3,40	-5,42 a -1,38	5,74	3,72 a 7,76	0,02	NS
PLn X (%)	18	1,15	-0,34 a 2,64	5,87	-4,72	-7,32 a -2,12	7,03	4,43 a 9,62	0,03	NS
An max (%)	18	1,54	-0,08 a 3,17	6,42	-4,88	-7,71 a -2,04	7,96	5,13 a 10,80	0,38	NS
PMn max (%)	18	0,91	-0,61 a 2,44	6,01	-5,10	-7,75 a -2,44	6,92	4,27 a 9,58	0,05	NS
PLn max (%)	18	1,54	0,04 a 3,03	5,90	-4,36	-6,96 a -1,75	7,43	4,83 a 10,04	0,03	NS

ES - Error sistemático; IC 95% - Intervalo de confianza al 95%; LDA - Límite de acuerdo; LDA Inf - Límite de acuerdo inferior; LDA sup - Límite de acuerdo superior; NS - No significativo; r - Regresión; p - Nivel de significación; X - Media.

Para finalizar, las tablas 5.4.4 y 5.4.5 muestran los resultados del Coeficiente de Correlación Intraclase (CCI), el Coeficiente de Correlación de Pearson (r de Pearson), el Error Estandar de la Medida (EEM), la mínima diferencia detectable (MDD) y el punto de corte (PDC) descrito por Jacobson y Truax<sup>163</sup>.

Tabla 5.4.4. Resultados del CCI, r de Pearson, EEM, MDD y PDC valores absolutos.

Variable	n	CCI	CCI IC %95	r Pearson	DE	EEM	PDC	MDD
A X (cm)	18	0,95	0,87 a 0,98	0,96	8,75	1,92	6,09	5,32
PM X (cm)	18	0,92	0,77 a 0,97	0,93	5,15	1,49	4,77	4,14
PL X (cm)	18	0,90	0,76 a 0,96	0,91	5,86	1,85	5,69	5,12
A max (cm)	18	0,94	0,84 a 0,98	0,96	8,82	2,10	6,54	5,81
PM max (cm)	18	0,88	0,70 a 0,95	0,88	5,24	1,85	5,80	5,12
PL max (cm)	18	0,89	0,71 a 0,96	0,92	5,88	1,92	5,98	5,33

CCI - Coeficiente de correlación intraclase; DE - Desviación estándar; IC 95% - Intervalo de confianza al 95%; EEM - Error estándar de la medida; PDC - Punto de corte; Max - Máxima; MDD - Mínima diferencia detectable; X - Media.

En primer lugar analizamos los resultados de los valores absolutos con la media y la máxima de los tres intentos (tabla 5.4.5). En este caso la fiabilidad relativa está por encima de 0,90 empleando la media, y por encima de 0,88 cuando se utiliza la máxima. La fiabilidad relativa es siempre mayor calculado a partir de la r de Pearson, que mediante el CCI. Tanto el CCI y la r de Pearson son mejores en todos los casos con la media con estimador a excepción de la línea PL, que obtiene una fiabilidad mejor con la máxima (0,01 puntos más). Los índices de fiabilidad

absoluta (EEM, PCD y MDD) también son mejores utilizando la media de los tres intentos, excepto en la línea PL.

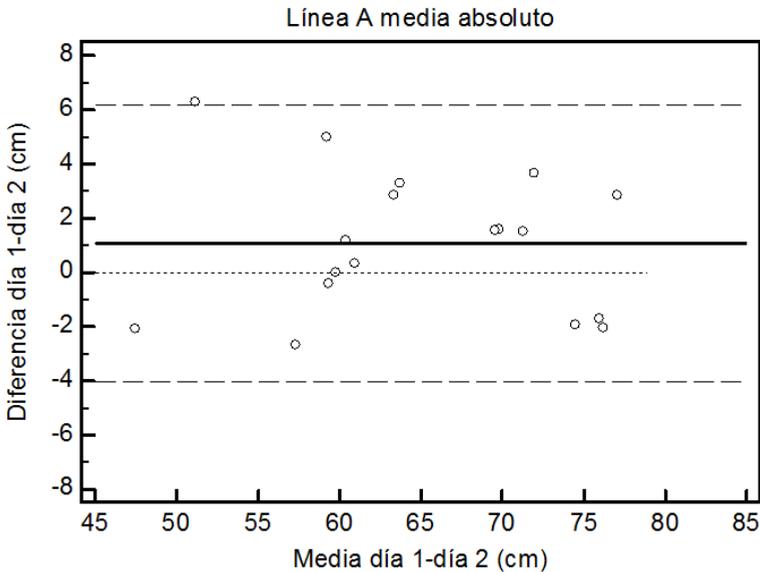
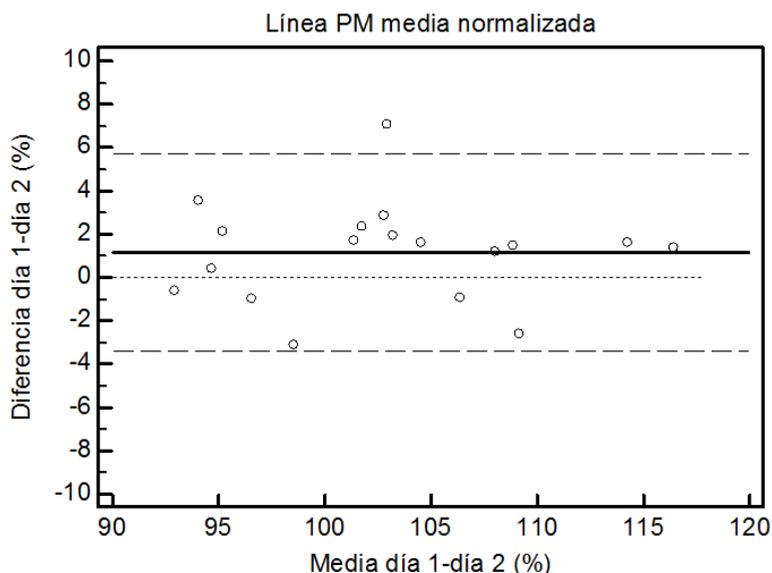


Figura 5.4.2. Gráfico Bland-Altman línea anterior empleando la media de los valores absolutos.  
 ---- Error sistemático (ES); - - - Límite de acuerdo (LDA) superior o inferior.

Siguiendo con los valores absolutos obtenidos en el *SEBT* (tabla 5.4.5), la línea que mejor fiabilidad relativa tiene es la línea A, tanto con la media como con la máxima. Sin embargo, la línea A es la que mayor DE tiene y en consecuencia observando los otros índices de fiabilidad absoluta, es la que mayor error de medición tiene. La línea PM es la que peor fiabilidad relativa tiene con la media pero la que mejor fiabilidad absoluta tiene al ser todos los índices de fiabilidad absoluta los más bajos. La línea PL es la que peor fiabilidad relativa reporta utilizando la media, pero en valores absolutos es la segunda mejor. Esto coincide con lo mostrado por los LDA que también son más alto en la línea A y los más bajos en la PM. Respecto a los resultados de la fiabilidad relativa, en el caso de utilizar la máxima como estimador seguido de la línea A, la segunda línea que mejor índice de fiabilidad reporta es la PM en el caso del CCI y la PL con r de Pearson. El EEM, PDC y MDD coinciden con el CCI, mostrando la línea PM la más fiable, y la línea A la menos fiable.



**Figura 5.4.3.** Gráfico Bland-Altman línea postero-medial empleando la media de los valores normalizados.

--- Error sistemático (ES); - - - Límite de acuerdo (LDA) superior o inferior.

En relación a los resultados de fiabilidad con la distancia normalizada en relación a la longitud de la pierna, los valores de fiabilidad relativa son iguales o superiores a 0,90 en todos los casos. Al igual que con valores absolutos, la  $r$  de Pearson es igual o superior al CCI. Utilizando la media como estimador el CCI y la  $r$  de Pearson son mejores que con la máxima. Los índices de fiabilidad absoluta también con la pierna normalizada coinciden con los LDA, al indicar que la fiabilidad es mejor con la media.

**Tabla 5.4.5.** Resultados del CCI,  $r$  de Pearson, EEM, MDD y PDC con valores normalizados.

Variable	n	CCI	CCI IC 95%	r Pearson	DE	EEM	PDC	MDD
<b>An X (%)</b>	18	0,95	0,86 a 0,98	0,95	9,71	2,25	7,02	6,25
<b>PMn X (%)</b>	18	0,93	0,81 a 0,98	0,94	6,86	1,78	5,69	4,92
<b>PLn X (%)</b>	18	0,92	0,80 a 0,97	0,93	7,68	2,18	6,96	6,05
<b>An max (%)</b>	18	0,94	0,83 a 0,98	0,95	9,81	2,46	7,46	6,83
<b>PMn max (%)</b>	18	0,92	0,77 a 0,96	0,91	7,07	2,17	6,84	6,02
<b>PLn max (%)</b>	18	0,90	0,74 a 0,96	0,92	7,45	2,30	7,38	6,36

CCI – Coeficiente de correlación intraclase; DE – Desviación estándar; IC 95% – Intervalo de confianza al 95%; EEM – Error estándar de la medida; PDC – Punto de corte; Max – Máxima; MDD – Mínima diferencia detectable; X – Media.

Respecto a los resultados de fiabilidad comparando las líneas, también encontramos que la línea A es la que mejor fiabilidad relativa obtiene con ambos estimadores. Aunque cuando observamos los EEM, el PDC y el MDD, es la que mayor error de medición tiene, y lo tanto menos fiable es. Respecto a las líneas posteriores utilizando la media, la PM es la segunda más fiable con CCI cuando se utiliza la media, y obtiene una índice igual a la PL con la máxima. Con los índices de fiabilidad absoluta utilizando la media, al igual que con los valores absolutos, la línea PM es la más fiable y la PL la siguiente más fiable. Cuando empleamos la máxima, la fiabilidad relativa muestra la línea PM la más fiable, y la A, la menos, según lo mostrado por el EEM, el PDC y la MDD.

Cuando comparamos los valores absolutos (tablas 5.4.3 y 5.4.5) con los normalizados (tablas 5.4.4 y 5.4.6), los índices de fiabilidad relativa son mejores cuando utilizamos el porcentaje de la distancia (es decir los valores normalizados).

Respecto a la comparación entre los límites de acuerdo obtenidos mediante el método propuesto por Bland y Altman y otros de los índices de fiabilidad absoluta empleados (EEM, MDD y PDC), cabe decir que en todos los análisis realizados, el MDD es el índice que más aproxima a los LDA. La MDD resulta de 0,01 a 0,37 cm mayor que el LDA, y de 0,01% a 0,46% cuando hablamos de valores normalizados. En cuanto al PDC, este valor también es mayor al LDA siendo la diferencia de 0,68 a 1,10 cm y de 0,98% a 1,48%.

**ESTUDIO DE CAMPO DE EL *STAR EXCURSION BALANCE TEST*:  
SINTESIS DE LOS RESULTADOS**

- **No existe error sistemático** en el *Star Excursion Balance Test*.
- **Los límites de acuerdo son muy elevados**, pero en general **inferiores** a los mostrados por la **literatura previa**.
- **No hay tendencia** en el *Star Excursion Balance Test* a que las **diferencias** entre cada par de mediciones **aumente en función de la magnitud** de la medida.
- Los **coeficientes de correlación intraclase** muestran que la **fiabilidad** intra-observadores del *Star Excursion Balance Test* es **muy buena**.
- La **mínima diferencia detectable** está muy **próxima** al **límite de acuerdo**.
- El **punto de corte** es ligeramente **superior** al **límite de acuerdo**.

## **5.5-RESULTADOS DEL ESTUDIO DE FIABILIDAD DE LA PDFAF Y LA PDFAE EMPLEANDO DIFERENTES MÉTODOS DE EVALUACIÓN**

Realizamos la recogida de datos de este estudio del 18-10-2010 al 5-11-2010 en PDFAF1, y del 30-05-2011 al 21-06-2011 en el PDFAF2, pudiéndose observar todos los datos en el anexo 5.5.1. Al inicio de la temporada 2010-2011, 33 jugadoras que estaban entrenando y jugando partidos con su equipo en el momento de comenzar el estudio, de las cuales 29 aceptaron tomar parte en el mismo. Dos participantes fueron descartadas, una por no haber tenido la menarquia, y otra por una enfermedad que no le permitía entrenar. Por lo tanto, 27 jugadoras iniciaron el estudio de fiabilidad (PDFAF1), y ninguna de ellas abandonó el mismo. Para el análisis final descartamos dos participantes (figura 5.5.1) por haberse superado el lapso de tiempo establecido *a priori* entre cada par de mediciones, a pesar de que comprobamos que este hecho no afectaba a los resultados. Las características de las 25 participantes que analizamos en el PDFAF1 se pueden observar en el anexo 5.5.2.

Al final de temporada 27 jugadoras estaban en activo en el momento de comenzar el estudio, de las cuales 21 aceptaron tomar parte en el mismo (PDFAF2). Las participantes que aceptaron tomar parte en la investigación, cumplieron los criterios de inclusión de este estudio, y no incurrieron en ninguno de los de exclusión (figura 5.5.2), finalizando todas ellas el estudio. Sus características se pueden ver en el anexo 5.5.2.

En total, 46 participantes tomaron parte en el estudio de fiabilidad de PDFAF (tabla 5.5.1). La edad media de la participantes fue de 16,26 años y la mediana de 16,00 años, en cuanto al IMC la media fue de 20,51 kg/m<sup>2</sup>, siendo la mediana prácticamente idéntica 20,45 kg/m<sup>2</sup>. El resto de variables sobre las características de las participantes como son la longitud de la piernas, la altura y el peso y los años practicando voleibol. La única excepción los minutos por semana dedicados a entrenar la fuerza neuromuscular, y los minutos de voleibol entrenados por semana. En cuanto al volumen de este tipo de entrenamiento fue de 47,83 min/semana, sin embargo la mediana fue 0,00 min/semana. La distribución en dicha variable no es normal puesto que había 24 jugadoras que no realizaban este tipo de entrenamiento, y las 22 que sí, lo realizaban todas con el mismo volumen por semana. Las jugadoras que realizaban este tipo de entrenamiento lo realizaban de manera voluntaria y en función de la disponibilidad horaria y correspondían siempre a los equipos de LV1 y LV2. Respecto a los minutos de entrenamiento de voleibol, los equipos LV1 y LV2 entrenaban tres veces por semana empleando dos horas por sesión, y el equipo de liga escolar empleaba dos días por semana

realizando un entrenamiento de 90 minutos. Es por ello que la distribución de esta variable presenta cierta heterogeneidad.

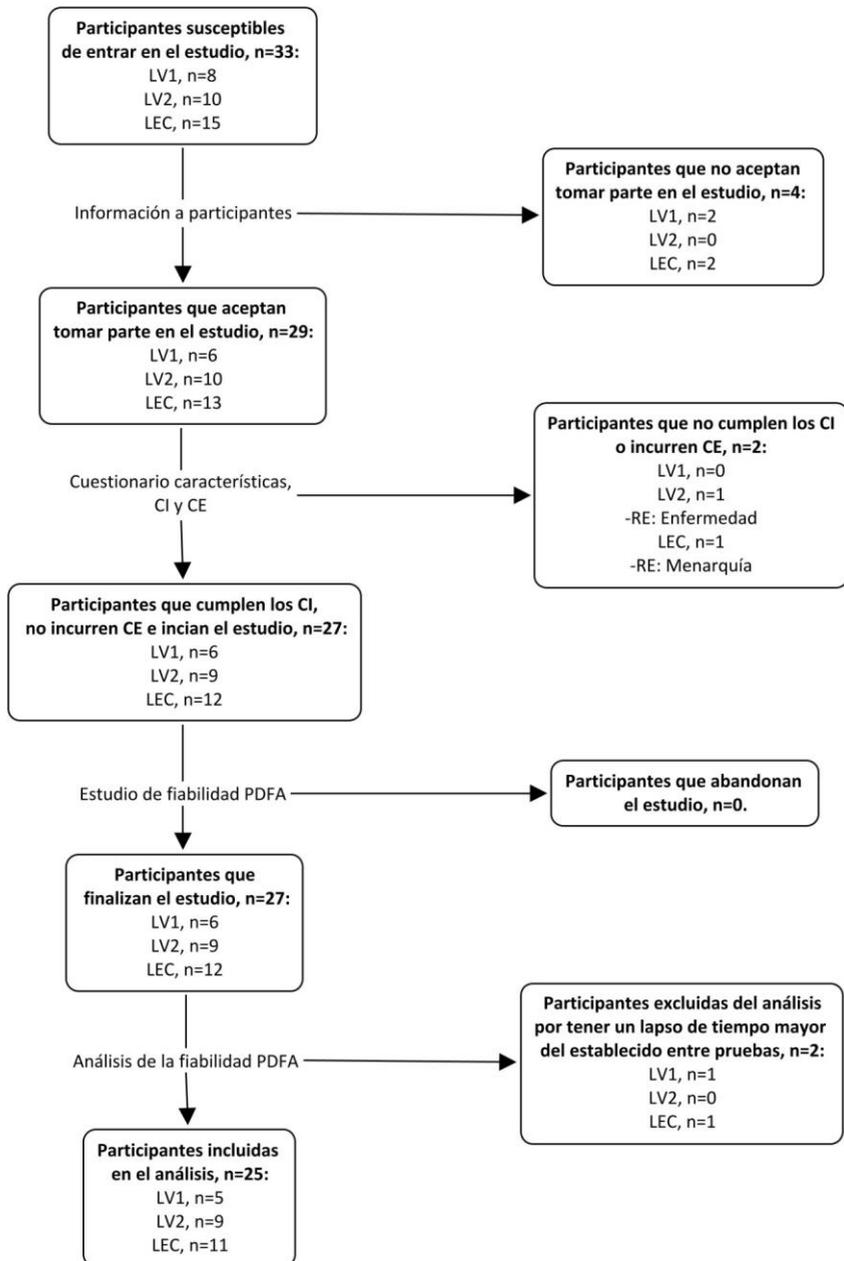


Figura 5.5.1. Diagrama de flujo de las participantes en el estudio PDFA1.

LV1 - Liga vasca primera división; LV2 - Liga vasca segunda división; LEC - Liga escolar cadete; CI - Criterios de inclusión; CE - Criterios de exclusión; RE - Razón de exclusión.

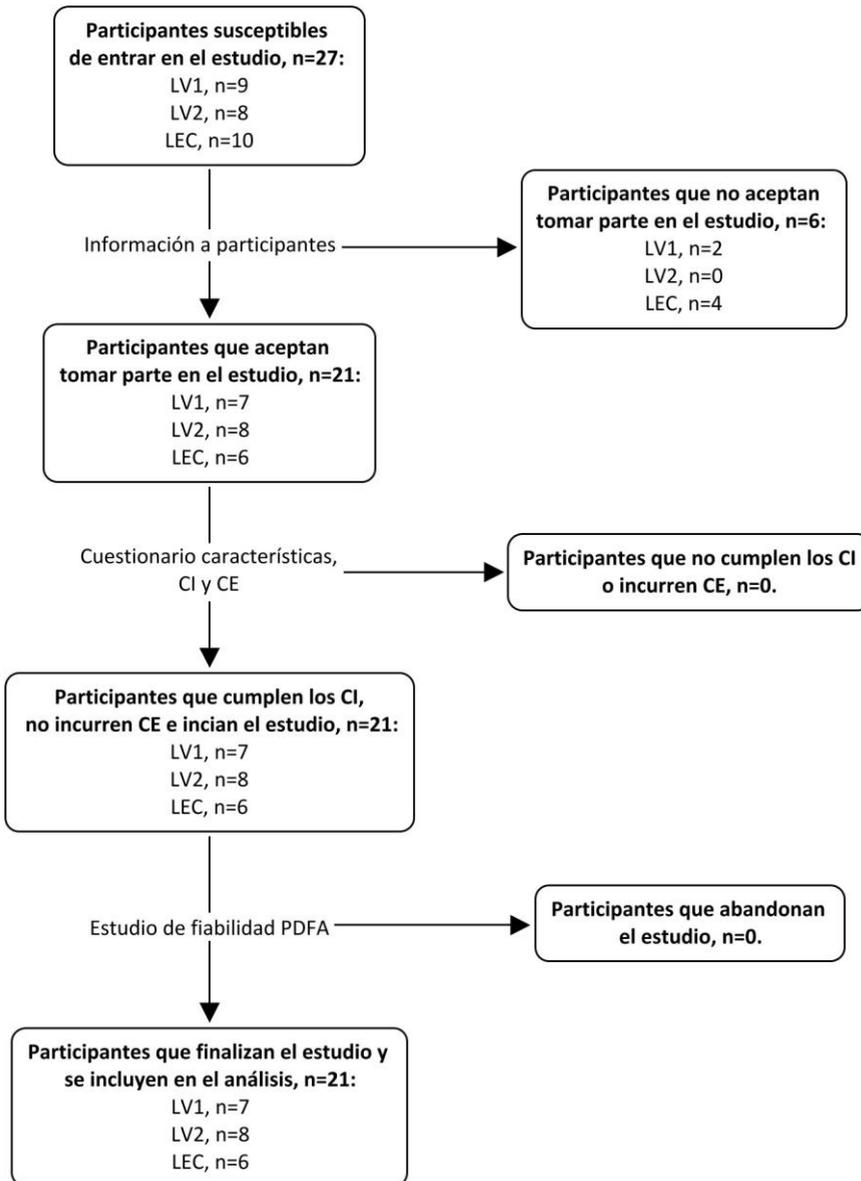


Figura 5.5.2. Diagrama de flujo de las participantes en el estudio PDFa2.

LV1 - Liga vasca primera división; LV2 - Liga vasca segunda división; LEC - Liga escolar cadete; CI - Criterios de inclusión; CE - Criterios de exclusión.

Respecto a las características de las participantes al inicio y final de temporada (PDFa1 y PDFa2), cabe decir que todas las variables evaluadas son muy similares entre ambos grupos (anexo 5.5.2), siendo la edad, los minutos por semana entrenados de fuerza y voleibol y la experiencia de las participantes ligeramente

mayores en el grupo en PDFA2. La causa de estas mínimas diferencias, es que al final de la temporada la muestra de jugadoras de liga escolar (las de menor edad, entrenamiento y experiencia) fue menor, descendiendo la número de 11 a 6 en PDFA2. La razón para la disminución de este grupo fue que habían dejado de entrenar por ser final de curso y estar en periodo de exámenes.

En cuanto a las características de las sesiones de evaluación la distribución también es normal. La separación media entre cada par de mediciones de fue 2,44 días, siendo la máxima de siete y la mínima de un día, con una mediana de dos días. La diferencia horaria media entre estas evaluaciones en los dos días evaluados fue de 15,56 minutos, y nunca superó los 50 minutos, siendo la mediana de 15,00 minutos.

Tabla 5.5.1 Características de las participantes del estudio fiabilidad PDFA.

Variable	n	Edad (años)	AVol (años)	Voleibol (min/sem)	Fuerza (min/sem)	Altura (cm)	Peso (kg)	IMC (kg/m <sup>2</sup> )	PD (cm)	PND (cm)	Tiempo ses (días)	Dif hora ses (min)
Media	46	16,3	5,2	282,4	47,8	165,7	56,5	20,5	86,5	86,6	2,4	15,6
DE	46	2,8	2,7	92,2	50,5	6,5	7,1	1,8	4,4	4,4	1,9	12,1
Mínima	46	13,0	1,0	0,0	0,0	153,0	41,1	17,2	79,0	79,0	1,0	0,0
Máxima	46	28,0	16,0	360,0	100,0	180,0	68,8	24,3	97,5	97,8	7,0	50,0
Mediana	46	16,0	5,0	360,0	0,0	164,3	57,5	20,5	85,1	85,1	2,0	15,0
Q1	46	14,0	4,0	180,0	0,0	162,0	52,2	19,4	83,7	83,9	1,0	5,3
Q3	46	17,0	5,8	360,0	100,0	168,9	61,3	21,8	89,1	88,7	3,0	18,8
RIQ	46	3,0	1,8	180,0	100,0	6,9	9,1	2,3	5,5	4,9	2,0	13,5

AVol - Número de años practicando voleibol; DE - Desviación Estándar; Tiempo ses - Separación entre cada par de sesiones de evaluación; Dif hora ses - Diferencia horaria entre cada par de sesiones de evaluación; Fuerza - Minutos de entrenamientos de fuerza semanales; IMC - Índice de masa corporal; PD - Longitud pierna dominante; PND - Longitud pierna no dominante; Q1 - Primer cuartil; Q3 - Tercer cuartil; RIQ - Rango intercuartil; Voleibol - Minutos de entrenamiento de voleibol semanales.

Para valorar la fiabilidad de las pruebas de dorsiflexión en apoyo empleamos el método propuesto por Bland y Altman (MBA), y posteriormente calculamos diferentes estimadores de fiabilidad empleados por otros autores<sup>121;125;126;130;134;135;143;171-173</sup> (CCI, EEM y MDD) que comprobaron la fiabilidad de las pruebas. Procedimos a calcular estos índices para poder comparar nuestros los resultados con los extraídos en la RS PDFA y los MA realizados. Además de manera análoga al estudio fiabilidad del *SEBT* también calculamos el PCD propuesto por Sabin et al. (2011)<sup>90</sup>. Como estimador de cada método de evaluación utilizamos la media de los tres intentos realizados de cada sesión con cada pierna siempre que esto fue posible, al igual que la mayoría (tabla 5.3.1) de los estudios que comprobaron la fiabilidad PDFA. En el caso de PDFA1, en 12 ocasiones (seis en DVf, seis en ADf, dos en DTe, dos en DDe, una en DHf y una en DDf), la media se realizó con dos intentos, puesto que en uno de los intentos no fue posible registrar la marca, al borrar la participante la marca realizada en la

línea cuando abandonaba la posición evaluada. En PDFAf2 no se perdió el registro de ninguna marca. En este trabajo analizamos ambas piernas de manera conjunta puesto que en un análisis previo observamos que ambas piernas obtenían errores de medición similares.

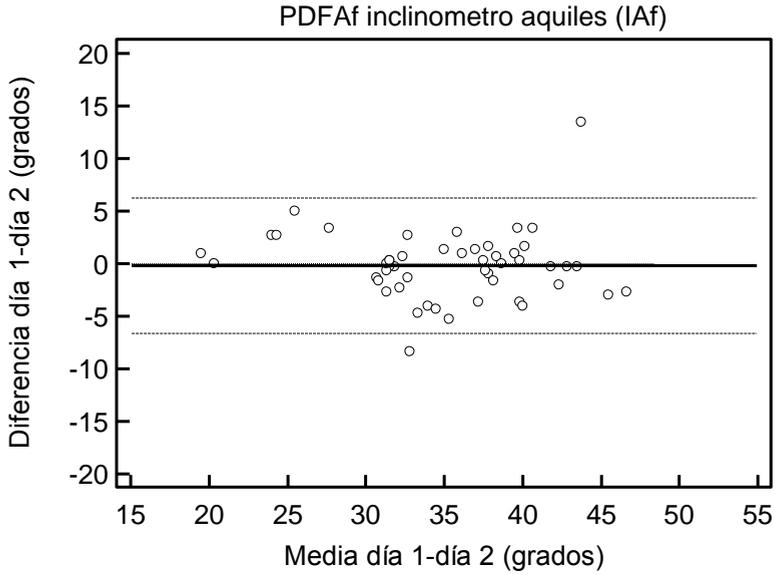
En la PDFAf, los resultados del análisis del MBA nos muestran (tabla 5.5.2 y anexo 5.5.3) que el error sistemático (ES) de la prueba es despreciable en todos los métodos empleados en para la prueba. Cuando la unidad de medición es cm, este error va de -0,12 cm a 0,23 cm, y cuando es en grados, de -0,19° a 0,41°. Respecto a los límites de acuerdo, estos son muy similares entre los métodos que evalúan la DF en cm (DHf, DDf y DVf), con valores de 1,78 cm a 2,42 cm. Atendiendo a los métodos que emplean la unidad de grado, cuando el inclinómetro es la herramienta de evaluación los LDA son de 6,41° para IAf y 6,05° para ITf, siendo inferior para el ADf con 3,01°. Estas diferencias hacen que, por ejemplo, cuando comparamos dos métodos de evaluación como el IAf y el ADf, los LDA inferior y superior estén más separados y que las participantes estén más dispersas en el método con un LDA más elevado (figura 5.5.3 y figura 5.5.4). En cuanto al análisis de regresión correspondiente a PDFAf, éste resulta no significativo en todas variables evaluadas.

Tabla 5.5.2. Resultados análisis de fiabilidad de la PDFAf con el MBA

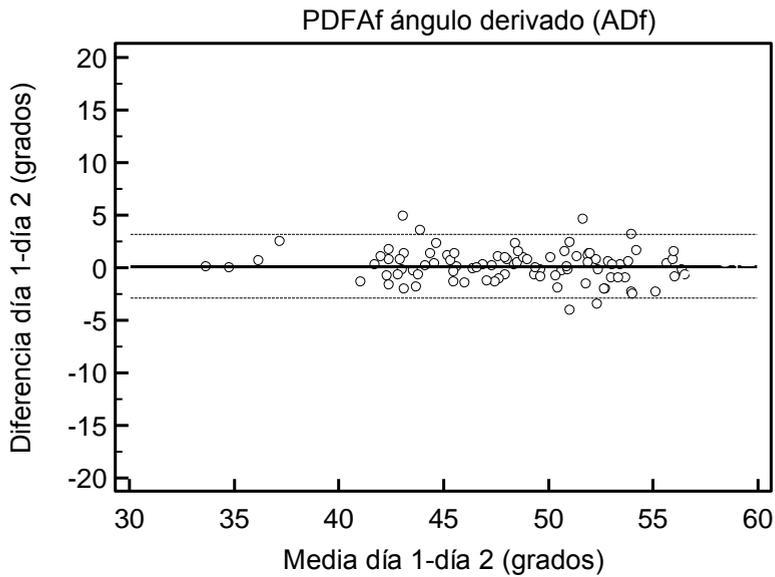
Variable	n	ES	ES IC 95%	LDA	LDA inf	LDA inf IC 95%	LDA sup	LDA sup IC 95%	r	p
DHf (cm)	92	0,07	-0,12 a 0,26	1,78	-1,71	-2,04 a -1,39	1,85	1,53 a 2,17	0,11	NS
DDf (cm)	50	0,23	-0,04 a 0,50	1,86	-1,63	-2,10 a -1,17	2,10	1,63 a 2,56	0,20	NS
DVf (cm)	92	-0,12	-0,38 a 0,13	2,42	-2,54	-2,98 a -2,10	2,29	1,85 a 2,73	0,11	NS
IAf (°)	50	-0,19	-1,12 a 0,74	6,41	-6,60	-8,20 a -5,00	6,22	4,62 a 7,82	0,05	NS
ITf (°)	50	0,41	-0,47 a 1,28	6,05	-5,64	-7,15 a -4,13	6,46	4,95 a 7,97	0,22	NS
ADf (°)	92	0,14	-0,18 a 0,46	3,01	-2,88	-3,42 a -2,33	3,15	2,61 a 3,70	0,14	NS

ES - Error sistemático; IC 95% - Intervalo de confianza al 95%; LDA - Límite de acuerdo; LDA Inf - Límite de acuerdo inferior; LDA sup - Límite de acuerdo superior; NS - No significativo; r - Regresión; p - Nivel de significación.

Respecto al análisis de fiabilidad con los indicadores utilizados por otros autores (tabla 5.5.3), cabe decir que el CCI en todas las variables es muy similar a la r de Pearson. De todos los índices de fiabilidad, el MDD es el que más se aproxima a los límites de acuerdo en todos los métodos de evaluación empleado. La MDD es 0,01 cm menor que los LDA en DHf y DVf. En el caso de la DDf la MDD 0,03 cm mayor que el LDA. En cuanto a las medidas en grados el MDD es de 0,01° a 0,08° menor que el LDA. Las diferencias entre el PDC respecto a los LDA varían de -0,15 cm a 0,19 cm y de -0,23° a 0,26°.



**Figura 5.5.3.** Gráfico Bland-Altman del inclinómetro ubicado en aquiles en la PDFAf.  
 ---- Error sistemático (ES); - - - Límite de acuerdo (LDA) superior o inferior.



**Figura 5.5.4.** Gráfico Bland-Altman del ángulo derivado en la PDFAf.  
 ---- Error sistemático (ES); - - - Límite de acuerdo (LDA) superior o inferior.

Los mayores índices de fiabilidad adimensionales en PDFAf son todos aquellos que emplean unidades en cm además del ADf, con valores de 0,95 y 0,96 tanto en ICC y como en la r de Pearson. La DF medida con el inclinómetro, obtiene unos índices inferiores, con valores de 0,86 a 0,88. Cuando los indicadores de grado de fiabilidad son EEM, PDC y MDD, las medidas en cm obtienen valores muy similares entre ellas, siendo ligeramente mayores en DVf. Entre las medidas proporcionadas en grados, la ADf obtiene el menor error de medición, con un valor de 3,00° en MDD, siendo este error superior a 6° en ITf y IAf.

Tabla 5.5.3. Resultados análisis de fiabilidad de la PDFAf con CCI, r de Pearson, EEM, PDC y MDD.

Variable	n	CCI	CCI IC 95%	r		DE	EEM	PDC	MDD
					Pearson				
DHf (cm)	92	0,96	0,94 a 0,97	0,96	3,17	0,64	1,84	1,77	
DDf (cm)	50	0,95	0,91 a 0,97	0,95	3,05	0,68	2,05	1,89	
DVf (cm)	92	0,96	0,93 a 0,97	0,96	4,14	0,87	2,27	2,41	
IAf (°)	50	0,87	0,78 a 0,92	0,86	6,24	2,28	6,18	6,33	
ITf (°)	50	0,87	0,79 a 0,93	0,88	6,11	2,17	6,31	6,02	
ADf (°)	92	0,95	0,93 a 0,97	0,96	5,04	1,08	3,12	3,00	

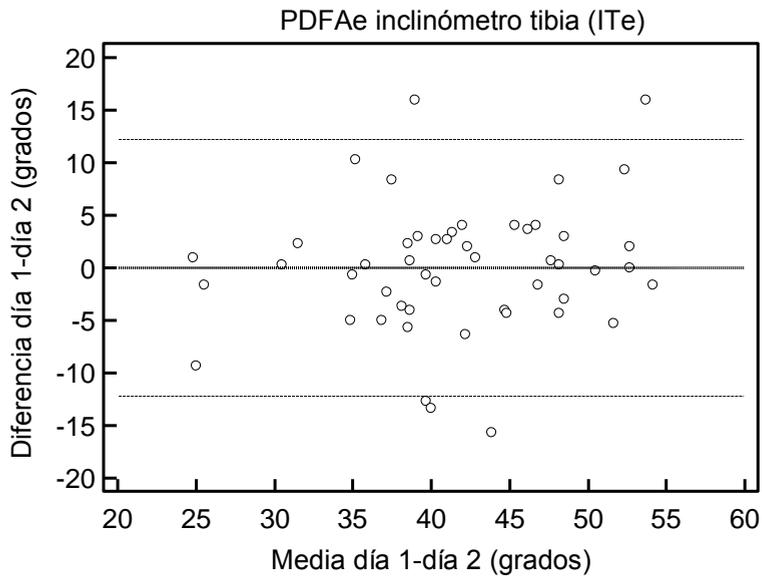
CCI – Coeficiente de correlación intraclase; DE – Desviación estándar; IC 95%- Intervalo de confianza al 95%; EEM - Error estándar de la medida; MDD – Mínima diferencia detectable; PDC - Punto de corte.

Respecto al PDFAe, los resultados del análisis de fiabilidad empleando el MBA (tabla 5.5.4 y anexo 5.5.3) muestran que la media de las diferencias (o ES) es muy próxima a 0 en todos los métodos de evaluación empleados, variando de -0,19 cm a 0,17 cm, y de -0,34° a 0,00°. En cuanto a los límites de acuerdo (LDA), en el caso de las medidas con unidades en cm, estas van de 2,34 cm (DVe) a 10,49 cm (DDe). En cuanto las medidas tomadas en grados, el LDA más reducido es ADe con 6,72°, y el más alto se obtiene en ITe con 12,21°, estando las participantes mucho más dispersas en el gráfico de Bland y Altman de ITe (figura 5.5.5) que en el de ADe (figura 5.5.6). En cuanto al análisis de regresión, ninguno de los análisis que realizamos resulta significativo.

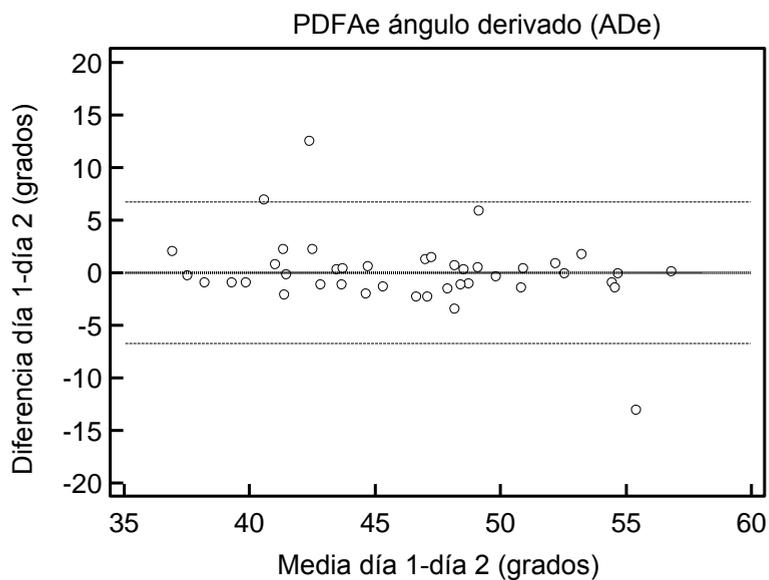
Tabla 5.5.4. Resultados análisis de fiabilidad de la PDFAe con el MBA..

Variable	n	ES	ES IC 95%	LDA	LDA inf	LDA inf IC 95%	LDA sup	LDA sup IC 95%	r	p
DTe (cm)	92	0,13	-0,86 a 1,12	9,38	-9,26	-10,96 a -7,56	9,51	7,89 a 11,21	0,05	NS
DRe (cm)	42	0,10	-0,98 a 1,17	6,74	-6,65	-8,49 a -4,80	6,84	5,00 a 8,68	0,18	NS
DHe (cm)	42	0,17	-0,56 a 0,90	4,56	-4,39	-5,54 a -3,14	4,74	3,48 a 5,99	0,27	NS*
DDe (cm)	50	-0,12	-1,64 a 1,40	10,49	-20,61	-13,23 a -8,00	10,37	7,75 a 12,99	0,10	NS
DVe (cm)	42	-0,19	-0,56 a 0,18	2,34	-2,53	-3,17 a -1,88	2,15	1,51 a 2,79	0,02	NS
IAe (°)	50	-0,34	-1,48 a 0,80	7,83	-8,17	-10,12 a -6,22	7,49	5,54 a 9,44	0,26	NS
ITe (°)	50	-0,01	-1,78 a 1,76	12,21	-12,22	-15,26 a -9,17	12,20	9,16 a 15,25	0,18	NS
ADe (°)	42	0,00	-1,07 a 1,07	6,72	-6,72	-8,55 a -4,88	6,72	4,88 a 8,55	0,28	NS

ES - Error sistemático; IC 95% - Intervalo de confianza al 95%; LDA - Límite de acuerdo; LDA Inf - Límite de acuerdo inferior; LDA sup - Límite de acuerdo superior; NS - No significativo; r - Regresión; p - Nivel de significación; \* - calculada excluyendo un outlier: Pierna dominante de la participante ID 9 en PDFA2 (anexo 5.5.1). El primer día obtuvo una media de 37,8 cm y el segundo 26,43 cm, lo que supone una media entre cada par de mediciones de 32,12 cm y una diferencia entre las mismas de 11,37 cm.



**Figura 5.5.5.** Gráfico Bland-Altman inclinómetro tibia en la PDFAe.  
 --- Error sistemático (ES); - - - Límite de acuerdo (LDA) superior o inferior.



**Figura 5.5.6.** Gráfico Bland-Altman del ángulo derivado en la PDFAe.  
 --- Error sistemático (ES); - - - Límite de acuerdo (LDA) superior o inferior.

En cuanto a los resultados de CCI,  $r$  de Pearson, EMM, PDC y MDD (tabla 5.5.5) de la PDFAe, observamos que al igual que en PDFAf, el CCI y  $r$  de Pearson son muy similares. Tanto la  $r$  de Pearson como el CCI obtienen valores muy parecidos en todos los métodos de evaluación empleados, siendo la DVe la variable que mayor índice de fiabilidad proporciona con 0,96. El resto de variables obtienen valores muy parejos que varían de 0,81 a 0,86 a excepción de ITe (0,69) y de DHe (CCI de 0,76 y  $r$  de Pearson de 0,78), que son inferiores. Respecto a las medidas dimensionales, volvemos a apreciar que el MDD es el valor que más se aproxima a los LDA. En cm el MDD es de 0,01 cm a 0,13 cm menor que el LDA y de 0,15° a 0,08° menor. En cuanto al PCD, este índice presenta una diferencia respecto al LDA de -0,36 cm a 0,39 cm y de -0,11° a 0,23°.

Tabla 5.5.5. Resultados análisis de la fiabilidad de la PDFAe con CCI,  $r$  de Pearson, EEM, PDC y MDD.

Variable	n	CCI	CCI IC 95%	$r$ Pearson	DE	EEM	PDC	MDD
DTe (cm)	92	0,84	0,77 a 0,89	0,84	8,40	3,36	9,42	9,32
DRe (cm)	42	0,82	0,68 a 0,90	0,82	5,58	2,40	6,86	6,64
DHe (cm)	42	0,76	0,60 a 0,86	0,78	3,33	1,63	4,20	4,51
DDe (cm)	50	0,81	0,69 a 0,89	0,81	8,63	3,74	10,21	10,36
DVe (cm)	42	0,96	0,93 a 0,98	0,96	4,17	0,84	2,73	2,33
IAe (°)	50	0,81	0,69 a 0,89	0,82	6,68	2,80	7,73	7,75
ITe (°)	50	0,69	0,51 a 0,81	0,69	7,83	4,35	12,44	12,06
ADe (°)	42	0,81	0,68 a 0,90	0,82	5,51	2,38	6,61	6,61

CCI – Coeficiente de correlación intraclase; IC 95% - Intervalo de confianza al 95%; EEM - Error estándar de la medida; MDD – Mínima diferencia detectable; PDC - Punto de corte.

En lo que respecta a las medidas de fiabilidad interpretables desde el punto de vista práctico, cabe decir que el menor error de medición en las variables medidas en cm corresponde a la DVe en todos los estimadores, seguido del DHe. En cuanto a los métodos de evaluación que emplean unidades en grados, el menor error de medición corresponde a ADe con un PDC y un MDD de 6,61°, estando próximo el IAe con 7,73° y 7,75° respectivamente. El ITe es la medida en grados que mayor error de medición reporta en todos los índices calculados en este trabajo.

Pasando a comparar las pruebas entre ellas, el CCI y la  $r$  de Pearson siempre resulta más elevada en las medidas tomadas en cm, respecto a las tomadas con grados, con la excepción de ADf y ADe que son similares a las evaluadas en cm. En cuanto a la comparación entre las variables parejas en las dos pruebas (DHf-DHe/DTe, DDf-DDe, IAF-IAe, ITf-ITe y ADf-ADe) encontramos que estos índices adimensionales resultan mejores en PDFAf que en PDFAe a excepción de DVf-DVe que son iguales. Lo mismo ocurre cuando evaluamos la fiabilidad con medidas dimensionales, el error de medición siempre es más reducido en PDFAf en comparación al PDFAe.

Para completar este análisis, comparamos las variables evaluadas tanto en PDFA1 como en PDFA2 (DHf, DVf, ADf y DTe) mediante el método de Bland y Altman (MBA) (anexo 5.5.4). El ES en todas las variables, tanto a principio como al final de temporada, es muy cercano a 0, con valores de -0,13 cm a 0,27cm, y de 0,22° en ADf del PDFA1 y de 0,04° en PDFA2. En cuanto a los LDA, estos son algo menores PDFA1 que en PDFA2, con una diferencia de 0,08 cm en DHf, de 0,39 cm en DVf y de 0,59° en ADf. La variable que mayor diferencia obtiene (2,38 cm) es la DTe, con un LDA de 10,56 cm PDFA1 y de 7,88 cm en PDFA2. La regresión en todas las variables exceptuando la DHf en PDFA1, resulta no significativa.

Para finalizar, en lo referente a los resultados de las PDFA, evaluamos el acuerdo entre métodos de evaluación dentro de cada una de las pruebas de DF mediante el MBA (anexo 5.5.5). Cabe destacar que el ES un muy elevado entre I Af-ITf (-14,24°), I Af-ADf (-12,49°) y I Ae-ITe (-10,63°), siendo el ES más pequeño en el acuerdo ITf-ADf (1,93°). Respecto a los límites de acuerdo estos varían de 4,24° (ITf-ADf) a 7,96° (I Af-ITf), siendo la regresión significativa en todos los casos a excepción de I Af-ITf.

### ESTUDIO DE CAMPO DE LAS PRUEBAS DE DORSIFLEXIÓN EN APOYO: SÍNTESIS DE LOS RESULTADOS

- **No existe error sistemático** en las pruebas de dorsiflexión en apoyo.
- Los **límites de acuerdo son reducidos** y en general **inferiores** a los mostrado por la **literatura previa**:
  - **El ángulo derivado** muestra los **límites de acuerdo más bajos** de todas los métodos de evaluación en grados.
- **No hay tendencia** en las PDFA a que las **diferencias** entre cada par de mediciones  **aumente en función de la magnitud** de la medida.
- Los **coeficientes de correlación intraclase** muestran una fiabilidad intra-observadores de las PDFA de **buena a muy buena** dependiendo del método de evaluación:
  - **El ángulo derivado** muestra un **coeficiente de correlación intraclase mas elevados** que otros métodos de evaluación en grados.
- La **mínima diferencia detectable** está muy **próxima** al **límite de acuerdo**.
- El **punto de corte** es ligeramente **superior** al **límite de acuerdo**.
- La prueba de dorsiflexión en apoyo con la rodilla flexionada **es más fiable que** con la rodilla en extensión en todos los análisis realizados.
- **No existe relación** entre las medidas de los resultados de los diferentes **métodos de evaluación** empleados.

# 6-DISCUSIÓN



En esta tesis hemos estudiado los factores de riesgo intrínsecos (FR) de las lesiones de tobillo en mujeres practicantes de actividad física mediante el método de revisión sistemática (RS). Además hemos comprobado la fiabilidad de dos pruebas que se emplean para valorar el estado anatómico-funcional del tobillo como predisponentes a sufrir una lesión de tobillo, a saber; el *Star Excursion Balance Test (SEBT)* y las pruebas de dorsiflexión en apoyo (PDFA). Para estudiar la fiabilidad estas pruebas, hemos empleado el método de RS que se completó con un estudio de campo que utilizó una muestra de jugadoras de voleibol.

La RS sobre los factores riesgo (FR) identificó 10 estudios que habían analizado algún FR en la población de interés. En total, la literatura previa estudió 103 FR diferentes dentro de diferentes áreas (características de las participantes, fuerza específica del tobillo, equilibrio...). Los métodos para evaluar los FR dentro de una misma área son muy diferentes, y en su mayoría complejos y sofisticados, habiéndose replicado únicamente 15 de ellos. De todos los rasgos estudiados, esta revisión no ha podido confirmar ni descartar ninguno de ellos ya que las pruebas científicas encontradas no permiten realizar un juicio definitivo al respecto.

En cuanto a la fiabilidad de las pruebas para valorar el estado anatómico-funcional del tobillo, las revisiones demostraron que ningún estudio había comprobado la precisión de las mismas empleando población activa de mujeres. En cuanto a la calidad de los estudios incluidos en ambas RS es reseñable mencionar que está fue escasa. Asimismo, los procedimientos para realizar tanto el *SEBT* como las PDFA, fueron muy dispares demostrando la falta de estandarización de la pruebas.

En cuanto a la *Star Excursion Balance Test (SEBT)*, tanto la RS como el estudio de campo con la población de interés, muestran que el *SEBT* es una prueba poco precisa, no siendo conveniente su empleo como método de predicción de lesiones, ni herramienta para el diagnóstico.

Respecto a las pruebas de dorsiflexión en apoyo, la RS estudio la prueba tanto con la rodilla flexionada (PDFAf) como con la rodilla en extensión (PDFAe). La revisión encontró que la literatura había empleado, además de diferentes procedimientos, métodos evaluación diversos como el goniómetro, inclinómetro, cinta métrica u otros aparatos. Este trabajo ha demostrado que el error de medición de estas pruebas, aplicando un procedimiento preciso y un método de evaluación muy sencillo como es el ángulo derivado de las distancias por razones trigonométricas (AD), es pequeño en comparación con otros métodos para evaluar las PDFA propuestos en la literatura previa. Es más, este método de evaluación es aplicable a ambas pruebas: PDFAf y PDFAe. Esta tesis también ha demostrado que no existe relación entre las medidas de resultado de los diferentes métodos de evaluación propuestos para las PDFA.

## **6.1-REVISIÓN SISTEMÁTICA SOBRE LOS FACTORES DE RIESGO DE LAS LESIONES DE TOBILLO**

A pesar de que las lesiones de tobillo son muy frecuentes entre la población activa, únicamente 10 trabajos han estudiado los FR en mujeres deportistas, tras una búsqueda que incluyó las bases de datos más importantes existentes en este área de conocimiento. Consideramos que este número es pequeño teniendo en cuenta la incidencia de la lesión y las consecuencias que produce. Asimismo solo el 60% de los trabajos tenían como objetivo estudiar los FR de para las lesiones de tobillo y únicamente dos<sup>40;169</sup> emplearon exclusivamente una muestra de mujeres. A pesar de ello, la calidad de los estudios que incluimos en la RS fue buena, presentado un riesgo de sesgo pequeño. No obstante cabe destacar que la mayoría de los estudios presentaban riesgo de sesgo en dos apartados principalmente: la pérdida diferencial de participantes y los factores de confusión. Respecto a la pérdida de participantes, es preciso señalar que a pesar que la mayoría de los trabajos describe correctamente la muestra analizada, en ninguno de ellos se han descrito los criterios de selección, los criterios de inclusión y exclusión correctamente, ni la pérdida de las participantes durante el reclutamiento de la muestra y su seguimiento. Es preciso que los estudio pronósticos acerca de los FR de lesión de tobillo describan esto correctamente para ayudar a los lectores a interpretar los resultados, y conocer si los resultados pueden estar sesgados en este sentido.

Atendiendo al riesgo de sesgo encontrado en la evaluación de los factores de confusión, es reseñable que únicamente el 50% de los estudios los tienen en cuenta. Los trabajos sólo toman en cuenta un factor de confusión, la exposición a la actividad física que es evaluado en algunos casos en el tiempo de exposición<sup>18;40;170</sup>, mientras otros cuentan el número de exposiciones<sup>44;55</sup>. Si bien es cierto que medir el tiempo de evaluación resulta muy laborioso, esta variable deberá ser evaluada preferentemente de esta manera<sup>175</sup> ya que el número de exposiciones resulta engañoso puesto que el tiempo que dura una exposición puede variar.

Además de la exposición a la actividad física, dos estudios, Willems (2004)<sup>18</sup> y Beynnon et al. (2001)<sup>55</sup>, consideran la lesión previa de tobillo o MI como factor de confusión, y no incluyen a participantes con lesión previa. Aunque esta RS muestra que este puede ser un FR, no consideramos adecuado el no incluir a este tipo de participantes, ya que no refleja la realidad de la población estudiada. Si examinamos la incidencia que obtienen estos estudios encontramos que el 19% para Beynnon et al. (2001)<sup>55</sup> y un 20% para Willems (2004)<sup>18</sup> de las participantes supuestamente sanas, se lesionan. Teniendo en cuenta que las lesiones de tobillo

no revisten mucha gravedad, y las participantes no abandonan la práctica deportiva por este motivo, es de esperar que la prevalencia en mujeres que alguna vez han sufrido alguna lesión en el tobillo sea elevada. Por lo que, la lesión previa deberá ser considerada en el diseño de los estudios realizando los ajustes oportunos en el análisis estadístico, pero no pensamos que sea adecuado descartar a participantes por esta causa. Aparte de lo mencionado anteriormente será preciso considerar otros factores que pueden distorsionar el efecto de un FR sobre la incidencia de las lesiones de tobillo, como son el uso de vendajes funcionales, ortosis o programas de prevención de lesiones que han demostrado reducen el riesgo de tobillo y del miembro inferior<sup>32;40;47;176-180</sup>.

Tras la revisión sistemática y el meta-análisis las pruebas científicas muestran que solo podemos afirmar con garantías que la altura, el peso y el IMC no son FR de riesgo, ya que la forma de evaluar estos FR es precisa, los estudios primarios han determinado que no predisponen a los atletas a sufrir lesión de tobillo, y el MA realizado en esta RS lo corrobora. En cuanto a la edad, consideramos que también puede descartarse ya que los estudios primarios de manera individual tampoco han encontrado que esté relacionado con un mayor riesgo de lesión. Así mismo, la medición de este FR resulta fiable, y a pesar de que no se ha podido realizar MA, atendiendo a la lógica no consideramos que pueda ser FR. A pesar de ello, es preciso seguir evaluando estas variables en los estudios longitudinales, ya que describen bien a la muestra y deberán ser tenidos en cuenta en los análisis sobre los FR.

Respecto a la historia de lesión previa, a pesar de que el MA muestra que el haber sufrido una lesión multiplica el riesgo de sufrir otra lesión de tobillo por 1,55, es preciso interpretar estos resultados con cautela por varias razones. En primer lugar, si analizamos los estudios de manera individual observamos que tres trabajos<sup>40;168</sup> no encuentran que el haber sufrido una lesión de tobillo predisponga a las atletas a sufrir un nuevo esguince, y uno sí<sup>170</sup>. Del mismo modo, el trabajo de Mei-Dan et al. (2004)<sup>169</sup> que no pudo ser incluido en el MA tampoco encuentra relación. Por otro lado, al observar el MA vemos que el trabajo de Steffen et al. (2007)<sup>170</sup> que sí muestra asociación entre la historia de lesión de tobillo y una nueva lesión, utiliza una muestra notablemente mayor que el resto de trabajos incluidos en el MA, por lo que su peso es muy grande en el MA. Asimismo Steffen et al. (2007)<sup>170</sup> emplean en el análisis ambas piernas, la lesionada y la control, por lo que sus resultados pueden estar sesgados<sup>181</sup>. También debemos de tener en cuenta que este trabajo muestra una pérdida diferencial de participantes importante, tal y como muestra el análisis de calidad (tabla 5.1.1).

De la misma manera, el riesgo relativo (RR) combinado no es muy elevado (figura 5.1.2), lo que tampoco nos permite determinar con garantías si finalmente este es

realmente un rasgo predictivo de lesión de tobillo o no. Además, el tipo de lesión registrada (*outcome*) difiere entre los estudios meta-analizados entre lesión previa esguince<sup>40;168</sup> y lesión de tobillo<sup>170</sup>, lo que muestra poca consistencia. Respecto al propio FR también hay discrepancias, algunos emplean esguinces de tobillo<sup>168</sup> y otros lesión de tobillo<sup>40;170</sup>. Por todo ello, no podemos confirmar si una lesión previa discrimina a las participantes con riesgo de lesión de las que no, por lo tanto éste FR deberá ser estudiado de manera rigurosa en futuras investigaciones.

El resto de 99 posibles factores de riesgo evaluados en los estudios incluidos en la RS, no los podemos ni confirmar y descartar como FR, ya que sólo han sido examinados en uno o dos estudios lo cual no es suficiente para realizar un juicio definitivo al respecto. A pesar de que 15 FR han sido replicados, encontramos falta de consistencia tanto en los resultados como en los métodos de evaluación empleados para evaluar una misma variable.

En cuanto a la controversia en los resultados en esta RS observamos tres casos. El primero de ellos fue en la variable del grupo de FR de propiocepción (propiocepción del tobillo a 15° en inversión cuando se proporcionan resultados absolutos). Esta variable se encuentra asociada al riesgo de lesión en el estudio de Payne et al. (1997)<sup>52</sup>, y no así en el de Willems (2004)<sup>18</sup>. En este caso, no pudimos realizar MA ya que Payne et al. (1997)<sup>52</sup> no proporcionaron resultados de manera analizable, por lo que dicho FR no puede ser descartado. En cuanto al RDM en EV del calcáneo sin apoyo, esta variable discriminó a población susceptible de lesionarse en el trabajo de Beynnon et al. (2001)<sup>55</sup>, pero no en el de Willems et al. (2001)<sup>18</sup>. Cuando realizamos el MA, encontramos que la diferencia estandarizada de medias (DEM) muestra que no existe asociación. Pero debemos ser prudentes a la hora interpretar el MA, ya que el estudio que no encuentra asociación tiene más peso en el análisis y además la heterogeneidad es la más alta de todas las observadas en los MA sobre FR. En cuanto a la fuerza, el MA muestra que la fuerza concéntrica del tobillo en dorsiflexión a 30°/s está moderadamente asociada al riesgo de lesión, pero los estudios de manera individual no encuentran tal asociación en el FR y la lesión.

Respecto a la inconsistencia de los métodos de evaluación, encontramos que en las seis variables sobre fuerza específica de tobillo que fueron incluidas en el MA, los dos estudios que las emplearon<sup>18;55</sup> muestran diferencias. En los movimientos de inversión (IN) y eversión (EV) ambos estudios colocan a las participantes decúbito supino, pero Beynnon et al. (2001)<sup>55</sup> flexiona las rodillas a 90°, mientras que Willems (2004)<sup>18</sup> realiza la prueba con a piernas extendidas y el respaldo reclinado a 45° respecto a la vertical. En los movimientos de dorsiflexión y flexión plantar, Beynnon et al. (2001)<sup>55</sup> se decanta por colocar a las participantes decúbito prono, y Willems (2004)<sup>18</sup> decúbito supino. En cuanto al estimador, ambos miden

el *peak torque*, pero discrepan en varios aspectos al respecto. Beynnon et al. (2001)<sup>55</sup> emplea la media de cinco intentos mientras que Willems (2004)<sup>18</sup> no lo define. Incluso esta autora, mide el torque a 5º grados de la IN o EV máxima y dan los resultados relativos al peso de las participantes. Sin embargo, Beynnon et al. (2001) no describe nada de esto, y ni siquiera describe las unidades. En cuanto a la variable de propiocepción replicada (propiocepción del tobillo a 15º en inversión cuando se proporcionan resultados absolutos) Willems (2004)<sup>18</sup> sienta a las participantes con la cadera a rodillas a 90º, con una flexión plantar de 15º. Sin embargo, Payne et al. (1997)<sup>52</sup> mantiene la cadera y las rodillas de las participantes flexionadas a 100º. Estas diferencias en los métodos de evaluación para evaluar una misma variable, es decir un mismo FR, son determinantes ya que no podemos saber si la falta de asociación, o la asociación encontrada es debida a la diferencia de métodos evaluación, o al FR en sí.

Cuando observamos los métodos que se emplean para evaluar los FR, estos resultan en su mayoría complejos e inverosímiles, lo que hace altamente complicada su réplica y estandarización. Encontramos, que por ejemplo, la mayoría de las variables de fuerza específica de tobillo se miden a una determinada velocidad medida en grados por segundo, en una posición del cuerpo con las articulaciones a un determinado grado de flexión empleando dinamómetros, goniómetros y personal cualificado para su evaluación. En consecuencia, estos ME son muy difíciles de reproducir entre estudios, además de requerir material específico y muchas veces caro, como son los dinamómetros, goniómetro y las máquinas específicas como *Neurocom Balance Master*<sup>18</sup> y *Neoretest System Neurocom*<sup>55</sup>. Así su uso en la investigación dentro de esta área esta tremendamente limitado, lo que se refleja cuando observamos los todos FR estudiados en este RS (anexos 5.1.3 y 5.1.4). Cabe destacar que en total se exploran al menos 103 FR diferentes y únicamente 15 fueron replicados, de los cuales cinco fueron la edad, peso, altura, IMC e historia de lesiones. Estos cinco FR son sencillos de evaluar, al no requerir material ni personal especializado para su evaluación, lo hace que su uso sea plausible.

Además de la complejidad para replicar los métodos de evaluación utilizados en general, los métodos empleados para valorar los FR presentan un precisión cuestionable. A pesar de que muchos de los trabajos incluidos en la RS citan a otros trabajos para demostrar la fiabilidad del método de evaluación que van a emplear o realizan su propio estudio previo de fiabilidad<sup>18</sup>, la fiabilidad de los mismos es cuestionable. Cuando analizamos los estudios citados que pretende demostrar la precisión de la medición, observamos que en todos los casos se ha empleado el CCI u otra medida no dimensional para determinar su fiabilidad.

Respecto al grupo de FR sobre rango de movimiento, alineación y flexibilidad, los tres estudios<sup>18;52;55</sup> que evalúan algún FR de este grupo, emplean en el goniómetro como herramienta de evaluación, la cual tiene una dudosa fiabilidad tal y como se describe en la introducción de esta tesis. Los dos trabajos que citan estudios de fiabilidad sobre los métodos utilizados, que emplean métodos adimensionales, Willems cita 2 estudios<sup>109;116</sup> y Beynnon otros tres<sup>109;182;183</sup>. En cuanto al RDM la primera metarsofalange del dedo del pie, que parece ser FR para Willems (2004)<sup>18</sup>, estiman ellos mismos la fiabilidad proporcionando únicamente el CCI.

En cuanto las variables fuerza del tobillo, Willems (2004)<sup>18</sup> cita el trabajo Karnofel et al. (1989)<sup>184</sup> que emplea la *r* de Pearson para comprobar la fiabilidad de todas las variables que comprueba para demostrar la precisión de la prueba. Por otro lado, Beynnon et al. (2001)<sup>55</sup> y Payne et al. (1997)<sup>52</sup> no comprueban la fiabilidad ni citan a otros autores que lo hagan, por lo que no conocemos la precisión de sus métodos evaluación en la variables sobre fuerza del tobillo. En cuanto al equilibrio, tampoco hemos podido comprobar con garantías que los métodos de evaluación empleados para evaluar las diferentes variables sean precisos.

Cuando los trabajos sobre FR citan a otros estudios de fiabilidad, no se toman en cuenta las especificidades del tipo de investigación como son los FR de riesgo de las lesiones de tobillo en mujeres deportistas. Ejemplo de ello son los FR sobre la laxitud y las cualidades físicas. Estas medidas son métodos comúnmente empleados, y tomados como "válidos". Pero desafortunadamente, no sabemos qué error de medición proporcionan, y si esta precisión es suficiente como para que las podamos considerar en estudios que pretenden discriminar a población con y sin riesgo de lesión. En el ámbito de investigación, los trabajos de fiabilidad presentan una calidad en general escasa<sup>147</sup>, como observamos en las revisiones sobre *SEBT* y las *PDFA*. La causa de esto, es que no se la da importancia a la precisión de mediciones cuando este hecho va determinar el tamaño del efecto que tiene método de evaluación para discriminar participantes debido a un rasgo. Es por ello que no podemos determinar si la falta de asociación es a la falta de precisión o a la falta de predictibilidad de ese FR<sup>185</sup>.

Aunque la fiabilidad de los métodos fuera adecuada y estuvieran estandarizados, lo cual facilitaría su réplica, éstas son en general muy complejas (anexo 5.1.3). Las variables evaluadas requieren material específico como son los goniómetros, dinamómetros, aparatos para medir la fuerza isocinética o maquinas específicas que muchas veces además resulta caro. Además, estos aparatos precisan personal formado. Todo ello hace muy difícil su uso en estudios longitudinales con intención de estudiar los FR de las mujeres. Es más, aunque dispusiéramos de recursos para realizar dichas investigaciones, su uso en la realidad deportiva femenina estaría muy limitado.

En conclusión, a pesar de que las lesiones de tobillo son muy frecuentes entre mujeres deportistas, únicamente 10 trabajos han estudiado los FR en esta población. En estos trabajos se han examinado 103 rasgos diferentes que pueden llevar a las deportistas a sufrirla. Los resultados de la revisión muestran que las características de las participantes como la edad, altura, IMC y peso no son FR, y que la historia de previa de lesiones en el tobillo deberá ser confirmada en futuras investigaciones. En lo que respecta al resto de variables analizadas, no podemos descartar ni confirmar ninguna como FR, ya que presentan una dudosa precisión y encontramos resultados contradictorios cuando se han podido evaluar en más de un trabajo.

## **6.2-STAR EXCURSION BALANCE TEST (SEBT)**

El *Star Excursion Balance Test (SEBT)* es una prueba que según la literatura consultada, es un método de evaluación del equilibrio<sup>65;66</sup> y se propone como alternativa a aparatos más sofisticados y caros. La revisión sistemática (RS) realizada en esta tesis ha identificado dos variables que miden algún aspecto del equilibrio y cinco de propiocepción, que es una cualidad ligada al equilibrio<sup>186</sup>, las cuales podían estar asociadas a un mayor riesgo de lesión de tobillo en mujeres.

En lo que respecta a los resultados de los estudios incluidos en la RS, todos los CCI proporcionados muestran una fiabilidad intra-observadores (FIA) de buena a muy buena<sup>148</sup> independientemente de la línea, estimador o de si los valores son absolutos o relativos a la longitud de la pierna (normalizados). Asimismo, 19 de 35 de los CCI analizados en las líneas estudias (A, PM y PL) obtienen índices por encima de 0,90, lo que significa que la fiabilidad es muy buena (tabla 1.1.1). Los CCI combinados (meta-análisis) demuestran una fiabilidad buena o muy buena en todas las líneas siendo la  $I^2$  adecuada. Nuestro estudio de fiabilidad sobre el *SEBT* obtiene unos CCI mejores que los logrados por los estudios previos así como todos combinados en el meta-análisis (MA). Con lo que si nos basamos exclusivamente en el CCI podríamos concluir que el *SEBT* tiene una FIA muy buena.

Sin embargo, si observamos los resultados dimensionales de los estudios previos (5.2.5-5.2.8), encontramos un error aleatorio muy elevado como para considerar la prueba fiable con errores de medición de 4,71 cm a 11,06 cm. El error del estudio de campo de esta tesis es más reducido, ya que encontramos que los límites de acuerdo (LDA) van de 3,88 cm a 5,45 cm, dependiendo de la línea y el estimador empleado. No obstante, y a pesar de que el *SEBT* no tiene error sistemático ni la diferencia entre días (mediciones) tiende a aumentar o disminuir en función de la magnitud de la medida, consideramos que la prueba no es fiable. Esto es debido a que sí que existe un importante error aleatorio en la prueba, lo que quiere decir que entra dentro de lo esperable que una participante obtenga una diferencia

entre mediciones dentro del error de medición. Por ejemplo, en la línea A empleando la media como estimador, si una participante logrará 61,3 cm tendríamos que asumir que ese valor es lo mismo que obtener cualquier otro entre 66,42 cm y 56,18 cm, ya que el LDA es de 5,11 cm. Lo que este LDA denota es que cualquier valor menor de 5,11 cm significa que no existe cambio en las participantes o entre participantes debido a una intervención, una lesión, patología o el estado de salud entre otras. Por lo tanto sólo habrá que considerar que hay un cambio en la prueba que puede indicar una patología o una mejora de las marcas fuera de este rango, lo cual nos parece un margen de error no asumible desde el punto de vista práctico.

Observando otros trabajos que ha estudiado la fiabilidad del *SEBT* y no incluidos en la revisión cabe decir que la fiabilidad es peor que la encontrada por el estudio de campo de esta tesis. El estudio realizado por Plisky et al. (2009)<sup>70</sup> que empleó una muestra de hombres deportistas y un aparato diseñado específicamente para realizar el *SEBT* (figura 1.5.2), encontramos CCI más bajos de 0,89 a 0,85 que este estudio, y MDD más elevados que en el presente trabajo (de 5,57 cm a 8,62 cm). Demura y Yamada (2010)<sup>187</sup> encontraron el *SEBT* fiable para hombres sanos con un CCI de 0,92 a 0,99 y Stocket y Barakatt (2005)<sup>188</sup> también determinaron la prueba fiables en población geriátrica de ambos sexos, con CCI de 0,92 a 0,95. Estos valores adimensionales son similares a los estudios primarios, al meta-análisis y a los aportados por el estudio de fiabilidad del *SEBT* de esta tesis (tabla 5.2.3).

Las pruebas científicas encontradas en otros trabajos que han tratado de investigar la validez del *SEBT*, muestran resultados contradictorios, confirmando que la precisión del *SEBT* no es suficiente como emplear la prueba como herramienta de diagnóstico. Concretamente, en el ámbito de lesiones deportivas Plisky et al. (2006) emplearon el *SEBT* en un estudio de cohortes para valorar el riesgo de lesión del miembro inferior. Concretamente en mujeres, el trabajo determinó que una diferencia mayor a 4 cm entre piernas suponía un riesgo 2,5 veces mayor de sufrir una lesión del miembro inferior. Atendiendo a los resultados dimensionales proporcionados por los estudios primarios y los logrados en el trabajo de fiabilidad de esta tesis, encontramos una conclusión paradójica: la mayoría de las participantes tendrían 2,5 veces más de riesgo de lesionarse sobre su misma pierna. La razón para esta contradicción se puede deber a que los resultados de los estudio de fiabilidad no sean aplicables a la población del estudio pronóstico de Plisky et al. (2006)<sup>58</sup>. No obstante, si observamos las características de la muestra de Plisky et al. (2006)<sup>58</sup> con la nuestra, encontramos que la población es muy parecida. Al realizar una lectura crítica del este estudio observamos que estos límites fueron establecidos de manera arbitraria, sin haber

empleado una metodología adecuada para determinar los puntos de corte que mejor discriminan una población sana de otra lesionada. Asimismo, Munro et al. (2010)<sup>72</sup> también determinaron que 4 cm entre líneas también era una diferencia muy pequeña para el error de medición que se observó en su estudio, confirmando lo que se defiende en esta tesis.

En el ámbito de las lesiones de tobillo, tratando de diferenciar pacientes con inestabilidad crónica de tobillo de los que no, Olmsted et al. (2002)<sup>76</sup>, Akbari et al. (2005)<sup>189</sup>, Gribble et al. (2004)<sup>85</sup> y Hertel et al. (2006)<sup>79</sup> encontraron diferencias significativas entre el tobillo lesionado y el sano, o entre participantes sanos y con inestabilidad. Sin embargo, Sefton et al. (2009)<sup>190</sup>, Martínez-Ramírez et al. (2010)<sup>75</sup> y Nakagawa y Hoffman (2004)<sup>91</sup> no diferenciaron el grupo con inestabilidad del grupo control empleando la distancia obtenida en *SEBT*.

En cuanto a las patologías de rodilla, Herrington et al. (2009)<sup>88</sup> encontraron diferencias significativas entre los participantes con rotura del ligamento cruzado anterior y control, al igual que Aminaka y Gribble (2008)<sup>74</sup> con participantes con síndrome de dolor patelofemoral. En cuanto a las intervenciones, Leavey et al.<sup>92</sup> no encontraron diferencias significativas en las marcas del *SEBT* entre el grupo control y el experimental después de un entrenamiento de fuerza y propiocepción de seis semanas.

Una RS reciente sobre el *SEBT*, de Gribble et al. (2012)<sup>66</sup>, muestran que la prueba sigue sin estandarizarse, y más aún, se sigue sosteniendo que la prueba es perfectamente fiable y válida como herramienta de diagnóstico a pesar de los valores interpretables que la literatura previa proporciona. Esta RS que incluye población mixta, calcula la *d-Cohen* para determinar el tamaño del efecto del *SEBT* como herramienta diagnóstica para diferenciar lesiones en el miembro inferior, intervenciones y en la predicción de lesiones, así mismo comprueba el efecto que el uso de ortosis y vendajes pueden tener sobre el *SEBT*. En cuanto a la inestabilidad crónica de tobillo, los 8 artículos incluidos que proporcionan 23 comparaciones entre grupos, muestran una estimación puntual moderada y todos los intervalos de confianza cruzan el 0, lo que determina que el *SEBT* no tiene la capacidad de diferenciar población con inestabilidad crónica de tobillo<sup>66</sup> de la que no, ni el tobillo sano del que padece inestabilidad crónica de tobillo. Sin embargo, diferencia muy bien a población con ligamento cruzado anterior roto o reconstruido de la que no, así como participantes con dolor patelofemoral de la sana.

Consideramos que las causas para esta falta de precisión de la prueba, se deben fundamentalmente a la variabilidad propia de la persona que realiza la prueba. A pesar de que en el presente estudio hemos tratado de controlar las fuentes de variabilidad al máximo realizando una descripción exhaustiva de cómo realizar la

prueba, la variabilidad intrínseca de los participantes resulta difícil de controlar siendo incluso a veces imposible<sup>146</sup>. Pero el *SEBT* es una prueba muy compleja y los movimientos que se requieren durante el *SEBT* son multiarticulares e implican a todo el cuerpo, por lo que son muchas las posibles causas que pueden producir tal variabilidad en la participante de un momento de medición a otro. Para que la prueba sea consistente, es necesario que ningún elemento que se emplea para realizar el control postural esté alterado de un momento a otro<sup>90</sup>. Además los participantes pueden utilizar diferentes técnicas para conseguir la máxima distancia en cada intento<sup>67</sup>, lo que puede producir cierta variabilidad las mediciones comparadas.

En cuanto a la fiabilidad inter-observadores (FIE) del *SEBT*, consideramos que al ser los límites de acuerdo tan elevados en la FIA, la FIE la será peor. Los resultados sobre fiabilidad inter-observadores que hemos observado en los dos trabajos que estudiado esta fiabilidad en la prueba<sup>65;68</sup>, son poco apropiados ya que comparan mediciones realizadas el mismo día. De esta manera el efecto de aprendizaje y efecto de memoria afectaran a la fiabilidad de la *SEBT*, además de no reflejar la realidad de aplicación de la prueba.

Respecto a los valores normalizados, estamos de acuerdo en que será necesario calcular la distancia en relación a la longitud de la pierna para que la prueba sea comparable entre participantes, ya que la longitud de la pierna es determinante con la distancia obtenida<sup>77</sup>. Pero, cabe decir que estos dan una visión distorsionada de la fiabilidad de la prueba ya que este valor es derivado de dos mediciones como son la longitud de la pierna y la distancia obtenida. Cada medición tendrá un error asociado a su propia medición, si se combinan dichas mediciones, no podremos saber cuál es el error real del *SEBT* y cuál es el error de la medición de la longitud de la pierna. Al no haber demostrado el *SEBT* suficiente precisión en valores absolutos, no merece hacer mención a lo encontrado en los valores normalizados.

En conclusión, desde el punto de vista práctico, y basándonos en lo observado en la RS y el estudio de campo realizado en esta tesis, el *SEBT* no es una herramienta precisa en mujeres deportistas y consideramos que tampoco lo es otro tipo de población. Si bien es cierto que otros estudios analizados en la RS sobre el *SEBT* (tabla 5.2.3) y otros que han comprobado la fiabilidad<sup>70;187;188</sup> han determinado que la prueba es fiable, todos los estudios han basado sus conclusiones en el CCI determinando que la prueba era precisa. A pesar de que algunos estudios han empleado métodos interpretables desde el punto de vista práctico (EEM, MDD, PDC o MBA) y han encontrado errores de mediciones más elevados que en presente trabajo, todos se han basado en el CCI. Es más, consideramos que el modelo ha podido ser escogido de manera arbitraria, ya que sólo el 50% lo define, y entre los que los definen existen discrepancias frente a un aparente mismo

diseño (anexos 5.2.4 y 5.2.5). Por lo tanto, pensamos que las diferencias entre trabajos son fruto de la escasa validez interna de los trabajos de fiabilidad sobre el *SEBT*, de la diferencia de análisis estadístico empleando en dichos trabajos y de la interpretación de los mismos. Es por ello que no recomendamos su uso como método para evaluar el riesgo de lesión en mujeres deportistas ni tampoco para emplear esta herramienta para el diagnóstico en otro tipo de población y ámbitos de aplicación.

### **6.3-PRUEBAS DE DORSIFLEXIÓN EN APOYO**

La revisión sistemática sobre las pruebas de dorsiflexión en apoyo (PDFA), muestra que existe gran variedad de procedimientos para realizar estas pruebas, y que además se emplean diferentes métodos para evaluarlas. Estos métodos proporcionan mediciones tanto en cm (DD, DX) como en grados (IA, IT, CR, GO, IL). Para realizar la evaluación se emplean herramientas de medición como cinta métricas, goniómetros, inclinómetros o aparatos específicos y más sofisticados. A pesar de ello, el método de evaluación más empleado es utilizar el inclinómetro ubicando éste en la tibia (IT).

En cuanto a los resultados sobre fiabilidad de los métodos de evaluación, al igual que en el *SEBT*, realizamos varios meta-análisis. En cuanto a estos resultados (figuras 5.3.3-5.3.6) es preciso analizarlos con cautela ya que en la mayoría la consistencia del CCI combinado ha sido débil. Era de esperar que la heterogeneidad de los MA realizados fuera elevada, ya que como hemos observado en los análisis de las características de los estudios existe una alta variabilidad entre trabajos, así como un alto riesgo de sesgo. Según la Colaboración Cochrane<sup>191</sup>, las fuentes de un alta heterogeneidad en los resultados puede ser debidas o bien a la variabilidad entre los participantes, métodos de evaluación (intervenciones) y *outcomes*, así como una importante diversidad metodológica; diferentes diseños de estudio y riesgo de sesgo, o a una combinación de ambas. En nuestra opinión, los estudios incluidos en la RS incurren en todas las razones para una  $I^2$  elevada.

Por todo ello ha sido laborioso encontrar las causas de la inconsistencia del CCI de manera coherente y sin caer en la arbitrariedad, ya que la descripción de los estudios es parca (anexos 5.3.4-5.3.5). Conscientes de esta diversidad realizamos un los meta-análisis empleando un modelo de efectos aleatorios, y analizamos lo métodos de evaluación en cm y en grados de manera separada en primer lugar. Si bien es cierto que algunos CCI combinados presentan una consistencia adecuada, es conveniente tener en cuenta que en algunos casos ha sido necesario extraer resultados de algunos trabajos incluidos en la revisión.

### 6.3.1-FIABILIDAD INTRA-OBSERVADORES DE LA PRUEBA DE DORSIFLEXIÓN EN APOYO CON LA RODILLA FLEXIONADA (PDFAF)

En lo referente a la fiabilidad, los MA (figura 5.3.2) muestran que la repetibilidad es mejor en los métodos que utilizan los grados como unidad de evaluación que aquellos que emplean los cm. Sin embargo, nuestro estudio de campo muestra una mejor fiabilidad de los métodos que emplean cm como unidad de evaluación (tabla 5.5.3). Las causas para esta diferencias se deben a que únicamente dos estudios han comprobado la FIA con unidades en cm, el de Bennell et al. (1998)<sup>121</sup> y el de Jones et al. (2005)<sup>135</sup>. Este último<sup>135</sup>, que emplea un aparato específico (figura 1.5.9) para valorar la PDFAf obtiene unos CCI notablemente más bajos (0,66 a 0,68), que Bennell et al. (1998) y que el resto de trabajos tanto de PDFAF como de PDFAe. Además, el peso en el MA de este estudio es muy alto (anexo 5.3.6), al ser su tamaño muestral elevado y proporcionar resultados de los tres observadores empleados. La heterogeneidad de los CCI combinados incluyendo el estudio de Jones et al. (2005)<sup>135</sup> es muy elevada, y cuando extraemos este trabajo del análisis, los CCI combinados presentan una consistencia adecuada. Una vez eliminados los resultados de Jones et al. (2005)<sup>135</sup>, la fiabilidad los métodos en cm es mejor que los métodos que emplean grados, coincidiendo con este trabajo.

Respecto a la literatura previa, los pocos estudios que como esta tesis han comparado varios métodos de evaluación simultáneamente, también coinciden en que la medición de la DF en cm es más fiables. Bennell et al. (1998)<sup>121</sup> que empleo el inclinómetro (ITf) y la distancia desde el primer dedo del pie hasta la pared (DDf), encontró una mejor fiabilidad del método de evaluación que empleo cm. En un estudio reciente<sup>192</sup>, no incluido en la RS, donde se compararon los métodos de evaluación ITf, DDf además del goniómetro, también se encontró la medición realizada en cm más fiable.

No obstante, el método que se emplea para valorar la DF en cm (distancia desde el primer dedo del pie hasta la pared) tienen el inconveniente que no pueden comparar participantes<sup>121</sup>, ya que a una misma dorsiflexión la distancia con un pie mayor será menor. Es por ello que para la predicción de lesiones no es un método adecuado, porque no permite comparar sujetos. Como alternativa a éste método en esta tesis planteamos el uso del ángulo derivado por razones trigonométricas (ADf) de las distancias horizontal (DHf) y vertical (DVf).

La ADf muestra un CCI similar a otros métodos planteados por la literatura previa y los observado en los MA independientemente de utilizar cm o grados como unidad de medición. Cuando los estudios reportan índices interpretables de desde un punto de vista práctico, la ADf en general muestra un error del método menor. Los

LDA proporcionados por Barton et al. (2005) empleando el IT van 4,45° a 6,35°, siendo los de esta tesis 3,01°. Respecto a la MDD, los valores reportados por Bennell et al. (1998) son similares (2,77° a 3,05°) al presente trabajo. Además en un reciente estudio donde se compara la precisión de la PDFAf medida mediante inclinómetro y goniómetro se encontraron MDD más elevados que la ADf propuesta por este trabajo<sup>192</sup> con errores del método de 3,7° empleando el ITf y de 5° a 7,7° con el goniómetro. Asimismo, la ADf dentro de nuestro estudio de campo es el método más preciso en todos los índices de fiabilidad calculados.

En consecuencia, por lo observado en la RS y en el estudio de campo, el ángulo derivado de las distancias por razones trigonométricas (ADf), ha demostrado mayor precisión que cualquier método de evaluación propuesto por la literatura científica existente. Asimismo, esta manera de evaluar la dorsiflexión en la PDFAf es muy sencilla, no requiere material específico como puede ser el inclinómetro y el goniómetro, y puede ser empleado por personal sin formación siguiendo un sencillo pero exhaustivo procedimiento para su realización.

### **6.3.2-FIABILIDAD INTRA-OBSERVADORES DE LA PRUEBA DE DORSIFLEXIÓN EN APOYO CON LA RODILLA EN EXTENSIÓN (PDFAE)**

En cuanto a la prueba de dorsiflexión en apoyo con la rodilla en extensión (PDFAE), observamos que todos los estudios han empleado medidas en grados empleando diferentes métodos de evaluación que requieren diferentes instrumentos. El inclinómetro ubicado en la tibia (ITe), es el método más extendido para valorar la PDFAE en los estudios incluidos en esta revisión.

El MA, que también incluye otros métodos de evaluación como el cristal con los ángulos marcados (figura 1.5.8), el goniómetro (GO) y el inclinómetro ubicado en la parte lateral de la tibia (IL), presenta un CCI combinado adecuado. En nuestro trabajo de campo, el método de evaluación ITe presenta unos CCI mucho menores que la literatura previa que ha empleado este mismo método, así como en comparación con los otros métodos de evaluación y el MA. Todas las medidas dimensionales (MDD y MBA) también demuestran peores resultados de nuestro estudio en comparación con la literatura previa. Sin embargo, el método de inclinómetro en el tendón de Aquiles (IAe), presenta una fiabilidad similar a la literatura previa en CCI y en medidas dimensionales. Esta diferencia puede ser debida a las variaciones que se pueden dar en el lugar de aplicación del inclinómetro durante la PDFAE. El IAe se aplicó a 7,5 cm de la parte más distal del calcáneo, muy cerca del eje de rotación del tobillo. La ITe, sin embargo de aplico mucho más alejado del tobillo, pudiendo variar la ubicación de la tibia de un momento de evaluación a otro, ya que en PDFAE1 no controlamos que la vertical

del borde supero-externo de la rótula cayera sobre la línea del suelo. Esto no afecta tanto a la fiabilidad de la IAe al ser su ubicación al eje de rotación, y las variaciones a este nivel ser menores. Por ello consideramos que si en PDFA1 se hubiera controlado la ubicación de la rodilla la fiabilidad del ITe e IAe serían similares.

La precisión del inclinómetro ubicado en el tendón de Aquiles (IAe) y el ángulo derivado de las distancias (ADe) muestra un CCI similares, sin embargo la ADe es 1º más precisa que la IAe observando los límites de acuerdo (tablas 5.5.4 y 5.5.5). Por lo tanto, la ADe es una medida más precisa que la IAe, además de ser más sencilla. En comparación con los métodos propuestos por la literatura previa el CCI es ligeramente menor pero cuando observamos las medidas interpretables desde el punto de vista práctico, la precisión de la ADe es mejor.

En los trabajos que presentan resultados de fiabilidad empleando el método propuesto por Bland y Altman<sup>130;171</sup> (figura 5.3.10) y que emplean el inclinómetro en la tibia (ITe) y el cristal con los ángulos marcados (CR), muestran unos límites de acuerdo (LDA) similares. Encontramos valores en general similares a nuestro trabajo (6,72º) de campo con un rango de 3,80º a 8,35º, siendo los límites de acuerdo en el 80% de los datos presentados, superior a 6º. Respecto a la mínima diferencia detectable (MDD) Krause et al. (2011) presenta valores 0,32º y 0,62º inferiores a los logrados por este estudio, lo cual nos parece una diferencia inapreciable. No obstante, existe una tendencia importante a medir más el primer día de evaluación que el segundo en los dos trabajos que realizan MBA, Barton et al. (2005)<sup>171</sup> y Muntenau et al. (2009)<sup>130</sup>. Esto es debido a que el error sistemático se presenta en 10 de los 11 resultados proporcionados por los estudios en negativo. Es más, el 63,63 % de los sesgo entre métodos presentados por Barton et al. (2005)<sup>171</sup> y Muntenau et al. (2009)<sup>130</sup> son mayores a 1º, lo cual desde el punto de vista práctico es una cuantía elevada. Sin embargo, los resultados de nuestro estudio muestran un error sistemático despreciable (-0,34º).

Por lo tanto, las pruebas científicas encontradas tanto en la RS como en el trabajo de campo, muestran que la ADe es un método de evaluación más fiable que los empleados por los estudios previos, al obtener un error aleatorio similar o menor que los estudios previos, pero ser carente de error sistemático. Asimismo, todas las variables mediante las que se obtiene el ADe, a saber; DTe, DRe, DHe y DVe, también han demostrado individualmente una fiabilidad adecuada en todos los análisis estadísticos sobre fiabilidad realizados. Además este método es también aplicable al PDFAf, siendo en esta prueba, indudablemente, este método el más fiable de todos los estudiados por la literatura previa, con lo que podemos evaluar ambas pruebas con métodos de evaluación análogos y precisos.

### 6.3.3.-FIABILIDAD INTER-OBSERVADORES DE LAS PRUEBA DE DORSIFLEXIÓN EN APOYO

Respecto a la fiabilidad inter-observadores en ambas pruebas, esta tesis la han estudiado empleando los métodos de revisión sistemática (RS) y meta-análisis (MA). Atendiendo a los resultados obtenidos, el CCI combinado es adecuado en ambas pruebas. Sin embargo, la alta heterogeneidad encontrada en los MA hace que sea necesario interpretar los resultados con cautela. Asimismo al comparar los MA de la FIA y la FIE, y los resultados de los estudios primarios que han comprobado ambas fiabilidades nos encontramos con resultados contradictorios.

Si observamos los MA realizados y analizamos los estudios primarios que han estudiado ambas fiabilidades<sup>121;125;126;130;135;171</sup> encontramos que algunos trabajo presentan una FIE mayor que la FIA o que las diferencias entre ambas fiabilidades son inapreciables, tanto con medidas dimensionales como adimensionales. Estos resultados son paradójicos, ya que lo lógico, es que la FIA sea siempre mejor que la FIE cuando las condiciones en la cuales se realizan las evaluaciones a comparar en el análisis de fiabilidad son idénticas. La causa lógica para esto, es que en la FIE se suman los errores de medición de las evaluaciones de cada observador que comparamos. Esto es, los errores de evaluación encontradas en la FIA por cada observador confluyen, con lo que el CCI debería ser mayor en FIA que en FIE, y los errores de medición menores. Sin embargo, en los trabajos incluidos en esta RS, las circunstancias donde se evalúan tanto la FIA como la FIE no son similares, lo que hace que nos encontremos en muchas ocasiones una mejor FIE que FIA, y hace que al mismo tiempo ambas fiabilidades no sean comparables.

En primer lugar, debemos de tener en cuenta que todos los trabajos que ha evaluado la FIA y la FIE<sup>121;125;126;130;135;171</sup>, la FIE siempre se ha evaluado comparando las evaluaciones de los diferentes observadores dentro de una misma sesión (anexos 5.3.4 y 5.3.5), es decir, dentro de un mismo día. Rabin y Kozol (2010)<sup>173</sup> que estudiaron exclusivamente la FIE, empleó una única sesión de evaluación. Sin embargo la FIA se evalúa en diferentes sesiones, con una separación de mínima de 7 días y una máxima de 21 días. Si bien cierto que para algunas pruebas, es indiferente que la prueba se haga seguida de otra, no ocurre lo mismo cuando realizamos pruebas con algún componente físico como la PDFA. En este caso en concreto, se pueden dar efectos de aprendizaje y de memoria, además de una modificación de la cualidad evaluada, como efecto del calentamiento.

En cuanto al efecto de memoria, al realizar la prueba en FIE seguida de la medición anterior, puede producir que los participantes inconscientemente reproduzcan siempre el mismo grado de flexión en todas las evaluaciones. Sin

embargo, en la FIA al compararse las evaluaciones distanciando estas con días, anula o al menos minimiza este efecto de memoria.

Respecto al efecto de aprendizaje, éste también puede afectar a la PDFA, ya que los estudios que analizado su precisión mediante el MBA<sup>130;135;171</sup> han encontrado en la FIA una importante tendencia a medir más en segunda sesión (figuras 5.3.6, 5.3.7 y 5.3.10). El error sistemático (ES) se muestra en negativo en 76,47% de la ocasiones (de un total de 17), siendo entre ellos el ES mayor a 1º en 53,85%. Por ello pensamos que los participantes de los estudios primarios pueden haber aprendido a hacer la prueba, ya que logran más DF el segundo día. A pesar de que en nuestro estudio no hemos detectado un efecto de aprendizaje (el ES no se muestra tendencia a mostrarse en negativo), este efecto puede ser el causante de haber afectado a los estudios haciendo que FIE sea mejor que la FIA. Las razones pueden ser debidas a la diferencias de método, u otras causas, pero no podemos concluir con certeza ya que únicamente dos estudios han realizado el MBA y la calidad de los mismos ha resultado escasa.

Por otro lado, se ha comprobado que la DF aumenta significativamente con de 5 a 30 minutos de estiramiento<sup>193</sup>. Por lo que si realizamos la prueba las PDFA seguidas unas de las otras, tal y como se comprueba la FIE en los estudios primarios incluidos en la RS, es de esperar que los participantes consigan una mayor DF con el segundo observador y los sucesivos. Barton et al. (2010)<sup>171</sup> en PDFAf, en la comparación de los observadores A-C y B-C presentan un error sistemático muy elevado:  $-2,35^\circ$  y  $-2,45^\circ$ . En PDFAe, los mismos observadores comparados, presentan un ES de  $-5,00^\circ$  y  $-5,30^\circ$  respectivamente que son muy elevados teniendo en cuenta lo observado en la RS y nuestro trabajo de campo. La causa para la esta presencia agudizada de ES puede ser debida este efecto del calentamiento, aunque no podemos ser concluyentes respecto a esto debido a la parquedad de información proporcionada por los autores, ya que no describen si aleatorizan o no a los observadores y tampoco conocemos su orden al evaluar a los participantes. Atendiendo a los resultados de los estudios primarios que realizan MBA, observamos que Munteanu et al. (2009) tiene un error sistemático de  $0,00^\circ$  pudiendo ser debido a que aleatorizan tanto a los observadores como a los método de evaluación y observadores. Es más, proporcionan datos de los cuatro observadores de manera conjunta, por lo que no podemos determinar si se da o no este efecto. También encontramos un tanto extraño que el error de medición en todos los métodos de evaluación logrados en este estudio<sup>130</sup> y días estudiados sea exactamente de  $0,00^\circ$ . En cuanto a Jones et al (2005)<sup>135</sup>, a diferencia de los otros dos trabajos, no hemos observado esa tendencia, aunque la escasa calidad del el estudio hace que no podamos confiar la validez de sus resultados.

En consecuencia, atendiendo a los resultados de RS debido a la escasa calidad y el diseño de los estudios, los resultados de la FIA y la FIE no son comparables. Asimismo, debido a las particularidades de las PDFA, el diseño de los estudios incluidos en esta RS para comprobar la fiabilidad inter-observadores es deficiente ya que se pueden dar efectos de aprendizaje, memoria y calentamiento. En conclusión, consideramos que sus resultados no son válidos y que la FIE de la PDFAe y PDFAf no está garantizada.

#### **6.3.4-COMPARACIÓN DE LA FIABILIDAD INTRA-OBSERVADORES DE LAS PRUEBAS DE DORSIFLEXIÓN EN APOYO**

Tanto la RS como el estudio de campo muestran que la prueba de dorsiflexión con la rodilla flexionada es más fiable que cuando se realiza con la rodilla en extensión. Era de esperar que la prueba de DF fuera más fiable que la PDFAf, ya que la dorsiflexión (DF) está mucho más controlada en la PDFAf, al obligar en la mayoría de los trabajos a tocar la pared que está en línea con el talón y primer dedo del pie, o a flexionar la rodilla en dirección al pie. Ésto en la PDFAe resulta más complejo de controlar, ya que en la estimación que se hace siempre de manera visual (o no se menciona) haciendo que la pantorrilla caiga sobre el pie. Si bien es cierto que en este trabajo se ha obligado a la rodilla de las participantes a caer sobre el pie de manera objetivable, y también se ha encontrado que la PDFAf es más fiable. Esto puede ser debido a que existen otros factores que pueden influir en la DF cuando esta se realiza en la PDFAe, como el grado de extensión de la rodilla en el momento de evaluar la DF. En la PDFAe es preciso realizar la prueba con la rodilla en extensión, y debido a las diferencias anatómicas entre participantes, es difícil de determinar si la participante mantiene completamente extendida la rodilla. Por ello, las diferencias entre las evaluaciones dentro de un participante puede variar debido a la influencia de un mayor o menor grado de extensión de la rodilla. Además la implicación de otras articulaciones por encima de la rodilla es mayor en PDFAe debido al procedimiento de la prueba, por lo que hay más factores propios de la variabilidad intrínseca de los participantes que pueden afectar al PDFAe que a la PDFAf.

A pesar de que la PDFAf es más precisa, en algunos casos será necesario realizar ambas pruebas, ya que la PDFAf y la PDFAe son medidas de DF complementarias. Realizando ambas podremos determinar si la DF se ve reducida por la acción del gemelo o no<sup>130</sup>. Eso sí, habrá tener en cuenta las diferencias del error del método de la pruebas a la hora de interpretar los resultados. Es más, algunos trabajos han encontrado que la DF con la rodilla en extensión es FR para las lesiones del MI<sup>194</sup>, con lo que, en algunos casos será preciso evaluar también la PDFAe.

### 6.3.5-RELACIÓN ENTRE LAS MEDIDAS DE RESULTADO

Respecto a los métodos de evaluación empleados en el trabajo de campo, cabe decir que la asociación encontrada entre los estudios ha resultado negativa (anexo 5.5.5). Esto es, los métodos de evaluación empleados en este estudio no pueden ser empleados de manera intercambiable, ya que además de haber encontrado un alto ES en entre los métodos, los LDA también son muy elevados, lo cual indica que no están asociados.

En cuanto a la relación entre las medidas tomadas con el inclinómetro, observamos que el ES entre los métodos (IAf-ITf e IAe-ITe) es muy elevado, y lo que es peor, el error aleatorio es muy alto. Por tanto, cuando evaluamos la DF con inclinómetro ubicando en la tibia (IT), no estamos midiendo el mismo fenómeno que cuando lo aplicamos en el tendón de Aquiles (IA).

En cuanto a la relación entre el ángulo derivado por razones trigonométricas (AD) y las medidas tomadas con el inclinómetro en la PDFAf, encontramos que las medidas tampoco están asociadas. Si bien es cierto que cuando el inclinómetro se aplica en la tibia el sesgo es menor ( $ES=1,29^\circ$ ) del mismo modo que LDA ( $4,24^\circ$ ) las diferencias son muy elevadas como para considerar que los métodos son intercambiables.

A pesar de que algunos estudios han planteado utilizar los métodos de evaluación de manera intercambiable<sup>121;122</sup>, Bennell et al. (1998)<sup>121</sup> y Hollman et al.(2008)<sup>122</sup>, consideramos que estos resultados pueden estar sesgados debido a la escasa calidad y adecuación del análisis estadístico empleado. En primer lugar, Bennell et al. (1998)<sup>121</sup> encontraron una correlación entre la DD y el ángulo entre la vertical y la tibia medido mediante IT en la PDFAf. Este método no es el adecuado para determinar el grado de acuerdo entre métodos de evaluación de manera fidedigna, pudiendo llegar a conclusiones erróneas<sup>150;154;155;195</sup>. Además las razones para una correlación positiva pueden ser debidas al pequeño tamaño muestral o la homogeneidad de la muestra. En cuanto al Hollman et al. (2008)<sup>122</sup>, estos autores encontraron una alta correlación obtenido mediante el goniómetro, con inclinómetro en la tibia y en el tendón de aquiles. Además del riesgo de sesgo que existe debida a la falta de información proporcionada al tratarse de un *abstract*, debemos de tener en cuenta al igual que Bennell et al. (1998) el análisis estadístico realizado no es el adecuado. Otro estudio, de Dickson et al. (2008)<sup>123</sup>, que empleó también la correlación encontró que existía asociación entre métodos, sugiriendo que debían considerarse como métodos o formas diferentes para evaluar la DF.

Por todo ello, suponemos que los otros métodos de evaluación empleados en los estudios primarios como son el goniómetro o el inclinómetro proporcionarían

análogamente errores de medición diferentes. En consecuencia, consideramos que parte de la heterogeneidad encontrada en los meta-análisis podría deberse también a este factor. Es más, consideramos que existirán variaciones en los resultados incluso cuando se emplea un mismo método de evaluación de diferente manera. Por ejemplo, cuando los estudios emplean el método de evaluación en la tibia, algunos estudios deciden colocarlos a 15 cm de la tuberosidad tibial sobre el borde anterior de la tibia, mientras que otros no lo definen o deciden aplicarlo en el centro de la tibia, lo cual es impreciso. Igualmente existirán diferencias en las variables evaluada, puesto que muchos trabajos no definen que en ángulo miden; si miden el rango de movimiento, el ángulo entre la vertical o la horizontal y la tibia u otro ángulo.

Como conclusión a este apartado, el estudio de asociación realizado demuestra que la ubicación de la propia herramienta, en este caso inclinómetro, así como el método de evaluación (IT, IA o AD), no mide el mismo fenómeno. A pesar de medir la dorsiflexión consideramos que miden aspectos diferentes y que de ninguna manera puede sustituirse un método por otro.

## **6.4-FORTALEZAS DE LA TESIS**

Esta tesis incluye tres revisiones sistemáticas que según nuestro conocimiento son las primeras de estas características y que sintetizan de manera crítica toda la información publicada respecto a los objetivos previamente establecidos. La metodología empleada en estas revisiones ha sido robusta. Las búsquedas se han realizado de manera exhaustiva y han sido adaptadas a las diferentes bases de datos con la colaboración de una experta documentalista. La selección de documentos y la valoración de la calidad se realizaron por pares. Asimismo las revisiones analizaron las características del método empleado por los estudios primarios así como los resultados derivados de los mismos, y realizamos meta-análisis siempre que fue posible. Estas tres revisiones han sido realizadas y descritas siguiendo las recomendaciones de expertos en el área<sup>191;196;197</sup>.

Esta tesis se completa con el examen de fiabilidad de las pruebas con estudios de campo, una para SEBT y otro para las PDFA. Estos trabajos fueron diseñados empleando una metodología de acorde a los estándares propuestos por expertos en comprobar la fiabilidad<sup>146;147;147;150;154;165;198</sup>, basándonos en los estudios previos y teniendo en cuenta las particularidades de cada prueba. Es por ello que la fiabilidad del *SEBT* y de la PDFA ha sido estudiada de manera rigurosa. Asimismo, la descripción de los métodos empleados ha sido realizada siguiendo las recomendaciones de expertos<sup>147;196;197</sup>. En cuanto a la PDFA, esta tesis ha propuesto un procedimiento para ambas PDFA empleando un método sencillo y preciso, como lo es ángulo derivado por razones trigonométricas (AD), que además

proporciona una precisión mejor a todos los métodos de evaluación observados en la literatura previa.

## **6.5-DEBILIDADES DE LA TESIS**

Las revisiones sistemáticas han estado limitadas a los idiomas inglés, castellano, francés, alemán, italiano y euskera. Sin embargo, esta limitación ha sido anecdótica puesto que solo excluimos un artículo en la RS del *SEBT*: estaba escrito en japonés. Otra debilidad de las revisiones es que la extracción de las características del método así como de los resultados de los estudios se realizó por una sola revisora.

En cuanto de los estudios de campo sobre fiabilidad de las prueba, las participantes fueron más jóvenes que los trabajos revisados. Aun así, no consideramos que pueda haber diferencias en la fiabilidad debida a la edad, ya que se incluyó la menarquia como criterio de inclusión. En cuanto a la aplicabilidad de las pruebas en la realidad de la población femenina, observamos las características de las participantes incluidas en la RS FR (tabla 5.1.2) son muy similares a las del trabajo de campo desarrollado por esta tesis. La edad media de las participantes de la RS fue de 17,27 años con una altura de 166,73 cm. Respecto al peso estas tenían una media 58,42 kg y presentaban un IMC máximo de 21,56 kg/m<sup>2</sup> y mínimo de 20 kg/m<sup>2</sup>. En cuanto al tiempo de actividad física practicada, no podemos sacar ninguna conclusión debido a que solo tres estudios de la RS describen esta variable. Por lo que las características del grupo donde hemos comprobado la fiabilidad del *SEBT* y PDFA son similares al tipo de población al que se le pretende aplicar las pruebas.

## **6.6-IMPLICACIONES PARA LA PRÁCTICA Y LA INVESTIGACIÓN**

En cuanto a las RS sobre FR , dadas las dificultades encontradas para llevarlas a cabo recomendamos a los investigadores el seguir recomendaciones STROBE<sup>199</sup> para reportar sus resultados en el estudio de los factores de riesgo de las lesiones. Asimismo, en el caso de los estudios de fiabilidad recomendamos seguir las indicaciones de GRASS<sup>147</sup>. En cuanto a los *abstract*, tanto en los estudios pronósticos como en los estudios de fiabilidad, pensamos que será útil incluir sexo de los participantes, la edad o el diseño del estudio, ya que en la mayoría de los casos ha sido necesario consultar el texto completo para determinar si el estudio cumplía con los criterios de inclusión. Todas estas medidas ayudarán mucho a los lectores a poder comprender los estudios y facilitará la síntesis de los mismos mediante los métodos de revisión sistemática (RS) y meta-análisis (MA).

En lo que respecta a la investigación sobre los factores de riesgo, tras lo observado en la revisión sistemática, será preciso realizar más investigaciones sobre los factores de riesgo de las lesiones de tobillo en mujeres deportistas. Consideramos que el número de estudios ( $n=10$ ) es muy reducido teniendo en cuenta la incidencia de las lesiones de tobillo en mujeres deportistas. Para el diseño de éstas recomendamos a los investigadores que tengan en cuenta los criterios propuestos por la iniciativa PROGRESS<sup>185;200-202</sup>, y lo expuesto en la revisión sistemática de la tesis acerca los métodos de evaluación de los FR. La revisión muestra que éstos son en general complejos y sofisticados, lo que supone una gran limitación para su uso en el ámbito de la investigación, así como en la práctica. Será por lo tanto preciso identificar métodos más sencillos para evaluar rasgos que discriminen a mujeres con riesgo de lesión, para poder realizar estudios longitudinales de una manera sencilla. La aplicabilidad de estos métodos será mucho mayor, y más aún si podemos incidir en ellos, es decir, si estos FR son modificables. No obstante, la investigación sobre los FR no modificables y más complejos de evaluar debe continuar. Uno de estos FR es el ciclo menstrual, ya que se ha observado que puede afectar a un mayor riesgo de lesión en ligamento cruzado anterior de rodilla<sup>203-206</sup>. Más concretamente, a nivel de tobillo se ha detectado que en ciertas fases del ciclo menstrual se producen más lesiones<sup>207</sup>.

Todos estos métodos de evaluación, obviando su nivel de complejidad, deben ser precisos para poder realizar con garantías los estudios pronósticos, que requieren un gran esfuerzo humano y económico. Para ello, en primer lugar será necesario realizar estudios de fiabilidad sobre los métodos de evaluación que pretendan ser empleados para discriminar población, teniendo en cuenta las circunstancias de aplicación. Para establecer la fiabilidad de las mediciones será necesario emplear un metodología rigurosa y repórtalos correctamente<sup>147</sup>. Asimismo, estos trabajos deberán emplear valores de precisión interpretables desde el punto de vista práctico, es decir en las mismas unidades en las cuales se van a realizar la mediciones. Así los investigadores podrán determinar su aplicabilidad en la población de interés teniendo en cuenta las características de la lesión estudiada así como las particularidades de dicho factor de riesgo. Para esto, será preciso emplear análisis estadísticos que proporcionen valores prácticos y fácilmente interpretables<sup>147</sup> para los investigadores como es el método propuesto por Bland y Altman<sup>154;155;165</sup>.

Respecto a la fiabilidad de las pruebas estudiadas en la tesis consideramos que el *SEBT* no es una prueba lo suficientemente precisa para emplearla como herramienta diagnóstico y recomendamos no seguir empleándola. En cuanto a las pruebas de dorsiflexión en apoyo (PDFA), consideramos que será preciso establecer la fiabilidad inter-observadores (FIE) del ángulo derivado por razones

trigonométricas (AD) en mujeres deportistas pero confiamos que la fiabilidad será mejor en comparación a otros métodos, ya que partimos de la base que la fiabilidad intra-observadores es mejor que los otros métodos. Respecto a su validez como herramienta diagnóstico será preciso determinar si la precisión del método ángulo derivado para evaluar las PDFA de la es suficiente como para discriminar población con riesgo o sin riesgo de lesión.

En cuanto a otras áreas de investigación y para la práctica consideramos que se deberá estudiar la fiabilidad de este método en otras poblaciones y con otro tipo de observadores. Pero al igual que en el punto anterior, confiamos que la fiabilidad también será adecuada, lo cual permitirá sustituir los métodos de evaluación en la PDFAf y PDF Ae por la del ángulo derivado (AD). El empleo de un método común y estandarizado a la par que preciso y accesible para todo tipo de población y profesionales, permitirá a la ciencia comparar resultados entre estudios y poblaciones. Hasta nuestro conocimiento se están empleando diferentes métodos para evaluar la DF que son en general complejos y de dudosa fiabilidad y limitada accesibilidad, además tal y como se demuestra en esta tesis, las medidas de resultado de los diferentes métodos no son intercambiables ni por lo tanto comparables. El uso de un método común para evaluar la dorsiflexión en apoyo permitirá acumular conocimiento de una manera más sencilla y eficiente, haciendo avanzar a la ciencia.

# 7- CONCLUSIONES



Las conclusiones derivadas de esta tesis, son las siguientes:

1. No se ha demostrado ningún factor de riesgo intrínseco que discrimine con garantías a las mujeres deportistas con riesgo de sufrir una lesión de tobillo de las que no.
2. Los estudios que han analizado la predisposición de las mujeres a sufrir una lesión de tobillo, han analizado FR mediante métodos complejos y de dudosa precisión.
3. El *Star Excursion Balance Test (SEBT)* muestra un error aleatorio elevado de 3,86 a 5,11 cm y no consideramos que sea una medida fiable para predecir el riesgo de lesión de tobillo en mujeres adultas practicantes de actividad física.
4. La prueba de dorsiflexión en apoyo con la rodilla flexionada (PDFAf) y evaluada con el método ángulo derivado de las distancias horizontal y vertical (ADf) tiene un límite de acuerdo de  $3,01^\circ$  y es una herramienta fiable para evaluar la DF en mujeres adultas practicantes de actividad física.
5. La prueba de dorsiflexión en apoyo con la rodilla en extensión (PDFAe) evaluada con el método de evaluación ángulo derivado de las distancias horizontal y vertical (ADe) tiene un límite de acuerdo  $6,01^\circ$  y es una herramienta fiable para evaluar la dorsiflexión en mujeres adultas practicantes de actividad física.
6. Las prueba de dorsiflexión en apoyo (PDFA) medidas mediante el ángulo derivado (AD) son más fiables que los métodos de evaluación propuestos previamente.
7. Los métodos de evaluación de las prueba de dorsiflexión en apoyo (PDFA) no están relacionados entre sí por lo que no se pueden emplear de manera intercambiable.



# 8-REFERENCIAS



- 1-Woodcock J, Franco OH, Orsini N, Roberts I. Non-vigorous physical activity and all-cause mortality: systematic review and meta-analysis of cohort studies. *Int J Epidemiol*. 2011;40(1):121-138.
- 2-Nocon M, Hiemann T, Müller-Riemenschneider F, Thalau F, Roll S, Willich SN. Association of physical activity with all-cause and cardiovascular mortality: a systematic review and meta-analysis. *Eur J Cardiovasc Prev Rehabil*. 2008;15(3):239-246.
- 3-Ruiz JR, Sui X, Lobelo F, Morrow Jr JR, Jackson AW, Sjöström M, et al. Association between muscular strength and mortality in men: prospective cohort study. *BMJ*. 2008;337(7661):92-101.
- 4-Ortega F, Silventoinen K, Tynelius P, Rasmussen F. Muscular strength in male adolescents and premature death: cohort study of one million participants. *BMJ*. 2012;345:e7279.
- 5-Bize R, Johnson JA, Plotnikoff RC. Physical activity level and health-related quality of life in the general adult population: a systematic review. *P Prev Med (Baltim)*. 2007;45(6):401-415.
- 6-Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. *CMAJ*. 2006;174(6):801-809.
- 7-Gordon BA, Benson AC, Bird SR, Fraser SF. Resistance training improves metabolic health in type 2 diabetes: a systematic review. *Diabetes Res Clin Pract*. 2009;83(2):157-175.
- 8-Umpierre D, Ribeiro PA, Kramer CK, Leitao CB, Zucatti AT, Azevedo MJ, et al. Physical Activity Advice Only or Structured Exercise Training and Association With HbA1c Levels in Type 2 Diabetes A Systematic Review and Meta-analysis. *JAMA*. 2011;305(17):1790-1799.
- 9-Keating SE, Hackett DA, George J, Johnson NA. Exercise and non-alcoholic fatty liver disease: a systematic review and meta-analysis. *J Hepatol*. 2012;57(1):157-666.
- 10-Tambalis K, Panagiotakos DB, Kavouras SA, Sidossis LS. Responses of blood lipids to aerobic, resistance, and combined aerobic with resistance exercise training: a systematic review of current evidence. *Angiology* 2009;60(5):614-632.

11-Baillet A, Vaillant M, Guinot M, Juvin R, Gaudin P. Efficacy of resistance exercises in rheumatoid arthritis: meta-analysis of randomized controlled trials. *Rheumatology (Sunnyvale)* 2012;51(3):519-527.

12-Lange AK, Vanwanseele B. Strength training for treatment of osteoarthritis of the knee: a systematic review. *Arthritis Care Res (Hoboken)* 2008;59(10):1488-1494.

13-Howe TE, Shea B, Dawson LJ, Downie F, Murray A, Ross C, et al. Exercise for preventing and treating osteoporosis in postmenopausal women. *Cochrane Database Syst Rev.* 2011;6;(7): CD000333. doi: 10.1002/14651858.CD000333.pub2.

14-O'Shea SD, Taylor NF, Paratz JD. Progressive Resistance Exercise Improves Muscle Strength and May Improve Elements of Performance of Daily Activities for People With COPDA Systematic Review. *Chest* 2009;136(5):1269-1283.

15-Mishra SI, Scherer RW, Geigle PM, Berlanstein DR, Topaloglu O, Gotay CC, et al. Exercise interventions on health-related quality of life for cancer survivors. *Cochrane Database Syst Rev.* 2012 15;8:CD007566. doi: 10.1002/14651858.CD007566.pub2.

16-Liu CJ, Latham NK. Progressive resistance strength training for improving physical function in older adults. *Cochrane Database Syst Rev.* 2009; 8;(3):CD002759. doi: 10.1002/14651858.CD002759.pub2. 3(3).

17-Bahr R, Krosshaug T. Understanding injury mechanisms: a key component of preventing injuries in sport. *Br J Sports Med.* 2005;39(6):324-329.

18-Willems TM. Intrinsic risk factors for sports injuries to the lower leg and ankle. [tesis Doctoral]. Belgium: Universiteit Gent; 2004.

19-Hootman J, Dick R, Agel J. Epidemiology of collegiate injuries for 15 sports: Summary and recommendations for injury prevention initiatives. *J Athl Train.* 2007;42(2):311-319.

20-Powell JW, Barber-Foss KD. Injury patterns in selected high school sports: a review of the 1995-1997 seasons. *J Athl Train.* 1999;34(3):277.

21-Soderman K, Alfredson H, Pietila T, Werner S. Risk factors for leg injuries in female soccer players: a prospective investigation during one out-door season. *Knee Surg Sports Traumatol Arthrosc.* 2001;9(5):313-321.

- 22-Agel J, Olson DE, Randall D, Arendt EA, Marshall SW, Sikka RS. Descriptive epidemiology of collegiate women's basketball injuries: National Collegiate Athletic Association Injury Surveillance System, 1988-1989 through 2003-2004. *J Athl Train* 2007 Jun;42(2):202-210.
- 23-Agel J, Palmieri-Smith R, Dick R, Wojtys E, Marshall S. Descriptive epidemiology of collegiate women's volleyball injuries: National Collegiate Athletic Association Injury Surveillance System, 1988-1989 through 2003-2004. *J Athl Train*. 2007;42(2):295-302.
- 24-Dick R, Putukian M, Agel J, Evans TA, Marshall SW. Descriptive epidemiology of collegiate women's soccer injuries: National Collegiate Athletic Association Injury Surveillance System, 1988-1989 through 2002-2003. *J Athl Train*. 2007;42(2):278-285.
- 25-Bahr R, Lian O, Barh IA. A twofold reduction in the incidence of acute ankle sprains in volleyball after the introduction of an injury prevention program: a prospective cohort study. *Scand J Med Sci Sports*. 1997;7(3):172-177.
- 26-Wedderkopp N, Kalltoft M, Lundgaard B, Rosendahl M, Froberg K. Prevention of injuries in young female players in European team handball. A prospective intervention study. *Scand J Med Sci Sports*. 1999;9(1):41-57.
- 27-Hewett TE, Lindenfeld TN, Riccobene JV, Noyes FR. The Effect of Neuromuscular Training on the Incidence of Knee Injury in Female Athletes A Prospective Study. *Am J Sports Med*. 1999;27(6):699-706.
- 28-Knowles SB, Marshall SW, Miller T, Spicer R, Bowling JM, Loomis D, et al. Cost of injuries from a prospective cohort study of North Carolina high school athletes. *Inj Prev*. 2007;13(6):416-421.
- 29-Van Mechelen W, Hlobil H, Kemper HC. Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. *Sports Med*. 1992;14(2):82-99.
- 30-Romero D, Tous J. Prevención de lesiones en el deporte: Claves para un rendimiento óptimo. 1ª ed. Madrid: Médica Panamericana; 2011.
- 31-Cos F, Cos M, Buenaventura L, Pruna R, Ekstrand J. Modelos de análisis para la prevención de lesiones en el deporte. Estudio epidemiológico de lesiones: el modelo Union of European Football Associations en el fútbol. *Apunts Med Esport*. 2010;45(166):95-102.

32-Parkkari J, Kujala UM, Kannus P. Is it possible to prevent sports injuries? Review of controlled clinical trials and recommendations for future work. *Sports Med.* 2001;31(14):985-995.

33-Rechel JA, Yard EE, Comstock RD. An epidemiologic comparison of high school sports injuries sustained in practice and competition. *J Athl Train.* 2008;43(2):197-204.

34-Deitch JR, Starkey C, Walters SL, Moseley JB. Injury Risk in Professional Basketball Players. *Am J Sports Med.* 2006;34(7):1077-1083.

35-Griffin L, Albohm M, Arendt E, Bahr R, Beynon B, Demaio M, et al. Understanding and preventing noncontact anterior cruciate ligament injuries: a review of the Hunt Valley II meeting, January 2005. *Am J Sports Med.* 2006;34(9):1512-1532.

36-Hewett T, Ford K, Hoogenboom B, Myer G. Understanding and preventing ACL injuries: Current biomechanical and epidemiologic considerations-update 2010. *N Am J Sports Phys Ther.* 2010;5(4):234-51.

37-Hewett T, Myer G, Ford K, Heidt RJ, Colosimo A, McLean S, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. *Am J Sports Med.* 2005;33(4):492-501.

38-Alentorn-Geli E, Myer GD, Silvers HJ, Samitier G, Romero D, Lázaro-Haro C, et al. Prevention of non-contact anterior cruciate ligament injuries in soccer players. Part 1: Mechanisms of injury and underlying risk factors. *Knee Surg Sports Traumatol Arthrosc.* 2009;17(7):705-729.

39-Alentorn-Geli E, Myer G, Silvers HJ, Samitier G, Romero D, Lázaro-Haro C, et al. Prevention of non-contact anterior cruciate ligament injuries in soccer players. Part 2: a review of prevention programs aimed to modify risk factors and to reduce injury rates. *Knee Surg Sports Traumatol Arthrosc.* 2009;17(8):859-879.

40-Kofotolis N, Kellis E. Ankle sprain injuries: A 2-year prospective cohort study in female Greek professional basketball players. *J Athl Train.* 2007;42(3):388-394.

41-Nelson AJ, Collins CL, Yard EE, Fields SK, Comstock RD. Ankle injuries among United States high school sports athletes, 2005-2006. *J Athl Train.* 2007;42(3):381-387.

- 42-Hickey GJ, Fricker PA, McDonald WA. Injuries to young elite female basketball payers over a six-year period. *Clin J Sport Med.* 1997;7(4):252-256.
- 43-Fong DT-P, Hong Y, Chan LK, Yung PS-H, Chan KM. A systematic review on ankle injury and ankle sprain in sports. *Sports Med.* 2007;37(1):73-94.
- 44- Waterman BR, Belmont J, Cameron KL, Deberardino TM, Owens BD. Epidemiology of ankle sprain at the United States Military Academy. *Am J Sports Med* 2010;38(4):797-803.
- 45-Baumhauer JF, Alosa DM, Reeström, Trevino S, Beynnon. A prospective study of ankle injury risk factors. *Am J Sports Med.* 1995;23(5):564-570.
- 46-McGuine TA, Greene JJ, Best T, Levenson G. Balance as a predictor of ankle injuries in high school basketball players. *Clin J Sport Med.* 2000;(10):239-244.
- 47-Beynnon BD, Murphy DF, Alosa DM. Predictive factors for lateral ankle sprains: a literature review. *J Athl Train.* 2002;37(4):376-380.
- 48-McKay GD, Goldie PA, Payne WR, Oakes BW. Ankle injuries in basketball: injury rate and risk factors. *Br J Sports Med.* 2001;35(2):103-108.
- 49-Hubbard TJ, Kramer LC, Denegar CR, Hertel J. Contributing Factors to Chronic Ankle Instability. *Foot Ankle Int.* 2007;28(3):343-354.
- 50-Renstrom P, Ljungqvist A, Arendt E, Beynnon B, Fukubayashi T, Garrett W, et al. Non-contact ACL injuries in female athletes: an International Olympic Committee current concepts statement. *Br J Sports Med.* 2008;42(6):394-412.
- 51-Meeuwisse WH. Assessing causation in sport injury: a multifactorial model. *Clin J Sport Med.* 1994;4(3):166-70.
- 52-Payne KA, Berg K, Latin RW. Ankle injuries and ankle strength, flexibility, and proprioception in college basketball players. *J Athl Train.* 1997;32(3):221-225.
- 53-Noronha M, Refshauge KM, Herbert RD, Kilbreath SL. Do voluntary strength, proprioception, range of motion, or postural sway predict occurrence of lateral ankle sprain? *Br J Sports Med.* 2008;40:824-828.
- 54-Willems TM, Witvrouw E, Delbaere K, Philippaerts R, De Bourdeaudhuij I, De Clercq D. Intrinsic risk factors for inversion ankle sprains in females-a prospective study. *Scand J Med Sci Sports* 2005;15(5):336-345.

55-Beynon BD, Renstrom PA, Alosa DM, Baumhauer JF, Vacek PM. Ankle ligament injury risk factors: A prospective study of college athletes. *J Orthop Res* 2001;19(2):213-220.

56-McGuine TA, Greene JJ, Best T, Levenson G. Balance as a predictor of ankle injuries in high school basketball players. *Clin J Sport Med* 2000;10(4):239-244.

57-Hubbard TJ, Kramer LC, Denegar CR, Hertel J. Contributing factors to chronic ankle instability. *Foot Ankle Int.* 2007;28(3):343-354.

58-Plisky PJ, Rauh MJ, Kaminski TW, Underwood FB. Star excursion balance test as a predictor of lower extremity injury in high school basketball players. *J Orthop Sports Phys Ther.* 2006;36(12):911-9.

59-Willems T, Wivtouw E, Delbaere K, De Cock A, De Clercq D. Relationship between gait biomechanics and inversion sprains: a prospective study of risk factors. *Gait Posture.* 2005 Jun 1;21(4):379-87.

60-Jovell A, Navarro-Rubio M. Evaluación de la evidencia científica. *Med Clin (Barcelona).* 1995;(105):740-3.

61-Levels of Evidence. *BJU International* 2010;105:155.

62-Greenhalgh T. How to read a paper: The basics of evidence-based medicine. 3<sup>rd</sup> ed. Oxford: Blackwell;2006.

63-Glasziou P, Haynes R. The paths from research to improved health outcomes. *Evid Based Med.* 2005;10:4-7.

64-Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ.* 2001;323(7306):224-228.

65-Hertel J, Miller SJ, Denegar CR. Intratester and intertester reliability during the star excursion balance tests. *J Sport Rehabil.* 2000;9(2):104-116.

66-Gribble P, Hertel J, Plisky P. Using the Star Excursion Balance Test to Assess Dynamic Postural Control Deficits and Outcomes in Lower Extremity Injury – A Literature and Systematic Review. *J Athl Train.* 2012;47(3):339-357.

67- Kinzey SJ, Armstrong CW. The reliability of the star-excursion test in assessing dynamic balance. *J Orthop Sports Phys Ther.* 1998;27(5):356-360.

- 68-Clark R, Saxion C, Cameron K, Gerber J. Associations between three clinical assessment tools for postural stability. *N Am J Sports Phys Ther.* 2010;5(3):122-130.
- 69-Earl JE, Hertel J. Lower-extremity muscle activation during the star excursion balance tests. *J Sport Rehabil.* 2001;10(2):93-104.
- 70- Plisky PJ, Gorman P, Butler RJ, Kieser K, Underwood F, Elkins B. The Reliability of an Instrumented Device for Measuring Components of the Star Excursion Balance Test. *N Am J Sports Phys Ther.* 2009;4(2):92-99.
- 71-Nagano Y, Fukano M, Itagaki K, Li S, Miyakawa S, Fukubayashi T. Influence of Lower Limb Clinical Physical Measurements of Female Athletes on Knee Motion During Continuous Jump Testing. *Open Sports Med J.* 2010;4:134-139.
- 72-Munro AG, Herrington LC. Between-session reliability of the star excursion balance test. *Phys Ther Sport.* 2010;11(4):128-132.
- 73-Gribble PA, Robinson RH, Hertel J, Denegar CR. The effects of gender and fatigue on dynamic postural control. *J Sport Rehabil* 2009;18(2):240-257.
- 74-Aminaka N, Gribble PA. Patellar taping, patellofemoral pain syndrome, lower extremity kinematics, and dynamic postural control. *J Athl Train.* 2008;43(1):21-28.
- 75-Martinez-Ramirez A, Lecumberri P, Gomez M, Izquierdo M. Wavelet analysis based on time-frequency information discriminate chronic ankle instability. *Clin Biomech (Bristol, Avon).* 2010;25(3):256-264.
- 76-Olmsted LC, Carcia CR, Hertel J, Shultz SJ. Efficacy of the star excursion balance tests in detecting reach deficits in subjects with chronic ankle instability. *J Athl Train.* 2002;37(4):501-506.
- 77-Gribble PA, Hertel J. Considerations for normalizing measures of the Star Excursion Balance Test. *Meas Phys Educ Exerc Sci.* 2003;7(2):89-100.
- 78-Cote KP, Brunet ME, Gansneder BM, Shultz SJ. Effects of pronated and supinated foot postures on static and dynamic postural stability. *J Athl Train.* 2005;40(1):41-46.

79- Hertel J, Braham RA, Hale SA, Olmsted-Kramer LC. Simplifying the star excursion balance test: Analyses of subjects with and without chronic ankle instability. *J Orthop Sports Phys Ther.* 2006;36(3):131-137.

80-Bressel E, Yonker JC, Kras J, Heath EM. Comparison of static and dynamic balance in female collegiate soccer, basketball, and gymnastics athletes. *J Athl Train.* 2007;42(1):42-6.

81-Hale SA, Hertel J, Olmsted-Kramer LC. The effect of a 4-week comprehensive rehabilitation program on postural control and lower extremity function in individuals with chronic ankle instability. *J Orthop Sports Phys Ther.* 2007;37(6):303-311.

82-Robinson R, Gribble P. Kinematic predictors of performance on the star excursion balance test. *J Sport Rehabil* 2008;17(4):347-357.

83-Hubbard T, Kramer L, Denegar C, Hertel J. Correlations among multiple measures of functional and mechanical instability in subjects with chronic ankle instability. *J Athl Train.* 2007;42(3):361-366.

84-Gribble P, Hertel J, Denegar C, Buckley W. Reliability and Validity of a 2-D Video Digitizing System During a Static and Dynamic Task. *J Sport Rehabil.* 2005;14(2):137-149.

85-Gribble PA, Hertel J, Denegar CR, Buckley WE. The effects of fatigue and chronic ankle instability on dynamic postural control. *J Athl Train.* 2004;39(4):321-329.

86-Lanning CL, Uhl TL, Ingram CL, Mattacola CG, English T, Newsom S. Baseline values of trunk endurance and hip strength in collegiate athletes. *J Athl Train.* 2006;41(4):427-434.

87-Robinson RH, Gribble PA. Support for a Reduction in the Number of Trials Needed for the Star Excursion Balance Test. *Arch Phys Med Rehabil* 2008;89(2):364-370.

88-Herrington L, Hatcher J, Hatcher A, McNicholas M. A comparison of Star Excursion Balance Test reach distances between ACL deficient patients and asymptomatic controls. *Knee* 2009;16(2):149-152.

- 89-Plisky PJ, Rauh MJ, Kamiski TW, Wajswelner H. Star Excursion Balance Test as a predictor of lower extremity injury in high school basketball players. *J of Orthopa Sports Physi Ther.* 2006;36(12):911-9.
- 90-Sabin M. Reliability and validity of the condition-modified Star Excursion Balance Test: Influence of concussion history. [Tesis doctoral]. Illinois: University of Illinois at Urbana-Champaign; 2011.
- 91-Nakagawa L, Hoffman M. Performance in static, dynamic, and clinical tests of postural control in individuals with recurrent ankle sprains. *J Sport Rehabil* 2004;13(3):255-268.
- 92-Leavey VJ, Sandrey MA, Dahmer G. Comparative effects of 6-week balance, gluteus medius strength, and combined programs on dynamic postural control. *J Sport Rehabil.* 2010;19(3):268-287.
- 93- Hale SA, Hertel J, Olmsted-Kramer LC. The effect of a 4-week comprehensive rehabilitation program on postural control and lower extremity function in individuals with chronic ankle instability. *J Orthop Sports Phys Ther.* 2007;37(6):303-311.
- 94-Leavey VJ. The comparative effects of a six-week balance training program, gluteus medius strength training program, and combined balance training/gluteus medius strength training program on dynamic postural control [Tesis doctoral]. Morgantown, West Virginia; 2006.
- 95-Sabin MJ, Ebersole KT, Price JW, Martindale AR. Relationship between the Star Excursion Balance Test And functional measures of strength, power, and agility. *J Athl Train* 2008 May 2;43(3(supplement)):S16.
- 96-Filipa A, Byrnes R, Paterno MV, Myer GD, Hewett TE. Neuromuscular training improves performance on the star excursion balance test in young female athletes. *J Orthop Sports Phys Ther.* 2010;40(9):551-558.
- 97-Dannelly BD, Otey SC, Croy T, Harrison B, Rynders CA, Hertel JN, et al. The effectiveness of traditional and sling exercise strength training in women. *J Strength Cond Res.* 2011;25(2):464-471.
- 98-Gribble PA, Tucker WS, White PA. Time-of-day influences on static and dynamic postural control. *J Athl Train.* 2007;42(1):35-41.

99-English RA, Malone T, Uhl TL. Expected performance values for the Star Excursion Balance Test and Four-Square Step Test in people 30 to 60 years old. *J Orthop Sports Phys Ther.* 2009;39(1):A38.

100-Sabin MJ, Ebersole KT, Martindale AR, Price JW, Broglio SP. Balance performance in male and female collegiate basketball athletes: Influence of testing surface. *J Strength Cond Res.* 2010;24(8):2073-2078.

101-Pope R, Herbert R, Kirwan J. Effects of ankle dorsiflexion range and pre-exercise calf muscle stretching on injury risk in Army recruits. *Aust J Physioter.* 1998;44(3):165-171.

102-Prilutsky BI, Zatsiorsky VM. Tendon action of two-joint muscles: transfer of mechanical energy between joints during jumping, landing, and running. *J Biomech.* 1994;27(1):25-34.

103- Fong C, Blackburn J, Norcross M, McGrath M, Padua D. Ankle-Dorsiflexion Range of Motion and Landing Biomechanics. *J of Athl Train.* 2011;46(1):5-10.

104-Kasuyama T, Sakamoto M, Nakazawa R. Ankle Joint Dorsiflexion Measurement Using the Deep Squatting Posture. *J Phys Ther Sci.* 2009;21(2):195-199.

105-Protopapadaki A, Drechsler WI, Cramp MC, Coutts FJ, Scott OM. Hip, knee, ankle kinematics and kinetics during stair ascent and descent in healthy young individuals. *Clin Biomech (Bristol, Avon).* 2007;22(2):203-210.

106-Vicenzino B, Branjerdporn M, Teys P, Jordan K. Initial changes in posterior talar glide and dorsiflexion of the ankle after mobilization with movement in individuals with recurrent ankle sprain. *J Orthop Sports Phys Ther.* 2006;36(7):464-471.

107-Backman L, Danielson P. Low range of ankle dorsiflexion predisposes for patellar tendinopathy in junior elite basketball players: A 1-year prospective study. *Am J Sports Med.* 2011 Dec;39(12):2626-2633.

108-Bagett B, Young G. Ankle Joint Dorsiflexion - Establishment of A Normal Range. *J Am Podiatr Med Assoc.* 1993;83(5):251-4.

109-Elveru RA, Rothstein JM, Lamb RL. Goniometric reliability in a clinical setting subtalar and ankle joint measurements. *Phys Ther.* 1988;68(5):672-627.

- 110-Jones R, Carter J, Moore P, Wills A. A study to determine the reliability of an ankle dorsiflexion weight-bearing device. *Physiotherapy* 2005;91(4):242-249.
- 111-Gatt A, Chockalingam N. Clinical assessment of ankle joint dorsiflexion. A review of measurement techniques. *J J Am Podiatr Med Assoc.*2011;101(1):59-69.
- 112-Martin R, McPoil T. Reliability of ankle goniometric measurements - A literature review. *J Am Podiatr Med Assoc.* 2005;95(6):564-572.
- 113- Hart JM, Cosby NL, Chinn L. Assesment of dorsiflexión range of motion using two weigth beargin devireces. *Athl Train Sport Health Care.* 2011;3(4):162-163.
- 114- Martin R, McPoil T. Reliability of ankle goniometric measurements - A literature review. *J Am Podiatr Med Assoc.* 2005;95(6):564-572.
- 115-Peeler JD, Anderson JE. Reliability limits of the modified Thomas test for assessing rectus femoris muscle flexibility about the knee joint. *J Athlet Train.* 2008;43(5):470-478.
- 116-Ekstrand J, Wiktorsson M, Oberg B, Gillquist J. Lower-Extremity Goniometric Measurements - A Study to Determine Their Reliability. *Arch Phys Med Rehabil.* 1982;63(4):171-175.
- 117-Lee KM, Chung CY, Kwon DG, Han HS, Choi IH, Park MS. Reliability of physical examination in the measurement of hip flexion contracture and correlation with gait parameters in cerebral palsy. *J Bone Joint Surg Am.* 2011;93(2):150-158.
- 118-Johanson M, Baer J, Hovermale H, Phouthavong P. Subtalar joint position during gastrocemiis stretching and ankle dorsiflexion range of motion. *J Athl Train.* 2008;43(2):172-178.
- 119-Calvo-Guisado M, Díaz-Borrego P, González-García de Velasco J, Fernández-Torrico J, Conejero-Casares J. Three techniques for measuring dorsal flexion of the ankle: Intra- and inter-observer reliability. *Rehabilitación* 2007;41(5):200-206.
- 120-Wilken J, Rao S, Estin M, Saltzman C, Yack H. A New Device for Assessing Ankle Dorsiflexion Motion: Reliability and Validity. *J Orthop Sports PhysTher.* 2011;41(4):274-279.
- 121-Bennell K, Talbot R, Wajswelner H, Techovanich W, Kelly D, Hall A. Intra-rater and inter-rater reliability of a weight-bearing lunge measure of ankle dorsiflexion. *Austr J Physiot.* 1998 Sep;44(3):175-80.

122-Hollman KM, Dickson D, Bronner S, Ojofeitimi S. Reliability and validity of functional ankle range of motion measurements in dancers. *J Orthop Sports Phys Ther.* 2008;38(1):A19-A20.

123-Dickson D, Hollman K, Bronner S, Ojofeitimi S. Reliability, validity and interpretation of a new way to measure ankle dorsiflexion. *J Orthop Sports Phys Ther.* 2008;38(1):A20.

124-Denegar C, Hertel J, Fonseca J. The effect of lateral ankle sprain on dorsiflexion range of motion, posterior talar glide, and joint laxity. *J Orthop Sports Phys Ther.* 2002;32(4):166-173.

125-Dinh N, V, Freeman H, Granger J, Wong S, Johanson M. Calf stretching in non-weight bearing versus weight bearing. *Int J Sports Med.* 2011;32(3):205-210.

126-Krause D, Cloud B, Forster L, Schrank J, Hollman J. Measurement of ankle dorsiflexion: a comparison of active and passive techniques in multiple positions. *J Sport Rehabil* 2011;20(3):333-344.

127- McPoil T, Cornwall M. The relationship between static lower extremity measurements and rearfoot motion during walking. *J Orthop Sports Phys Ther.* 1996;24(5):309-14.

128- Baggett BD, Young G. Ankle Joint Dorsiflexion - Establishment of A Normal Range. *J Am Podiatr Med Assoc.* 1993;83(5):251-4.

129-Rabin A, Kozol Z. Weightbearing and Nonweightbearing Ankle Dorsiflexion Range of Motion Are We Measuring the Same Thing? *J Am Podiatr Med Assoc.* 2012;102(5):406-11.

130-Munteanu SE, Strawhorn AB, Landorf KB, Bird AR, Murley GS. A weightbearing technique for the measurement of ankle joint dorsiflexion with the knee extended is reliable. *J Sci Med Sport* 2009;12(1):54-59.

131- Bennell K, Khan KM, Matthews B, De Gruyter M, Cook E, Holzer K, et al. Hip and ankle range of motion and hip muscle strength in young novice female ballet dancers and controls. *Br J Sports Med.* 1999;33(5):340-346.

132-Bennell K, Khan K, Matthews B, Singleton C. Changes in hip and ankle range of motion and hip muscle strength in 8-11 year old novice female ballet dancers and controls: a 12 month follow up study. *Br J Sports Med.* 2001;35(1):54-59.

- 133-Burns J, Crosbie J. Weight bearing ankle dorsiflexion range of motion in idiopathic pes cavus compared to normal and pes planus feet. *Foot (Edinb)* 2005;15(2):91-94.
- 134-Nitz JC, Choy NL. The relationship between ankle dorsiflexion range, falls and activity level in women aged 40 to 80 years. *New Zealand Journal of Physiotherapy* 2004 Nov;32(3):121-125.
- 135-Jones R, Carter J, Moore P, Wills A. A study to determine the reliability of an ankle dorsiflexion weight-bearing device. *Physiotherapy* 2005;91(4):242-249.
- 136-Montgomery L, Nelson F, Norton F, Deuster J. Orthopedic History and Examination in the Etiology of Overuse Injuries. *Med Sci Sports Exerc.* 1989;21(3):237-243.
- 137-Spink M, Menz H. Efficacy of a multifaceted podiatry intervention to improve balance and prevent falls in older people: study protocol for a randomised trial. *BMC Geriatr.* 2008;8:30.
- 138-Soper C, Reid D, Hume PA. Reliable passive ankle range of motion measures correlate to ankle motion achieved during ergometer rowing. *Phys Ther Sport.* 2004;5(2):75-83.
- 139-Venturni C, André A, Prates Aguilar B, Giacomelli B. Reliability of two evaluation methods of active range of motion in the ankle of healthy individuals. *Acta Fisiát.* 2006;13(1):39-43.
- 140- Kasuyama T, Sakamoto M, Nakazawa R, Kawagoe M, Kato K. Normative Data and Comparison of Measurement Ankle Dorsiflexion Methods. *Rigakuryoho Kagaku* 2008;23(6):741-745.
- 141-Gabbe B, Finch C, Wajswelner H, Bennell K. Predictors of lower extremity injuries at the community level of Australian football. *Clin J Sport Med.* 2004;14(2):56-63.
- 142-Gabbe B, Bennell K, Finch C, Wajswelner H, Orchard J. Predictors of hamstring injury at the elite level of Australian football. *Scand J Med Sci Sports.* 2006;16(1):7-13.
- 143-Burne SG, Khan KM, Boudville PB, Mallet RJ, Newman PM, Steinman LJ, et al. Risk factors associated with exertional medial tibial pain: a 12 month prospective clinical study. *Br J Sports Med* 2004;38(4):441-445.

144-Menz HB, Tiedemann A, Mun-San Kwan M, Latt MD, Sherrington C, Lord SR. Reliability of clinical tests of foot and ankle characteristics in older people. *J Am Podiatr Med Assoc.* 2003;93(5):380-7.

145-Hulley SB, Cummings SR, Browner W, Grady D, Newman T. *Designing in clinical research.* Philadelphia: Lippincott Williams & Wilkins; 2007.

146-Bland M. *An Introduction to medical statistics.* 2<sup>nd</sup> ed. Oxford: Oxford University Press; 1996.

147- Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64(1):96-106.

148-Pita Fernandez S, Pértengas Díaz S. La fiabilidad de las mediciones clínicas: el análisis de la concordancia para variables numéricas. *Cuadernos de Atención Primaria* 2003;10(4):290-296.

149-Müller R, Büttler P. A critical discussion of intraclass correlation coefficients. *Stat Med.* 1994;13:2465-2476.

150-Bland J, Altman D. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol.* 2003;22(1):85-93.

151-Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol bull.* 1979;86(2):420-428.

152-Denegar C, Ball D. Assessing reliability and precision of measurement: an introduction to Intraclass Correlation and Standard Error of Measurement. *J Sport Rehabil.* 1993;2:35-42.

(153) Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26(4):217-238.

154-Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983;307-17.

155-Bland J, Altman D. Statistical-Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet* 1986;8;1(8476):307-10.

- 156-Moller M, Attermann J, Myklebust G, Wedderkopp N. Injury risk in Danish youth and senior elite handball using a new SMS text messages approach. *Br J Sports Med.* 2012;46(7):531-537.
- 157-Le Gall F, Carling C, Reilly T. Biological maturity and injury in elite youth football. *Scand J Med Sci Sports.* 2007;17:564-72.
- 158-World Health Organization [Sede Web]. Disponible en : <Http://www.who.int/mediacentre/factsheets/fs334/en/index.html>. [acceso el 20 de marzo de 2012].
- 159-Hayden JA, Côtè P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med.* 2006;144(6):427.
- 160- Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
- 161- Cabello, J.B. por CASPe. Plantilla para ayudarte a entender un Estudio de Diagnóstico. En: CASPe. Guías CASPe de Lectura Crítica de la Literatura Médica. Alicante: CASPe; 2005. Cuaderno I. p.22-25.
- 162- Bland J, Altman D. Comparing Methods of Measurement - Why Plotting Difference Against Standard Method Is Misleading. *Lancet* 1995;346(8982):1085-1087.
- 163-Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991;59(1):12-19.
- 164-Ekstrand J, Wiktorsson M, Oberg B, Gillquist J. Lower-Extremity Goniometric Measurements - A Study to Determine Their Reliability. *Arch Phys Med Rehabil.* 1982;63(4):171-175.
- 165-Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8(2):135-150.
- 166-DuRant RH, Pendergrast RA, Seymore C, Gaillard G, Donner J. Findings from the preparticipation athletic examination and athletic injuries. *Am J Dis Child.* 1992;146(1):85-91.

167-Krivickas LS, Feinberg JH. Lower extremity injuries in college athletes: relation between ligamentous laxity and lower extremity muscle tightness. *Arch Phys Med Rehabil.* 1996;77(11):1139-1143.

168-Faude O, Junge A, Kindermann W, Dvorak J. Risk factors for injuries in elite female soccer players. *Br J Sports Med* 2006;40(9):785-790.

169-Mei-Dan O, Kahn G, Zeev A, Rubin A, Constantini N, Even A, et al. The medial longitudinal arch as a possible risk factor for ankle sprains: A prospective study in 83 female infantry recruits. *Foot Ankle Int* 2004;26(2):180-183.

170- Steffen K, Myklebust G, Andersen TE, Holme I, Bahr R. Self-reported injury history and lower limb function as risk factors for injuries in female youth soccer. *Am J Sports Med.* 2008;36(4):700-708.

171-Barton CJ, Bonanno D, Levinger P, Menz HB. Foot and ankle characteristics in patellofemoral pain syndrome: a case control and reliability study. *J Orthop Sports Phys Ther.* 2010 May;40(5):286-296.

172-Crossley KM, Thancanamootoo K, Metcalf BR, Cook JL, Purdam CR, Warden SJ. Clinical features of patellar tendinopathy and their implications for rehabilitation. *J Orthop Res.* 2007;25(9):1164-1175.

173-Rabin A, Kozol Z. Measures of range of motion and strength among healthy women with differing quality of lower extremity movement during the lateral step-down test. *J Orthop Sports Phys Ther.* 2010;40(12):792-800.

174-Munteanu SE, Strawhorn AB, Landorf KB, Bird AR, Murley GS. A weightbearing technique for the measurement of ankle joint dorsiflexion with the knee extended is reliable. *J Sci Med Sport.* 2009;12(1):54-59.

175-Fuller CW, Ekstrand J, Junge A, Andersen TE, Barh R, Dvorak J, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Br J Sports Med.* 2007;40(3):193-201.

176-Pasanen K, Parkkari J, Pasanen M, Hiilloskorpi H, Makinen T, Jarvinen M, et al. Neuromuscular training and the risk of leg injuries in female floorball players: cluster randomised controlled study. *BMJ* 2008;337:a295.

177-Handoll HH, Rowe BH, Quinn KM, De Bie R. Interventions for preventing ankle ligament injuries. *Cochrane Database Syst Rev.* 2001;3(3): CD000018.

- 178-Olsen O-E, Myklebust G, Engebretsen L, Holme I, Bahr R. Exercises to prevent lower limb injuries in youth sports: cluster randomised controlled trial. *Br J Sports Med.* 2005;330(7489):449.
- 179-Soligard T, Myklebust G, Steffen K, Holme I, Silvers H, Bizzini M, et al. Comprehensive warm-up programme to prevent injuries in young female footballers: Cluster randomised controlled trial. *BMJ* 2008;337:a2469.
- 180-Verhagen EALM, Bay K. Optimising ankle sprain prevention: a critical review and practical appraisal of the literature. *Br J Sports Med.* 2010;44(15):1082-1088.
- 181-Menz HB. Two feet, or one person? Problems associated with statistical analysis of paired data in foot and ankle medicine. *Foot* 2004;14(1):2-5.
- 182- Baumhauer JF, Alosa DM, Renstrom PAF, Trevino S, Beynnon B. Test-retest reliability of ankle injury risk factors. *Am J Sports Med.* 1995;23(5):571-574.
- 183-Smith-Oricchio K, Harris BA. Interrater reliability of subtalar neutral, calcaneal inversion and eversion. *J Orthop Sports Phys Ther* 1990;12(1):10-15.
- 184-Karnofel H, Wilkinson K, Lentell G. Reliability of isokinetic muscle testing at the ankle. *T J Orthop Sports Phys Ther.* 1989;11(4):150-154.
- 185- Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis research strategy (PROGRESS) 2: Prognostic factor research. *PLoS Med.* 2013;10(2):e1001380.
- 186-Witchalls J, Blanch P, Waddington G, Adams R. Intrinsic functional deficits associated with increased risk of ankle injuries: a systematic review with meta-analysis. *Br J Sports Med.* 2011;46:515-523.
- 187- Demura S, Yamada T. Proposal for a practical star excursion balance test using three trials with four directions. *Sport Sci Health.* 2010;6(1):1-8.
- 188-Stockert B, Barakatt E. Test-retest reliability of the Star Excursion Balance Test in a geriatric population. *J Geriatr Phys Ther.* 2005;28(3):116.
- 189-Akbari M, Karimi H, Farahini H, Faghihzadeh S. Balance problems after unilateral lateral ankle sprains. *J Rehabil Res Dev.* 2006;43(7):819-824.

180-Sefton JM, Hicks-Little CA, Hubbard TJ, Clemens MG, Yengo CM, Koceja DM, et al. Sensorimotor function as a predictor of chronic ankle instability. *Clin Biomech.* 2009;24(5):451-8.

191- Higgins J, Green S. *Manual Cochrane para revisiones sistematizadas de intervenciones.* Versión 5.1.0. The Cochrane Collaboration; 2011.

192-Konor MM, Morton SK, Eckerson JM, Grindstaff TL. Reliability of three measures of ankle dorsiflexion range of motion. *Int J Sports Phys Ther* 2013;7(3):279-287.

193-Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]. York (UK): Centre for Reviews and Dissemination; 2013 (3): Does stretching increase ankle dorsiflexion range of motion: a systematic review.

194-Malliaras P, Cook J, Kent P. Reduced ankle dorsiflexion range may increase the risk of patellar tendon injury among volleyball players. *J Sci Med Sport.* 2006;9(4):304-309.

195- Bland J, Altman D. Measurement error and correlation coefficients. *BMJ* 1996;313(7048):41-42.

196-Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *Ann Intern Med.* 2009;151(4):W65-W94.

197-Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med.* 2009;151(4):264-269.

198-Kottner J, Raeder K, Halfens R, Dassen T. A systematic review of interrater reliability of pressure ulcer classification systems. *J Clin Nurs* 2009;18(3):315-336.

199- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Epidemiology* 2007;18(6):805-835.

200- Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ* 2013;346: e5595.

- 201-Hingorani AD, van Windt DIA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ* 2013;346:e5793.
- 202- Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Med.* 2013;10(2):e1001381.
- 202-Hewett T, Myer G, Ford K. Anterior cruciate ligament injuries in female athletes: Part1, mechanisms and risk factors. *Am J Sports Med.* 2006;34(2):298-311.
- 204-Hewett T. Neuromuscular and hormonal factors associated with knee injuries in female athletes. Strategies for intervention. *Sports Med.* 2000;29(5):313-327.
- 205-Liu S, al-Shaikh R, Panossian V, Yang R, Nelson S, Soleiman N, et al. Primary immunolocalization of estrogen and progesterone target cells in the human anterior cruciate ligament. *J Orthop Res.* 1996;14(4):526-533.
- 206- Slauterbeck JR, Fuzie SF, Smith MP, Clark RJ, Xu KT, Starch DW, et al. The Menstrual Cycle, Sex Hormones, and Anterior Cruciate Ligament Injury. *J Athl Train.* 2002;37(3):275-280.
- 207- Ingraham SJ. A theoretical model for the implications of the relationship of relaxin on injury in female athletes [Tesis doctoral]. Minnesota: University of Minnesota; 2002.