A systematic review of the effect of virtual reality exercise on balance, strength, pain, and function in people with functional ankle instability

Abdullah Al-Alshehi[1], Abdullah Alyami[2], Hammad Alhasan[1], Minju Kim[1], Daniel T.P. Fong[1*]

Affiliations

1. National Centre for Sport and Exercise Medicine, School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK

2. Department/School of physiotherapy, Najran University, Saudi Arabia

Corresponding author*

Daniel T.P. Fong

Postal address: National Centre for Sport and Exercise Medicine, School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK

ORCID

Abdullah Al-Alshehi, N/A

Abdullah Alyami, N/A

Hammad Alhasan, N/A

Minju Kim, N/A

Abstract

Background: Virtual reality has been effective in treating patients with ankle functional instability (AFI). It has become a popular alternative to traditional lower limb rehabilitation. However, no systematic review (SR) has yet been conducted to investigate virtual reality effectiveness to AFI.

Amis: To evaluate the effectiveness and significance of virtual reality for patients with AFI in comparison with any intervention or no treatment.

Methods: Systematic review for Randomise Control Trail's (RCTs) and single-case report concerning virtual reality and AFI was conducted across six bibliographic databases for dates up to and including July 2020. RCT quality was assessed using the Cochrane assessment risk-of-bias tool, while Joanna Briggs Institute tool for case report. Statistical differences between the virtual reality and traditional rehabilitation or control groups were also determined.

Results: Of the 2176 studies reviewed, fifteen studies were reviewed with a bias quality as followed: four RCT's (high) and ten RCT's (moderate), while single-case report (high). Most of these studies (n=12) showed significance differences in balance, strength, pain and function between virtual reality group and other intervention or control groups (P> 0.05). however, three of the studies showed no significant differences.

Conclusion: This review demonstrates moderate evidence that virtual reality is more effective than other interventions in treating AFI, although there are limitation and bias in the studies reviewed. Future high-quality reviews of large-sample studies are needed to investigate all of the important outcome measures together, namely balance, strength, function, and pain.

1.1 Introduction

Ankle-sprain is a pervasive injury associated with physical activity in both subadults and adults (1,2). Among individuals engaging in physical activity, a 10-44% of sustained injuries are ankle-sprains (1,3), with one or more ankle-sprains being experienced by 42-70% of this population (4,5). A certain chronic symptom (6–8) and functional ankle instability (FAI) have respectively been reported at follow-up by 32-74% and 32-47% of individuals with a history of ankle-sprain. For the United Kingdom, statistics indicate that about 302,000 individuals visit accident and emergency departments with new ankle-sprains, and about 42,000 appear with severe new ankle-sprains (9). Furthermore, statistics reveal that the annual cost of treatment of every case of severe ankle-sprain differs depending on the choice of intervention, such as boots (£3 million), braces (£1.5 million), below knee plaster (£0.6 million), and Tubigrip (£0.1 million) (9). One common classification system for ankle-sprain characterises severity as mild (grade I), moderate (grade II), and severe (grade III) (10).

FAI is associated with occupational impairment in around 6% of cases (11), with the length of impairment varying between 9 months to 6.5 years in 13-15% of cases (11,12). The initial definition of FAI was provided by Freeman (1965) (13), who characterised the condition as a repetitive occurrence of ankle-sprain or a sense of “giving way”(14). Otherwise referred to as recurring ankle-sprain, chronic ankle-instability (CAI) or unstable-ankle, there is a consensus among authors about the main manifestations of FAI, namely a sense of ankle-joint-instability (15–17) or a subjective perception of lack of ankle-stability caused by ankle proprioceptive and neuromuscular dysfunction (18). From the perspective of the skeleton, lower limb rotation on the weight-bearing foot may become difficult when FAI involves the subtalar and talocrural joints, having an impact on the momentum force and ground reaction force on the subtalar joint, alongside adaptive changes to standing posture or gait (19). Meanwhile, from the perspective of muscles, the evertor muscles are essential for supporting body weight loading and for generating external

loading on the ankle-joint in the context of weight bearing. Disruption of loading stability results in a rapid increase in inverting torque acting on the muscles, which in turn can cause the ankle to "give way" (18).

With regard to loss of balance and muscle strength, the multifactorial character of FAI was first researched by Tropp and colleagues (19,20), who identified not only primary FAI-related risk-factors, but also possible secondary risk-factors, including regulation of posture and movement coordination. Improper regulation of proprioception, balance, and muscular strength have been suggested by a number of authors as potential causes of FAI conditions (17,20,21). Furthermore, another study (22) have suggested that clinicians should have knowledge of every available form of treatment in order to achieve optimal rehabilitation outcomes, as ankle-dysfunction can manifest in a number of ways in cases of FAI. Lateral-ankle-sprain often engenders CAI and surgical intervention may be necessary if symptoms do not subside and the external ankle ligaments sustain elongation or tearing (23).

The usefulness of virtual-reality for rehabilitation purposes has been considered in many studies (24,25). Virtual-reality has had promising applications in science and medication since the 1990s, including intercession conveyance (26). For instance, it has been suggested that motor learning could be improved through virtual-reality-assisted intervention (VRAI) whereby patients can engage in entertaining functional tasks within real-world simulations as well as provide feedback (27,28). Given its features, VRAI has been extensively assessed regarding its relevance for rehabilitation purposes. Indeed, numerous peer-reviewed studies have advocated that the benefits of VRAI justify its use over traditional rehabilitation interventions (TRI) (24,25,29,30). Moreover, several studies confirmed the effectiveness of a new strategy involving the use of the Wii-Fit-game console as VRAI (31,32). Produced by Nintendo Inc. (Kyoto, Japan), the Wii-Fit-game console is affordable and highly popular, so it has broad availability (32). This game console has a range of features, including the Wii-Fit-Plus with a balance board capable of detecting users' motions for conversion of a particular action in the game, as well as offering real-

time visual-feedback. Therefore, it can be a useful instrument for improving muscle strength. The feedback can be used to enhance the capacity for sensory integration and to control muscle strength more effectively (32). As far as the current author is aware, no SR of this topic has been carried out so far. Hence, this study seeks to fill this knowledge gap literature by examining the effect of virtual-reality games on balance, functional stability, pain and strength in cases of FAI.

Methodology

2.1 Systematic Reviews Definition

One SR (33) advise that a SR explicitly summarises the effects of a healthcare intervention in an up-to-date and reproducible way, while other SR (34) state that the risk-of-bias in SR can be moderated through a peer review protocol conducted by two researchers with specific clinical expertise.

2.2 Systematic Review Strengths

A SR effectively answers a clinical question by exploring the results of previous studies and understanding their motivation. Wong, (2007) (35) explain that transparency runs throughout a SR, enabling the reader to understand the decisions made. Also, stresses the SR's effectiveness in producing an overall, rather than individually-focused, summary (33).

2.3 Systematic Review Limitations

A SR (33) indicate that irregular results may emerge when reviewers misunderstand the research design. In addition, a SR (36) suggest that the quality-of-primary studies fundamentally impacts a SR's reliability. Results that argue against the null hypothesis are more frequently published than those which agree with it, potentially leading to publication bias and unreliable results (33,37). Meanwhile, a SR may include clinical trial data that is inappropriate for general clinical practise (38). However, two

studies (33,39) point out that many SRs only investigate English language journals, leading to language bias.

2.4 Eligibility Criteria

The eligibility criteria for this study were agreed with the project supervisor and designed to explore the effect of virtual-reality on strength, balance, function and pain for FAI. This SR was targeted toward experimental studies, which were included if they met the inclusion and exclusion criteria. To answer the research questions, the criteria ensured that eligible studies included patients who had used virtual-reality to improve ankle-stability in clinical practice.

All published studies were considered eligible based on PICO, while database searches were not limited by human subjects, date, language or gender. Inclusion criteria were (1) studies with participants aged ≤15- 64 years old, since one study (40) advise that patients aged 15-19 are the peak age for ankle-sprain, as well as the black matter and of dopamine seen in the striatum toward the end of the sixth decade of life due to the dopaminergic neurons level decreasing up to 40–50% (41); (2) evidence of ankle instability; (3) had previous ankle-sprain once or more in the last year; (4) studies that included a control group; (5) studies of interventions aimed at AFI or ankle-sprain, including case report studies and RCT. Exclusion criteria included participants who (1) had previous ankle surgery; (2) had neurological issues; (3) were taking part in other exercise-based rehabilitation programmes; (4) had musculoskeletal or similar disorders. Table 2.1 summarises the inclusion and exclusion criteria.

Table 2.1. Inclusion & exclusion criteria

Inclusion Criteria
- Population: adolescent (≥ 15-64 years old).
- Intervention: virtual reality or other .
- Comparator: any intervention or control group.
- Outcome: balance, strength, function and pain or any valid outcome tool/measurement
- Study design: randomised controlled trial, and case report.
Exclusion Criteria
1. Studies of VR in other patient populations.
2. Abstracts and conference proceedings.
3. Not full text.

2.5 Search Strategy

A health librarian confirmed the search strategy, which was based on PICO in line with the review's aims and objectives. The search terms were developed according to previous reviews and studies that explored the effectiveness of virtual-reality in treating FAI (22,42–47). The search terms were approved by the project supervisor, and in each database, both free text and controlled vocabulary were searched, including subject headings and MeSH terms such as "ankle sprain", "ankle instability", "virtual reality" "balance", "strength", "pain" and relevant alternatives or synonyms (Table 2.2).

Using the EBSCO Host interface, this review searched electronic databases including the CINAHL, Cochrane Library , MEDLINE, PsycINFO, AMED and SPORTDiscus, with the time zone set according to the date of the first study published in (2014).

Table 2.2. summarises the search strategy, while the results are in Appendix I.

Search terms	MeSH/subject headings
Ankle instability	Chronic ankle instability, functional ankle instability, ankle injury, and lateral ankle sprain
Virtual reality	Exergames, active video games, Nintendo Wii, Xbox Kinect, Wii Fit, augmented, visual feedback, biofeedback and virtual reality.
Balance	Balance, strength, pain, proprioception, and strengthening exercises,

2.6 Study Selection

The records retrieved from the search were exported to Microsoft Word, while three assessors were used to reach a consensus with the supervisor, one to translate from Korean to English language and the other two for the literature inclusion.

2.6.1 Screening

To exclude ineligible studies based on the selection criteria, titles and abstracts were screened first. If results were inconclusive, then full-text articles were assessed to determine eligibility according to the inclusion and exclusion criteria. The robustness of this process was assured by independent reviews by the second and third assessors and discussion took place to ensure agreement. All included studies' references were screened for further eligible studies before determining the final list of eligible full-text articles for inclusion in the systematic review.

2.7 Data Extraction

Data extraction tools based on JBI-SUMARI were used to collect the primary study data, while the data tool came from the Joanna Briggs Institute (Appendix 2). The first author extracted the data, which includes study design, inclusion and exclusion criteria, participant demographics and details of interventions and outcomes.

2.8 Quality Appraisal

It is critical to assess the quality of included studies in SR; for RCTs and case reports, this study used the Cochrane assessment of bias tool (Appendix 3), while Joanna Briggs Institute (JBI) tool was used just for case reports (48). The Cochrane risk of bias tool assesses the quality of internal validity, covers seven bias domains and has been refined by Cochrane review groups (49).

Meanwhile, case reports include unique findings or novel occurrences of disease and the JBI is the only appropriate tool for such descriptive studies (Appendix 4). Independent not-for-profit JBI has

developed appraisal checklists that evaluate the appropriateness, effectiveness, feasibility and meaningfulness of healthcare interventions (50). Hallgren, (2012) (51) indicate that when two reviewers share ratings, the inter-rater reliability is excellent. In this study, the primary author with two assessors rated the studies before coming to a consensus following discussion.

2.9 Data Analysis

The effect-size estimates the extent that effectiveness differs among experimental and control groups (52). Cohen's d was used to calculate this; d=0.2 indicates a small effect size; d=0.5 a medium effect-size, d=0.8 a large effect-size; d=2 a strong effect-size (52), for equation please see 2.1.

$$\text{Cohen's d} = \frac{(x^1 - x^2)}{S^{average}}$$

And

$$S^{average} = \frac{(s^1 - s^2)}{2}$$

Where; d= effect size, X^1= experimental group, X^2= control group, S^1= experimental group mean standard division and S^2= control group mean standard division.

Equation 2.1: Cohen's equation for effect size

Results

3.1 Selection of Studies

After the original search for articles yielded 2176 results, and a further 1129 duplicate articles were removed, a number of exclusion criteria were applied to further narrow down the selection. From the remaining 1047, 1018 were deemed inappropriate after reviewing the title and the abstract, which were dropped from the list because they: contained irrelevant topics; were review articles; were based on studies of other populations. In turn, 240 were removed because they did not examine the

effectiveness of virtual-reality on the FAI, were unpublished, or were unavailable pre-publication from the authors.

The 28 remaining articles underwent deeper analysis, and 15 were excluded because of a lack of suitability as they did not have a control group and because they did not address AFI. A further reason was that the interventions they referred to were supplementary additions. A total of 13 studies were selected for inclusion according to the formulated inclusion and exclusion criteria. In addition, there were around 400 references included in the reference lists of those 13 included studies, which were screened and identified with the same procedure as previously performed. Two additional studies were identified that met the selection criteria.

Thus, 15 studies were selected for the following analysis, synthesis and review. The flow diagram (Figure.3.1) provides a summary of the aforementioned literature selection process.

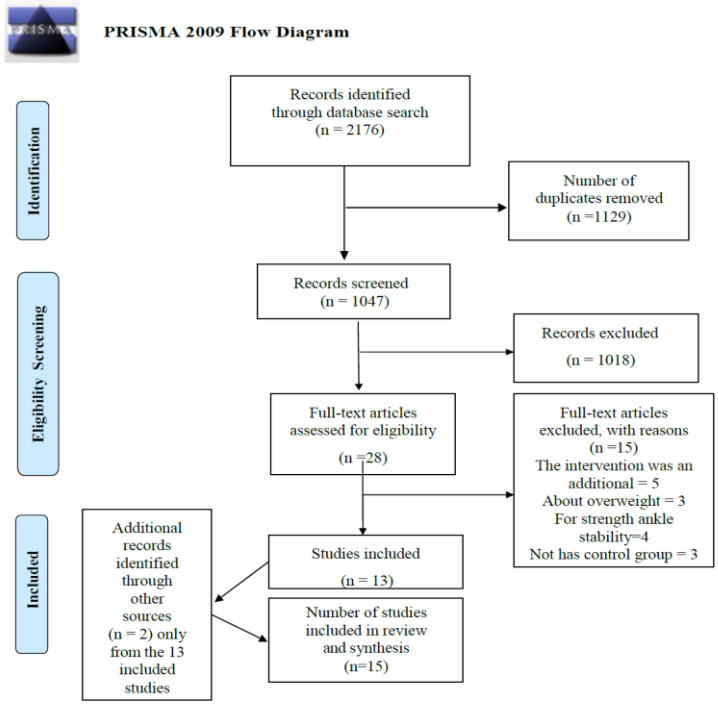

Figure 3.1. Flow diagram of electronic database search and study selection

3.2 Characteristics of Included Studies

Table 3.2 Outline of the characteristics of included studies.

3.3 Setting of Studies

A total of fifteen included studies have been selected. They were all published between 2014 and 2020 and include fourteen RCTs and single-case-report. Of the fourteen RCTs, ten were from Korea (22,43–47,53–56), two from Switzerland (57,58), one from Iran (59), and one from Greece (42). The case-report was from the United States (60).

3.4 Characteristics of Participants

A total of 572 subjects were included in the selected RCTs and 2 subjects were included in the case report. All participants suffered from AFI or ankle-sprain. The ankle-sprains were chronic and the recurrent incidents happened more than once (n=13). Two studies did not have more than one ankle-sprain involved (57,58). The mean age of subjects recruited in the studies ranged from 15-64. The percentage of female to male subjects varied between studies, with a majority of male subjects in the RCTs, while two females in the case-report.

3.5 Characteristics of Interventions

Virtual-reality was used across the studies in this review, involving the Nintendo Wii Fit , the Xbox (Microsoft, Redmond, WA, USA) and head-mounted-device, with games that focused on balance (Table 3.1) . Treatment was either completed under supervision or with a home rehabilitation program. The duration of interventions ranged from 20 minutes to 30 minutes and from 2-3 per week, lasting for 2-10 weeks. However, one RCT study did not state the details of treatment duration (59).

Table 3.1. Game type

Studies	Devices	Games type
(43–47,55,56,59–61).	Nintendo Wii Fit	Slalom skiing Slalom snowboarding Penguin sliding Table-tilt balance. Bubble Tightrope walking
(42)	Xbox (Microsoft, Redmond, WA, USA)	Rally ball Reflex ridge River rush 2000 leaks
(55)	Head -mounted device (VR box, Max, China)	Virtual reality funambulism(Yi Wang, China) Virtual reality traffic bike racing(Babloo Games, United Arab Emirates) Subway Surf Race Virtual reality 2017 (Daisy Daisy, USA)

3.6 Characteristics of Comparators

All included studies (n=15) utilising virtual-reality as the experimental group, whilst strength exercise or traditional therapy as the control group.

3.7 Characteristics of Outcome Measures and Effect Size

All studies referred to in this review (n=15) used different outcome measures. Moreover, none of them included all four outcome measures together (balance, strength, function and pain). All the included studies used reliable and valid outcome measurement tools. All outcome measures are shown in table 3.2. For the effect-size of all outcome measures, see Appendix 6.

3.9 Outcome Assessment Tools and Effect Size

In terms of balance tools, 9 of the 15 included studies used different measurement tools such as the biodex balance system (BSS), bio-rescue, Y-balance, G-jump, star excursion balance test (SEBT), error scoring system

(ESS), and biodex isokinetic dynamometer (BID). They all showed significant differences between groups with a multiple effect-size. The BBS, Bio-rescue, Y-balance and BID measurements found a large effect-size d >0.8 (22,43, 44, 46, 48, 54, 56, 57). One RCT study used G-jump, showing a medium effect-size d>0.5 (56). Furthermore, the SEBT and ESS measurements were used by one-case report study (60), with no indication of effect-size because they used a diagram.

In terms of functional tests, 7 of the 15 included studies used different measurement tools such as functional performance tests (FPT), foot and ankle ability measure (FAAM), Cumberland ankle instability tool (CAIT), temporal-spatial gait parameters (TGP) and kinematic gait parameters (KGP). Some studies (n=5) showed significant differences (P<0.05) (43,46,47,59,60). However, two studies (57,58), showed no significant differences between groups. The FPT, CAIT and KGB measurement found a large effect-size of d>0.8 (44,47,48, 59. 60), while TGP measurement showed medium effect-size d >0.5 (59). The FAAM measurement showed a small effect-size d>0.2 in one RCT study (57). However, in single-case-report study (60), the effect-size was not available, because they used a diagram.

In terms of strength tools, 3 of the 15 included studies used the BID measurement tool, with two RCT's of the studies finding a significant improvement (54,55). However, one RCT research (44) found no significant differences. All three studies showed a large effect-size of d>0.8. In terms of pain tools, 2 of the 15 included studies used the visual analog scale (VAS) measurement and discovered significant improvement in one-case-report (60). The effect-size was not available because the author used a diagram. However, in on RCT study (57), there was no significant improvement ($P \geq 0.292$) and the effect-size was small d>0.2.

3.10 Follow Up Period

Two RCTs conducted a follow-up after 6 months and 1 year respectively, which were longest follow ups (57,58). However, twelve RCTs (22,42–47,53–56,59) and one-case report did not conduct a lengthy follow-

up after treatment therapy, as they only measured within the baseline and at the end of the treatment.

3.11 Missing Data/Dropouts

Twelve RCT's studies and one-case-report did not have any dropouts. However, dropouts occurred in two RCTs (57,58), with explanations given

Table 3.2. Findings of included studies, viewing outcome measures, and brief description.

Authors	Outcomes	Outcome measurement	Brief description
1- (45) 2- (22) 3- (42)	Balance	Biodex Balance System	A- All studies 1, 2, 3, 4, 5, 6, 7, 8 ,9 and 10 in the all measurements showed a significant difference between and within groups. B- Studies 1, 2 and 6 measured static and dynamic ranges from 1 to 8 degree for anterior-posterior and medial-lateral movement. C- Studies 3 and 7 measured the limit of stability. However, study 3 has an additional measure which was the stability index. D- Studies 4 and 5 measure the surface area and the whole path length. However, study 4 has an additional measure which was the limit of stability. E- Study 7 uses only a measure of the limit of stability. F- Study 9 used the tool to measure ankle proprioception. G- Study 10 used two measurement for balance.
4- (56) 5- (53) 6- (55) 7- (47)		Bio-rescue	
8- (55)		Y-balance	
9- (43)		Biodex isokinetic Dynamometer	
10- (60)		Star excursion balance test. Error scoring system.	
1- (44) 2- (55) 3- (54)	Strength	Biodex isokinetic Dynamometer	A- Study 1 showed no significant differences, however within groups shoed significant differences. B- Studies 2 and 3 showed significant differences between and within groups, C- Studies 1, 2 and 3 measure the ankle strength for plantarflexion, dorsiflexion, inversion and eversion.
1- (57)		Visual analog	A- Studies 1 and 2 showed significant

2-	(60)	Pain	scale (VAS)	improvement within groups. However, there were no significant differences between the groups in study 1, while study 2 showed significant differences.
1-	(59)	Movement impairment	Functional Performance Tests	A- Studies 2 and 8 showed no significant differences, however within groups showed significant differences. B- Studies 1,3, 4, 5, 6 and 7 showed significant improvement between and within groups. C- All different measurement tools used for studies 1-8 were to measure ankle function.
2-	(57)		Foot and ankle ability measure	
3-	(60)			
4-	(43)		Cumberland ankle instability tool (CAIT)	
5-	(47)			
6-	(46)			
7-	(56)		G-jump	
8-	(58)		Temporal-spatial gait parameters. Kinematic gait parameters.	

3.12 Quality Assessment of Included Studies

3.12.1 Risk of Bias in Included RCT's

The results of the quality assessment for the included RCTs are shown in Table 3.4 and Figure 3.2.

Random Sequence Generation

Three studies had a low risk-of-bias as they reported random sequence generation, with two studies using an electronically generated block randomization list (57,58). The other used a sealed envelope with an enclosed number (42). Eleven RCT's studies (22,43,44,46,47,53–56,59,62) had an unclear risk-of-bias as they reported that participants were randomized with no information of the procedure of randomization.

Allocation Concealment

Two RCT's (57,58) studies had a low risk-of-bias as the concealment procedures were stated clearly. The twelve RCT's remaining studies (22,42–44,46,47,53–56,59,62) had an unclear risk-of-bias as there was no allocation concealment reported by any of the studies.

Blinding of Participants and Personnel

Two RCT's studies (57,58) had high risk-of-bias as the patients were aware of the treatment. All twelve RCT's studies (22,42–44,46,47,53–

56,59,62) had an unclear risk-of-bias as they did not report any information regarding this.

Blinding of Outcome Assessment

A total of three RCT's studies (42,57,58) had a low risk-of-bias. The remaining eleven RCT's studies (22,42–44,46,47,53–56,59,62) had an unclear risk-of-bias as there was insufficient information to make a judgment.

Incomplete Outcome Data

Twelve RCT's studies (22,42–44,46,47,53–56,59,62) had a low risk-of-bias as no missing data was reported. However, two RCT's studies (57,58) had a high risk-of-bias.

Selective Reporting

Twelve included studies (22,42–47,53–56,59) had a low risk-of-bias as the authors reported outcome measurement and the P value, whereas the remaining two RCT's studies (57,58) had an unclear risk-of-bias, as there was insufficient information with which to make a judgment.

Other Bias

Nine studies had a low risk-of-bias (42,43,45,53–58), whereas the remaining five studies had an unclear risk-of-bias(44,46,47,54,59).

3.12.2 JBI Scale in The Case Report

The outcome ratings for JBI are present in Table3.3. There was one-case-report study (60), which had a high score 7/8. The domain of item 8, 'adverse events (harm) or unanticipated events' was considered to have been 'no', while others were 'yes'.

Table 3.3. characteristics of JBI

Major Components	(60)
1.patient's demographic characteristics	Yes
2.patient's history	Yes
3.current clinical condition of the patient	Yes
4.diagnostic tests and results	Yes
5.the intervention or treatment procedure	Yes
6.post-intervention condition	Yes
7.adverse events (harm) or unanticipated events	No
8.provide takeaway lessons	Yes
Total number of 'yes'	7/8

3.13 Summary of Results

Overall, the results showed that the virtual-reality group had statistically significant improvement ($p< 0.05$) in most of the outcomes compared to the other group in the majority of the included studies (n=12) (22,42,43,45–47,53–56,59,60).

Nonetheless, three RCT's studies provided contrary evidence that demonstrated that there were no significant differences ($p >0.05$) between the virtual-reality group and control group using two measurements: functional and pain in two RCT's studies (57,58), while the other study used strength measurement (44). The most used measure was the balance outcome, which showed significant differences with large effect-size where RCTs were employed (n=9) (22,42,43,45–47,53,55,56) and for the single-case-report no effect-size available, because they used diagram (60). Pain outcome was only used in one RCT study (57) and one-case-report (60).

Because the control selection of bias was relatively not well conducted before the trials took place, the RCTs under review were deemed to be moderate risk-of-bias. It is acknowledged that the 'blinding' of patients and researchers is integral to the decrease of bias within such research, but this presents seemingly insurmountable challenges in terms of maintaining overall blindness throughout the entire process.

To summarise, the overall quality assessment of the RCTs (n=14) under review shows some concerns (moderate quality), with one-case-report (60) scoring high on the JBI scale at 7/8, indicating a low risk-of-bias.

Table 3.4. Results of the risk of bias table according to the review authors' judgements about risk of each bias item for included studies

Author	Random sequence generation	Allocation concealment	Blinding of participants and personal	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other bias
(44)	?	?	?	?	+	+	?
(54)	?	?	?	?	+	+	?
(43)	?	?	?	?	+	+	+
(59)	?	?	?	?	+	+	?
(45)	?	?	?	?	+	+	+
(42)	?	?	?	+	+	+	+
(57)	+	+	—	+	—	?	+
(47)	?	?	?	?	+	+	?
(55)	?	?	?	?	+	+	+
(61)	?	?	?	?	+	+	+
(22)	+	?	?	?	+	+	+
(58)	+	+	—	+	—	?	+
(53)	?	?	?	?	+	+	+
(46)	?	?	?	?	+	+	?

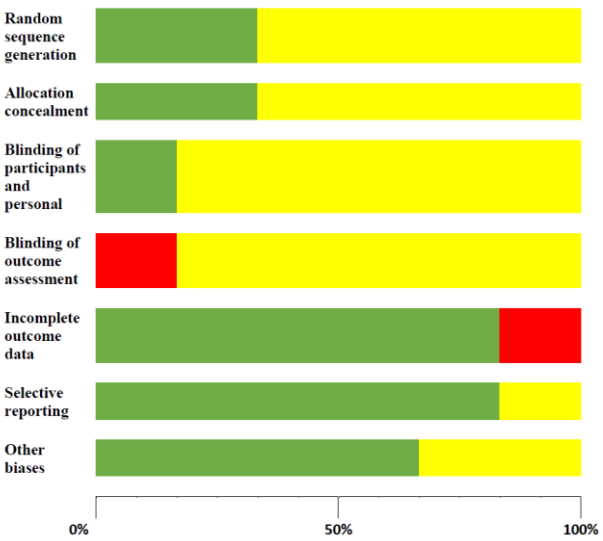

Figure 3.2. The risk of bias summary: review authors' judgements about risk of bias for each item, presented as percentages across all included studies.

+	Low risk of bias	?	Unclear of bias	—	High risk of bias

Discussion

4.1 Summary of Principle Findings

The present review aimed to systematically synthesise existing evidence in order to uncover the advantages of virtual-reality with a goal to improve AFI. A total of fifteen studies satisfied the inclusion criteria and were included in this SR. All included trials investigated the effectiveness of virtual-reality for the AFI.

In the literature, the virtual-reality program used for the treatment of the FAI involved one of the following devices; Nintendo Wii Fit , Xbox (Microsoft, Redmond, WA, USA) and head -mounted device. The Nintendo Wii Fit is the most commonly applied treatment across studies in this SR. All included studies have investigated virtual-reality for AFI, regardless of device. There were thirteen studies which used the Nintendo Wii Fit (22,43–47,53–55,57–60), one-study which used head-mounted-device (55) and one-study which used Xbox Kinetic (42).

The intended results of this SR were balance, strength, function and pain, with one or more of these evaluated in the studies examined. The current section discusses the SR results and the consequences of these outcomes. In turn, the study's limitations are discussed, and suggestions for further research on the topic at hand are discussed.

The results in all twelve studies showed that the virtual-reality group showed a statistically significant improvement ($p< 0.05$) in most of the outcomes when compared to the other (control) group in the majority of the included studies (n=12). However, three RCT's studies found no significant differences ($p>0.05$) (44,57,58)

Overall, regardless of the virtual-reality used in the included studies, this review has generated moderate evidence to suggest that significant differences exist between virtual-reality and other intervention in the

provision of balance, strength, pain and functional improvement. A total of fifteen studies support this finding, including five high-quality papers (22,42,57,58,60) and ten moderate-quality papers

Furthermore, the outcomes of this SR showed that virtual-reality presented major disparities across groups. Virtual-reality is seemingly superior to strength exercise, traditional therapy, or no intervention. Of the fourteen RCTs studies examined, and these all supported the notion of a different effect-size with the majority of large effect-size. In addition, one-case-report was of high-quality, but the effect-size was not available because of the diagram used.

With regard to balance measurements, all included RCT's studies used different measurements; the BBS tool used in two high-quality studies (22,42), while the Bio-rescue, Y-balance and BID used moderate-quality RCT's studies (44, 46, 47,48,54,56), and all showed a large effect-size. The G-jump measure tools were used by one moderate-quality study and showed a medium effect-size (56). The BESS and SEBT measure tools used by one-case-report study showed significant differences but the effect-size was not available, because they used a diagram (60). The reader should take into consideration that different studies used different measurement tools for balance and that could be a source of bias due to the fact that different measurement tools can demonstrate different levels of reliability and validity (63). In respect of other included studies (n=6) (44,46,54,57–59), balance was not included in their assessments, which was questionable because this outcome is considered a very important measurement in such populations (64).

In terms of strength measurement, there were only three RCT (44,54,55) studies measured strength and all of them used BID. Two RCT's studies showed significant differences (54,55). However, the third RCT study (44) showed no significant differences, all three studies RCT's showed large effect-size. The reason for the differences in results between this study and the other two studies that showed significant differences could be attributed to the intensity of exercise program, as this was determined by different physiotherapists in the studies. In respect of other

included studies (n=12) (22,42,59,60,43,45–47,53,56–58), strength measure was not included in their assessments, which was questionable because this outcome is considered as a very important measurement in such populations (65).

With regard to function outcomes measures, the CAIT and FPT measurement tool used by four moderate-quality RCT's studies showed significant differences; all four RCT's studies showed a large effect-size (44,47,48, 60). However, the FAAM, TGP and KGP measures tool used by two high-quality RCT's studies (58, 59) that showed no significant differences with large effect-size in TGP tool, small effect size in FAAM tool and medium effect-size in KGP tool. In contrast, the single-case-report (60) showed significant differences and the effect-size was not available, due to the author using a diagram. The reason that the two RCT's studies (57,58) showed no significant differences could be due to the patients in these two studies being asked to practice the treatment at home without physiotherapist supervision, which may influence the effects of the treatment. Furthermore, the outcomes could be due to the dropout rate as the two RCT's studies (57,58) had missing data, while the other studies did not report missing data. The reader should take into consideration that different studies used different measurement tools for functionality and this may lead to bias.

In respect of other included studies (n=8) (22,42,44,45,53–56), function was not included in their assessments, which was questionable because the World Health Organisation (WHO) recommended function as an outcome measure in such populations (66).

In terms of the last outcome measure (pain), only two of the included studies involved this outcome measure , with both using VAS and being high-quality studies. One RCT study (57) showed no significant improvement with a small effect-size. However, single-case-report study by (60) showed significant differences with no availability for the effect-size due to the author using a diagram. The reason for the differences in results between the two studies could be attributed to the differences in research design and to the dropout rate in the RCT study (57), which reported

missing data. In respect of other included studies (n=13) (22,42,58–60,43–47,53,54,56), pain was not included in their assessments.

Referring back to the rationale of this SR, which was designed to investigate the effectiveness of virtual-reality for patients with AFI, a previous SR (67) conducted in this field looked at the effectiveness of virtual reality for orthopaedic rehabilitation in general. However, this SR did not include all previous studies for AFI. This SR is different since it has included all previous studies which investigated virtual-reality for AFI exclusively.

All of the included studies provided an informative abstract. Contents were well organised and placed under appropriate headings. Additionally, objectives were clearly stated, and the authors provided an overview of the methodology they employed, including clearly identifying populations, results and conclusions (68). Furthermore, the majority of the authors of the included studies (n=14) (22,42–44,46,47,53–59,62) conducted RCT's in an effort to minimise errors, reduce bias and achieve good-quality results. The one-case report study (60) scored high-quality. Eleven included studies clearly explained that they had obtained ethical approval and informed consent, but in four studies (46,53,55,56,60) this issue was not addressed. Moreover, it is important that researchers obtain both ethical approval and informed consent prior to conducting their research (69).

The sample-size calculation is a vital component of research methodology from both an ethical and scientific perspective (70). Therefore, it is necessary that the sample-size used in a study be justified through statistical power-calculation. As Moore et al. (1998) (71) explain, the effectiveness of a particular intervention on a small number of people may not be generalizable to a larger population. Unfortunately, the majority of the included studies (n=13) did not calculate statistical power prior to conducting their studies, meaning that their sample-size may have been too small (69). However, two studies calculate statistical power (57,58). In addition, a lack of statistical power-calculation may introduce a type II experimental error. This error can lead to incorrect reporting of significant

variation between control and treatment groups, which in turn can result in the rejection of valid hypotheses (72).

It is important that the spectrum of patients in a study represent the full population of patients for whom the health intervention is designed. Nonetheless, the spectrum bias may be introduced by differences in demographic and clinical features (73). The studies included in the SR all exclusively include patients suffering from FAI. To explain the patient variation that appears in the included studies, it is necessary that all of the studies clearly indicate the age and gender of the patients involved. Notably, all included studies have more men than women, which may cause gender bias (74).

It is important when conducting research to perform initial assessments and to list baseline characteristics in order to provide a clear image of participants, levels of severity and differences between individuals in each group (75). All researchers in this SR carried out patient assessments and baseline recording prior to the study.

4.2 Strengths of Review

Firstly, to the best of this author's knowledge, the current SR is the first of its kind investigating the effectiveness of virtual-reality for AFI. Secondly, based on the design of this SR, specific questions were defined, and the relevant protocols were produced. The thorough searching strategy used was clearly and robustly conducted, and discussed with the supervisor and librarian. The eligibility criteria for studies were clearly defined as well. One of the strengths was also that two reviewers independently screened and assessed the quality of the included studies, and a health librarian was used to validate the search strategy. Any disagreement related to the study's suitability was resolved by discussion between researcher and supervisor until agreement was achieved.

Lastly, the other strength of this SR is that effect-sizes have been calculated. This is useful when trying to assist therapists consider the effectiveness of interventions and not only depending on the statistical significance of outcomes when making decisions. However, one-case-report

(60) study was not able to calculate effect-size due to the unavailable mean and standard deviation data. Most studies believed that virtual reality has a moderate to high effectiveness for the AFI.

4.3 Limitations of Review

One of the limitations of this SR was that it included a low-quality study (case-report) in the synthesis and summary of evidence. Another limitation is that the overall quality assessment of included studies showed moderate results. This was most commonly due to the randomised bias, allocation concealment and evidence for generalisation. Also, the sample-size in some of the included studies was relatively small. The majority of the included studies (n=13) (22,42–47,53–56,59,60), did not conduct power-calculation, and they did not complete long follow ups.

4.4 Clinical Implications

There are several clinical implications in relation to this study. Firstly, this SR provides clear evidence for clinicians regarding the effectiveness of virtual-reality for AFI. It also provides comprehensive knowledge about the condition as well as virtual reality. A further important implication is that this SR has critically appraised the current evidence about virtual-reality for the treatment of AFI. Finally, based on the results of this SR, clinicians should include all four outcome measures discussed when virtual-reality is selected for AFI, in order to achieve best results.

4.5 Recommendations for Future Research

Further research should involve more RCTs with a larger sample-size and higher-quality, in order to effectively assess the effectiveness of virtual-reality on AFI, thus investigating its benefits in clinical trials more efficiently. These studies should include all important outcome measures for AFI populations, which are balance, strength and functional. Finally, there is also a need to determine the most suitable form of virtual-reality to be offered to patients as standard.

5.1 Conclusion

The virtual-reality group showed a statistically significant improvement in most of the outcomes compared to the other group in the majority of the included studies (n=12). However, three RCT's studies found no significant differences (44,57,58). Studies varied in terms of quality and most studies were of moderate-quality, but the author observed that there was no relationship between study quality and results, therefore virtual-reality can improve FAI. In most studies, patients considered the virtual-reality experience to offer benefits, and to be effective.

Overall, regardless of the virtual-reality used in the included studies, this review has generated moderate evidence to suggest that there were significant differences between virtual-reality and other treatments in the provision of balance, strength, pain and functional improvement. However, clinicians and researchers should take into consideration the variation in quality of the included studies. Future studies are recommended to investigate the effect of virtual-reality on participation level with large sample-sizes and an adequate follow-up period.

References

1. Vuurberg G, Hoorntje A, Wink LM, Van Der Doelen BFW, Van Den Bekerom MP, Dekker R, et al. Diagnosis, treatment and prevention of ankle sprains: update of an evidence-based clinical guideline. Br J Sports Med. 2018;52(15):956.
2. Junge A, Dvorak J. Injuries in female football players in top-level international tournaments. Br J Sports Med. 2007;41(suppl 1):i3–7.
3. Ekstrand J, Tropp H. The incidence of ankle sprains in soccer. Foot Ankle. 1990;11(1):41–4.
4. Larsen E, Jensen PK, Jensen PR. Long-term outcome of knee and ankle injuries in elite football. Scand J Med Sci Sports. 1999;9(5):285–9.
5. Smith RW, Reischl SF. Treatment of ankle sprains in young athletes. Am J Sports Med. 1986;14(6):465–71.
6. Anandacoomarasamy A, Barnsley L. Long term outcomes of inversion ankle injuries. Br J Sports Med. 2005;39(3):e14–e14.
7. Braun BL. Effects of ankle sprain in a general clinic population 6 to 18 months after medical evaluation. Arch Fam Med. 1999;8(2):143.
8. Konradsen L, Bech L, Ehrenbjerg M, Nickelsen T. Seven years follow-up after ankle inversion trauma. Scand J Med Sci Sports. 2002;12(3):129–35.
9. Bridgman SA, Clement D, Downing A, Walley G, Phair I, Maffulli N. Population based epidemiology of ankle sprains attending accident and emergency units in the West Midlands of England, and a survey of UK practice for severe ankle sprains. Emerg Med J. 2003;20(6):508–10.
10. Nyska M, Mann G. The unstable ankle. Human Kinetics; 2002.
11. Verhagen RAW, De Keizer G, Van Dijk CN. Long-term follow-up of inversion trauma of the ankle. Arch Orthop Trauma Surg. 1995;114(2):92–6.
12. Schaap GR, De Keizer G, Marti K. Inversion trauma of the ankle. Arch Orthop Trauma Surg. 1989;108(5):273–5.
13. Freeman MAR. Instability of the foot affer injuries to the lateral ligament of the ankle. J Bone Joint Surg Br. 1965;47(4):669–77.

14. Freeman MAR, Dean MRE, Hanham IWF. The etiology and prevention of functional instability of the foot. J Bone Joint Surg Br. 1965;47(4):678–85.
15. Kaminski TW, Hartsell HD. Factors contributing to chronic ankle instability: a strength perspective. J Athl Train. 2002;37(4):394.
16. Ross SE, Guskiewicz KM. Examination of static and dynamic postural stability in individuals with functionally stable and unstable ankles. Clin J Sport Med. 2004;14(6):332–8.
17. Willems T, Witvrouw E, Verstuyft J, Vaes P, De Clercq D. Proprioception and muscle strength in subjects with a history of ankle sprains and chronic instability. J Athl Train. 2002;37(4):487.
18. Tropp H. Commentary: functional ankle instability revisited. J Athl Train. 2002;37(4):512.
19. Tropp H, Odenrick P, Gillquist J. Stabilometry recordings in functional and mechanical instability of the ankle joint. Int J Sports Med. 1985;6(03):180–2.
20. Tropp H. Pronator muscle weakness in functional instability of the ankle joint. Int J Sports Med. 1986;7(05):291–4.
21. Fu ASN, Hui-Chan CWY. Ankle joint proprioception and postural control in basketball players with bilateral ankle sprains. Am J Sports Med. 2005;33(8):1174–82.
22. Kim KJ, Heo M. Comparison of virtual reality exercise versus conventional exercise on balance in patients with functional ankle instability: A randomized controlled trial. J Back Musculoskelet Rehabil. 2019;32(6):905–11.
23. de Vries JS, Krips R, Sierevelt IN, Blankevoort L, Van Dijk CN. Interventions for treating chronic ankle instability. Cochrane Database Syst Rev. 2011;(8).
24. Bryanton C, Bosse J, Brien M, Mclean J, McCormick A, Sveistrup H. Feasibility, motivation, and selective motor control: virtual reality compared to conventional home exercise in children with cerebral palsy. Cyberpsychology Behav. 2006;9(2):123–8.

25. Rizzo A "Skip", Kim GJ. A SWOT analysis of the field of virtual reality rehabilitation and therapy. Presence Teleoperators Virtual Environ. 2005;14(2):119–46.
26. Dascal J, Reid M, IsHak WW, Spiegel B, Recacho J, Rosen B, et al. Virtual reality and medical inpatients: A systematic review of randomized, controlled trials. Innov Clin Neurosci. 2017;14(1–2):14.
27. Corbetta D, Imeri F, Gatti R. Rehabilitation that incorporates virtual reality is more effective than standard rehabilitation for improving walking speed, balance and mobility after stroke: a systematic review. J Physiother. 2015;61(3):117–24.
28. Sveistrup H. Motor rehabilitation using virtual reality. J Neuroeng Rehabil. 2004;1(1):10.
29. Herzog W, Longino D, Clark A. The role of muscles in joint adaptation and degeneration. Langenbeck's Arch Surg. 2003;388(5):305–15.
30. Lohse KR, Hilderman CGE, Cheung KL, Tatla S, Van der Loos HFM. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PLoS One. 2014;9(3):e93318.
31. Herz NB, Mehta SH, Sethi KD, Jackson P, Hall P, Morgan JC. Nintendo Wii rehabilitation ("Wii-hab") provides benefits in Parkinson's disease. Parkinsonism Relat Disord. 2013;19(11):1039–42.
32. Liao Y-Y, Yang Y-R, Wu Y-R, Wang R-Y. Virtual reality-based Wii fit training in improving muscle strength, sensory integration ability, and walking abilities in patients with Parkinson's disease: a randomized control trial. Int J Gerontol. 2015;9(4):190–5.
33. Egger M, Davey-Smith G, Altman D. Systematic reviews in health care: meta-analysis in context. John Wiley & Sons; 2008.
34. Hemingway P, Brereton N. What is a systematic review. What is. 2009;1–8.
35. Wong LP. Summarizing research findings: systematic review and meta-analysis. Malaysian Fam physician Off J Acad Fam Physicians Malaysia. 2007;2(1):8.

36. Garg AX, Hackam D, Tonelli M. Systematic review and meta-analysis: when one study is just not enough. Clin J Am Soc Nephrol. 2008;3(1):253–60.
37. Cooper H, Booth K, Gill G. Using combined research methods for exploring diabetes patient education. Patient Educ Couns. 2003;51(1):45–52.
38. Rees K, Ebrahim S. Promises and problems of systematic reviews. Hear Drug. 2001;1(5):247–8.
39. Gerrish K, Lacey A. The research process in nursing. John Wiley & Sons; 2010.
40. Herzog MM, Kerr ZY, Marshall SW, Wikstrom EA. Epidemiology of ankle sprains and chronic ankle instability. J Athl Train. 2019;54(6):603–10.
41. Kowalska M, Owecki M, Prendecki M, Wize K, Nowakowska J, Kozubski W, et al. Aging and neurological diseases. In: Senescence-Physiology or Pathology. IntechOpen; 2017.
42. Vernadakis N, Derri V, Tsitskari E, Antoniou P. The effect of Xbox Kinect intervention on balance ability for previously injured young competitive male athletes: A preliminary study. Phys Ther Sport. 2014;15(3):148–55.
43. Kim KJ, Jun H ju. Effects of Virtual Reality Programs on Proprioception and Instability of Functional Ankle Instability. J Int Acad Phys Ther Res. 2015;6(2):891–5.
44. Kim K, Choi B, Lim W. The efficacy of virtual reality assisted versus traditional rehabilitation intervention on individuals with functional ankle instability: a pilot randomized controlled trial. Disabil Rehabil Assist Technol [Internet]. 2019;14(3):276–80. Available from: https://doi.org/10.1080/17483107.2018.1429501
45. Kim K-J, Heo M. Effects of virtual reality programs on balance in functional ankle instability. J Phys Ther Sci. 2015;27(10):3097–101.
46. Nam S-M, Lee D-Y. Effects of Visual Feedback Closed Kinetic Chain Exercise on the Lower Limb Muscles Activity and ankle Instability in Adult men with Chronic Ankle Instability. Korean Soc Phys Med.

2019;14(1):131–8.
47. Nam SM, Kim K, Lee DY. Effects of visual feedback balance training on the balance and ankle instability in adult men with functional ankle instability. J Phys Ther Sci. 2018;30(1):113–5.
48. Vardell E, Malloy M. Joanna briggs institute: an evidence-based practice database. Med Ref Serv Q. 2013;32(4):434–42.
49. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. Cochrane bias methods group; cochrane statistical methods group. Cochrane Collab tool Assess risk bias randomised trials BMJ. 2011;343(7829):d5928.
50. Ma L-L, Wang Y-Y, Yang Z-H, Huang D, Weng H, Zeng X-T. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res. 2020;7(1):1–11.
51. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol. 2012;8(1):23.
52. Law MC, MacDermid J. Evidence-based rehabilitation: A guide to practice. Slack Incorporated; 2008.
53. Kim J-H, Uhm Y-H. Effect of Ankle Stabilization Training Using Biofeedback on Balance Ability and Lower Limb Muscle Activity in Football Players with Functional Ankle Instability. J Korean Phys Ther. 2016;28(3):189–94.
54. Kim KJ, Jun HJ, Heo M. Effects of Nintendo Wii Fit Plus training on ankle strength with functional ankle instability. J Phys Ther Sci. 2015;27(11):3381–5.
55. Baek J-S, Kim Y-J, Kim H-J, Park J-H, Lee N-R, Lee B-R, et al. The Effect of Unstable Supporting Exercise in Young Adults with Functional Ankle Instability when Training with a Virtual {Reality-Head} Mounted Display System. PNF Mov. 2019;17(1):81–92.
56. Yang D-J, Park S-K, Uhm Y-H. Impact of Virtual Reality Based Neuromuscular Postural Control Fusion Training on Balance Ability and Jump Performance of Soccer Players with Functional Ankle Instability. J Digit Converg. 2016;14(11):357–67.

57. Punt IM, Ziltener JL, Monnin D, Allet L. Wii Fit™ exercise therapy for the rehabilitation of ankle sprains: Its effect compared with physical therapy or no functional exercises at all. Scand J Med Sci Sports. 2016;26(7):816–23.
58. Punt IM, Armand S, Ziltener JL, Allet L. Effect of Wii Fit™ exercise therapy on gait parameters in ankle sprain patients: A randomized controlled trial. Gait Posture [Internet]. 2017;58:52–8. Available from: http://dx.doi.org/10.1016/j.gaitpost.2017.06.284
59. Mohammadi N, Hadian MR, Olyaei G. The Effects of Wii Fit Plus Training on Functional Ability in Athletes with Functional Ankle Instability. Sport Orthop Traumatol [Internet]. 2020;36(1):52–9. Available from: https://doi.org/10.1016/j.orthtr.2020.01.003
60. Arvinen-Barrow M, Maresh N, Earl-Boehm J. Functional Outcomes and Psychological Benefits of Active Video Games in the Rehabilitation of Lateral Ankle Sprains: A Case Report. J Sport Rehabil. 2020;29(2):213–24.
61. Kim J-H, Uhm Y-H. Effect of ankle stabilization training using biofeedback on balance ability and lower limb muscle activity in football players with functional ankle instability. J Korean Phys Ther. 2016;28(3):189–94.
62. Kim KJ, ju Jun H. Effects of virtual reality programs on balance in functional ankle instability. J Int Acad Phys Ther Res. 2015;6(2):891–5.
63. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. Bmj. 2010;340.
64. Ross SE, Guskiewicz KM, Gross MT, Yu B. Assessment tools for identifying functional limitations associated with functional ankle instability. J Athl Train. 2008;43(1):44–50.
65. Kim K-J, Kim Y-E, Jun H-J, Lee J-S, Ji S-H, Ji S-G, et al. Which treatment is more effective for functional ankle instability: strengthening or combined muscle strengthening and proprioceptive exercises? J Phys Ther Sci. 2014;26(3):385–8.
66. Organisation WH. International classification of functioning, disability and health. Geneva World Haelth Organ 200l. 2001;

67. Gumaa M, Youssef AR. Is Virtual Reality Effective in Orthopedic Rehabilitation? A Systematic Review and Meta-Analysis. Phys Ther. 2019;99(10):1304–25.
68. Kuyper BJ. Bringing up scientists in the art of critiquing research. Bioscience. 1991;41(4):248–50.
69. Hendry C. Research Methods in Health. Investigating Health and Health Services. Nurse Res. 2002;10(2):81–3.
70. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001;134(8):663–94.
71. Moore RA, Gavaghan D, Tramer MR, Collins SL, McQuay HJ. Size is everything–large amounts of information are needed to overcome random effects in estimating direction and magnitude of treatment effects. Pain. 1998;78(3):209–16.
72. Polgar S, Thomas S. Introduction to research in the health sciences. Churchill Livingstone. London Published by RDC Group Limited; 2000.
73. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3(1):25.
74. Holdcroft A. Gender bias in research: how does it affect evidence based medicine? SAGE Publications Sage UK: London, England; 2007.
75. Al-Jundi A, Sakka S. Critical appraisal of clinical research. J Clin diagnostic Res JCDR. 2017;11(5):JE01.