



Most ankle sprain research is either false or clinically unimportant: A 30-year audit of Randomized Controlled Trials

Bleakley, C. M., Matthews, M., & Smoliga, J. (2020). Most ankle sprain research is either false or clinically unimportant: A 30-year audit of Randomized Controlled Trials. *Journal of Sport and Health Science*. <https://doi.org/10.1016/j.jshs.2020.11.002>

[Link to publication record in Ulster University Research Portal](#)

Published in:

Journal of Sport and Health Science

Publication Status:

Published online: 11/11/2020

Document Version

Author Accepted version

General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk.

1 **Most ankle sprain research is either false or clinically unimportant: A 30-year audit**
2 **of Randomized Controlled Trials**

3

4 Bleakley CM PhD ¹

5 Matthews M PhD ²

6 Smoliga JM PhD ³

7

8 ¹School of Health Science, Ulster University, County Antrim, Northern Ireland

9 ²School of Sport, Ulster University, County Antrim, Northern Ireland

10 ³Department of Physical Therapy, Congdon School of Health Science, High Point
11 University, North Carolina, USA

12

13 **Corresponding author**

14 Dr Chris Bleakley

15 Room 1D117

16 School of Health Science,

17 Ulster University,

18 County Antrim,

19 Northern Ireland

20 Tel: +442890366025

21 c.bleakley@ulster.ac.uk

22

23

24 **Abstract**

25 Lateral ankle sprain (LAS) is the most common musculoskeletal injury. Although clinical
26 research in this field is growing, there is a broader concern that clinical trial outcomes are
27 often false and fail to translate into patient benefits. The aim of this review was to audit
28 30 years of experimental research related to LAS management (n=74 RCT) and to
29 determine if reports of treatment effectiveness could be validated beyond statistical
30 certainty. Seventy-seven percent of trials reported positive treatment effects but there
31 was a high risk of false discovery. Most trials were unregistered and relied solely on
32 statistical significance, or lack of statistical significance, rather than interpreting key
33 measures of minimum clinical importance (eg. minimal detectable change, minimal
34 clinically important difference). Future clinical trials must adopt higher standards of
35 reporting and data interpretation. This includes consideration of the ethical responsibility
36 to preregister their research; and interpretation of clinical outcomes beyond statistical
37 significance.

38

39

40

41

42

43

44

45

46

47

48 **Background**

49 Lateral ankle sprain (LAS) is the most prevalent musculoskeletal injury in physically active
50 populations.¹ Although often considered innocuous, LAS has the highest re-injury rate
51 across all lower limb musculoskeletal injuries,² and the annual costs associated with
52 sports-related ankle sprain in the Netherlands is estimated at €187,200,000.³ LAS also
53 occurs frequently in the general population, with large cohorts suffering chronic
54 problems;⁴ indeed, 30⁵-75%⁶ develop a clinical condition known as chronic ankle
55 instability (CAI), characterized by recurrent injury and self-reported instability.⁵ The long-
56 term costs associated with LAS and CAI are significant^{7 8} and relate to lower quality of
57 life,⁹ physical inactivity⁴ and an increased risk of post-traumatic ankle osteoarthritis.^{5 10-}
58 12,13

59 Randomized controlled trials (RCTs) are currently considered to be the gold standard
60 methodology for determining treatment superiority.¹⁴ The first RCT involving acute LAS
61 was published in 1972.¹⁵ The Physiotherapy Evidence Database (PEDro) now archives
62 over 150 RCTs involving patients with LAS or CAI, and a 2017 meta-evaluation¹⁶ in this
63 field included 46 systematic reviews. Having access to high volumes of experimental
64 research should improve the quality of healthcare, but there is much concern that many
65 clinical trial outcomes are either false^{17,18} or they fail to translate into clinical benefits for
66 patients.¹⁹ False discovery in science (eg. erroneously claiming a treatment is effective)
67 often occurs due to over reliance on frequentist reasoning and *p*-value thresholds;²⁰ a
68 problem further compounded by unplanned multiple testing, selected reporting, and
69 confirmation bias.²¹

70 Recently we introduced a four-point checklist (FAIR), which aims to validate experimental
71 research beyond statistical certainty.²² The checklist assesses the following criteria:
72 **False Positive Risk (FPR)**, which is ‘the probability of observing a statistically significant
73 *p*-value and declaring that an effect is real, when it is not’.²³ **A priori registration**, which
74 is essential for controlling the ‘degrees of freedom’ researchers have during data analysis
75 and reporting,²¹ **thereby reducing the risk of false positive findings. Clinical Importance**,
76 whereby the magnitude of treatment effect is compared to relevant minimal detectable
77 change (MDC) and minimal clinically important difference (MCID)²⁴ data. And finally,
78 **Replication**, which should underpin all scientific discovery.

79 Evidence based health care relies on the production of valid experimental data that
80 translates into clinical benefits. This review examines the validity of conclusions from 30
81 years of clinical trials into one of the most common musculoskeletal injuries - LAS and
82 CAI. Our primary objective was to examine the extent to which reports of treatment
83 effectiveness in this field, could be validated beyond statistical certainty. The FAIR
84 checklist²² was applied, with higher validity placed on trials presenting with: low false
85 positive risk; pre-registration; treatment effect magnitudes which exceeded relevant MDC
86 and MCID values; and the corroboration of treatment effectiveness through independent
87 replication.

88

89 **Methods**

90 *Trial selection*

91 Review methods aligned with PRISMA.²⁵ Electronic searching was undertaken
92 independently by two authors (CB, MM) on MEDLINE, and the Physiotherapy Evidence

93 Database (PEDro).^{26 27} In MEDLINE we undertook a broad search strategy based on
94 MeSH terms (ankle AND randomized controlled trial) and we used the PEDro search
95 interface to run three separate searches for clinical trials using the terms ‘ankle sprain’,
96 ‘chronic ankle instability’, and ‘CAI’. Citation tracking was also undertaken using a recent
97 meta-evaluation.¹⁶ To be eligible for inclusion, **trials** must have met the following criteria:
98 a randomized controlled design; participants with LAS and/or CAI managed with at least
99 one conservative treatment intervention; assessment of at least one clinically relevant
100 outcome measure (eg. pain, function, range of motion, strength, balance). **Trials** were
101 excluded if they **involved** any surgical intervention. No restrictions were placed on injury
102 severity, participant demographics or follow-up duration. We did not include RCTs using:
103 >2 treatment arms, equivalency or non-inferiority **trials**, pilot **trials** or **trials** published prior
104 to 1990. Any disagreements in trial selection were resolved through consensus with a
105 third reviewer (JS).

106

107 *Data extraction and analysis*

108 PICO (population, intervention, comparison, outcome) characteristics were extracted
109 from the full text of all eligible trials, in addition to aims and hypothesis, **n** participants,
110 follow-up time points, and the total number of **between-group** statistical comparisons
111 undertaken. Included trials were then classified as being either statistically significant or
112 null. A statistically significant trial was defined as a trial having a *p*-value less than 0.05 in
113 the trial results tab for any clinical outcome.²⁸ We also calculated the proportion of
114 between-group comparisons that resulted in statistically significant findings within each
115 individual trial, and whether they were recorded in primary or secondary outcome

116 measures. When trials included multiple outcome measures but did not clearly specify a
117 'primary' outcome, the primary outcome was determined by the authors based on the
118 nature of the research question and the following definition of a primary outcome 'a
119 specific key measurement(s) or observation(s) used to measure the effect of experimental
120 variables in a trial.'²⁹ The FAIR checklist²² was applied as follows:

121

122 False Positive Risk

123 Calculation of FPR followed methods used in a previous research audit in this field.³⁰ FPR
124 calculation is a special case of Bayesian analysis. It allows the p-value to be
125 supplemented by a single number that gives a much better idea of the strength of the
126 evidence than a p-value alone.²³ We calculated FPR for all trials reporting a statistically
127 significant finding from their primary outcome. All FPR calculations were performed using
128 the False Positive Risk Web Calculator (version 1.5) using the following data: the *n* of
129 participants in each group; a relevant *p*-value; and the corresponding effect size (Hedges
130 *g*).³¹ Further details of the analysis script and simulated examples of FPR calculations
131 can be found in Colquhoun's recent articles.^{20 23} If a trial reported a *p*-value threshold
132 such as $p < 0.05$, rather than an exact *p*-value, we assumed that the *p*-value was one
133 decimal place below the threshold value (e.g. $p < 0.05$ was inputted as 0.049). The
134 calculation of FPR also requires an estimation of the prior probability that there is a real
135 effect [P(H1)] for a given treatment. In all trials, we initially assumed that P(H1) was 0.5 –
136 ie. treatment interventions had a 50:50 chance of a (positive) real effect before the
137 experiment was done.^{18 20} In all cases FPR estimations were calculated using the p-
138 equals method, as our aim was to interpret a single *p*-value from a single experiment

139 (rather than trying to estimate the long term error rate).³¹ Descriptive statistics were used
140 to determine the median FPR and the number (%) of statistically significant p -values
141 associated with FPR less than 5%.

142

143 A Priori trial registration

144 We determined the number (%) of eligible trials reporting preregistration; defined as the
145 trial protocol being publicly available within a trial registry (e.g.ClinicalTrials.gov) prior to
146 the initiation of participant recruitment. In a secondary analysis, we used odds ratios
147 (ORs) and 95% confidence intervals (95% CI) to determine whether the likelihood of
148 reporting a statistically significant outcome was influenced by *a priori* trial registration.

149

150 Clinical Importance

151 Initially, we determined the number (%) of trials that referenced or reported MDC and/or
152 MCID values within the full text manuscript. When enough data were available, we
153 calculated the mean differences (MD) and 95% confidence intervals (CI) for each clinical
154 outcome, where $MD = \text{mean}_{\text{experimental}} - \text{mean}_{\text{control}}$. MD (95% CI) data were then
155 compared to corresponding MDC and MCID data. If a trial did not report MDC or MCID
156 data for a particular outcome, we searched the literature for relevant figures and inputted
157 them. MDC was set at confidence levels of 95% and considered to be ‘the amount of
158 change that must be observed before it is considered above the bounds of measurement
159 error’.³² MCID was considered to be ‘the smallest change that would be important to
160 patients’, and could have been quantified by externally referenced (anchor) or internally
161 referenced (distribution) methods.³³

162

163 Replication

164 PICO criteria were compared across trials. If possible, homogeneous trials were sub
165 grouped and their trial effects (magnitude and direction) were compared to screen for
166 successful replication.

167

168

169 **Results**

170 We screened 1098 titles and abstracts (937 from Medline and 161 from PEDro), with 169
171 selected for full-text review. n=74 RCTs were eligible for inclusion (Supplemental data 1),
172 with the remainder (n=95) excluded (>2 treatment arms (n=45); no clinical outcomes
173 (n=9); non RCT (n=8); non English language (n=8) surgical intervention (n=7); non
174 inferiority / equivalency (n=5), non-ankle sprain/CAI (n=5); other (n=8) (Figure 1). Trials
175 included participants with either LAS (n=53 trials) or CAI (n=21 trials). In most trials, the
176 primary intervention involved external supports (n=30), exercise intervention (n=18),
177 pharmacotherapy (n=14) manual therapy or electro-physical agents (n=11). The mean
178 sample size was n=85.1 (SD=96.8; range 13-522) and 50% (37/74) reported using *a priori*
179 sample size calculation. Most sample size estimations included alpha (Type 1 error) and
180 beta (Type 2 error) levels of 5% and 20% respectively, with the average effect size
181 estimated at 0.7 (SD=0.45) a priori.

182 **Insert Figure 1 here.**

183

184 Twenty-three percent (17/74) of RCTs were classed as null (no treatment effects
185 reported). The remaining 77% (57/74) reported statistically significant findings from at
186 least one outcome measure. We extracted an aggregate of 966 p -values relating to
187 between-group statistical comparisons involving primary or secondary outcomes, of
188 which 35.4% (342/966) were statistically significant ($p < 0.05$) (Figure 2A). Most statistically
189 significant findings were derived from secondary outcomes, with just 17% (58/342)
190 derived from primary outcome measures (Figure 2B). Out of the 966 p -values reported in
191 the literature, only 11 (1%) represented statistically significant findings in a primary
192 outcome measure reported from a pre-registered trial (Figure 2C). (Supplemental data 2)

193 **Insert Figure 2 here**

194

195 False positive risk

196 Enough data were available to calculate effect sizes and FPR in 68% of trials (39/57)
197 reporting significant effects ($p < 0.05$) in their primary outcome. FPR is summarized in
198 Figure 3; the median FPR was 14% (range 0.6 to 100%) and 28% of trials (11/39) had
199 FPR less than 5%. (also see Supplemental data 3)

200 **Insert Figure 3**

201

202 A Priori trial registration

203 Only 19% (14/74) of trials were preregistered. The average number of between-group
204 comparisons reported across registered and unregistered trials was similar [12.8 (SD 9.0)
205 vs 13.3 (SD 10.9) respectively], however unregistered trials were more likely to report p -
206 values less than 0.05 (OR=1.7 Cis: 1.2 to 2.4; $p = .004$).

207

208 Clinical importance

209 Of the 57 trials reporting statistical significance, only 9% (5/57) made any reference to
210 either MDC or MCID values. In a further 16 trials, we were able to extract relevant MDC
211 and/or MCID values extracted from the existing literature, for the following outcomes
212 measures: Foot and ankle outcome measure (FAAM);^{34 35} Cumberland ankle instability
213 tool (CAIT);³⁶ Lower extremity functional scale (LEFS);³⁷ isometric / isokinetic ankle
214 strength;^{38 39} limb circumference / swelling;^{40 41} range of motion;^{38 42} postural control;²⁷
215 pain⁴³. Effect magnitudes (MD) exceeded the respective MDC or MCID values in 12 and
216 7 trials respectively. Effect magnitudes exceeded both MDC and MCID in just 3 trials (also
217 see Supplemental data 3)

218

219 Replication

220 Figure 4 summarizes the number of trials meeting **more** than one of the FAIR criteria.
221 Three trials were both pre-registered and reported a low FPR (<5%), and one of the pre-
222 registered trials also reported a clinically important effect. No trial met all the following
223 conditions: preregistered; low false positive risk (<5%); clear evidence that the magnitude
224 of treatment effect exceeded both MDC and MCID values. There were no instances when
225 a positive treatment effect was independently replicated.

226 **Insert Figure 4 here**

227

228 **Discussion**

229 There is concern that a large proportion of scientific research is based on false positive,
230 non-replicable conclusions.¹⁷ Strategies known to reduce the risk of false discovery
231 include: mandatory trial registration;²¹ false positive risk calculation,²⁰ and use of MDC
232 and MCID values to determine if reported treatment magnitudes are clinically
233 meaningful.^{22 24} There is a dearth of empirical meta-research investigating the credibility
234 of research practices in SEM research. Recent audits have highlighted a high propensity
235 for questionable research practices (eg. HARKing, cherry picking, p-hacking) in high
236 impact SEM journals;⁴⁴ and we have previously found a high risk of false positive claims
237 in the sports physiotherapy literature.³⁰ This is the first piece of meta-research using a
238 saturation of RCTs from a single field of musculoskeletal medicine. n=74 trials met our
239 inclusion criteria, with 77% reporting statistically significant findings from at least one
240 outcome measure. However, in most trials, data interpretation was limited to all or nothing
241 Null Hypothesis Significance Testing, and most positive conclusions could not be
242 validated beyond statistical certainty.

243 Only 19% of trials in the LAS/CAI research literature were preregistered. Trial registration
244 is now required as a condition of ethical approval,⁴⁵ and audits of clinical trials undertaken
245 in other fields of medicine (cardiology, rheumatology, and gastroenterology), show better
246 adherence to current guidelines.⁴⁶ One of our key findings was that unregistered trials
247 were 70% more likely to report statistical significance (OR=1.7 Cis: 1.2-2.4) compared to
248 those that were registered *a priori*. Unregistered trials typically carry a higher risk of false
249 discovery due to: significance seeking, selective reporting of outcomes,⁴⁷ or HARKing
250 (hypothesizing after the results are known).²¹ In contrast, preregistration helps to control
251 the 'degrees of freedom' a researcher has during data analysis and reporting,²¹ reducing

252 such risks. A related finding was that out of the 342 statistically significant p -values
253 (<0.05) reported across trials, only 11 were generated from primary outcomes within pre-
254 registered trials. Consequently, the vast majority of statistically significant findings within
255 the LAS/CAI evidence base, are derived from secondary outcomes in unregistered trials,
256 and should therefore be considered exploratory or hypothesis generating.²¹

257
258 Measures of minimum clinical importance, (MDC and MCID) are increasingly recognized
259 as important thresholds for evaluating the efficacy of an intervention. However, the
260 reporting of clinical significance is poor in RCTs involving patients with LAS or CAI, with
261 just 9% of trials, referring to MDC or MCID data. After extracting MDC and MCID for
262 clinical outcomes relating to pain, function, instability, strength and swelling, we were able
263 to examine clinical efficacy in 21 trials; however, the results were disappointing with 50%
264 of trials recording treatment effects which could not be differentiated from measurement
265 error. Furthermore, in most trials, the treatment effects did not exceed relevant MCID
266 figures, and are therefore unlikely to be considered important by patients with LAS and
267 CAI. An initial audit⁴⁸ of interventional research in the sports medicine literature, found
268 that MDC or MCID was considered in 53% and 40% of trials respectively. However, a
269 much larger audit of orthopaedic literature, found that only 7.5% of clinical science articles
270 made reference to MCID,²⁴

271
272 It is expected that musculoskeletal injuries are managed from an evidence-based
273 perspective, whereby the best available evidence is integrated with patient preference,
274 clinical expertise, and the clinical context. As RCTs represent the gold standard

275 methodology for determining treatment superiority, they have a considerable influence on
276 the relevance of adopting an evidence-based framework when treating patients with LAS
277 or CAI. Our results raise fundamental questions about the current value of **evidence-**
278 **based practice** in this field and clarify that future clinical trials must adopt higher standards
279 of reporting and data interpretation. Interestingly, there is a lack of robust clinical
280 interpretation in other fields of medicine,⁴⁹ and continuing to rely solely on NHST, not only
281 wastes research funding, but erodes credibility and slows down scientific progress.⁵⁰
282 Although NHST remains an important step for determining treatment effectiveness, it is
283 most efficient in the context of long-run repeated testing.⁵⁰ We support the idea that *p*-
284 values are supplemented with a formal estimation of the false positive risk^{18 31} which
285 represents “the probability, in the light of the *p*-value that you observe, you declare that
286 an effect is real, when in fact, it isn’t.”²³ Although it is often assumed that the FPR is equal
287 to the reported *p*-value, they are different constructs and often vary considerably. Indeed,
288 our audits shows that the median FPR associated with statistically significant findings
289 (*p*<0.05) was 14% (range 0.6-100%), and only 27% of trials had a FPR lower than 5%.
290 **These figures suggest** that statistical significance alone is not a solid foundation for
291 determining treatment effect, particularly when it is based on binary thresholds (*p*<0.05).

292

293 **Limitations**

294 Higher validity was assumed under the following conditions: derived from registered trials;
295 low false positive risk; treatment effects exceeding MDC and MCID values. This is not an
296 exhaustive list and we did not fully consider false discoveries relating to multiple treatment
297 arms, the analysis of multiple outcomes, or multiple analyses of the same outcome at

298 different times.⁵¹ We acknowledge although preregistration increases the transparency
299 and validity of trial conclusions, it is not a [cure-all](#) for efficient and accurate dissemination.
300 Audits of [clinicaltrials.gov](#) show that approximately 20% of registered trials disseminate
301 their results within 1 year of completion,⁵² with others highlighting quite a high risk of
302 discordance between the original registry data and the published data.⁵³

303 We must also consider that our FPR calculations were based on assumptions that the
304 prior probability of effect was 50%, but it is likely that some trials were underpinned by
305 more extreme hypotheses. [In previous data simulations,²⁸ we have shown that](#) a positive
306 conclusion from an optimistic research question (*i.e.* a higher prior probability) is likely to
307 be correct; whereas an unlikely hypotheses (where researchers are driven by pursuit of
308 novelty) will have a much higher risk of false-positive reporting. [Alternatives to FPR have](#)
309 [been discussed by Colquhoun.²³ Perhaps the most clinically intuitive option is use of a](#)
310 [reverse Bayesian approach,⁵⁴ where the observed p-value is used to calculate the prior](#)
311 [probability required to achieve a specific or minimal false positive risk \(eg. 5%\). This then](#)
312 [allows the researcher to determine whether the calculated prior is plausible or not.³⁰](#)
313 Finally, many latent constructs influence false discovery; this includes a scientific culture
314 which places most value on statistically significant findings or novel discoveries.²¹

315

316 **Conclusion**

317 There is a high risk of false positive discovery in a core field of musculoskeletal research.
318 A key concern is that most of the research in this field remains unregistered, and relies
319 [solely](#) on statistical significance, or lack of statistical significance, rather than interpreting
320 the magnitude of change. Researchers must consider the ethical responsibility to

321 preregister their research; and their interpretation of clinical outcomes must evolve
322 beyond statistical significance.

323

324 **Author Contributions**

325 CB and JS conceived of the presented idea. CB and MM planned and undertook the review.

326 CB and JS extracted data. CB undertook much of the analysis and JS verified the
327 analytical methods.

328 All authors discussed the results and contributed to the final manuscript.

329

330 **Competing interests**

331 Authors have no competing interests to declare

332

333 **References**

334

- 335 1. Gribble PA, Bleakley CM, Caulfield BM, et al. 2016 consensus statement of the International Ankle
336 Consortium: prevalence, impact and long-term consequences of lateral ankle sprains. *Br J Sports*
337 *Med* 2016;50(24):1493-95. doi: 10.1136/bjsports-2016-096188
- 338 2. Hootman JM, Dick R, Agel J. Epidemiology of collegiate injuries for 15 sports: summary and
339 recommendations for injury prevention initiatives. *J Athl Train* 2007;42(2):311-9.
- 340 3. Hupperets MD, Verhagen EA, Heymans MW, et al. Potential savings of a program to prevent ankle
341 sprain recurrence: economic evaluation of a randomized controlled trial. *Am J Sports Med*
342 2010;38(11):2194-200. doi: 10.1177/0363546510373470
- 343 4. Hiller CE, Nightingale EJ, Raymond J, et al. Prevalence and impact of chronic musculoskeletal ankle
344 disorders in the community. *Arch Phys Med Rehabil* 2012;93(10):1801-7. doi:
345 10.1016/j.apmr.2012.04.023
- 346 5. Gribble PA, Bleakley CM, Caulfield BM, et al. Evidence review for the 2016 International Ankle
347 Consortium consensus statement on the prevalence, impact and long-term consequences of
348 lateral ankle sprains. *Br J Sports Med* 2016;50(24):1496-505. doi: 10.1136/bjsports-2016-096189
- 349 6. Anandacoomarasamy A, Barnsley L. Long term outcomes of inversion ankle injuries. *Br J Sports Med*
350 2005;39(3):e14; discussion e14. doi: 10.1136/bjism.2004.011676
- 351 7. Waterman BR, Owens BD, Davey S, et al. The epidemiology of ankle sprains in the United States. *J*
352 *Bone Joint Surg Am* 2010;92(13):2279-84. doi: 10.2106/JBJS.I.01537

353 8. Knowles SB, Marshall SW, Miller T, et al. Cost of injuries from a prospective cohort study of North
354 Carolina high school athletes. *Inj Prev* 2007;13(6):416-21. doi: 10.1136/ip.2006.014720
355 9. Arnold BL, Wright CJ, Ross SE. Functional ankle instability and health-related quality of life. *J Athl Train*
356 2011;46(6):634-41.
357 10. Valderrabano V, Hintermann B, Horisberger M, et al. Ligamentous posttraumatic ankle
358 osteoarthritis. *Am J Sports Med* 2006;34(4):612-20. doi: 10.1177/0363546505281813
359 11. Hintermann B, Boss A, Schafer D. Arthroscopic findings in patients with chronic ankle instability. *Am J*
360 *Sports Med* 2002;30(3):402-9. doi: 10.1177/03635465020300031601
361 12. Hashimoto T, Inokuchi S. A kinematic study of ankle joint instability due to rupture of the lateral
362 ligaments. *Foot Ankle Int* 1997;18(11):729-34. doi: Doi 10.1177/107110079701801109
363 13. Wikstrom EA, Hubbard-Turner T, McKeon PO. Understanding and treating lateral ankle sprains and
364 their consequences: a constraints-based approach. *Sports Med* 2013;43(6):385-93. doi:
365 10.1007/s40279-013-0043-z
366 14. Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials
367 important? *BMJ* 1998;316(7126):201. doi: 10.1136/bmj.316.7126.201
368 15. Wilson DH. Treatment of soft-tissue injuries by pulsed electrical energy. *Br Med J* 1972;2(5808):269-
369 70.
370 16. Doherty C, Bleakley C, Delahunt E, et al. Treatment and prevention of acute and recurrent ankle
371 sprain: an overview of systematic reviews with meta-analysis. *Br J Sports Med* 2017;51(2):113-
372 25. doi: 10.1136/bjsports-2016-096178
373 17. Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2(8):e124. doi:
374 10.1371/journal.pmed.0020124
375 18. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. *R*
376 *Soc Open Sci* 2014;1(3):140216. doi: 10.1098/rsos.140216
377 19. Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for
378 patients. *Trials* 2017;18(1):122. doi: 10.1186/s13063-017-1870-2
379 20. Colquhoun D. The reproducibility of research and the misinterpretation of p-values. *R Soc Open Sci*
380 2017;4(12):171085. doi: 10.1098/rsos.171085
381 21. Forstmeier W, Wagenmakers EJ, Parker TH. Detecting and avoiding likely false-positive findings - a
382 practical guide. *Biol Rev Camb Philos Soc* 2017;92(4):1941-68. doi: 10.1111/brv.12315
383 22. Bleakley C, Smoliga JM. Validating new discoveries in sports medicine: we need FAIR play beyond p
384 values. *Br J Sports Med* 2020 doi: 10.1136/bjsports-2019-101797 [published Online First:
385 2020/06/26]
386 23. Colquhoun D. The False Positive Risk: A Proposal Concerning What to Do About *p*-Values. *The*
387 *American Statistician* 2019;73:192-201.
388 24. Copay AG, Eyberg B, Chung AS, et al. Minimum Clinically Important Difference: Current Trends in the
389 Orthopaedic Literature, Part II: Lower Extremity: A Systematic Review. *JBJS Rev* 2018;6(9):e2.
390 doi: 10.2106/JBJS.RVW.17.00160
391 25. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-
392 analyses: the PRISMA statement. *BMJ* 2009;339:b2535. doi: 10.1136/bmj.b2535 [published
393 Online First: 2009/07/21]
394 26. Kamper SJ, Moseley AM, Herbert RD, et al. 15 years of tracking physiotherapy evidence on PEDro,
395 where are we now? *Br J Sports Med* 2015;49(14):907-9. doi: 10.1136/bjsports-2014-094468
396 27. Plisky PJ, Gorman PP, Butler RJ, et al. The reliability of an instrumented device for measuring
397 components of the star excursion balance test. *N Am J Sports Phys Ther* 2009;4(2):92-9.
398 28. Ramagopalan SV, Skingsley AP, Handunnetthi L, et al. Funding source and primary outcome changes
399 in clinical trials registered on ClinicalTrials.gov are associated with the reporting of a statistically

- 400 significant primary outcome: a cross-sectional study. *F1000Res* 2015;4:80. doi:
401 10.12688/f1000research.6312.2
- 402 29. Ramagopalan S, Skingsley AP, Handunnetthi L, et al. Prevalence of primary outcome changes in
403 clinical trials registered on ClinicalTrials.gov: a cross-sectional study. *F1000Res* 2014;3:77. doi:
404 10.12688/f1000research.3784.1
- 405 30. Bleakley C, Reijgers J, Smoliga JM. Many High-Quality Randomized Controlled Trials in Sports Physical
406 Therapy Are Making False-Positive Claims of Treatment Effect: A Systematic Survey. *J Orthop*
407 *Sports Phys Ther* 2020;50(2):104-09. doi: 10.2519/jospt.2020.9264
- 408 31. Longstaff C, Colquhoun D. <http://fpr-calc.ucl.ac.uk/>. [accessed 01-02-2019].
- 409 32. Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of
410 responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal
411 Clinically Important Difference. *J Rheumatol* 2001;28(2):400-5.
- 412 33. King MT. A point of minimal important difference (MID): a critique of terminology and methods.
413 *Expert Rev Pharmacoecon Outcomes Res* 2011;11(2):171-84. doi: 10.1586/erp.11.9
- 414 34. Martin RL, Irrgang JJ. A survey of self-reported outcome instruments for the foot and ankle. *J Orthop*
415 *Sports Phys Ther* 2007;37(2):72-84. doi: 10.2519/jospt.2007.2403
- 416 35. Eechaute C, Vaes P, Van Aerschot L, et al. The clinimetric qualities of patient-assessed instruments
417 for measuring chronic ankle instability: a systematic review. *BMC Musculoskelet Disord*
418 2007;8:6. doi: 10.1186/1471-2474-8-6
- 419 36. Wright CJ, Linens SW, Cain MS. Establishing the Minimal Clinical Important Difference and Minimal
420 Detectable Change for the Cumberland Ankle Instability Tool. *Arch Phys Med Rehabil*
421 2017;98(9):1806-11. doi: 10.1016/j.apmr.2017.01.003
- 422 37. Alcock GK SP. Validation of the Lower Extremity Functional Scale on Athletic Subjects with Ankle
423 Sprains *Physiother Canada* 2002;Fall 233-40.
- 424 38. Fraser JJ, Koldenhoven RM, Saliba SA, et al. Reliability of Ankle-Foot Morphology, Mobility, Strength,
425 and Motor Performance Measures. *Int J Sports Phys Ther* 2017;12(7):1134-49.
- 426 39. Sekir U, Yildiz Y, Hazneci B, et al. Reliability of a functional test battery evaluating functionality,
427 proprioception, and strength in recreational athletes with functional ankle instability. *Eur J Phys*
428 *Rehabil Med* 2008;44(4):407-15.
- 429 40. Devoogdt N, Cavaggion C, Van der Gucht E, et al. Reliability, Validity, and Feasibility of Water
430 Displacement Method, Figure-of-Eight Method, and Circumference Measurements in
431 Determination of Ankle and Foot Edema. *Lymphat Res Biol* 2019 doi: 10.1089/lrb.2018.0045
- 432 41. Rohner-Spengler M, Mannion AF, Babst R. Reliability and minimal detectable change for the figure-
433 of-eight-20 method of, measurement of ankle edema. *J Orthop Sports Phys Ther* 2007;37(4):199-
434 205. doi: 10.2519/jospt.2007.2371
- 435 42. Searle A, Spink MJ, Chuter VH. Weight bearing versus non-weight bearing ankle dorsiflexion
436 measurement in people with diabetes: a cross sectional study. *BMC Musculoskelet Disord*
437 2018;19(1):183. doi: 10.1186/s12891-018-2113-8
- 438 43. Alghadir AH, Anwer S, Iqbal A, et al. Test-retest reliability, validity, and minimum detectable change
439 of visual analog, numerical rating, and verbal rating scales for measurement of osteoarthritic
440 knee pain. *J Pain Res* 2018;11:851-56. doi: 10.2147/JPR.S158847
- 441 44. Büttner F, Toomey E, McClean S, et al. Are questionable research practices facilitating new
442 discoveries in sport and exercise medicine? The proportion of supported hypotheses is
443 implausibly high. *Br J Sports Med* 2020 doi: 10.1136/bjsports-2019-101863 [published Online
444 First: 2020/07/22]
- 445 45. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical
446 research involving human subjects. *JAMA* 2013;310(20):2191-4. doi: 10.1001/jama.2013.281053

447 46. Mathieu S, Boutron I, Moher D, et al. Comparison of registered and published primary outcomes in
448 randomized controlled trials. *JAMA* 2009;302(9):977-84. doi: 10.1001/jama.2009.1242
449 47. John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with
450 incentives for truth telling. *Psychol Sci* 2012;23(5):524-32. doi: 10.1177/0956797611430953
451 48. Nwachukwu BU, Runyon RS, Kahlenberg CA, et al. How are we measuring clinically important
452 outcome for operative treatments in sports medicine? *Phys Sportsmed* 2017;45(2):159-64. doi:
453 10.1080/00913847.2017.1292108
454 49. Cocks K, King MT, Velikova G, et al. Quality, interpretation and presentation of European
455 Organisation for Research and Treatment of Cancer quality of life questionnaire core 30 data in
456 randomised controlled trials. *Eur J Cancer* 2008;44(13):1793-8. doi: 10.1016/j.ejca.2008.05.008
457 50. Szucs D, Ioannidis JPA. When Null Hypothesis Significance Testing Is Unsuitable for Research: A
458 Reassessment. *Front Hum Neurosci* 2017;11:390. doi: 10.3389/fnhum.2017.00390
459 51. Li G, Taljaard M, Van den Heuvel ER, et al. An introduction to multiplicity issues in clinical trials: the
460 what, why, when and how. *Int J Epidemiol* 2017;46(2):746-55. doi: 10.1093/ije/dyw320
461 52. Prayle AP, Hurley MN, Smyth AR. Compliance with mandatory reporting of clinical trial results on
462 ClinicalTrials.gov: cross sectional study. *BMJ* 2012;344:d7373. doi: 10.1136/bmj.d7373
463 53. Goldacre B, Drysdale H, Dale A, et al. COMPare: a prospective cohort study correcting and
464 monitoring 58 misreported trials in real time. *Trials* 2019;20(1):118. doi: 10.1186/s13063-019-
465 3173-2
466 54. Matthews R. Why should clinicians care about Bayesian methods? *J Stat Plan Inference* 2001;94:43-
467 58. doi: doi:10.1016/S0378-3758(00)00232-9

468

469

470

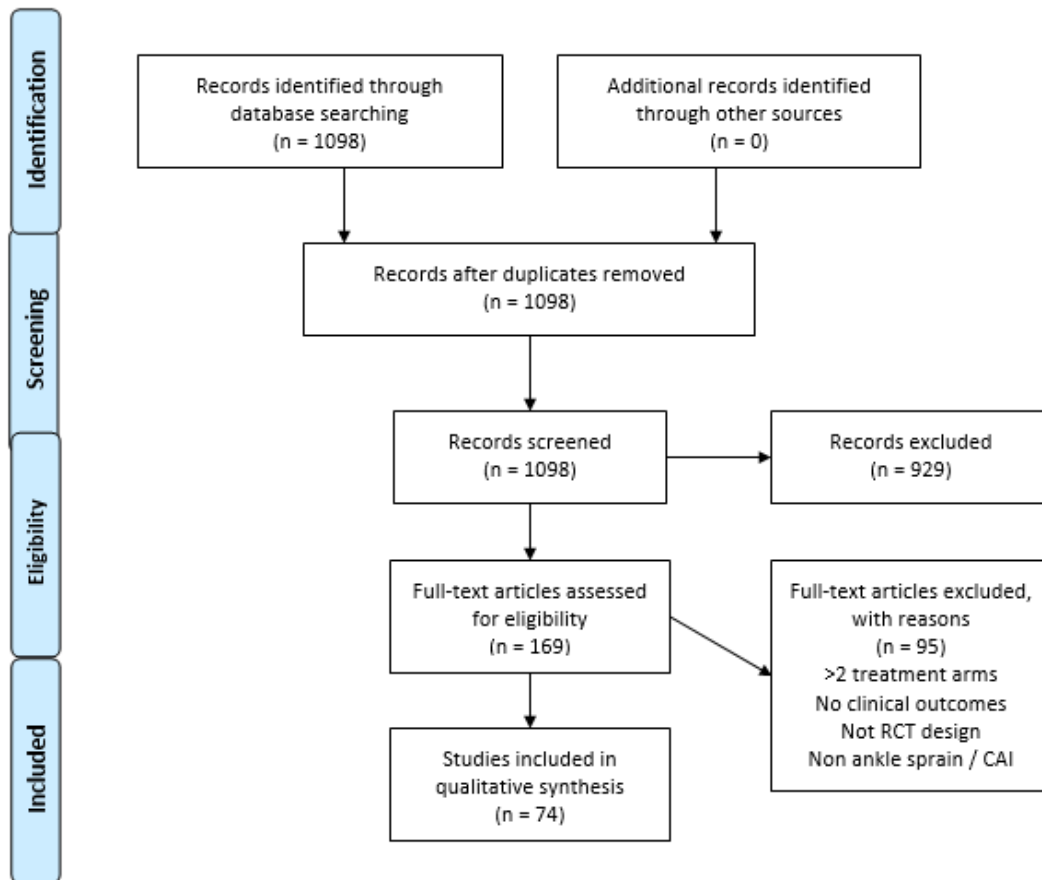
471

472

473

474 **Figure 1**

475 Flow diagram summarizing trial selection



476

477

478

479

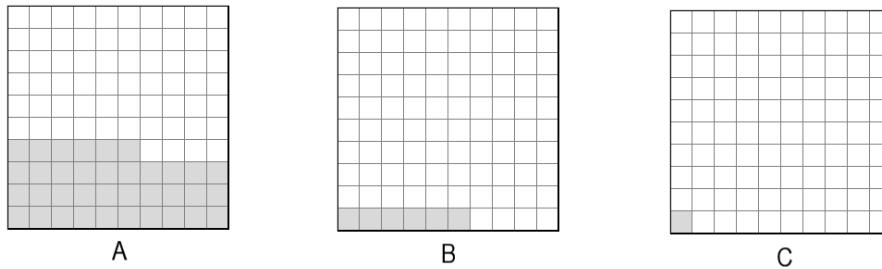
480

481

482

483 **Figure 2**
484 Area plots subgrouping p -values (n=966) by: level of significance (A), primary outcomes
485 (B) and pre-registration (C)

486



487

488 **Figure 2 footnote**

489 Each square represents ~10 p -values generated from between-group comparisons.

490 White squares = No statistical significance ($p > 0.05$)

491 Shaded squares represent:

492 A). Statistically significant – primary or secondary outcomes

493 B). Statistically significant - primary outcomes only, any trial

494 C). Statistically significant - primary outcomes, pre-registered trials only

495

496

497

498

499

500

501

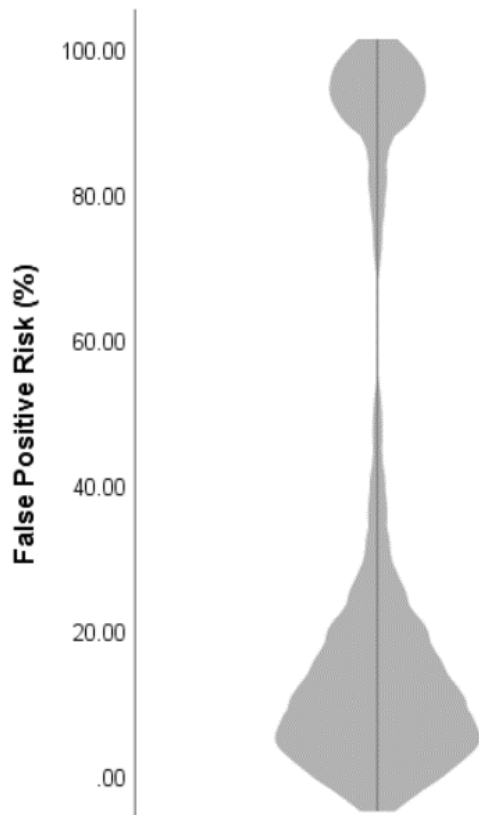
502

503

504

505 **Figure 3**

506 Violin plot summarizing False Positive Risk in trials reporting significant ($p < 0.05$) effects
507 in their primary outcome



508

509

510

511

512

513

514

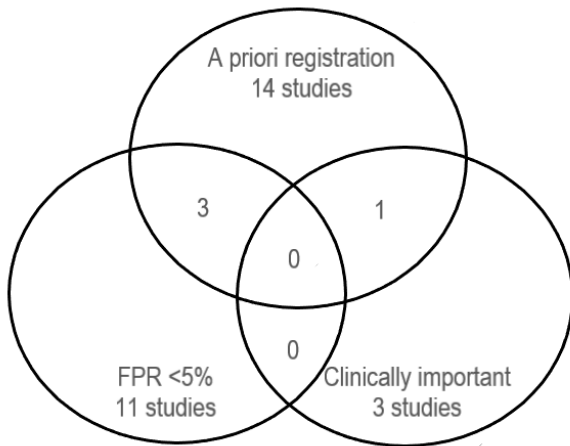
515

516

517 **Figure 4**

518 Venn diagram illustrating N trials meeting one than one FAIR criteria

519



520

521

522

523

524