

Should this systematic review and meta-analysis change my practice? Part 2: exploring the role of the comparator, diversity, risk of bias and confidence

Mervyn J Travers,^{1,2} Myles Calder Murphy,^{1,3}
James Robert Debenham,¹ Paola Chivers,⁴ Max K Bulsara,⁴
Matthew K Bagg,^{5,6} Thorvaldur Skulli Palsson,⁷ William Gibson¹

INTRODUCTION

Systematic review and meta-analyses (SRMA) represent the highest quality of evidence of the effects of interventions and have the capacity to usefully inform clinical decisions, reduce research waste and direct policy making.¹⁻⁴ Interventional SRMA often amalgamate the results of multiple randomised controlled trials (RCTs) to determine the magnitude of the estimate of the effect for a chosen intervention and to examine variation among study outcomes (heterogeneity).⁵ As such they represent the highest level of evidence to inform clinicians on the usefulness of a given treatment. Of course, the robustness of SRMA findings is dependent on the quality of the trials that have been pooled. The aim of this two-part Educational Editorial Series is to outline a process of scrutiny and analysis for SRMA to facilitate answering the important question 'Should this SRMA change my practice?'. An understanding of key features of SRMA which have been outlined in related BJSM Educational Editorials is needed by clinicians in order to make this decision.^{2,6-10} Here, we present a worked example for clinicians illustrating some

important considerations when reading an interventional SRMA. Specifically, we highlight the need to consider the pooling of comparators, clinical diversity, comparisons to active control, risk of bias and confidence in the results when interpreting interventional SRMA findings.

THE EXAMPLE SRMA

The example SRMA investigated the efficacy of platelet-rich plasma (PRP) injections for tendinopathy.¹¹ Based on data extracted from 16 RCTs (see online supplementary file 1 for a list of references), the authors reported a moderate treatment effect in favour of PRP with a standard mean difference (SMD) of 0.47 (95% CI 0.22 to 0.72, $p < 0.001$). Part 1 of this Educational Editorial series demonstrated issues of methodological reporting, transparency and reproducibility which undermined the trustworthiness of this result and we presented a revised estimate of treatment effect of PRP for tendinopathy (figure 1). Our result was derived from an incomplete dataset. Nonetheless, it suggested a moderate treatment effect in favour of PRP injections similar to the original authors'. Next, we re-examine this result considering the *choice of comparator(s)*.

COMPARED TO WHAT? CONSIDERING THE ROLE OF THE CONTROL INTERVENTION

The selection of an appropriate comparator is important in sports medicine research as we know that some interventions can negatively impact pain and function. Ideally, a SR with pairwise comparison MA (pairwise=single comparison between two interventions) would inform the clinician as to whether PRP is preferable to a single comparator (e.g. corticosteroid injection) for a given condition (e.g. Achilles tendinopathy) at a common time point of interest (e.g. 12 weeks post-injection). RCTs investigating PRP injections

for tendinopathy commonly use a range of control injections (saline, local anaesthetic or glucocorticoid) as the comparator. Accordingly, the treatment effect of PRP injections must be interpreted in light of the comparator employed in individual trials. The example SRMA included RCTs investigating PRP versus *various* control injections (saline, dry needling, local anaesthetic agents and corticosteroids) at *multiple sites* (Achilles, patella, shoulder and elbow) and at *a range of time points* post-intervention (3, 6 and 12 months).¹¹ For SRs using pairwise MA, *homogeneity of the comparator* across the included trials is a critical element of the study design. The above control injections most likely have differing underlying effect sizes based on their location and nature as well as differing temporal responses. Comparing PRP to such a heterogeneous group of interventions does not provide a clear estimate of the effect of PRP injection to any one of the comparators. On this basis alone, it is arguable that synthesis in a pairwise comparison is not appropriate at all because it falsely presumes homogeneity of the comparators. Network meta-analysis (NMA) is a more appropriate methodology in situations where multiple comparators are involved. NMA represents all comparisons between interventions (in this case, PRP and the various comparators) in a network and produces effect estimates for each of these comparisons, provided strong assumptions are justified.¹²⁻¹⁴

It is not surprising to observe significant statistical heterogeneity in the example pairwise comparison, given the difference in the comparator across trials. Heterogeneity may be defined as the proportion of variability in the results of individual studies (above that expected by chance) and is outlined in further detail in a recent BJSM Editorial.⁷ See box 1 for a brief description of evidence for heterogeneity reported in our forest plot and we refer the interested reader to the cited references.¹⁵⁻¹⁷ We now comment on the other contributors to observed heterogeneity in pairwise meta-analysis, namely, clinical diversity (differences in participant characteristics, interventions or outcomes between studies) or methodological diversity (differences in design or risk of bias between studies).

CLINICAL DIVERSITY

Two arguments could be made regarding the clinical variability between the studies included in the example SRMA. If the SRMA only included studies where, for

¹School of Physiotherapy, University of Notre Dame Australia, Fremantle, Western Australia, Australia

²School of Physiotherapy and Exercise Science, Curtin University, Perth, Western Australia, Australia

³Sports Science Sports Medicine Department, Western Australian Cricket Association, Perth, Western Australia, Australia

⁴Institute for Health Research, University of Notre Dame Australia, Fremantle, Western Australia, Australia

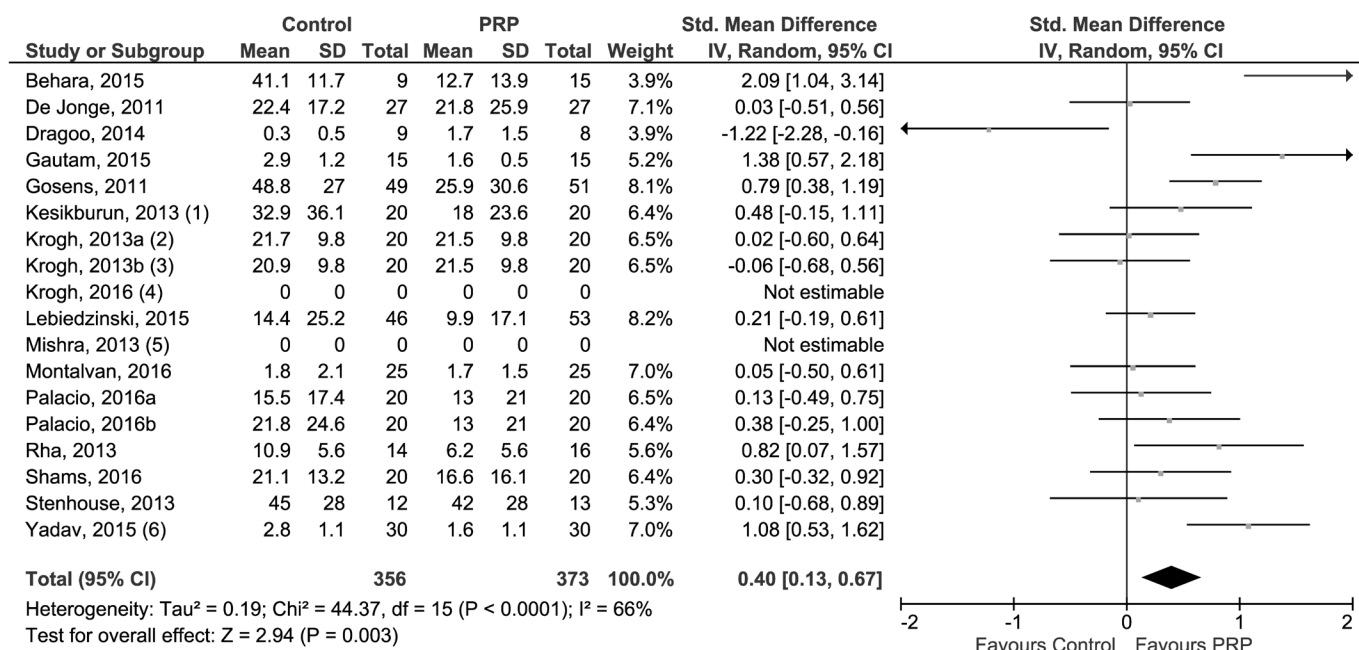
⁵Pain Research, Education and Management Program, Neuroscience Research Australia (NeuRA), Sydney, New South Wales, Australia

⁶UNSW Prince of Wales Clinical School, Randwick, New South Wales, Australia

⁷Laboratory for Musculoskeletal Pain and Motor Control, Center for Sensory-Motor Interaction (SMI), Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

Correspondence to Dr Mervyn J Travers, School of Physiotherapy, University of Notre Dame Australia, Fremantle, WA 6160, Australia; mervyn.travers@nd.edu.au

Discussion



Footnotes

- (1) Received additional data from author
- (2) Baseline variance carried forward
- (3) Baseline variance carried forward
- (4) Unclear on pain outcome measure reported
- (5) No raw pain data provided
- (6) Calculated variance from p-value

Figure 1 Our reproduction of PRP versus control on tendinopathy pain with 'Not Estimable' comparisons removed. Not Estimable=insufficient data to generate standard mean difference; Total=sample size. Statistics reporting heterogeneity, see Box 1 for brief explanations. PRP, platelet-rich plasma; SMD, standardised mean difference.

example, female competitive middle-distance runners with a 3 month history of Achilles tendinopathy received either saline or PRP and were followed up at 12 weeks post-intervention, then the clinician may be provided with a robust estimate of treatment effect that is only applicable

Box 1 A brief definition of the evidence of heterogeneity reported on a forest plot

Statistical evidence of heterogeneity is represented by several values reported in the forest plot:

1. χ^2 is a statistic that comes from a 'yes/no' test for the presence of heterogeneity.¹⁶
2. I^2 is derived from the χ^2 test and represents the proportion of variability across studies that is not due to sampling error.¹⁵
3. τ^2 is the variance of the random-effects distribution. Further discussion of τ^2 and the related implications for selection of fixed or random effects models for meta-analysis are out of the scope of this paper.¹⁷

to that specific population. However, including more varied RCTs in a SRMA yields a result that may be generalisable to a broader range of patients if it is plausible to believe that the outcomes of other wider patient populations would not be influenced by the clinical diversity. Figure 2 illustrates the clinical diversity of the example SRMA. We encourage clinicians to consider whether a pooled estimate of treatment effect from this clinically diverse set of RCTs can usefully inform their practice. Similarly, might pooling the outcomes from clinically diverse studies risk misinforming their practice? Further, among the clinical diversity in the example SRMA, there is a particular intervention with a known treatment effect that warrants further examination.

COMPARISONS WITH ACTIVE CONTROL

Of the control injections included in the example SRMA, corticosteroid injections have a known treatment effect which may bias the pooled estimate of effect. For example, PRP injections have previously outperformed corticosteroid injections for tendinopathy at the lateral elbow at 6 months and 1 year follow-up points.

However, such results should be considered in light of a known temporal effect of corticosteroids for this condition at those specific follow-up points.^{18–20} For example, while a short-term improvement may be observed at 12 weeks, this is countered by a significant deterioration in pain and function at both 6 and 12 months post-corticosteroid injection when compared with PRP in patients with lateral elbow tendinopathy (tennis elbow).¹⁸ This represents a 21% reduction in the relative risk of overall improvement 1 year after corticosteroid injection in these patients when compared with no intervention (relative risk 0.79, 95% CI 0.69 to 0.90).¹⁸ Clearly, this suboptimal medium to long-term outcome at the elbow may lead to erroneous conclusions and misguided management choices for people with that specific tendinopathy. Conversely, there appears to be positive treatment effects reported for corticosteroids for gluteal tendinopathy and plantar fascia pain.^{21 22}

Pooling an active treatment intervention alongside the other control comparisons in a MA may yield confounded estimates of the effect. The example SRMA included

Table 1 Patient and study design characteristics

Study	Treatment period	Sample size		Female gender (%)		Age (year)		Tendinopathy	Minimum symptom duration (month)	Pain outcome	Follow-up (month)
		PRP	Control	PRP	Control	PRP	Control				
Behera <i>et al</i> ¹³	2011–2011	15	10	80	50	38	37	Lateral epicondylar	3	VAS	12
de Jonge <i>et al</i> ¹⁴	2008–2009	27	27	52	52	49	50	Achilles	2	VISA-A	12
Dragoo <i>et al</i> ¹⁵	2009–2012	A10	13	11	0	28	40	Patellar	1.5*	VAS	6
Gautam <i>et al</i> ¹⁶	2011–2012	15	15	—	—	—	—	Lateral epicondylar	6	VAS	6
Gosens <i>et al</i> ¹⁷	2006–2008	51	49	52	56	47	47	Lateral epicondylar	6	VAS	24†
Kesikburun <i>et al</i> ¹⁸	2011–2011	20	20	65	70	46	51	Rotator cuff	3	VAS	12
Krogh <i>et al</i> ¹⁹	2009–2010	20	20	55	55	48	45	Lateral epicondylar	3	PRTEE	3
Krogh <i>et al</i> ⁴⁰	2009–2010	20	20	55	45	48	44	Lateral epicondylar	3	PRTEE	3
Krogh <i>et al</i> ²⁰	2009–2011	12	12	42	50	47	52	Achilles	6	VAS	3
Lebiedziński <i>et al</i> ²¹	2009–2011	64	56	47	74	47	54	Lateral epicondylar	1.5	DASH	12
Mishra <i>et al</i> ²²	2006–2011	112	113	—	—	48	47	Lateral epicondylar	3	VAS	6
Montalvan <i>et al</i> ²³	2010–2014	25	25	32	32	47	46	Lateral epicondylar	—§	VAS	12
Palacio <i>et al</i> ²⁴	2012–2014	20	20	—	—	47	48	Lateral epicondylar	—	PRTEE	6¶
Palacio <i>et al</i> ²⁴	2012–2014	20	20	—	—	47	46	Lateral epicondylar	—	PRTEE	6¶
Rha <i>et al</i> ²⁵	2010–2011	20	19	55	58	52	54	Rotator cuff	6	SPADI pain	6
Shams <i>et al</i> ²⁶	2013–2015	20	20	50	45	52	50	Rotator cuff	3	ASES	6
Stenhouse <i>et al</i> ²⁷	2010–2012	15	13	47	62	53	48	Lateral epicondylar	6	VAS	6
Yadav <i>et al</i> ²⁸	2012–2014	30	30	67	77	37	37	Lateral epicondylar	1	VAS	3

‘—’ indicates missing data.

*Patients failed to respond to ≥6 weeks physical therapy; total symptom duration not reported.

†Data extracted through 12 months only for meta-analysis per systematic review methods (ie, data extraction at 3, 6 and 12 months); however, total symptom duration not reported.

‡Study includes same PRP group and different control groups. PRP group sample size adjusted in meta-analysis based on number of groups.

§Maximum symptom duration was 3 months.

¶Data extracted through 3 months only for meta-analysis due to implausible reported 6-month outcomes.

ASES, American Shoulder and Elbow Surgeons; DASH, Disabilities of the Arm, Shoulder and Hand score; PRP, platelet-rich plasma; PRTEE, Patient-Rated Tennis Elbow Evaluation; SPADI, Shoulder Pain and Disability Index; VAS, visual analogue scale; VISA-A, Victorian Institute of Sport Assessment-Achilles questionnaire.

Figure 2 Pain outcomes included in example systematic review and meta-analysis. Reproduced from Miller LE, *et al*. BMJ Open Sport Exerc Med 2017;3:e000237. doi:10.1136/bmjsem-2017-000237 (CC BY-NC 4.0). See: <http://creativecommons.org/licenses/by-nc/4.0/>.

seven RCTs comparing PRP to corticosteroid injection at different anatomical locations and time points. Thus, we conducted a sensitivity analysis to explore the treatment effect size for PRP when corticosteroid comparators were removed. This reduced the SMD to 0.25 (95% CI -0.16 to 0.65, $p=0.23$) which *negates the*

reported *treatment effect for PRP*. Thus, when an active ‘control’ intervention was removed from the control comparator in the meta-analysis, we derive a markedly different conclusion than that reported in the original papers MA (figure 2). This exercise is not intended to suggest a true estimate of treatment effect of PRP for

tendinopathy, rather it demonstrates the effect that an inappropriate control may have on the findings of an interventional SRMA. It may have been appropriate for the authors of the example paper to present subgroup analyses informed by the known effects of the different comparators. A SRMA by Laudy *et al* (2015)

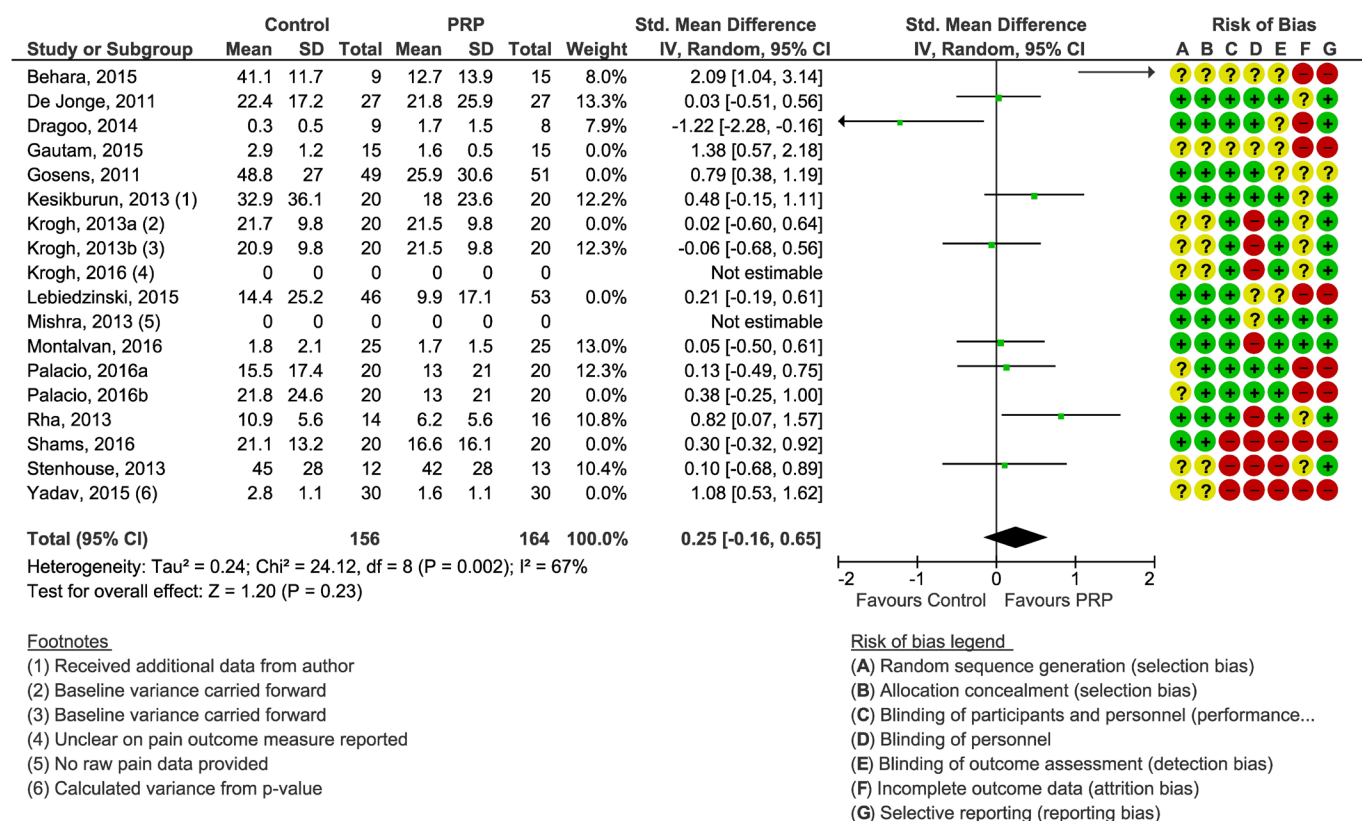


Figure 3 Our reproduction of PRP versus control on tendinopathy pain with corticosteroid comparators removed. Note: Not Estimable=insufficient data to generate standard mean difference; Total=sample size. Note: Risk of bias domains as evaluated by example paper authors presented for all studies included in the example paper; +=low risk of bias; ?=Some concerns re risk of bias; -=high risk of bias. PRP, platelet-rich plasma; SMD, standardised mean difference.

published within the BJSM on the efficacy of PRP for osteoarthritis is a nice example of how different comparators received separate analysis and were not pooled into one single control group.²³

RISK OF BIAS AND CONFIDENCE

Risk of Bias (RoB) represents the risk of over/underestimation of the true intervention effect in study results. Thus, if a study is at high risk of bias, clinicians should consider its findings likely to be associated with an increased possibility of error in the estimate of effect. This extends to the effect estimate from MA when studies at high risk of bias are included in the analysis. Features of studies that may imply bias are usually formally assessed and reported in a SRMA and in the example paper these domains were assessed using the Cochrane Collaboration's tool for assessing risk of bias.²⁴ We encourage readers to familiarise themselves with these sources of bias to be better able to judge the quality of RCTs and subsequent SRMA.

Most of the included studies were judged by the original authors to be at uncertain or high RoB across multiple domains. In light of this, clinicians are encouraged to critically judge the effects

reported in each RCT contributing to a MA. In this case, most of the treatment effects reported in the included RCTs are unlikely to represent the true effect of the intervention. The RoB assessment can be observed next to the estimate of the effect for treatment for each study in figure 3 and may give greater insight into the confidence the reader should have in each individual treatment effect and in turn the pooled treatment effect.

Judgement of confidence in a pooled estimate of effect is a recommended methodological step for SRMA. The Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach suggests downgrading the confidence in the estimate of treatment effect based on study limitations (including risk of bias), inconsistency of results (including heterogeneity), indirectness of evidence (including inappropriate comparator, varying time points and varying body parts), imprecision (including small sample sizes) and publication bias.²⁵⁻²⁷ Though not performed in the example SRMA, use of this approach would have enhanced the methodological rigour of the example SRMA and given greater clarity for clinicians regarding the confidence readers

might have in the efficacy of PRP injections for tendinopathy. Thus, clinicians are encouraged to consider the above features of the body of evidence pooled in interventional SRMA before deciding on a treatment choice.

HOW SHOULD THIS BE INTERPRETED BY THE END USER?

This two-part Educational Editorial series encourages clinicians to read systematic reviews and meta-analyses with a critical and forensic eye. Here, we provided a worked example to highlight for clinicians the need to consider pooling of comparators, clinical diversity, comparisons to active control, risk of bias and confidence in the results when interpreting interventional SRMA findings. Clinicians are well placed to make such judgements in the context of their specific patient(s) when deciding 'Should this SRMA change my practice?'

Acknowledgements Benedict Martin Wand provided advice, direction and oversight to the planning of this project.

Contributors All authors contributed to the inception of the research idea. MJT, MCM and MKB performed the independent data extraction. WG provided

oversight to the data extraction and associated disputes. MJT, WG and MCM performed the statistical analysis with oversight from PC and MB. All authors contributed to drafting/review of manuscript. MKB = Matthew K Bagg; MB = Max K Bulsara

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bjsports-2018-099959>).



To cite Travers MJ, Murphy MC, Debenham JR, *et al.* *Br J Sports Med* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bjsports-2018-099959

Accepted 18 March 2019



► <https://doi.org/10.1136/bjsports-2018-099958>

Br J Sports Med 2019;0:1–5.

doi:10.1136/bjsports-2018-099959

REFERENCES

- Annane D, Jaeschke R, Guyatt G. Are systematic reviews and meta-analyses still useful research? Yes. *Intensive Care Med* 2018;44:512–4.
- Ardern CL. Systematic review hacks for the sports and exercise clinician: five essential methodological elements. *Br J Sports Med* 2016;50:447–9.
- Møller MH, Ioannidis JPA, Darmon M. Are systematic reviews and meta-analyses still useful research? We are not sure. *Intensive Care Med* 2018;44:518–20.
- Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs* 2003;12:77–84.
- Gurevitch J, Koricheva J, Nakagawa S, *et al.* Meta-analysis and the science of research synthesis. *Nature* 2018;555:175–82.
- Bero L. Getting the systematic review basics right helps clinical practice: 4 common pitfalls for systematic review authors to avoid. *Br J Sports Med* 2019;53.
- Grindem H, Mansournia MA, Øiestad BE, *et al.* Was it a good idea to combine the studies? Why clinicians should care about heterogeneity when Making decisions based on systematic reviews. *Br J Sports Med* 2019;53:399–401.
- Mansournia MA, Altman DG. Invited commentary: methodological issues in the design and analysis of randomised trials. *Br J Sports Med* 2018;52:553–5.
- Verhagen E, Stovitz SD, Mansournia MA, *et al.* BJSME educational editorials: methods matter. *Br J Sports Med* 2018;52:1159–60.
- Weir A, Rabia S, Ardern C. Trusting systematic reviews and meta-analyses: All that glitters is not gold! *Br J Sports Med* 2016;50:1100–1.
- Miller LE, Parrish WR, Roides B, *et al.* Efficacy of platelet-rich plasma injections for symptomatic tendinopathy: systematic review and meta-analysis of randomised injection-controlled trials. *BMJ Open Sport Exerc Med* 2017;3.
- Bagg MK, Salanti G, McAuley JH. Comparing interventions with network meta-analysis. *J Physiother* 2018;64:128–32.
- Efthimiou O, Debray TPA, van Valkenhoef G, *et al.* GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods* 2016;7:236–63.
- Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods* 2012;3:80–97.
- Borenstein M, Higgins JPT, Hedges LV, *et al.* Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Res Synth Methods* 2017;8:5–18.
- Higgins JPT, Deeks JJ, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, eds. *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration, 2011.
- Nikolakopoulou A, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. *Evid Based Ment Health* 2014;17:53–7.
- Coombes BK, Bisset L, Vicenzino B. Efficacy and safety of corticosteroid injections and other injections for management of tendinopathy: a systematic review of randomised controlled trials. *Lancet* 2010;376:1751–67.
- Gosens T, Peerbooms JC, van Laar W, *et al.* Ongoing positive effect of platelet-rich plasma versus corticosteroid injection in lateral epicondylitis: a double-blind randomized controlled trial with 2-year follow-up. *Am J Sports Med* 2011;39:1200–8.
- Scott A, Khan KM. Corticosteroids: short-term gain for long-term pain? *Lancet* 2010;376:1714–5.
- Johannsen FE, Herzog RB, Malmgaard-Clausen NM, *et al.* Corticosteroid injection is the best treatment in plantar fasciitis if combined with controlled training. *Knee Surg Sports Traumatol Arthrosc* 2019;27:5–12.
- Mellor R, Bennell K, Grimaldi A, *et al.* Education plus exercise versus corticosteroid injection use versus a wait and see approach on global outcome and pain from gluteal tendinopathy: prospective, single blinded, randomised clinical trial. *BMJ* 2018;361.
- Laudy ABM, Bakker EWP, Rekers M, *et al.* Efficacy of platelet-rich plasma injections in osteoarthritis of the knee: a systematic review and meta-analysis. *Br J Sports Med* 2015;49:657–72.
- Higgins JPT, Altman DG, Gøtzsche PC, *et al.* The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343.
- Ryan R, Hill S. How to grade the quality of the evidence, 2016. Cochrane consumers and communication group. Available: <http://cccrp.cochrane.org/author-resources>
- Guyatt GH, Oxman AD, Vist GE, *et al.* Grade: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- Guyatt GH, Oxman AD, Kunz R, *et al.* What is "quality of evidence" and why is it important to clinicians? *BMJ* 2008;336:995–8.