















Meta-analysis prediction intervals are under reported in sport and exercise medicine

David N. Borg^{1,2,3}  | Franco M. Impellizzeri⁴  | Samantha J. Borg¹  |
 Kate P. Hutchins²  | Ian B. Stewart²  | Tamara Jones⁵  | Brenton J. Baguley⁶  |
 Lucas B. R. Orssatto^{2,6}  | Aaron J. E. Bach^{7,8}  | John O. Osborne⁹  |
 Benjamin S. McMaster²  | Robert L. Buhmann¹⁰  | Joshua J. Bon^{3,11}  |
 Adrian G. Barnett¹ 

¹Australian Centre for Health Services Innovation (AusHSI), School of Public Health and Social Work, Queensland University of Technology, Brisbane, Queensland, Australia

²School of Exercise and Nutrition Sciences, Queensland University of Technology, Brisbane, Queensland, Australia

³Centre for Data Science, Queensland University of Technology, Brisbane, Queensland, Australia

⁴School of Sport, Exercise and Rehabilitation, University of Technology Sydney, Sydney, New South Wales, Australia

⁵Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Victoria, Australia

⁶Institute for Physical Activity and Nutrition (IPAN), School of Exercise and Nutrition Sciences, Deakin University, Burwood, Victoria, Australia

⁷School of Health Sciences and Social Work, Griffith University, Gold Coast, Queensland, Australia

⁸Cities Research Institute, Griffith University, Gold Coast, Queensland, Australia

⁹School of Sport Sciences, UiT The Arctic University of Norway, Tromsø, Norway

¹⁰School of Health, University of Sunshine Coast, Sippy Downs, Queensland, Australia

¹¹School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia

Correspondence

David N. Borg, Australian Centre for Health Services Innovation (AusHSI), School of Public Health and Social Work, Queensland University of Technology, Brisbane, QLD, Australia.
 Email: dn.borg@qut.edu.au

Abstract

Aim: Prediction intervals are a useful measure of uncertainty for meta-analyses that capture the likely effect size of a new (similar) study based on the included studies. In comparison, confidence intervals reflect the uncertainty around the point estimate but provide an incomplete summary of the underlying heterogeneity in the meta-analysis. This study aimed to estimate (i) the proportion of meta-analysis studies that report a prediction interval in sports medicine; and (ii) the proportion of studies with a discrepancy between the reported confidence interval and a calculated prediction interval.

Methods: We screened, at random, 1500 meta-analysis studies published between 2012 and 2022 in highly ranked sports medicine and medical journals. Articles that used a random effect meta-analysis model were included in the study. We randomly selected one meta-analysis from each article to extract data from, which included the number of estimates, the pooled effect, and the confidence and prediction interval.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Scandinavian Journal of Medicine & Science In Sports* published by John Wiley & Sons Ltd.

Results: Of the 1500 articles screened, 866 (514 from sports medicine) used a random effect model. The probability of a prediction interval being reported in sports medicine was 1.7% (95% CI=0.9%, 3.3%). In medicine the probability was 3.9% (95% CI=2.4%, 6.6%). A prediction interval was able to be calculated for 220 sports medicine studies. For 60% of these studies, there was a discrepancy in study findings between the reported confidence interval and the calculated prediction interval. Prediction intervals were 3.4 times wider than confidence intervals.

Conclusion: Very few meta-analyses report prediction intervals and hence are prone to missing the impact of between-study heterogeneity on the overall conclusions. The widespread misinterpretation of random effect meta-analyses could mean that potentially harmful treatments, or those lacking a sufficient evidence base, are being used in practice. Authors, reviewers, and editors should be aware of the importance of prediction intervals.

KEYWORDS

confidence interval, forest plot, heterogeneity, random effects model, review

1 | INTRODUCTION

There are big incentives for researchers to publish a meta-analysis, namely, they appear relatively easy to undertake compared with empirical studies, are often highly cited, and can influence practice. In reality, conducting a meta-analysis requires significant time, methodological skills and statistical expertise.¹⁻³ Nonetheless, the aforementioned incentives have contributed to a rapid increase in published meta-analyses in the past decade,⁴ facilitated by improved software accessibility and accompanied resources (e.g., “Doing Meta-Analysis in R”⁵).

A decision that should be made before any analysis is selecting a meta-analysis model to combine study effect sizes.^{3,6} Expected between-study heterogeneity is an important consideration for model selection.⁷ Causes of heterogeneity in sport and exercise medicine (hereafter “sports medicine”) include studying different participant groups or employing different exercise modes. Random effect models are commonly used to combine estimates from studies investigating different populations, where heterogeneity is expected.^{6,8} Another relatively common meta-analysis model is the fixed effect model, which is more suited to structured trials, often undertaken by drug companies, where all estimates are assumed to share the same *true* effect.⁹ In sports medicine, pooled estimates from random effect models are often mistakenly interpreted as an overall (*true*) effect like in a fixed effect model,¹⁰ ignoring that random effect models estimate the underlying *mean* effect.⁷

Once modeled, the source and pattern of heterogeneity should be investigated,⁸ including via forest plots and the

exploration of subgroups in the data, where appropriate and outlined in a pre-determined protocol. Heterogeneity should also be considered when interpreting the results. Yet, evidence from medical research indicates that heterogeneity is not commonly considered in meta-analysis conclusions.¹¹ It is more common for researchers to report and focus on the point estimate of the (average) treatment effect, along with its 95% confidence interval, including when using random effect models. However, while the confidence interval reflects the uncertainty around the point estimate (or the range of effects compatible with the data),¹² it provides an incomplete summary of the underlying heterogeneity.¹³

A practical way to consider heterogeneity when estimating effects is via a prediction interval (see “In Depth”).¹³ A prediction interval can be defined as the interval within which the effect size of a new study would fall, if the study was selected at random from the same population as those in the studies included in the meta-analysis.⁸ This aligns with the interest of most researchers and practitioners, who generally want to draw inferences about the potential effects of interventions when implemented in future settings, to inform (clinical) decision-making and policies.¹⁴ One way to think about the prediction interval is that it should capture the likely effect if the treatment were applied in practice in a “sufficiently similar” population and setting.⁷ This, however, assumes that practice in the real world is the same as practice in the study, which may not be the case. For example, due to differences from the study exclusion criteria. In the presence of between-study heterogeneity, prediction intervals are wider than confidence intervals, and therefore, study conclusions may differ if

based on the prediction interval rather than the confidence interval.¹³ For example, a confidence interval shows a “significant” benefit to employing a treatment that does not include zero, whereas the wider prediction interval is more uncertain and does include zero (see Figure 1; replicated from figure 2, subgroup 1.5.1 in¹⁵).

It is reasonable to assume that meta-analysis prediction intervals are overlooked in sports medicine research in a similar way to medicine.^{11,16} No previous study has examined the reporting or use of prediction intervals in sports medicine. In this study, we aimed to estimate: (i) the proportion of studies that report a prediction interval when using a random effect meta-analysis approach; and (ii) the proportion of studies where there was a discrepancy between the confidence interval and prediction interval. We hypothesized that: (i) the proportion of sports medicine studies that reported a prediction interval would be very low, and less than medicine; and (ii) a high proportion of studies would have a discrepancy in the conclusions if based on the prediction interval rather than the confidence interval.

In Depth: Prediction intervals from a random effect meta-analysis.

Calculating prediction intervals around an overall effect requires both the estimated between-study heterogeneity variance and the standard error of the pooled effect.

Prediction intervals are calculated as follows⁸:

$$SDPI = \sqrt{SE_{\hat{\mu}}^2 + \hat{\tau}^2}$$

$$\hat{\mu} \pm t_{K-2}(1 - \alpha/2) \times SDPI$$

where:

$\hat{\mu}$ = the pooled effect

$t_{K-2}(1 - \alpha/2)$ = right tail $\alpha/2$ quantile of a t -distribution with $K-2$ degrees of freedom, where K is the number of estimates in the analysis

$SE_{\hat{\mu}}$ = the standard error of the pooled effect

$\hat{\tau}^2$ = estimated between-study heterogeneity variance

$SDPI$ = standard deviation of the prediction interval

2 | METHODS

This was an observational study. Our target population was any published meta-analysis in the field of sport and exercise medicine that used a random effect model. We randomly sampled meta-analyses from 10 highly ranked sport and exercise medicine journals (according to *Scimago*) published between 2012 and 2022. Random effect meta-analyses published in eight highly ranked medical journals were used as comparator for reporting practices. All selected journals were indexed in *MEDLINE* and are listed in Data S1. No ethical approval for the study was required as we exclusively used publicly available data.

2.1 | Sample

Articles that contained “meta-analysis” in the title or abstract were identified through *PubMed* (Data S1). Using this search criterion is reasonable, as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guidelines state that “meta-analysis” should be included in the title and abstract.¹⁷ A total of 2281 articles across sports medicine ($n=1214$) and medicine ($n=1067$) were identified. We randomly selected 750 articles from each field for full-text screening, using the “sample_n” function in the R package *dplyr*.¹⁸ The total sample size (i.e., 1500 articles) was not determined using a power calculation, but was instead chosen for practical reasons, based on the number of authors available to extract data.

2.2 | Data extraction

To be included, articles must have used a random effect meta-analysis model. We extracted data from only one meta-analysis per study to avoid within-study correlation. When multiple meta-analyses were reported in the same article, Google’s random number generator was used to select the meta-analysis to be included. The following variables were extracted the selected meta-analysis:

- The type of effect size (i.e., mean difference, standardized mean difference, odds ratio, hazard ratio, or risk ratio)
- The number of estimates in the meta-analysis
- The overall pooled effect
- The upper and lower confidence limits of the pooled effect, and the level of confidence (e.g., 95%)
- Whether a forest plot was presented (true, false)
- The percentage of total variability due to between-study heterogeneity, I^2
- The between-study heterogeneity variance, τ^2
- Whether a prediction interval was reported (true, false)
- The upper and lower prediction interval, and the level of interval (e.g., 95%)

If no overall pooled effect was reported, data were extracted from the subgroup with the most estimates. When a summary effect from both a fixed and random effect model was reported, the estimate from the random effect model was extracted.

Data extraction accuracy checks were performed on 20% of articles ($n=300$), by a second, independent reviewer. The accuracy of article eligibility judgments was

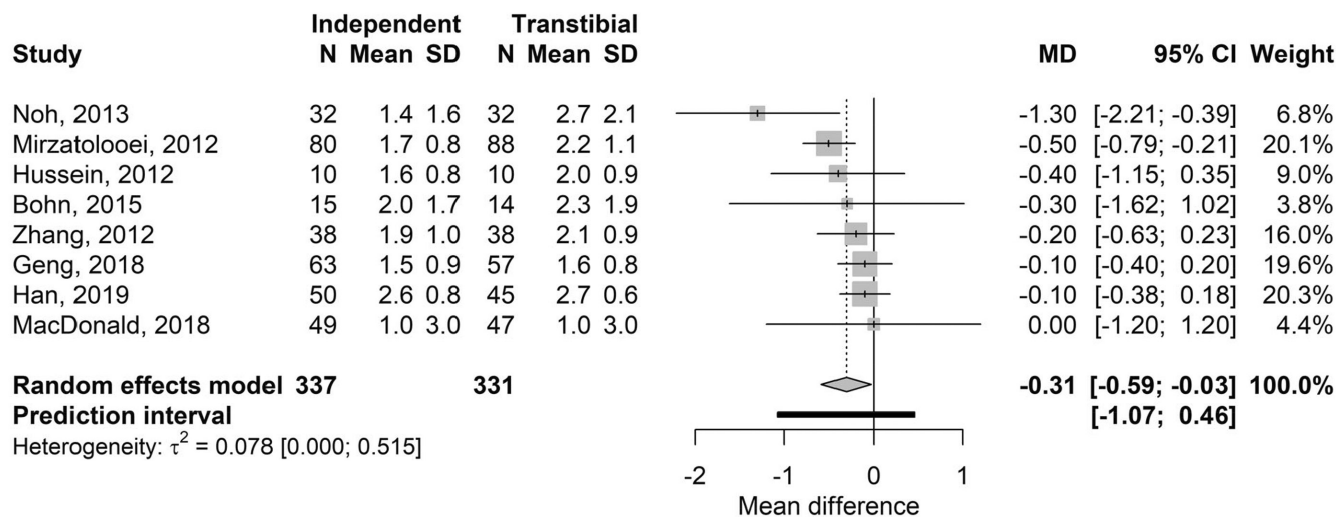


FIGURE 1 Replicated forest plot of differential laxity between the index and uninjured knee after anterior cruciate ligament reconstruction using two different femoral tunnel drilling techniques (independent and transtibial). Note that the 95% confidence interval (width of gray diamond) of the pooled estimate does not include the null value of zero, whereas the prediction interval (horizontal black line) does, spanning both negative and positive values. The forest plot was replicated using the Sidik-Jonkman method to estimate the between study variance (τ^2). CI, confidence interval; MD, mean difference; SD, standard deviation.

98.7% (95% CI = 96.6% to 99.6%). The accuracy of extracted data (based on 164 studies) ranged from 97% to 100% (Data S2).

All extracted data were checked for boundary violations, i.e., if the point estimate was outside the confidence interval, or if the ratio estimates included negative values, which is impossible.

2.3 | Data analysis

The analysis had several steps. First, the characteristics of studies were summarized using descriptive statistics. This included plotting the distribution of I^2 and τ^2 values. Second, we estimated the proportion of studies in sports medicine that reported a prediction interval, using medicine as a comparison. If reported, we determined whether the prediction interval was considered in study conclusions, and whether there was a discrepancy between the prediction interval and confidence interval. Third, where possible, we calculated a prediction interval for each meta-analysis, to estimate the proportion of studies where there was a discrepancy in conclusions between the prediction interval and the confidence interval.

Prediction interval reporting in sports medicine was compared to medicine using a generalized linear model with a binomial response distribution. The model included *discipline* (levels: medicine, sports medicine) as a fixed effect. In line with our directional hypothesis (sports medicine less than medicine), we report the measure of effect (odds ratio) with a 90% confidence interval. To look

for recent improvements in practice, the proportion of studies that reported a prediction interval was plotted by year of publication.

When a prediction interval was reported, the study's abstract, results and discussion were examined to determine whether the prediction interval was considered in the overall conclusions. We identified studies with a discrepancy in the conclusion drawn from the prediction interval compared to the confidence interval, using the criteria of whether each interval included the null effect (i.e., a ratio of 1, or a mean or standardized mean difference of 0). For example, if the prediction interval for a standardized mean difference ranged from -0.2 to 1.1 but the confidence interval ranged from 0.3 to 0.6 , we concluded there was a discrepancy between the intervals, as only the prediction interval included the null effect (i.e., 0). While we caution against a simplistic dichotomous interpretation, this interpretation is consistent with practices in the published literature.¹⁹ To demonstrate how much wider prediction intervals can be compared with confidence intervals, we calculated the ratio of the prediction interval to the confidence interval, with values greater than 1 indicating a wider prediction interval.

We calculated a prediction interval for studies that reported the mean pooled effect and τ^2 using the equation in the "In Depth" box. The standard error for effect sizes on the observed or standardized scale were calculated using the equation: $\text{standard error} = (u - l) / (2 \times z)$, where "u" is the upper confidence limit, "l" is the lower confidence limit, and "z" is the Z-value from the standard normal distribution corresponding to the specified confidence

interval, which is approximately 1.96 for a 95% interval.²⁰ Standard errors for ratios were calculated on the log-scale, i.e., standard error = $(\log(u) - \log(l)) / (2 \times z)$.²⁰ We then determined the proportion of studies where there was a discrepancy between the reported confidence interval and the calculated prediction interval. Confidence intervals for proportions were calculated using the Clopper–Pearson method for the binomial distribution²¹ in the R package *binom*.²²

All analyses were undertaken in R.²³ No values were imputed for missing observations. The data and R code are available from <https://www.doi.org/10.5281/zenodo.7783823>.

3 | RESULTS

3.1 | Summary

Fifty-eight percent (866/1500) of the articles screened met the inclusion criteria. Of the articles included, 514 (59%) were from sports medicine. We found that nine out of 514 studies in sports medicine reported a prediction interval when using a random effect meta-analysis. This level of reporting was lower than medicine, where reporting was higher (14 in 352 studies) but still very poor.

Nine sports medicine studies reported a prediction interval, and three of these studies based their conclusions on the prediction interval. We were able to calculate a prediction interval for 220 sports medicine studies. There was a discrepancy between the reported confidence interval and calculated prediction interval for 121 (60%) of these studies, with the prediction interval including the null value (i.e., no difference), but the confidence interval excluding the null value.

3.2 | Included studies

We screened a total of 1500 articles for eligibility. Forty-nine articles were excluded because they did not include

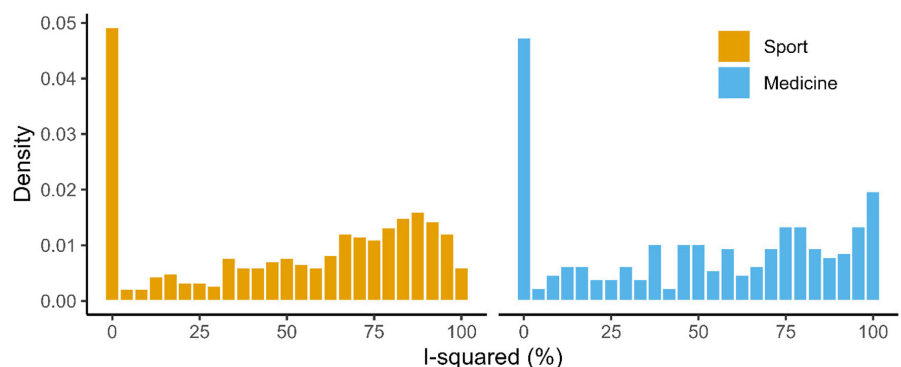
a meta-analysis. These 49 articles were systematic or narrative reviews ($n=33$), comments, corrections, or editorials ($n=8$), infographics ($n=5$), interventions ($n=1$), methods ($n=1$), or economic analyses ($n=1$). Of the 1451 studies, 872 used a random effects model. We removed six articles: three ratio results where the pooled effect was less than zero, and three articles because the pooled effect was outside the confidence interval. Therefore, 866 studies were included, with 514 from sports medicine, and 352 from medicine. Data S1 provides a list of the screened and included articles from the journals searched.

3.3 | Characteristics of included meta-analyses

In sports medicine, the median number of estimates in each meta-analysis was 10 (first to third quartile = 6–19). The standardized mean difference was commonly reported (49%), followed by the mean difference (35%) and a hazard, odds, or risk ratio (16%). Nearly all studies reported a 95% confidence interval (99%), with only two studies reporting a 90% confidence interval. Most sports medicine studies included a forest plot in the main paper (90%) and reported an I^2 value (84%), with reporting of τ^2 less common (44%). Figure 2 shows the distribution of reported I^2 values, and Figure 3 shows the distribution of τ^2 values. About 5% of studies in sports medicine (and medicine) had an I^2 value of 0 (Figure 2). An I^2 estimate of 0 is typically accompanied by large uncertainty (e.g., a wide 95% CI)—which is also the case for estimates of τ^2 .¹³ Further, a limitation of the I^2 is that small imprecise studies with varying treatment effects can yield an I^2 value of 0.¹³ As such, while there was a “spike” of studies with an I^2 of 0, it is unlikely that the *true* I^2 value of these studies was exactly 0.

In medicine, the median number of estimates in each meta-analysis was 12 (first to third quartile = 7–20). Ratios were commonly reported (68%), followed by the mean difference (24%) and standardized mean difference (8%). Almost all studies reported a 95% confidence interval

FIGURE 2 The distribution of I^2 values reported by sports medicine and medicine studies, published between 2012 and 2022, using random effect meta-analysis. The I^2 indicates the percentage of total variability due to between-study heterogeneity. The histogram bin widths of 4% were used.



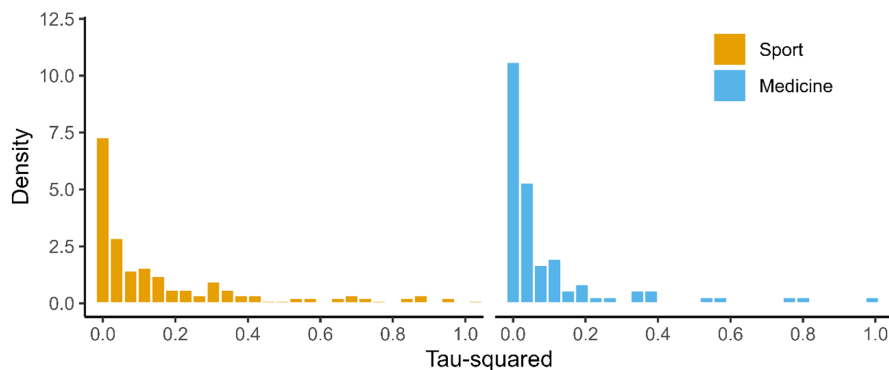


FIGURE 3 The distribution of τ^2 values reported by sports medicine and medicine studies, published between 2012 and 2022, using random effect meta-analysis. Values of τ^2 indicate the between-study heterogeneity variance. Note that some τ^2 values were very small (e.g., $\tau^2=0.0022$) and so may be binned with 0 in the plot. There were 49 (22%) τ^2 values of 0 in sports medicine, and 18 (19%) τ^2 values of 0 in medicine. Histogram bin widths of 0.04 were used. The x-axis has been restricted to values between 0 and 1, which excludes 52 values from sports medicine, and 10 values from medicine.

(99%). One study reported a 99% confidence interval. Most medical studies included a forest plot in the main paper (85%) and reported an I^2 value (87%; Figure 2), with reporting of τ^2 again less common (27%; Figure 3).

3.4 | Prediction interval reporting

Prediction interval reporting was extremely low in both disciplines. In sports medicine, only nine of the 514 studies reported a prediction interval. The probability of a prediction interval being reported was 1.7% (95% CI=0.9%, 3.3%). In medicine, 14 of the 352 studies reported a prediction interval. The probability of a prediction interval being reported in medicine was 3.9% (95% CI=2.4%, 6.6%). The odds of a prediction interval being reported in sports medicine were on average 57% lower than medicine (odds ratio=0.43, 90% CI=0.21, 0.87). However, there was considerable uncertainty in these odds, with the 90% CI compatible with the odds being lower by only 13.2%. Although prediction interval reporting was very low between 2012 and 2022, there was some indication of an incremental improvement in sports medicine since 2019 (Figure 4). Nonetheless, with a reporting rate in sports medicine of 7.7% in 2021, there are still significant improvements to be made.

3.5 | Studies reporting a prediction interval

Twenty-three studies reported a prediction interval (Data S3). Nine sports medicine studies reported a prediction interval but only three studies considered the prediction interval when interpreting their results. Five studies had a discrepancy between the confidence interval and

prediction interval. Based on the median, prediction intervals were 3.4 times wider than the confidence interval (first to third quartile=2.7–4.0). Three studies had a discrepancy between the confidence interval and prediction interval and based their conclusions on the confidence interval.

Fourteen medical studies reported a prediction interval, and 10 considered the interval when interpreting their results (Data S3). Seven studies reported a discrepancy between the confidence interval and prediction interval. Prediction intervals were (median) 4.2 times wider than the confidence interval (first to third quartile=2.8–5.0). Only one study had a discrepancy between the confidence interval and prediction interval but based their conclusions on the confidence interval.

3.6 | Discrepancy between confidence intervals and prediction intervals

Prediction intervals were able to be calculated for 314 studies (Table 1). Nine studies ($n=7$ from sports medicine) reported a τ^2 value but a prediction interval could not be calculated as these studies only included two estimates (see “In-Depth” box). In sports medicine, 121 of 220 studies (55%, 95% CI=48%, 62%) had a discrepancy between the reported confidence interval and the calculated prediction interval, suggesting that the study conclusions would differ if based on the prediction interval, rather than the confidence interval—as it should be if the aim is to provide information about implementation. Prediction intervals were (on median) 3.4 times wider than the confidence intervals (first to third quartile=2.2–4.7). In medicine, there was a discrepancy between the reported confidence interval and the calculated prediction interval for 29 of 94 studies (31%, 95% CI=22%, 41%). Based on the median,

FIGURE 4 The proportion of random effect meta-analysis in sports medicine and medicine reporting a prediction interval by publication year.

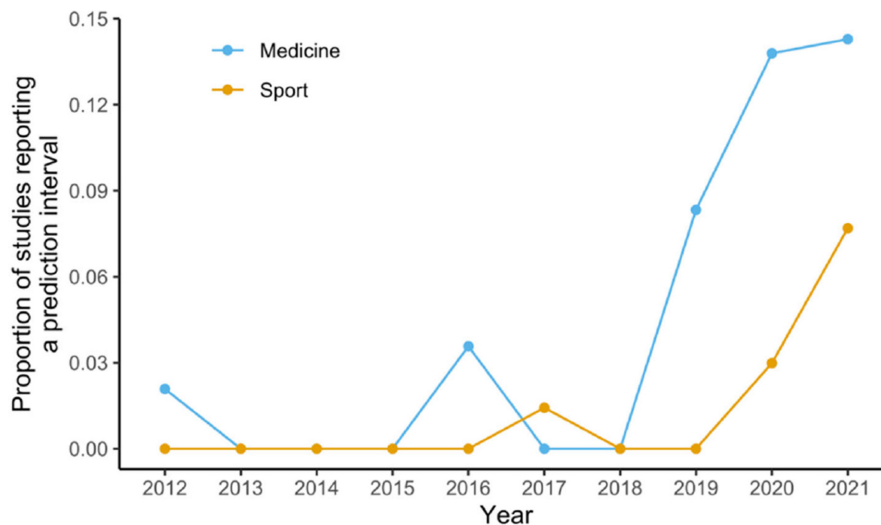


TABLE 1 The percentage of studies that excluded the null hypothesis value, for the 314 studies for which a prediction interval could be reconstructed because a τ^2 value was reported.

Interval	Value for τ^2 when reported as 0	Sports medicine $n = 220$	Medicine $n = 94$
Confidence	NA	75%	95%
Prediction	0	20%	64%
Prediction	0.005	17%	63%

Note: As a sensitivity analysis we also calculated prediction intervals for studies where $\tau^2 = 0$ using $\tau^2 = 0.005$, because most τ^2 values were presented on forest plots where values were rounded to two decimal places.

prediction intervals in medicine were 2.6 times wider than confidence intervals (first to third quartile = 1.6–4.2).

4 | DISCUSSION

Our study shows that there is a widespread failure of sports medicine meta-analyses to consider between-study heterogeneity when communicating estimates for future applications. Ignoring the prediction interval may result in the use of treatments that lack sufficient evidence, could create excessive and unsupported expectations, and in some contexts, may have negative effects that could cause harm to athletes or patients. Prediction intervals were much wider than confidence intervals (median of 3.4 times wider), meaning results may be less “positive”, and therefore, less appealing to researchers or journals.²⁴ The mandating of prediction interval reporting through journal policies is likely required to ensure widespread behavior change. Changes to the PRISMA reporting guidelines¹⁷ and to the default settings of meta-analysis software may help improve reporting rates.

Despite previous efforts to highlight the importance of prediction intervals^{8,11} and calls for greater reporting,¹³ our results show that researchers overlook and underuse prediction intervals. The common misinterpretation of

random effect meta-analyses results can be added to the growing list of problems with meta-analyses, such as confusing the standard error with the standard deviation and ignoring correlated estimates.^{1,25} No previous study has examined prediction interval reporting in sports medicine meaning no comparisons can be made. However, our findings are similar to the adjacent field of medicine, where an examination of 44 random effect meta-analyses over a decade ago found that no study reported a prediction interval.¹¹ The lack of improvement in reporting is concerning.

A prediction interval was able to be calculated for 220 of the 514 sports medicine studies (Table 1). We found that for 55% of these 220 studies, the prediction interval covered the null, but the confidence interval did not. This discrepancy indicates that, if these results are used to inform decisions, in a future study or application, the *true* effect could be negligible or in the opposite direction. A previous study of 65 931 meta-analyses on a range of healthcare-related topics from the Cochrane Database of Systematic Reviews found that the prediction interval from 12.8% of meta-analyses included the null but the confidence interval did not.²⁶ The four times higher prevalence in sports medicine can be explained by the presence of greater heterogeneity compared to studies from the Cochrane Database of Systematic Reviews. To illustrate the greater heterogeneity, of the

220 calculated prediction intervals 80% were at least two times wider than the confidence intervals compared to 26.2% of the meta-analyses from the Cochrane Database of Systematic Reviews.²⁶

The distribution of I^2 (Figure 2) and τ^2 (Figure 3) values shows that heterogeneity is often present in sports medicine. While it is encouraging that some indicators of heterogeneity, such as forest plots and I^2 values, were included in most articles, these are less practical ways of interpreting heterogeneity compared to a prediction interval.^{7,8} There seems an overreliance (and over-interpretation) on I^2 as an indicator of heterogeneity, whose limitations to assess heterogeneity are well known.^{27,28} Given the importance of meta-analyses, their perceived value among practitioners, and therefore their potential to influence guidelines and practice, journals need to mandate the reporting of prediction intervals. A journal reporting mandate should help ensure that important heterogeneity is not ignored when communicating results from meta-analyses—results that may be used to provide recommendations about future implementation, or to inform clinical decision making.^{8,13}

One way to help improve reporting rates could be to amend reporting guidelines. Item 20b of the PRISMA reporting guidelines, under “Synthesis of results”, states researchers should “Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g., confidence/credible interval) and measures of statistical heterogeneity”.¹⁷ This could be extended to include the reporting of prediction intervals when a random effect meta-analysis model is used. Another way to improve reporting could be to set the default of all meta-analysis software to include a prediction interval, for example, to be shown on a forest plot. While these strategies may help reporting rates, they will not directly improve researchers use of the prediction interval when communicating meta-analysis results. Such change is reliant on a greater awareness of prediction intervals and their importance by researchers, journal editors, and reviewers.

An assumption of the prediction interval calculation described in the “In-Depth” box is that the underlying heterogeneity follows a normal distribution.²⁹ When the number of effect sizes in a meta-analysis is small this may reduce the ability to estimate the underlying distribution. About 44% of meta-analysis in sports medicine had less than 10 estimates, with 21% having five or less. Even when the sample is small, heterogeneity cannot be ignored and should be transparently acknowledged, with conclusions consistent with this limitation. Researchers may consider calculating a non-parametric prediction interval when sample sizes are small, or when the assumption of normally distributed heterogeneity is unrealistic. See³⁰ for

examples and implementation. A spreadsheet to calculate a prediction interval nonparametrically can be found in Supplement 5 in.²⁹ There are active discussions about how intervals should be calculated.³¹ We encourage researchers to be aware of these discussions, and their relevance to sports medicine, in particular, the coverage of intervals when sample sizes are small, and heterogeneity is low.

The potential time burden of manually screening articles for prediction intervals could be reduced by means of automated tools.³² This would require the development of an algorithm capable of identifying prediction intervals using digitization methods, as most intervals are reported on forest plots, which are images. An automated process could send a note to the journal editor when authors submit a random effects meta-analysis without a prediction interval. Future research should focus on developing an automated prediction interval screening tool that could be implemented by journals, in similar way to other automated checks.³²

4.1 | Limitations

We were unable to calculate a non-parametric prediction interval for studies included in our analysis, as we did not extract study level estimates from each meta-analysis.²⁹ Given the small sample sizes in sports medicine, calculating a prediction interval non-parametrically may have been more appropriate for some studies. Data extraction error rates in this study were low (0%–3%). Nonetheless, there may be a small number of records with errors in our analysis (Data S2). Given the extremely low proportion of studies reporting prediction intervals, any data extraction errors would have no impact on our substantive conclusions. In our analysis and discussions, we did not address or consider risk of bias issues, which are an important when interpreting meta-analysis results, including prediction intervals. Our results are likely representative of practices across all journals in the field of sports medicine.

Another, more conceptual, potential limitation is that it could be argued that low reporting is due to a larger interest in sports medicine in presenting the current evidence of average effects, rather than informing future applications. Although this cannot be discounted, we consider this quite unlikely given that the interest of investigating the effect of interventions is typically to understand whether they can be used in practice. It is possible that a small proportion of the 98.3% of studies that did not report prediction intervals could be interested in only presenting current evidence without any interest in future applications. Regardless, the number of meta-analyses not reporting prediction intervals is still excessive and concerning.

5 | CONCLUSION

The widespread misinterpretation of random effect meta-analysis (i.e., pooled estimates being interpreted as if they were from a fixed effect model), the lack of consideration of the between-study heterogeneity, and the failure to report prediction intervals could mean that treatments lacking a sufficient evidence base, or that are potentially harmful, are being used and recommended in practice. It may also mean that excessive expectations of treatment effectiveness are being created. Journals need to mandate prediction interval reporting through their submission policy to force change within the field. Authors, reviewers, and journal editors should be aware of the importance of prediction intervals.

6 | PERSPECTIVE

Prediction intervals are a useful measure (from a decision-making perspective) of uncertainty for meta-analyses that capture the likely effect size of a new (similar) study based on the included studies. Our study found that very few meta-analysis studies in sports medicine report prediction intervals and hence are prone to missing the impact of between-study heterogeneity on the overall conclusions. The widespread misinterpretation of random effect meta-analyses could mean that potentially harmful treatments, or those lacking a sufficient evidence base, are being used in practice. It may also mean that excessive expectations of treatment effectiveness are being created. Journals need to mandate prediction interval reporting through their submission policy to force change within the field. Authors, reviewers, and editors should be aware of the importance of prediction intervals.

ACKNOWLEDGEMENTS

Open access publishing facilitated by Queensland University of Technology, as part of the Wiley - Queensland University of Technology agreement via the Council of Australian University Librarians.

CONFLICTS OF INTEREST STATEMENT

The authors declare no conflicts of interest.

FUNDING INFORMATION

No funding was awarded for this project.

DATA AVAILABILITY STATEMENT

The data and R code are available at <https://www.doi.org/10.5281/zenodo.7783823>.

ORCID

David N. Borg  <https://orcid.org/0000-0002-0152-571X>
 Franco M. Impellizzeri  <https://orcid.org/0000-0002-1703-2573>
 Samantha J. Borg  <https://orcid.org/0000-0003-1376-5320>
 Kate P. Hutchins  <https://orcid.org/0000-0001-8490-7846>
 Ian B. Stewart  <https://orcid.org/0000-0003-1741-718X>
 Tamara Jones  <https://orcid.org/0000-0002-4854-0968>
 Brenton J. Baguley  <https://orcid.org/0000-0003-3974-8107>
 Lucas B. R. Orssatto  <https://orcid.org/0000-0003-3788-3700>
 Aaron J. E. Bach  <https://orcid.org/0000-0002-5581-5018>
 John O. Osborne  <https://orcid.org/0000-0001-8681-8521>
 Benjamin S. McMaster  <https://orcid.org/0000-0003-3639-9538>
 Robert L. Buhmann  <https://orcid.org/0000-0002-7984-4446>
 Joshua J. Bon  <https://orcid.org/0000-0003-2313-2949>
 Adrian G. Barnett  <https://orcid.org/0000-0001-6339-0374>

REFERENCES

1. Kadlec D, Sainani KL, Nimphius S. With great power comes great responsibility: common errors in meta-analyses and meta-regressions in strength & conditioning research. *Sports Med.* 2023;53:313-325.
2. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. Meta-analysis: pitfalls and hints. *Heart Lung Vessels.* 2013;5(4):219-225.
3. Higgins JP, Thomas J, Chandler J, et al. Chapter 10 "Analysing data and undertaking meta-analyses". *Cochrane Handbook for Systematic Reviews of Interventions.* John Wiley & Sons; 2019.
4. Hoffmann F, Allers K, Rombey T, et al. Nearly 80 systematic reviews were published each day: observational study on trends in epidemiology and reporting over the years 2000-2019. *J Clin Epidemiol.* 2021;138:1-11.
5. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. *Doing Meta-Analysis with R: A Hands-on Guide.* Chapman & Hall/CRC Press; 2021.
6. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis.* John Wiley & Sons; 2021.
7. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc A Stat Soc.* 2009;172:137-159.
8. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ.* 2011;342:964-967.
9. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods.* 2010;1:97-111.
10. Impellizzeri FM, McCall A, van Smeden M. Why methods matter in a meta-analysis: a reappraisal showed inconclusive

- injury preventive effect of Nordic hamstring exercise. *J Clin Epidemiol.* 2021;140:111-124.
11. Riley RD, Gates S, Neilson J, Alfirevic Z. Statistical methods can be improved within Cochrane pregnancy and childbirth reviews. *J Clin Epidemiol.* 2011;64:608-618.
 12. Amrhein V, Greenland S. Discuss practical importance of results based on interval estimates and p -value functions, not only on point estimates and null p -values. *J Inf Technol.* 2022;37:316-320.
 13. Int'Hout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open.* 2016;6:e010247.
 14. Lavis JN, Davies HT, Gruen RL, Walshe K, Farquhar CM. Working within and beyond the Cochrane collaboration to make systematic reviews more useful to healthcare managers and policy makers. *Health Policy.* 2006;1:21-33.
 15. Cuzzolin M, Previtali D, Delcogliano M, Filardo G, Candrian C, Grassi A. Independent versus transtibial drilling in anterior cruciate ligament reconstruction: a meta-analysis with meta-regression. *Orthop J Sports Med.* 2021;9:23259671211015616.
 16. Sainani KL, Borg DN, Caldwell AR, et al. Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. *Br J Sports Med.* 2021;55:118-122.
 17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev.* 2021;10:1-11.
 18. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. R package version 1.0.8. 2022 <https://cran.r-project.org/web/packages/dplyr/index.html>
 19. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature.* 2019;567:305-307.
 20. Altman DG, Bland JM. How to obtain the P value from a confidence interval. *Br Med J.* 2011;343:d2304.
 21. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika.* 1934;26:404-413.
 22. Dorai-Raj S. binom: Binomial confidence intervals for several parameterizations. R package version. 2014 <https://cran.r-project.org/web/packages/binom/index.html>
 23. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing; 2020 <https://www.R-project.org/>
 24. Borg DN, Barnett AG, Caldwell AR, White NM, Stewart IB. The bias for statistical significance in sport and exercise medicine. *J Sci Med Sport.* 2023;26:164-168.
 25. Borg DN, O'Brien JL. Comment on "effects of heat acclimation and acclimatisation on maximal aerobic capacity compared to exercise alone in both thermoneutral and hot environments: a meta-analysis and meta-regression". *Sports Med.* 2022;52:1715-1718.
 26. Al Amer FM, Lin L. Empirical assessment of prediction intervals in Cochrane meta-analyses. *Eur J Clin Investig.* 2021;51:e13524.
 27. von Hippel PT. The heterogeneity statistic I^2 can be biased in small meta-analyses. *BMC Med Res Methodol.* 2015;15:1-8.
 28. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Med Res Methodol.* 2008;8:1-9.
 29. Wang C-C, Lee W-C. A simple method to estimate prediction intervals and predictive distributions: summarizing meta-analyses beyond means and confidence intervals. *Res Synth Methods.* 2019;10:255-266.
 30. Nagashima K, Noma H, Furukawa TA. Prediction intervals for random-effects meta-analysis: a confidence distribution approach. *Stat Methods Med Res.* 2019;28:1689-1702.
 31. Partlett C, Riley RD. Random effects meta-analysis: coverage performance of 95% confidence and prediction intervals following REML estimation. *Stat Med.* 2017;36:301-317.
 32. Schulz R, Barnett A, Bernard R, et al. Is the future of peer review automated? *BMC Res Notes.* 2022;15:1-5.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Borg DN, Impellizzeri FM, Borg SJ, et al. Meta-analysis prediction intervals are under reported in sport and exercise medicine. *Scand J Med Sci Sports.* 2024;34:e14603. doi:[10.1111/sms.14603](https://doi.org/10.1111/sms.14603)