Review IJSM-07-2024-10699-re

General Comments

The article uses a semi-automated screening approach to describe the reporting of data processing strategies for gas exchange measurements in the published literature. The relevance of this meta-research article is clear, but a recent key reference with a comparable aim is missing (Nolte 2023). The text-mining approach allows for a large sample size of studies, but may require validation. The presentation of the methods and results could be improved, and some important aspects should be discussed.

Specific Comments

(Continuous line numbering would be helpful in the source document)

Abstract

I believe it is important to mention in the abstract that you used automatic screening procedures to assist your review. This is both a strength and a weakness of your research, but it is certainly a quite unique feature that warrants inclusion in the abstract (and possibly even in the title).

Please indicate at some point that you are providing CIs/margins of error here.

P1 19 Please state that the 1078 articles are a random subset of the larger sample.

Introduction

I generally like the reasoning and style of the introduction. Some minor comments:

P1 49 Please give a reference for this claim. Is this not what you are essentially trying to investigate?

P2 9 This is true as long as respiratory rates are lower than 60 per min, which is of course common in most test settings. However, during exhaustion, trained athletes may also reach rates of >60, in which case one-second interpolation would widen the confidence interval.

P2 13 You are missing a key reference here (Nolte 2023). They also investigated the reporting of data filtering and interpolation. However, their sample size was also small compared to your work.

P2 20 The low reporting rate of data processing methods is not only something that can be observed anecdotally but has also been demonstrated in published meta-research (Midgley 2007, Nolte 2023).

Methods:

You are missing the section ‘Protocol and registration’ from the PRISMA-ScR here (see the sources you cited).

Please provide the full RegExs in the supplemental material.

As one of the main reasons for your research here is to improve the replicability and reproducibility of other studies, I highly encourage you to make your own research reproducible by making your Python and R scripts publicly available.

You are using a semi-automated screening and data extracting procedure. Have you performed any validation for this on a subset of articles? If so, what were the performance metrics?

P2 42 Please provide a reason for not using a date restriction. Others may argue that processing strategies may have changes over time with changes in technology and education. This is also something you should discuss in the discussion section.

P2 47 Please provide the search dates for your data base searches.

P2 49 From what I can see, you provide the search strategies for all three sources in the supplemental materials?

P2 56 In my opinion, the number of excluded articles belongs in the results section. Perhaps you could write something as 'Articles that could not be accessed through our library or via unpaywall were excluded from the analysis' here.

P3 24 Please add more context on the random forest classifier(s?). How large was your training set and did you have a test set?

P3 15 Not all readers may be familiar with the term RegExs.

P3 26 Did you manually review every article flagged ineligible?

P3 32 Some metabolic carts offer both a BBB and mixing chamber mode. How did you deal with these?

P3 51 Did you also read the full-text if no snipped could be found?

P4 9 How did you generate the random sample?

P4 13 Thank you for providing a precision-based sample size calculation, something I have not seen often so far. I could fully reproduce your results. Please include an explanation how you determined your input parameters (margin of error and assumed proportion).

P5 26 How often was this the case and how did you consider this in your analysis when given proportions?

P6 34 You are calculating binomial confidence intervals, presumably using the classic Wald method. Please specify the method used and explain the reason for choosing it. Especially since the Wald interval has rather poor properties and some serious limitations that also apply to your results, e.g. overshoot (Brown 2001).

Results

P6 43 I am a bit surprised by the relatively low number of initial search results (21981). Given that you had no date restrictions, I would have expected a larger number of studies. Is there anything in your search strategy that may limit the number of studies retrieved?

P6 52 Thanks for providing the raw data of your research. I recommend uploading your data to a data repository with a persistent DOI instead of using Google Sheet links, which may become inaccessible in the future.

Discussion

As you are given recommendations on reporting of the strategies, please discuss them in light of the recently published reporting recommendations (Nolte 2023).

P8 11 This is even more important for individuals that are less fit: If evaluations for operations are based on cardiopulmonary test parameters (Johnson 1998).

P8 21 The influence of data processing on VT determination is indeed an interesting topic. I look forward to seeing this research in the future.

P8 30 In the absence of a recommendation for outlier removal, please discuss the strategies you have found in the literature. On what basis were these strategies likely chosen? And how 'good' might the +/- 4 SD strategy be? There are some important aspects to discuss (e.g., do you need outlier removal if you use filtering instead of averaging?, is there a potential value of using simultaneous confidence bands [Liebl 2023] or Bayesian credible intervals?)

P8 35 Discuss your reporting prevalence for outlier removal and interpolation with the prevalence found by Nolte (2023).

P8 41 Well argued. However, I am wondering how important the confidence interval shrinkage is to current published research. I am not aware of many articles providing confidence intervals for individual level gas exchange parameters (often, rather the point estimate is treated as the true value and confidence intervals are only constructed around the group mean, i.e., reflecting sampling uncertainty on the population level).

P8 46 'Though the effect of interpolation is likely small, …'

P9 12 Write the first author name out here

P9 14 These conversion formulas only work on the group level. When it comes to the individual patient/athlete, misclassifications can still occur because the effect of data processing strategies can vary on the individual level.

P9 25 Please discuss your finding of averaging strategies reporting with the results found in other research (Midgley 2007, Robergs 2010, Nolte 2023).

P9 27 Robergs (2010) recommends a digital filter, and the other strategies only as a substitute if researchers are unable to apply a filter. Also, he recommends a 30s-rolling average instead of the 30s-bin average you found to be most prevalent in the literature.

P9 29 Please discuss the prevalence of data processing strategies you have found in the research. Why were some common despite not recommended? Why were some uncommon despite recommendations?

P9 40 A way to deal with word limits may be to put detailed methodology in the supplemental material of a research article. However, this is something that could not be detected by your method.

P9 41 An even better approach for ensuring reproducibility than just reporting is to make data and code available. There is some existing open source software for analysis of gas exchange data you could point to here.

P9 45 I wonder how you would construct a larger sample, as you had no date and test parameter limitation in your search.

P9 55 You could discuss this in more detail. E.g., would you expect that data processing does differ based on the parameter of interest in the literature. And more importantly, should it differ?

P10 5 You can add a limitation regarding ambiguity in processing strategy reporting (and therefore your data extraction). I suppose there were many cases in which the exact data processing strategy was not clearly defined in the articles and you had to imply the original authors' meaning.

Figures

Fig 1: This figure rather confuses me instead of clarifying your research. I do not think it is necessary for your article.

Fig 2: Please use a PRISMA flowchart for this purpose. Also provide how many results you obtained from which source and how many duplicates you removed. You could also include the additional step of the random sample in the flowchart for clarity.

Fig 3/4: Maybe combine these two into one figure with two subfigures.

Tab 1: I find this table rather redundant given the result section and Fig 4.

References

Brown et al. (2001). Stat Sci 10.1214/ss/1009213286

Liebl & Reimherr (2023). J R Stat Soc Series B Stat Methodol 10.1093/jrsssb/qkad026

Nolte et al. (2023). Sports Med 10.1007/s40279-023-01903-3