PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Comparing the reporting and conduct quality of exercise and pharmacological randomized controlled trials: A systematic review
AUTHORS	Adams, Scott; McMillan, Julia; Salline, Kirsten; Lavery, Jessica; Moskowitz, Chaya; Matsoukas, Konstantina; Chen, Maggie; Santa Mina, Daniel; Scott, Jessica; Jones, Lee

VERSION 1 – REVIEW

REVIEWER	Hernandez-Martinez , Alba Universidad de Almería, Sport sciences
REVIEW RETURNED	11-Mar-2021

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript. The authors present a manuscript in which they reviewed the quality reporting of exercise and pharmacological randomized controlled trials. The manuscript is well-written, however, the objective of the research is not well addressed because of the gap in the literature is not justified. Furthermore, the methodology used for the comparison of studies is not suitable. The gap in the literature is not properly justified. In the introduction, the authors do not describe that there is an actual problem in the quality of the report of the exercise RCT. The authors do not provide an appropriate justification for the matching criteria between exercise and pharmacological RCT. The match between RCT according to reported criteria (journal, cohort, sample size, number of study sites) is insufficient to ensure an appropriate comparison. The authors have evaluated the quality reporting of exercise RCTs, however, the did not consider any journal within the sports sciences area. This issue might have biased the exercise RCT studies selection. Furthermore, it is described that this research is a review including RCT, however, the authors use a measurement tool to assess the methodological quality of systematic reviews (AMSTAR). Study selection criteria as impact factor >15 or the period of 2008-2018 are not justified. There is too much information in the supplementary material which makes it difficult to follow the flow of the manuscript.

REVIEWER	Saxton, John M Northumbria University
REVIEW RETURNED	20-Mar-2021

GENERAL COMMENTS	This manuscript presents a bold attempt to compare the reporting and conduct quality of exercise and pharmacological RCTs by

	assessing factors associated with quality separately. This is an important topic as it can help to raise awareness of weaknesses in the reporting of exercise trial data with a view to raising standards. This systematic review describes an extensive amount of work in which 48 “matched” exercise and pharmacological RCTs were compared, with the conclusion that research reporting and conduct quality are inferior in exercise trials. Notable findings were the relative lack of reporting of adverse events in exercise RCTs and evidence of improved quality scores (generally) in more recent studies. My main constructive criticism of this work is conceptual. While it is clear that an excellent level of rigour is demonstrated in the identification of studies, data extraction and use of validated tools to assess quality, the extent to which adequate matching was achieved between exercise and pharmacological RCTs is a key question that needs careful consideration and, in my view, some further discussion. For example, a stronger rationale needs to be provided for restricting the search to medical journals with an impact factor ≥15 (and ±5 impact factor points), as acceptance into such journals is influenced by factors other than study quality – what about publication bias and the preference for accepting studies with positive results which may be particularly relevant for exercise studies published in high impact factor medical journals? A comparison of the proportion of studies reporting a positive result for the primary outcome in each domain would be worthwhile in this respect. The tolerance on the sample size quality criterion (30%) seems quite large and requires further justification. More generally, of the four matching criteria defined, studies only had to be matched on two of the criteria to be included – why do the authors feel this is adequate for a robust comparison? Although on average, three matching criteria were used, in how many comparisons were only two criteria used? Additionally, because the tools for assessing the quality of research reporting and research conduct are designed (to more or less degree) for specific types of studies, to what extent were the authors able to make comparisons on the basis of important criteria that are common to both types of design? Could this be further elaborated on? There are some other more minor issues that could be considered to improve the overall quality of the manuscript. I think there is some tendency in the results section and discussion to present/discuss data for specific quality indices (e.g. adherence to protocol, etc.) for exercise studies without presenting/discussing the same parameter for pharmacological trials. Although this detail may be provided in the extensive tables and appendices, it would help the reader if such comparisons were more clearly presented in the main text. In particular, the reporting of adherence (and associated challenges) is so fundamental to this research question that I was surprised not to see a deeper consideration and discussion of it in the manuscript – while is it stated that compliance to the planned exercise regimen was missing or incomplete in 90% of exercise RCTs, the issue of compliance to pharmacological interventions (and validity of such measures) seems inadequately addressed. In summary, my view is that this is a very important research question that could help to improve the quality of reporting exercise RCTs. While the authors have robustly endeavoured to

	address this issue, I think a deeper consideration of the issues raised above would improve the contribution this work can make to the literature.

VERSION 1 – AUTHOR RESPONSE

Response to Reviewer 1

Item 1. Thank you for the opportunity to review this manuscript. The authors present a manuscript in which they reviewed the quality reporting of exercise and pharmacological randomized controlled trials. The manuscript is well-written, however, the objective of the research is not well addressed because of the gap in the literature is not justified. Furthermore, the methodology used for the comparison of studies is not suitable.

Response 1. We thank the Reviewer for their constructive feedback in suggesting how to improve the content of our manuscript. We agree with the Reviewer that the gap in the literature that our review was addressing was not clearly articulated in the original submission. The quality of reporting and conduct in RCTs testing the effects of numerous pharmacological and non-pharmacological interventions is well documented. However, despite national and international guidelines, comparatively few reviews have evaluated the quality of reporting and conduct in exercise RCTs. Indeed, to our knowledge, arguably only five similar reviews have been conducted to date. In addition, all reviews to date have been limited by focus on one disease area, small sample sizes, and use of single or select items from several reporting quality metrics. Furthermore, it has been difficult to judge the quality of exercise RCTs reported in the current available reviews because comparisons of reporting quality to similar non-exercise RCTs has not been performed. To address these gaps, we reviewed a comparatively large number of trials from diverse populations using multiple comprehensive research reporting and conduct assessment guidelines/tools and contextualized our findings via direct comparison with similar non-exercise trials.

We previously discussed the limitations of these previous reviews (please see Discussion: p. 13, paragraph 2). We have also added the following content to our Introduction (p. 4, paragraph 2)…

> "Numerous reviews have evaluated reporting quality and conduct of medical (e.g. surgical,[11] medical device[12] and pharmacological[13] interventions) RCTs. Only a few previous systematic reviews have assessed the quality of exercise RCT reporting and conduct.[14-18] However, these reviews were limited in scope (e.g., did not use comprehensive guidelines like CONSORT and Cochrane ROB; included a small number of trials) and incompletely reported key aspects of study methods (e.g., item rating criteria, reviewer training). Also, to our knowledge, no exercise reviews have contextualized their findings via direct comparison with trials in other research disciplines."

Item 3. The authors do not provide an appropriate justification for the matching criteria between exercise and pharmacological RCT. The match between RCT according to reported criteria (journal, population, sample size, number of study sites) is insufficient to ensure an appropriate comparison.

Response 3. We agree that the justification of our matching criteria could be improved. To address this concern, we have now added justifications for each of the defined matching criteria (please see below). We believe that matching based on these criterions represents a reasonable, balanced and reproducible approach to matching RCTs between disciplines, particularly in the absence of any best practice guidelines to inform this process. We also note that this concern is related to several expressed by Reviewer #2 (Items 2, 4, 5 and 6). The content that was added to the Main Document (Study Selection, Matching, Data Extraction and Additional Sources: p. 6, paragraph 2) and Supplementary Methods 3 (Section 5: RCT Matching) to address this specific concern is as follows:

> "Matching criteria for exercise and pharmacological therapy RCTs included: (1) publishing journal (±5 impact factor points according to the 2016 Journal Citation Reports [Clarivate Analytics, formerly ISI Web of Knowledge]), (2) study population (sharing similar disease characteristics), (3) study sample size (±30% difference in study sample size), and (4) number of study sites (single vs. multi-site). These specific matching criteria were selected to establish impartial comparison between exercise and pharmacological RCTs. The 'publishing journal' criteria was selected because studies published within the same journal should, in theory, be held to similar reporting standards. If no direct match could be identified within the same journal, we used an investigator-defined cut-off of ±5 impact factor points to find alternate matches because impact factor has been shown to be associated with RCT reporting and methodological quality.[29,30] The 'study population' criteria was chosen to account for differences in the research methods and standards across specific clinical populations and specialties. If no direct population match could be identified, we considered closely related populations. For example, for trials among patients with cardiac diseases, cardiomyopathy or heart failure were considered surrogates. We selected the 'study sample size' and 'number of study sites' as criterion to control for differences in the methods (eg, human and physical resources, infrastructure) used to conduct smaller versus larger trials. To this end, an investigator-defined cut-point of a 30% difference in sample size was used to match RCTs of similar scale and logistical complexity."

We also added a related limitation (p. 17) as follows:

> "We controlled for differences in the numbers of evaluable and applicable items across the reporting quality guidelines and used four matching criteria to control the influence of differences in (1) journal editorial standards and policies, (2) population-specific research methods and standards, and (3) the methods, resources, and infrastructure required to conduct smaller vs. larger trials. Future research could be strengthened by the establishment of standardized matching criteria to facilitate comparisons between branches of biomedical research."

Item 4. The authors have evaluated the quality reporting of exercise RCTs, however, the did not consider any journal within the sports sciences area. This issue might have biased the exercise RCT studies selection.

Response 4. We intentionally limited our search to journals that published both pharmacological and exercise RCTs to help ensure that the reporting and publication standards were consistently applied to both types of studies and strengthen the validity of our comparison between the study types. We acknowledge there are many other journals that publish either exercise or pharmacological studies, but including studies published in disparate journals with different reporting and publication standards would not allow the comparison between different intervention approaches, which is a major strength of our review. Nevertheless, we agree that not including journals within the sports sciences would likely have biased our exercise RCT selection. We have noted this concern in our limitations (p. 16, paragraph 2), as follows:

> "Relatedly, the exclusion of exercise RCTs published within sports science journals may underestimate the quality of exercise studies. Nevertheless, we felt it was necessary to selectively draw from this subset of journals given they are most likely to publish RCTs of both intervention types and endorse and enforce reporting quality guidelines[23-25] to impartially compare and contextualize our findings."

Item 5. Furthermore, it is described that this research is a review including RCT, however, the authors use a measurement tool to assess the methodological quality of systematic reviews (AMSTAR).

Response 5. We respectfully draw the Reviewer's attention to the following paper: https://www.bmj.com/content/358/bmj.j4008 . The AMSTAR tool is used to evaluate the quality of systematic reviews of randomized and non-randomized trials. Our study is a systematic review of randomized trials. Therefore, we endeavoured to meet the standards of a tool that is designed to evaluate the quality of our work. On the other hand, the PRISMA checklist defines the minimum standard for the reporting and conduct of systematic reviews of randomized trials. Given the nature of our study (i.e. critiquing the reporting quality and conduct of studies), we believed it was important to meet the highest possible standards for the reporting and conduct of systematic reviews (defined by PRISMA and AMSTAR guidelines).

Item 6. Study selection criteria as impact factor >15 or the period of 2008-2018 are not justified.

Response 6. We thank the Reviewer for allowing us to clarify these points.

Re: Impact factor – We agree with the Reviewer that the impact factor restriction needs to be better justified. To address this issue, we have expanded upon our previous justification within the Main Document (Search Strategy, p. 5, end of paragraph 2) and Supplementary Methods 3 (Section 2, paragraph 1). Specifically, we stated:

> "We purposefully restricted our search to medical journals with impact factors ≥15 because journals with higher impact factors are more likely to endorse and enforce reporting quality guidelines (Kunath et al., 2012; Samaan et al., 2013; Mills et al., 2005) and publish both exercise and pharmacological RCTs – leading to a more balanced foundation for comparison between study types. This impact factor-based restriction is also consistent with the methods from similar reviews of medical, psychosocial, and behavioural RCTs (Hoffmann et al., BMJ, 2013; Mills et al., Br J Clin Pharmacol, 2004; Mills et al., Contemp Clin Trials, 2005; Khan et al., JAMA Netw Open, 2019; Pandis et al., J Dent, 2010; Grant et al., PLoS One, 2013; Ghimire et al., Trials, 2012)."

Re: Search period – We agree with the Reviewer that this aspect of our Methods should be better justified within the Main Document. Notably, it is common for reviews of this nature to restrict their search criteria to a small publication period, within select journals, and involving targeted patient populations (e.g., Akl et al., 2012, BMJ, 344:e2809; Flemming et al., 2014, J Clin Epidemiol, 67 (2014) 754e759; Grant et al., 2013, PLoS ONE 8(5): e65442; Howard et al., 2017, PLoS ONE 12(7): e0180986; Peron et al., 2012, J Natl Cancer Inst 2012;104:982–989; Ritchie et al., 2020, Int J Pharm Pract 28:220-232). By comparison, our search included RCTs published over a 10-year period, within 45 journals with impact factors >15, and did not exclude on the basis of patient population – representing a far more comprehensive search and inclusion compared to similar reviews. To address this, we added the related content from Supplementary Methods 3 to the Main Document (Search Strategy: p. 5, paragraph 2). Specifically, we state:

> "The search was restricted to trials published between January 1st 2008 (the year the CONSORT extension for Non-Pharmacologic Treatments (CONSORT-NPT) was first published[22]) and the search date (March 8th, 2018)."

Item 7. There is too much information in the supplementary material which makes it difficult to follow the flow of the manuscript

Response 7. We agree with the Reviewer that the amount of supplementary material included for this review is considerable and the multiple references to it in the Main Document disrupts the flow of the manuscript. We have attempted to address these concerns in two ways. First, we have added general references to the supplementary materials at the beginning of the Methods (p. 5, paragraph 1) and Results (p. 10, paragraph 3) sections. Second, we removed

most of the references to specific sections of the supplementary materials from the body of the manuscript.

Methods text: “Full study methods are provided within Supplementary Methods 3-7 and Supplementary Table 1.”

Results text: “See Supplementary Tables 2-12 for full study characteristics and results.”

Response to Reviewer 2

Item 1. This manuscript presents a bold attempt to compare the reporting and conduct quality of exercise and pharmacological RCTs by assessing factors associated with quality separately. This is an important topic as it can help to raise awareness of weaknesses in the reporting of exercise trial data with a view to raising standards. This systematic review describes an extensive amount of work in which 48 “matched” exercise and pharmacological RCTs were compared, with the conclusion that research reporting and conduct quality are inferior in exercise trials. Notable findings were the relative lack of reporting of adverse events in exercise RCTs and evidence of improved quality scores (generally) in more recent studies. My main constructive criticism of this work is conceptual. While it is clear that an excellent level of rigour is demonstrated in the identification of studies, data extraction and use of validated tools to assess quality, the extent to which adequate matching was achieved between exercise and pharmacological RCTs is a key question that needs careful consideration and, in my view, some further discussion.

Response 1. We thank the Reviewer for their supportive appraisal of our work. We also thank the Reviewer for the opportunity to clarify our methods and the potential limitations of our work.

Item 2. For example, a stronger rationale needs to be provided for restricting the search to medical journals with an impact factor ≥15 (and ±5 impact factor points), as acceptance into such journals is influenced by factors other than study quality – what about publication bias and the preference for accepting studies with positive results which may be particularly relevant for exercise studies published in high impact factor medical journals? A comparison of the proportion of studies reporting a positive result for the primary outcome in each domain would be worthwhile in this respect.

Response 2. We thank the Reviewer for allowing us to clarify these issues.

Re: Impact factor – We agree that a stronger rationale for our impact factor criteria is needed. To this end, we have expanded our previous justification within the Main Document (Search Strategy: p. 5, end of paragraph 1) and Supplementary Methods 3 (Section 2, paragraph 1). Specifically, we state:

> "We purposefully restricted our search to medical journals with impact factors ≥15 because journals with higher impact factors are more likely to endorse and enforce reporting quality guidelines (Kunath et al., 2012; Samaan et al., 2013; Mills et al., 2005) and publish both exercise and pharmacological RCTs – leading to a more balanced foundation for comparison between study types. This impact factor-based restriction is also consistent with the methods from similar reviews of medical, psychosocial, and behavioural RCTs (Hoffmann et al., BMJ, 2013; Mills et al., Br J Clin Pharmacol, 2004; Mills et al., Contemp Clin Trials, 2005; Khan et al., JAMA Netw Open, 2019; Pandis et al., J Dent, 2010; Grant et al., PLoS One, 2013; Ghimire et al., Trials, 2012)."

Re: Publication bias – We agree with the Reviewer that any RCT reporting positive trial findings (including exercise RCTs) may be preferentially viewed and, therein, more likely to be accepted for publication. To directly address this concern, we compared the proportion of trials reporting a positive finding for the primary outcome. Specifically, 34/48 (71%) of exercise RCTs and 31/48 (65%) of pharmacological RCTs reported positive trial outcomes. This data supports the conclusion that publication bias may be similarly present for both interventions. We have now provided this context in the Main Document (p. 11, paragraph 2), wherein we state:

> "Overall, 34 of 48 exercise RCTs (71%) and 31 of 48 pharmacological RCTs (65%) reported positive primary outcomes."

Item 3. The tolerance on the sample size quality criterion (30%) seems quite large and requires further justification.

Response 3. We agree and thank the Reviewer for raising this important point. While this is an investigator defined cut-point, we contend that studies with sample sizes within 30% of each other (e.g. 25±8 participants; 75±23 participants; 300±90 participants) require a similar degree of planning, administration, and infrastructure to conduct and may encounter similar methodological issues that could impact their rigour. To address this concern, we have added the following justification to our methods (p. 7, paragraph 1):

> "We selected the 'study sample size' and 'number of study sites' as criterion to control for important differences in the methods (e.g. human and physical resources, infrastructure) used to conduct smaller vs. larger trials. To this end, an investigator-defined cut-point of a 30% difference in sample size was used to match RCTs of similar scale and logistical complexity."

Item 4. More generally, of the four matching criteria defined, studies only had to be matched on two of the criteria to be included – why do the authors feel this is adequate for a robust comparison? Although on average, three matching criteria were used, in how many comparisons were only two criteria used?

Response 4. This is an excellent point. Thank you for raising it. Our overarching approach attempted to strike a balance between maintaining external and internal validity. Only 13 of 48 (27%) matches were based on 100% matching and an additional 18 of 48 (38%) matches were based on 75% matching across the four criteria (total 65% of included studies). Our concern was that a review comparing only 62 matched exercise and pharmacological RCTs may have limited generalizability; so, we opted to expand our sample to include an additional 17 pairs of studies based on 50% matching across the four criteria.

Unfortunately, a sub-analysis among studies matched on 100% of criteria to confirm whether the overall pattern of results holds is not feasible or informative given there are only 13 studies in this group. However, to partially address this concern, we conducted one-way ANOVAs to compare the mean differences in CONSORT, CONSORT-Harms, and Intervention scores between studies matched on 100%, 75%, and 50% of criteria. There were no differences between any of the groups. We have now added this information to the Main Document (p. 10, paragraph 3; p. 12, paragraph 2) and Supplementary Table 3, as below:

> Results (p. 10): "Overall, 13 pairs of exercise and pharmacological RCTs were matched on 100% of our four matching criteria, 18 pairs of RCTs were matched on 75%, and 17 pairs of RCTs were matched on 50% (average 3 out of 4 criteria; 75%)."

> Results (p. 11): "Exploratory one-way ANOVAs did not indicate a difference in reporting quality outcomes between exercise and pharmacological RCTs matched on 50%, 75%, or 100% of the matching criteria."

Supplementary Table 3

Reporting Outcome	Criteria Matched	Number of Matched Studies	Difference Between Matched Studies	Standard Deviation	95% CI	p-value
CONSORT	100%	13	-14.8	17.4	-25.3 to -4.3	.97
	75%	18	-15.1	18.4	-24.2 to -5.9	

	50%	17	-13.7	19.3	-23.6 to -3.8	
CONSORT-Harms	100%	13	-19.7	29.0	-37.2 to -2.1	.85
	75%	18	-12.9	40.3	-32.9 to 7.2	
	50%	17	-17.7	32.7	-34.5 to -0.9	
Intervention	100%	13	-7.7	26.0	-23.4 to 8.0	.53
	75%	18	-14.8	22.1	-25.8 to -3.9	
	50%	17	-5.9	25.7	-19.1 to 7.3	

Item 5. Additionally, because the tools for assessing the quality of research reporting and research conduct are designed (to more or less degree) for specific types of studies, to what extent were the authors able to make comparisons on the basis of important criteria that are common to both types of design? Could this be further elaborated on?

Response 5. Another excellent point. All CONSORT-Harms items are applicable to both study types, whereas 37 items from CONSORT / CONSORT-NPT and six items from TIDieR guidelines are applicable to exercise and pharmacological RCT. We have now performed additional analyses on these 'core items' to address this concern. The previously reported differences for CONSORT-Harms between exercise and pharmacological trials stand as all items were applicable. However, we found that the mean overall reporting of 'core items' from CONSORT / CONSORT-NPT and TIDieR between exercise and pharmacological RCTs were not significantly different. We further report that several key 'core items' remain sub-optimally reported for exercise RCTs. We have now supplemented our objectives (Main Document, p. 4, paragraph 3), methods (Main Document, p. 7, paragraph 2; p. 8, paragraph 1), results (Main Document, p. 12, paragraph 3), and reframed our discussion (Main Document, p. 13, paragraph 2; p. 16, paragraph 1) to provide additional context to this end, as below:

> Objectives (p. 4): "We also compared the quality of research reporting and conduct from exercise RCTs to matched RCTs of pharmacological therapies (a well-established field of biomedical research with a long history of adopting RCT methods[19]) using (1) the complete guidelines and (2) only key items from the guidelines (ie, those generally applicable to both intervention types) to provide context for our findings."

Methods (p. 7): “Each trial was evaluated on two sets of criteria: (1) quality of research reporting and (2) quality of research conduct using complete standardized inventories and/or key items from these inventories, as needed.”

Methods (p. 8): “Therefore, intervention reporting for pharmacological interventions was assessed using six key items from TIDieR (including intervention length, modality, location, frequency, dose, and adherence). Exercise dose consisted of session intensity and duration (aerobic and resistance interventions) as well as the number of sets and repetitions (resistance interventions only). Exercise RCT reporting was also re-evaluated using just the 37 items from the CONSORT guidelines that are common to both intervention types.[7]”

Results (p. 12):

“**Comparison of Key Items**

Thirty-seven of 52 CONSORT items, all ten CONSORT-Harms items, and six of 16 TIDieR items were considered key items. Median reporting scores for the key items from CONSORT and TIDieR were not significantly different between exercise and pharmacological RCTs; whereas, reporting scores for CONSORT-Harms was significantly lower for exercise RCTs (Table 2). Compared to pharmacological RCTs, exercise RCTs had lower reporting of key study methods (e.g. blinding after group assignment [60% vs. 98%], balanced discussion of harms vs. benefits [39% vs. 66%], intervention modality [39% vs. 66%], intervention dose [50% vs. 98%], and complete intervention descriptions [0% vs. 67%]).”

Discussion (p. 13): “Our findings demonstrate that the quality of exercise therapy RCT reporting and conduct is suboptimal according to all complete guidelines and inventories used in this study and is inferior to RCTs of pharmacological therapy. However, the mean overall reporting quality for RCT methods and interventions, but not harms, was similar between intervention types when considering key items within the respective guidelines.”

Discussion (p. 15): “There were no differences observed in mean overall reporting quality when comparing exercise and pharmacological RCTs according to key items from the CONSORT guidelines; however, the reporting of several critical individual items was suboptimal within exercise RCTs (e.g. complete intervention descriptions, intervention dose, blinding status). Our findings provide important direction to improve the completeness and rigor of exercise trial reporting.”

Item 6. There are some other more minor issues that could be considered to improve the overall quality of the manuscript. I think there is some tendency in the results section and discussion to present/discuss data for specific quality indices (e.g. adherence to protocol, etc.) for exercise studies without presenting/discussing the same parameter for pharmacological trials. Although this detail may be provided in the extensive tables and appendices, it would help the reader if such comparisons were more clearly presented in the main text. In particular, the reporting of adherence (and associated challenges) is so fundamental to this research question that I was surprised not to see a deeper consideration and discussion of it in the manuscript – while is it stated that compliance to the planned exercise regimen was missing or incomplete in 90% of exercise RCTs, the issue of compliance to pharmacological interventions (and validity of such measures) seems inadequately addressed.

Response 6. We agree with the Reviewer that including a more balanced discussion of the findings across study types would strengthen the manuscript. We have now added content to this end in two sections of our discussion (p. 15, paragraph 1; p. 14, paragraph 2). Specifically, we added:

> "In our study, information on patient compliance to the planned exercise regimen as well as the expertise of the individuals implementing the intervention was missing or incomplete in >90% of trials; fundamental details pertaining to dose of prescribed exercise were also missing in 50% trials." "By contrast, pharmacological intervention compliance was similarly missing in ~80% of trials; however, prescribed pharmacotherapy dose was only missing in 2% of studies."

> "Our study extends these findings by demonstrating that harms-related monitoring and reporting were missing or incompletely reported in ≥75% of exercise RCTs; and, relatedly, >50% of articles failed to provide a balanced discussion of risks to benefits for the tested interventions." "In contrast, a related assessment of 325 chemotherapy trials reported a mean CONSORT-Harms score of 63%,[139] compared to mean harms scores of 36% (exercise RCTs) and 57% (pharmacological RCTs) in our study."

VERSION 2 – REVIEW

REVIEWER	Hernandez-Martinez , Alba Universidad de Almería, Sport sciences
REVIEW RETURNED	05-Jun-2021

GENERAL COMMENTS	I thank the authors for their responses and modifications to the manuscript.