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)	Depressive symptoms and the General Health of Retired Professional Footballers compared to the General Population in the United Kingdom: A Case Control Study
AUTHORS	Fernandes, Gwen; Parekh, Sanjay; Moses, Jonathan; Fuller, Colin; Scammell, Brigitte; Batt, Mark; Zhang, Weiya; Doherty, Michael

VERSION 1 - REVIEW

REVIEWER	Paul Gorczynski University of Portsmouth, UK
REVIEW RETURNED	19-Mar-2019

GENERAL COMMENTS	Overall, the paper addresses a major area of interest in professional sport. The authors need to make a stronger case as to why another cross sectional study on mental health in professional sport is needed, especially one that uses a series of self-report scales. Unfortunately, the paper does have a number of major limitations and, at times, lacks clarity and specificity. Perhaps most importantly, are the authors discussing diagnoses of generalized anxiety disorder or anxiety symptoms? Same when referencing depression. Depressive symptoms or diagnoses of major depressive disorder? Additionally, the paper lacks a definition of wellbeing, which is a major problem given this is a focus of the paper. The authors may wish to examine the following paper for clarity: http://www.internationaljournalofwellbeing.org/index.php/ijow/article/viewFile/89/238?origin=publicati Dodge, R., Daly, A., Huyton, J., & Sanders, L. (2012). The challenge of defining wellbeing. International Journal of Wellbeing, 2(3), 222-235. doi:10.5502/ijw.v2i3.4 Lastly, there is no definition of professional retired footballers. There is a lack of clarity as to who the target population actually was. Furthermore, more details are needed as to how the general population and how they were recruited for this study. Below are further comments on different aspects of the study: Methods:

	Please explain the rationale for the exclusion criteria. Why were individuals with serious health conditions, both mental and physical, excluded from the study? Their inclusion may have enriched your findings. Why was the general population limited to men over the age of 40 years? Also, how was gender handled in this study (e.g., self identifying?). Anxiety and Depression The HADS is a self-report scale to examine psychological distress. The scale has two subscales, anxiety and depression. The scale does not render a diagnosis of either GAD or MDD. To say that in the study anxiety and depression were determined by the scale is misleading as it is not capable of doing that. At most, the authors may report that they were exploring anxiety symptoms and depressive symptoms or symptoms of psychological distress. As such, I do not believe this method could actually fulfil objective number 1 (i.e. determine the prevalence of depression or anxiety). Discussion The authors should revisit their discussion and add more information about the following items: - application of findings, what is actually going to be done with these findings? - limitations, why were women excluded? why was no information on sexuality, race, ethnicity captured? These are major factors that are associated with mental health outcomes. - future research, be specific. How do we move on from cross sectional studies to interventions? What is it we need to do?

REVIEWER	Alexandra Parker Institute for Health and Sport, Victoria University Australia
REVIEW RETURNED	01-Apr-2019

GENERAL COMMENTS	This is a well-conducted case control study that met the objectives of determining the prevalence of mental health and general health concerns in retired professional footballers compared to general population controls. As the authors described, this study addresses some of the identified gaps in the evidence base of the mental health and wellbeing of professional sports-people. The following minor revisions are suggested: 1) Description of anxiety and depression in abstract: This phrased a little awkwardly, perhaps state that scores greater than 11 indicate probable caseness? 2) Description of SF-12 composites: Ensure this is referred to as "physical and mental health" component scores throughout the manuscript. 3) Description of anxiety and depression measure in methods: Best to rephrase as indicating likely caseness. Clinical interview is the gold-standard for definitive diagnosis according to DSM 5 criteria. 4) Quality of life measures in methods: should be described as "mental health component score". 5) Comorbidities measure in methods: sexual dysfunction would be an important consideration for this age group and perhaps not

	likely to be generated in open text responses. Perhaps consider inclusion within limitations? 6) Discussion: Consideration of selection of cut-off scores and their interpretation. Goutterbarge et al 2015 defined the presence of anxiety or depression as 2 or above on the GHQ-12 (rather than 3+ as indicating presence of disorder) . This may be comparable to a 'borderline' score of 8-10 on the HADS. Consider reporting the number of participants in current sample that scored 8-10 on HADS for comparison purposes within Results sections and included in the Discussion. 7) Discussion: Consider including risks associated with with analgesic medication use. Despite cross-sectional nature of study, retired footballers may be at greater risk of harm from the common side effects associated with pain medication use. 8) Discussion: Future directions could also include measurement of concussion history given its association with mental health concerns in elite athletes. Limitations and future directions were appropriately acknowledged and considered.

REVIEWER	David Webner, MD Crozer Keystone Health System, USA Philadelphia Union FC, USA
REVIEW RETURNED	02-Apr-2019

GENERAL COMMENTS	Syntax: On page 4, in the introduction paragraph, on line 12, the sentence beginning with "Retired professional..." is a very long, run-on sentence that is confusing. The authors need to retool this sentence so that it becomes more easily readable and understandable. In the next paragraph the first line should add the words "in footballers," after the words "high-profile cases of depression..." I do have an issue with the power calculation. It appears to me that some of the "non-significant findings" could have become statistically significant if the pool of interviewed footballers was larger. I think this reads more appropriately as a pilot study -- not a comprehensive study. I would also like to know if during the study period whether any of the participants died. That might also be an interesting long term outcome for further studies. Also, I'm not exactly sure which population the controls and the footballers came from. Would stratifying this with respect to income level be helpful? Finally, I did not see any references comparing this data set with the comprehensive US data set in American Football players. While they are clearly a collision sports and European football is a contact sport, the data from the NFL is rich with respect to sheer size. I think a sentence or to remarking on this data set would be appropriate.

REVIEWER	Michael Gaetz University of the Fraser Valley Canada
REVIEW RETURNED	07-Apr-2019

GENERAL COMMENTS	Overall impression: This is outstanding work. I have made a minor comment on Methods but other than that, a comprehensive, timely, well-planned and executed work. Congratulations. Abstract: No comments. Introduction: Very well-written. Methods: Page 7, Lines 33-35: You state: "The difference in SF versions was due to logistical reasons as the control postal questionnaire was more challenged by length restrictions than the second follow-up questionnaire to ex-footballers." Challenged by what? It is not clear to the reader why you chose one form over the other for your samples. Please clearly state why you chose to use the SF-12 for your control sample and the SF-36 for your football sample. The rationale provided by Jenkinson eta al. 1997 would be helpful here. Results: No comments. Discussion: No comments.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Paul Gorczynski

Institution and Country: University of Portsmouth, UK

Please state any competing interests or state 'None declared': None declared

Overall, the paper addresses a major area of interest in professional sport. The authors need to make a stronger case as to why another cross sectional study on mental health in professional sport is needed, especially one that uses a series of self-report scales. Unfortunately, the paper does have a

number of major limitations and, at times, lacks clarity and specificity. Perhaps most importantly, are the authors discussing diagnoses of generalized anxiety disorder or anxiety symptoms? Same when referencing depression. Depressive symptoms or diagnoses of major depressive disorder? Additionally, the paper lacks a definition of wellbeing, which is a major problem given this is a focus of the paper. The authors may wish to examine the following paper for clarity:

http://www.internationaljournalofwellbeing.org/index.php/ijow/article/viewFile/89/238?origin=publicati

Dodge, R., Daly, A., Huyton, J., & Sanders, L. (2012). The challenge of defining wellbeing. International Journal of Wellbeing, 2(3), 222-235. doi:10.5502/ijw.v2i3.4

We take on board the reviewer's comment on the clarity needed around our use of the Hospital Anxiety and Depression Scale as opposed to other measures. This was a postal questionnaire sent to both ex-professional footballers and control participants from the general population and therefore clinical diagnoses were simply not possible. However, we have now updated our methods section to reflect that whilst the DSM-5 may have been more definitive, the HADS has been shown to compare consistently and positively with the DSM.

We have also included a caveat in the discussion section of the manuscript to acknowledge that whilst HADS may have distinct advantages, it may in fact only be capturing the symptoms of anxiety and depression in a non-psychiatric population as opposed to a more definitive diagnosis.

Lastly, we have considered Dodge's publication above and have decided to alter the term well-being in our manuscript to what we actually measured and discussed – quality of life using the SF questionnaires. We appreciate some of the confusion may have come from the use of the terms 'wellness' or 'well-being' and we hope we have tightened the manuscript now by removing this.

Lastly, there is no definition of professional retired footballers. There is a lack of clarity as to who the target population actually was. Furthermore, more details are needed as to how the general population and how they were recruited for this study.

We thank the reviewer for this comment. We have now explicitly described our recruitment strategy for the study. This information was referred to from a previous publication and we used it's citation but we have now updated our methods for clarity.

Below are further comments on different aspects of the study:

Methods:

Please explain the rationale for the exclusion criteria. Why were individuals with serious health conditions, both mental and physical, excluded from the study? Their inclusion may have enriched

your findings. Why was the general population limited to men over the age of 40 years? Also, how was gender handled in this study (e.g., self identifying?).

The authors excluded participants with known terminal illness, severe psychiatric illness or dementia or any other condition considered by their GP to make them unsuitable for receiving a postal questionnaire as a) they may not be able to consent to taking part in the study, b) they may not be able to complete the questionnaire, and c) this was an stipulation of the participating GP practices who sent the questionnaires to their practice patients.

The original general sample population sample (Knee Pain in the Community Study) included both men and women (n=9507). However, as our population of ex-professional footballers were all male (professional football particularly played 20-30 years ago was a largely male dominated sport with only an all-male English Football League), we did not think it appropriate to compare the entire KPIC sample. We therefore included only male responders to this sub-study as a suitable comparison group for the male ex-professional footballers who were all over the age of 40 years. As this study was embedded into a much larger project on knee pain and knee osteoarthritis (which tends to affect an ageing population more), the cut-off of 40 years was established. As a result, our sample consists of only participants over this age threshold.

Anxiety and Depression

The HADS is a self-report scale to examine psychological distress. The scale has two subscales, anxiety and depression. The scale does not render a diagnosis of either GAD or MDD. To say that in the study anxiety and depression were determined by the scale is misleading as it is not capable of doing that. At most, the authors may report that they were exploring anxiety symptoms and depressive symptoms or symptoms of psychological distress. As such, I do not believe this method could actually fulfil objective number 1 (i.e. determine the prevalence of depression or anxiety).

We take the point of the reviewer that the HADS may be more suitable to identifying symptoms of anxiety and depression as opposed to providing a definitive clinical diagnosis. However, given the resources available and the format with which the questionnaires were sent (post), inclusion of HADS was the best option to identifying potential cases of anxiety and depression in both our case and control population. Secondly, we used a particularly stringent cut-off of scores of 11 and over to determine caseness of anxiety and depression in order to maximise sensitivity of this tool. This particular cut off was defined on the basis of psychiatric rations of anxiety and depression disorders – we have included this comment with reference in our methods section. Lastly, the authors could have made the case for the use of the 12-item general health questionnaires (GHQ) which has been used in previous research on anxiety and depression in retired professional footballers however, studies have shown that HADS displays better sensitivity and specificity in detecting depression compared to the GHQ.

Discussion

The authors should revisit their discussion and add more information about the following items:

- application of findings, what is actually going to be done with these findings?

The authorship team believe that these findings are relevant given the limited evidence on mental health outcomes in retired professional UK footballers compared to a general population control. Secondly, existing evidence and anecdotal case reports relay research that footballers are particularly depressed and anxious individuals compared to the general population (Gouttebarge et al., 2015) and this may not be accurate as is evidence by our findings. The focus should be on therefore identifying the risk as well as the preventative risk factors of our work – the focus on optimism or resilience in dealing with traumatic injuries or chronic pain would be a step in the right direction. We have now included a statement on this in our concluding paragraph alongside the call for healthcare providers and club management to focus on improving pain management strategies in these groups of ex-athletes.

- limitations, why were women excluded? why was no information on sexuality, race, ethnicity captured? These are major factors that are associated with mental health outcomes.

As explained previously, all participants in this study were male as the ex-professional footballers recruited were from a database of all male professionals (details stored by the Professional Footballers Association and individual league clubs in the English Football League). We did not ask about specific questions on sexuality in this demographic. We did ask about ethnicity and the majority of ex-professional footballers were white-Caucasian (98%) whilst 2% were Black or of African American decent.

- future research, be specific. How do we move on from cross sectional studies to interventions? What is it we need to do?

We appreciate the reviewer's comments on furthering the research in this area and in order to appropriate and targeted interventions to be developed, we need to identify what truly makes ex-footballers unique or different or perhaps better able to cope with significant traumatic injuries or chronic widespread pain. These are beyond the scope of the current project.

Reviewer: 2

Reviewer Name: Alexandra Parker

Institution and Country: Institute for Health and Sport, Victoria University

Australia

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

This is a well-conducted case control study that met the objectives of determining the prevalence of mental health and general health concerns in retired professional footballers compared to general population controls. As the authors described, this study addresses some of the identified gaps in the evidence base of the mental health and wellbeing of professional sports-people.

Thank you for your comments.

The following minor revisions are suggested:

1) Description of anxiety and depression in abstract: This phrased a little awkwardly, perhaps state that scores greater than 11 indicate probable caseness?

We have now amended this sentence.

2) Description of SF-12 composites: Ensure this is referred to as "physical and mental health" component scores throughout the manuscript.

We have now ensured this is consistent throughout the manuscript

3) Description of anxiety and depression measure in methods: Best to rephrase as indicating likely caseness. Clinical interview is the gold-standard for definitive diagnosis according to DSM 5 criteria.

We have now amended this section based on the comments of Review 1 and discussed the gold-standard as the DSM 5 with added comparisons of HADS performance to the DSM 5.

4) Quality of life measures in methods: should be described as "mental health component score".

We have now amended this.

5) Comorbidities measure in methods: sexual dysfunction would be an important consideration for this age group and perhaps not likely to be generated in open text responses. Perhaps consider inclusion within limitations?

This is a valid comment given the literature in this area. However, whilst we had an open text box for comorbidities, sexual dysfunction was not an outcome openly admitted to which is unsurprising given this population of elite male athletes. However, it is also worth highlighting to the Reviewer that we did undertake PPI for the project and any questions that generated discomfort were the reason that footballers opted not to complete the questionnaire, not to complete the questionnaire in the right detail or indeed, return the questionnaire to the unit. It is indeed a balance of asking the right questions in the correct manner and perhaps clinical interviews would better lend themselves to this type of data collection as opposed to postal questionnaires.

6) Discussion: Consideration of selection of cut-off scores and their interpretation. Goutterbarge et al 2015 defined the presence of anxiety or depression as 2 or above on the GHQ-12 (rather than 3+ as indicating presence of disorder) . This may be comparable to a 'borderline' score of 8-10 on the HADS. Consider reporting the number of participants in current sample that scored 8-10 on HADS for comparison purposes within Results sections and included in the Discussion.

I value this comment but respectfully would counter that the purpose of this work is to compare the rates of anxiety and depression in ex-professional footballers with a general population comparison

group in the United Kingdom. Goutterbarge's study is the only other comparison in this research area but is still of limited comparative value particularly of those with a 'moderate' caseness score in a study that had small sample sizes (n=70-149), did not include a general population control group and were drawn from different population groups (Finland, France, Norway, Sweden and Spain) with variations in mental health service provision or economic stability (Reibling et al., 2017).

Reibling N, Beckfield J, Huijts T, et al. Depressed during the depression: has the economic crisis affected mental health inequalities in Europe? Findings from the European Social Survey (2014) special module on the determinants of health. European journal of public health 2017;27(suppl_1):47-54. doi: 10.1093/eurpub/ckw225 [published Online First: 2017/03/30]

7) Discussion: Consider including risks associated with with analgesic medication use. Despite cross-sectional nature of study, retired footballers may be at greater risk of harm from the common side effects associated with pain medication use.

This may be true – the ex-footballers in this study do report greater overall body pain and have higher use of all types of analgesic medications. However, due to the nature of the study design, being a cross-sectional method, we would not be able to determine causality from this work.

8) Discussion: Future directions could also include measurement of concussion history given its association with mental health concerns in elite athletes.

This is an excellent point and in fact, the next project we are working on with this group of elite athletes is a study on neurocognitive functioning in relation to professional football practices and the potential impact that has had on their quality of life. This has only recently received funding (December 2018) so we are excited to investigate this aspect of health and well-being of our ex-professional footballers. We have now included this in our future work section.

Limitations and future directions were appropriately acknowledged and considered.

Thank you again for your valuable comments and feedback.

Reviewer: 3

Reviewer Name: David Webner, MD

Institution and Country: Crozer Keystone Health System, USA

Philadelphia Union FC, USA

Please state any competing interests or state 'None declared': N/A

Please leave your comments for the authors below

Syntax: On page 4, in the introduction paragraph, on line 12, the sentence beginning with "Retired professional..." is a very long, run-on sentence that is confusing. The authors need to retool this sentence so that it becomes more easily readable and understandable.

We have now amended this.

In the next paragraph the first line should add the words "in footballers," after the words "high-profile cases of depression..."

We have now amended this.

I do have an issue with the power calculation. It appears to me that some of the "non-significant findings" could have become statistically significant if the pool of interviewed footballers was larger. I think this reads more appropriately as a pilot study -- not a comprehensive study.

We carried out a sample size calculation based on the prevalence of 12.6% depression of community-based adults and used these to determine the number of ex-footballers and controls required. Although we would have ideally had high numbers (we have been told by the Professional Footballer's Association that there are unofficially 26,000 ex professional footballers in the UK), the study was limited to the availability of databases provided by the PFA and league clubs who could help with the recruitment.

I would also like to know if during the study period whether any of the participants died. That might also be an interesting long term outcome for further studies.

As of the end of the study, no ex-professional footballers or control participants had died at the time of our analysis however as this is a cohort that ranges in age from 40 to 94, this may no longer be the case. Any future work, such as studies on neurodegenerative diseases or other long-term health outcomes would need to consider this valid point.

Also, I'm not exactly sure which population the controls and the footballers came from. Would stratifying this with respect to income level be helpful?

Based on this comment and that of a previous reviewer, we have now described in detail where the population of control men and ex footballers were recruited from in the methods section of the manuscript. We unfortunately did not collect any income level data in our study.

Finally, I did not see any references comparing this data set with the comprehensive US data set in American Football players. While they are clearly a collision sports and European football is a contact sport, the data from the NFL is rich with respect to sheer size. I think a sentence or to remarking on this data set would be appropriate.

American Football is a complete collision sport whilst European football is a non-collision one. It would not be entirely fair to compare such distinct sporting populations despite them all being at an elite athlete level. Furthermore, the NFL studies that are being referred to are selected on the basis of concussion history whereas participants in this study were recruited regardless of injury or any other

health outcomes (Didehbani et al., 2013). The comparisons of higher depression rates would be invalid given the distinct differences in recruitment strategies as well as population level factors that may influence depression and mental health and well-being (e.g provision of primary health care services for mental health, differences in income levels, differences in environmental factors, etc).

Didehbani N, Munro Cullum C, Mansinghani S, Conover H, Hart J Jr. Depressive symptoms and concussions in aging retired NFL players. Arch Clin Neuropsychol. 2013;28(5):418–424. doi:10.1093/arclin/act028

Reviewer: 4

Reviewer Name: Michael Gaetz

Institution and Country: University of the Fraser Valley

Canada

Please state any competing interests or state ‘None declared’: None declared.

Please leave your comments for the authors below

This is outstanding work (and I rarely say that). Please see attached.

Thank you for your comments.

Abstract: No comments. Introduction: Very well-written.

Methods: Page 7, Lines 33-35: You state: “The difference in SF versions was due to logistical reasons as the control postal questionnaire was more challenged by length restrictions than the second follow-up questionnaire to ex-footballers.” Challenged by what? It is not clear to the reader why you chose one form over the other for your samples. Please clearly state why you chose to use the SF-12 for your control sample and the SF-36 for your football sample. The rationale provided by Jenkinson eta al. 1997 would be helpful here.

We take the reviewer’s comments and have now improved this section of our methodology by explaining the factors that led us to take our decision. We have also used the Jenkinson paper as providing the rationale behind our decision.

Results: No comments.

Discussion: No comments.

VERSION 2 – REVIEW

REVIEWER	Paul Gorczynski University of Portsmouth
REVIEW RETURNED	07-Jun-2019

GENERAL COMMENTS	The authors have done a good job of addressing the comments I have raised. But, I still stick with my point on the HADS and that what was explored was anxiety/depressive symptoms rather than diagnoses of anxiety or depression. I fully understand the nature of the study and the various constraints that were experienced. But diagnoses were not obtained.

REVIEWER	Alexandra Parker Institute for Health and Sport, Victoria University, Australia
REVIEW RETURNED	26-May-2019

GENERAL COMMENTS	The authors have responded appropriately to the reviewers' comments and, therefore, the recommendation is to accept the manuscript for publication.

REVIEWER	David Webner, MD Crozer Keystone Health System
REVIEW RETURNED	05-Jun-2019

GENERAL COMMENTS	Thank you for the revisions. Good luck.