PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Physical activity and sport participation among adolescents: associations with mental health in different age groups. Results from the Young-HUNT Study, a cross sectional survey
AUTHORS	Guddal, Maren; Stensland, Synne; Småstuen, Milada; Bakke Johnsen, Marianne; Zwart, John-Anker; Storheim, Kjersti

VERSION 1 - REVIEW

REVIEWER	Shuichi Suetani Queensland Centre for Mental Health Research Australia
REVIEW RETURNED	10-Jan-2019

GENERAL COMMENTS	Thank you very much for the opportunity to review this paper. Overall, I thought that the paper was well-written, and likely to be of much interest to the readers of the Journal. I have one major consideration: - Could you please provide a justification for not doing the whole population analysis. I understand that the gender analysis was done to compare boys and girls, but if you wanted to see there is a difference in different age groups, I would have thought that it would be worth comparing different age groups using both boys and girls, rather than separately. Some minor suggestions include: - Page 2, Abstract, under Results section - second to last line, CI[027-0.79] is missing "." - Page 2, Abstract, under Results, last line - how is team sport participation related to better mental health status? - Page 3, Article summary - I'm not sure if interpersonal violence is really a potential confounder between PA and mental health - Page 6 - I am not sure if I read this properly, but why was jogging/walking not defined as a sport? - Page 14 - sport participation and mental health, second paragraph - association was found ONLY with girls, not ESPECIALLY among girls, if I've read the results right.

	- Page 18 - our results according to... , second paragraph starting with "The current findings are..." (i) the use of the term "positive relationship" is confusing. If PA is high, mental health status is low, so it's not really a positive relationship in a mathematical sense (I get what you mean by this though). I think words like inverse or reciprocal have been used in the past - you may think of a better terminology there. (ii) In the same paragraph, you talk about longitudinal studies - the current study is a cross-sectional study, I don't think the findings can be compared to previous longitudinal studies. Page 20, paragraph starting "This study contributes to..." re: concept of "double health burden" - your study did not examine physical health parameters. This is a nice concept, and should be kept in discussion, but the wording around it should be clearer. i.e, your findings did not support this double burden. - Also, the word drop-out should be either one word or two (drop out). I prefer drop-out, but it doesn't really matter to me. Just be consistent throughout the manuscript please.

REVIEWER	Mark Beauchamp The University of British Columbia, Canada
REVIEW RETURNED	10-Jan-2019

GENERAL COMMENTS	In this study the authors sought to examine the relations between physical activity (PA) and sport participation in relation to various indices of mental health among adolescents. The study is based on data derived from a large cohort of Norwegian adolescents. The sample size is a notable strength, as is the broad question to examine whether participation in certain sports/activities is associated with improvements in markers of adolescent mental health. Balanced against these strengths I also have some concerns that particularly relate to various methodological aspects of the study. The Editor can decide on the extent to which these concerns are problematic. My observations are highlighted below: MAJOR CONCERNS 1. Other than the limitation highlighted by the authors that the study is cross-sectional in nature, all of the measures are also based on self-report questionnaire assessments. This increases the likelihood of common method bias/variance to a non-trivial extent especially when all of the (self-report) measures are collected at exactly the same point in time (see Podsakoff et al, 2012, Annual Review of Psychology). 2. I have a couple of concerns with the authors' assessment of PA. Specifically, a. Physical activity was reported to have been assessed using the WHO HBSC measure of physical activity. However, based on the information presented by the authors (page 6, para 1) there are a few substantive concerns. In particular, there appears to have been a notable departure from the way this measure was initially reported/validated. Specifically, in the original HBSC assessment (related to physical activity conducted outside of school) the item is prefaced by "Outside of school hours…. " and not "Not during the average school day…". In fact, I'm unclear what is being assessed

	by the wording “Not during the average school day…”. Does this refer to PA on the weekend, in holidays, on an atypical school day? Regardless, by virtue of the authors’ reworking of this stem has resulted in an item that fundamentally changes what is being assessed (it no longer appears to assess PA ‘outside of school hours’ as per the original HBSC procedures. b. In addition, in the HBSC, the assessment of PA outside of school appears to be assessed through a 2-item measure (see https://www.uib.no/sites/w3.uib.no/files/attachments/hbsc_external_study_protocol_2009-10.pdf). However, in this study, the authors appear to have culled one of those two items, resulting in a one-item measure with no known psychometric properties (and/or evidence for reliability/validity). 3. I also have some concerns with the authors’ operationalization of sport participation. Specifically, a. The most substantive concern corresponds to the authors’ categorization that ≥ 1 of sport per week is considered to reflect ‘active participation’. How was this cut point decided on? This seems rather arbitrary, and is not also not evident how this reflects ‘active’ sport participation. I reviewed what appeared to be the study’s protocol document that was included as an appendix, and no reference is made to this operationalization, and my concern here is that this cut-point reflects a post hoc decision (based on data mining/exploration) rather than an a priori decision. Indeed, from a substantive perspective, I’d have thought it reasonable to expect/hypothesize (based on prior research) that greater involvement in certain sports (e.g., team sports, as per Chekroud et al, 2018, Lancet Psychiatry) might be related to improved mental health symptomology. As such, why was the cut-point for sport participation just one or more times per week and not 2 or more, or 3 or more? No justification is/was provided, and participating in sport once per week would seem like a low bar for ‘active’ participation. b. On a more minor level, the authors reported 3 categories of sport participation based on 0, <1, and ≥ 1 times per week. Would <1 not be the same as 0? As such, one of these categories would seem to be redundant. c. In a point that aligns with 3a above, the groupings of sports/activities also seem arbitrary and in fact rather inconsistent (see page 6, para 2). For example, ‘running’ is included as an endurance sport, but later ‘jogging’ is excluded from the analyses entirely. Isn’t jogging the same thing as running? Also, how was the categorization ‘technical sports’ decided upon? Aren’t sports that appear elsewhere (e.g., soccer) technical, and also why would sports like skiing and snowboarding be considered technical sports and not ‘extreme sports’? Finally, why eliminate ‘walking’ from your analyses? This seems strange, especially given that walking represents one of the most common types of physical activities. Would it not have made sense to include walking and make comparisons between walking and other sports/activities? d. In sum, the operationalization of the predictor variables (for PA as per point 2a above, and for sport participation as per point 3a) in this study appear problematic. When these assessment procedures are considered along with the cross-sectional design, with all the assessments conducted through self-report, and the absence of a priori hypotheses (and the absence of information about hypotheses and the assessment/operationalizations within the protocol document/appendix) leads me to suspect that many of

	the reported findings may well be less than robust and likely spurious. 4. I have several concerns about the operationalization of the criterion/dependent measures in the study. Specifically, a. Why dichotomize measures of all of your mental health measures into high versus low rather than using the full range of score. It is widely recognized that categorization is unnecessary for statistical analysis and it has some serious drawbacks (see Altman, 2006, BMJ). b. The authors appear to have used abbreviated/short versions of some of the instruments. For example, rather than use the 10-item Rosenberg Self-esteem scale, a 4-item measure was used. Please provide evidence of the reliability/validity of measures derived from this shortened instrument based on data from adolescents (as per the sample from this study). The same goes for the other abbreviated measures as well (e.g., short version of SCL-5). c. General well-being appears to have been assessed using a one-item measure with no known psychometric properties. What is the source of this one-item instrument? Please provide the relevant citation. If general well-being was a focal measure, why not use a multi-item questionnaire, with known reliability/validity evidence? 5. The authors used an alpha level of $p < .05$ for all of their analyses. Given the extensive number of analyses conducted in the study, was there any consideration to account for (i.e., minimize) familywise error through some correction to the p values used? 6. Why were no a priori hypotheses used? In light of the fact that previous research has provided links between sport participation and mental health outcomes (e.g., Chekroud et al, 2018, Lancet Psychiatry, along with other studies cited in this paper such as Sabiston et al., 2016, JSEP) I'd have thought that a sufficient basis would have existed to map out some well-considered hypotheses. The study is very exploratory in nature. MINOR CONCERNS 1. In the introduction (page 4, para 2) the authors cite previous research linking PA and sports participation during adolescence to lifelong PA and well-being and use very causal language (see reference to 'influence'). However, the studies that the authors cite do not appear to have used experimental/causal designs and so the use of causal language is not well justified. 2. It is unclear if the measure of interpersonal violence (page 8, para 2) is based on a published instrument with known psychometric properties or was developed for this study. 3. The authors reported missing data (13% in girls, 15% in boys) for the PDS scores. Imputation is well justified if the data are missing at random (MAR) or missing completely at random (MCAR). It may be prudent to report the patterns of missing data (based on Little's Chi square test), and not just the amount of missing data, as a means of justifying the imputation procedures that were used.

REVIEWER	Elaine McMahon National Suicide Research Foundation, School of Public Health, University College Cork, Ireland
REVIEW RETURNED	30-Jan-2019

GENERAL COMMENTS	This manuscript makes a worthwhile contribution to the literature on physical activity and mental health in adolescents. Its strengths include a large sample size across a broad age range as part of the Young-HUNT study in Norway and a number of relevant mental health and activity/sport participation indicators. The analytical approach is appropriate and the findings are well presented and discussed. There are a few issues which I feel should be addressed however. The authors correctly say in the Introduction that mental ill-health commonly has onset in adolescence, with prevalence of mental ill-health increasing with age in adolescence. However, in Table 1 we see that in two of the three mental health measures examined in this study, prevalence decreases or stays broadly the same with increasing age (low self-esteem and low life satisfaction). Only psychological distress is more prevalent in the older group. I think this reflects the fact that self-esteem and life satisfaction, although useful indicators in and of themselves, are not proxy measures for aspects of mental health such as depressive symptoms or anxiety levels which increase throughout adolescence. This should be discussed as a limitation. The authors correctly emphasise the important links between PA and self-esteem in the Discussion, it should just be noted that depression and anxiety were not assessed. The percentage of participants engaging in sports, and in particular team sports, is very high by international standards. This could be discussed and reasons suggested, eg Norwegian school system prioritising the provision of sports, active community groups etc. As the authors mention, the benefits of high PA and of team sports are more striking in the older group. This requires further discussion as it is an interesting finding. An examination of the potential moderating role of pubertal stage in associations between PA and mental health measures would be very informative. Such a moderating role of pubertal stage may partly explain the stronger associations in the older group. The inclusion of the assessment of pubertal stage is a strength of the study but it hasn't been used to examine more closely some of the effects which appear to be age-specific. Introduction and Methods: the authors should mention that the recommendation of 60 minutes per day and the item used to assess PA both refer to moderate-to-vigorous activity. This terminology should be used to clarify the intensity of activity being examined. The term "mental health problem" is used throughout the manuscript. Some view this as an inappropriate phrase, preferring instead "mental ill-health". Table 1 includes some inaccurate labels, with Mean [SD] where n (%) should be for some variables. The terms "wellbeing" and "life satisfaction" are used interchangeably throughout the manuscript. I believe that the survey item used assessed life satisfaction which is distinct from

	wellbeing which generally reflects the absence of significant symptoms of mental ill-health. It would be worthwhile to note that the “High PA” group in fact are still falling short of the recommended levels of activity. They are more active than their peers but daily activity is still very rare. The Abstract should describe how PA was assessed.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1

Shuichi Suetani

Queensland Centre for Mental Health Research Australia

Thank you very much for the opportunity to review this paper. Overall, I thought that the paper was well-written, and likely to be of much interest to the readers of the Journal.

Responses to Reviewer #1's comments

1. I have one major consideration: Could you please provide a justification for not doing the whole population analysis. I understand that the gender analysis was done to compare boys and girls, but if you wanted to see there is a difference in different age groups, I would have thought that it would be worth comparing different age groups using both boys and girls, rather than separately.

Response: Thank you so much for the comments and your constructive feedback. We acknowledge the value of the suggestion to compare different age groups without stratifying on gender. However, given that PA levels and participation in organized sports have previously been shown to be lower among girls than boys (Baldursdottir et al., 2017; McMahon et al., 2017; Vilhjalmsson & Kristjansdottir 2003), we wanted to test these relationships according to gender and adolescent age groups. Further, regarding publication and dissemination of the findings, we believe it is most appropriate and meaningful to present the results for boys and girls separately. Further, as can be seen from the presented statistical analyses, the effect of PA was different for boys and girls, e.g. there was a statistically significant interaction which was an additional reason for us to present the stratified analyses. The rationale for the gender analysis has been emphasized in the Introduction, page 5:

“The aim of this study was…. Based on our knowledge and the literature, we anticipated that the associations between PA, sport participation and mental health measures would differ between boys and girls across adolescence”.

2. Some minor suggestions include:

- Page 2, Abstract, under Results section - second to last line, CI [027-0.79] is missing "."

Response: Thank you for noticing. "." has been inserted on page 2.

3. Page 2, Abstract, under Results, last line - how is team sport participation related to better mental health status?

Response: Thank you for pointing out this lack of clarity. Information has been added under Results in the abstract, page 2:

"… Physically active adolescents and participants in team sports had higher self-esteem and life satisfaction. …. Team sport participation was associated with reduced odds of psychological distress in senior high school girls".

4. Page 3, Article summary - I'm not sure if interpersonal violence is really a potential confounder between PA and mental health

Response: We chose to include interpersonal violence as a potential confounder between PA and mental health as these exposures may heavily impact on both PA behaviors and mental health. Long-term planning, coping and hope for the future may be hampered following exposure to violence, and exposure to violence is found to be strongly associated with the onset of psychological distress (McLaughlin et al., 2012). Adolescents exposed to interpersonal violence may also find it particularly difficult to maintain a healthy lifestyle, such as motivating, scheduling and completing PA after school (Stensland et al., 2105; Hughes et al., 2017). Justification for including interpersonal violence as a potential confounder has been elaborated on in the discussion, Strengths and limitations, page 18:

"Another strength is that we were able to adjust for a variety of possible confounders, as well as including exposures to interpersonal violence as these exposures may have an impact on both PA behaviours and mental health 31-33".

5. Page 6 - I am not sure if I read this properly, but why was jogging/walking not defined as a sport?

Response: "Jogging/walking" was not taken into consideration in the present study as these are not organized activities and are often components of other sports (warm-up routines). We therefore decided to examine participation in sports (individual and team sports) as exposures regardless of participants' status of jogging/walking. Information to clarify this issue has been added, page 6:

"Jogging/walking was not defined as an organized sport, and responses to this variable were not included as part of the sport participation exposure".

6. Page 14 - sport participation and mental health, second paragraph - association was found ONLY with girls, not ESPECIALLY among girls, if I've read the results right.

Response: Thank you, we realize that this is written inaccurately. The sentence has been re-written to clarify these results, page 15:

“Participation in team sports was associated with reduced odds of low life satisfaction, among all girls and among junior high school boys (Table 4)”.

7. Page 18 - our results according to... , second paragraph starting with "The current findings are..."

(i) the use of the term "positive relationship" is confusing. If PA is high, mental health status is low, so it's not really a positive relationship in a mathematical sense (I get what you mean by this though). I think words like inverse or reciprocal have been used in the past - you may think of a better terminology there.

(ii) In the same paragraph, you talk about longitudinal studies - the current study is a cross-sectional study, I don't think the findings can be compared to previous longitudinal studies.

Response: i) Thank you for this remark, we see that the term "positive relationship" may be confusing. The terminology has been changed, page 20:

“The current findings are in line with previous studies reporting associations between adolescents’ PA and mental health, including lower likelihood of depressive symptoms 1 11 13 41 42, as well as greater well-being 13 and higher self-esteem among those who are physically active 1 42”.

ii) We agree that our findings should not be compared to previous longitudinal studies. The sentence has been re-written, page 20:

“Longitudinal studies also indicate that PA may protect against the development of depression”.

8. Page 20, paragraph starting "This study contributes to..." re: concept of "double health burden" - your study did not examine physical health parameters. This is a nice concept, and should be kept in discussion, but the wording around it should be clearer. i.e, your findings did not support this double burden.

Response: We agree and have deleted the statement about our findings supporting this concept on page 22.

9. Also, the word drop-out should be either one word or two (drop out). I prefer drop-out, but it doesn't really matter to me. Just be consistent throughout the manuscript please.

Response: Thank you for the correction. We have gone through the manuscript and changed them all to drop-out.

Reviewer #2

Reviewer Name: Mark Beauchamp

Institution and Country: The University of British Columbia, Canada Please state any competing interests or state ‘None declared’: None declared

In this study the authors sought to examine the relations between physical activity (PA) and sport participation in relation to various indices of mental health among adolescents. The study is based on data derived from a large cohort of Norwegian adolescents. The sample size is a notable strength, as is the broad question to examine whether participation in certain sports/activities is associated with improvements in markers of adolescent mental health. Balanced against these strengths I also have some concerns that particularly relate to various methodological aspects of the study. The Editor can decide on the extent to which these concerns are problematic. My observations are highlighted below:

Responses to Reviewer #2's comments

1. MAJOR CONCERNS.Other than the limitation highlighted by the authors that the study is cross-sectional in nature, all of the measures are also based on self-report questionnaire assessments. This increases the likelihood of common method bias/variance to a non-trivial extent especially when all of the (self-report) measures are collected at exactly the same point in time (see Podsakoff et al, 2012, Annual Review of Psychology).

Response: Thank you for your careful review of our paper and for the helpful comments, corrections and suggestions.

We recognize the shortcomings of the cross-sectional design and the self-reported assessments. However, in order to access adolescents’ thoughts and feelings, as well as lifestyle choices (PA), surveys are one of the better options when gathering data from a large group. The availability of objective biological tests to measure mental health is limited, and it is difficult to monitor these outcomes at a population level. Obtaining information about maltreatment and psychological health is particularly challenging, and anonymized self-reporting may help adolescents to respond more honestly about these issues. A strength is that we use validated measures to assess psychological distress (Tambs et al., 1993; Strand et al., 2003), self-esteem (Isooma et al., 2013) and life satisfaction (Cheung & Lucas, 2014). Further, participants in the young-HUNT study had assistance if they did not understand the questions.

Regarding assessments of physical activity, it would be preferable to use objective methods, like accelerometers, to minimize the likelihood of information bias. However, a strength of the current study is the inclusion of questions about different types of sport participation, in addition to the HBSC PA assessment, which is lacking in other population-based studies. This provided us with the opportunity to elaborate and deepen our understanding of how PA and different types of sport participation can be related to various dimensions of mental health.

2. I have a couple of concerns with the authors' assessment of PA. Specifically,

a. Physical activity was reported to have been assessed using the WHO HBSC measure of physical activity. However, based on the information presented by the authors (page 6, para 1) there are a few substantive concerns. In particular, there appears to have been a notable departure from the way this measure was initially reported/validated. Specifically, in the original HBSC assessment (related to physical activity conducted outside of school) the item is prefaced by "Outside of school hours…. " and not "Not during the average school day…". In fact, I'm unclear what is being assessed by the wording "Not during the average school day…". Does this refer to PA on the weekend, in holidays, on an atypical school day? Regardless, by virtue of the authors' reworking of this stem has resulted in an item that fundamentally changes what is being assessed (it no longer appears to assess PA 'outside of school hours' as per the original HBSC procedures.

Response: Thank you for clarifying this very important issue. Physical activity has been assessed using the WHO HBSC measure of physical activity, as per the original HBSC procedures (translated to Norwegian). The frequency question has not been described precisely enough in the method section. We are sorry for the inaccuracy that led to this confusion. Corrections have been made on page 6:

"Outside school hours: How often do you usually exercise in your free time so much that you get out of breath or sweat?".

2 b. In addition, in the HBSC, the assessment of PA outside of school appears to be assessed through a 2-item measure (see https://www.uib.no/sites/w3.uib.no/files/attachments/hbsc_external_study_protocol_2009-10.pdf). However, in this study, the authors appear to have culled one of those two items, resulting in a one-item measure with no known psychometric properties (and/or evidence for reliability/validity).

Response: Thank you for the reminder to elaborate on the evidence of reliability and validity for

the HBSC assessment of PA. The questions from the HBSC questionnaire have been found to hold acceptable reliability and validity in adolescent samples (Rangul et al., 2008; Booth et al., 2001).

More specifically, Rangul et al. (2008) found that the HBSC questionnaire had substantial reliability (interclass reliability 0.71) and was an acceptable instrument for measuring cardiorespiratory fitness (VO2peak). The frequency question had a higher correlation with VO2peak than the duration question. A possible explanation suggested for the differences in the dimensions (frequency and

duration) is that the frequency question (days/week) is easier to estimate more precisely (and easier to remember) and is a rougher estimate than hours per week. Single-item measures to assess adolescents' PA level have been used in other large surveys (Currie et al., 2012; Inchley et al., 2016), and a similar single-item PA measure has been found to have comparable validity and reliability to the comprehensive Oxford Physical Activity Questionnaire (OPAQ) and accelerometer output (Scott et al., 2015). However, to highlight the shortcomings, we have made the following changes:

- The article from Booth et al. (2001) examining the reliability and validity of the HBSC PA assessment has been added as a reference, page 6.

- Limitations regarding the use of the 1-item HBSC assessment of PA have been included in the section Strengths and limitations, page 18-19: "We have used a single item measure to assess PA,… However, the WHO HBSC question on frequency of PA used in this study has been found to hold acceptable reliability and validity in adolescent samples 19 20".

- We have clarified the intensity of the item used to assess PA in Methods, page 6:

"The level of intensity during exercise where you breathe heavily and/or sweat refers to moderate to vigorous activity".

3.I also have some concerns with the authors' operationalization of sport participation. Specifically,

a. The most substantive concern corresponds to the authors' categorization that ≥ 1 of sport per week is considered to reflect 'active participation'. How was this cut point decided on? This seems rather arbitrary, and is not also not evident how this reflects 'active' sport participation. I reviewed what appeared to be the study's protocol document that was included as an appendix, and no reference is made to this operationalization, and my concern here is that this cut-point reflects a post hoc decision (based on data mining/exploration) rather than an a priori decision. Indeed, from a substantive perspective, I'd have thought it reasonable to expect/hypothesize (based on prior research) that greater involvement in certain sports (e.g., team sports, as per Chekroud et al, 2018, Lancet Psychiatry) might be related to improved mental health symptomology. As such, why was the cut-point for sport participation just one or more times per week and not 2 or more, or 3 or more? No justification is/was provided, and participating in sport once per week would seem like a low bar for 'active' participation.

Response: We recognize the shortcomings of these exposure variables and agree that the dichotomization provides a "rough" measure of sports participation. The choice of the cut-off point at ≥ 1 day/week for active sport participation was an a priori decision. This cut-off has been used in other studies examining associations between sport participation and mental health (Eime et al., 2013), and provides more information compared to using very crude yes/no categories.

Due to data limitations we had insufficient statistical power to model the sport exposures in more detail. In the responses for sports participation in the Young-HUNT3 questionnaire, "several times a week" was the highest frequency option. We therefore did not have the opportunity to distinguish between those performing sports a few days a week from those doing sports every day. Since very high degree of sports participation is found to be a risk factor for poor well-being (Merglen et al., 2014) and is associated with overtraining and increased likelihood of depression (Winsley & Matos, 2011),

we would like to argue that the category ‘several times a week’ applies to those training two-three times/week as well as those who train too often. We therefore considered the cut-off ≥ 1 time/week as the most appropriate in this study.

- Limitations regarding the categorization of ≥ 1 time/week as “active sport participation” have been included in Strengths and limitations, page 18-19:

“We have used a single item measure to assess PA, and the variable used to describe sport participation exposure provides a crude measure of frequency of sport participation”.

3 b. On a more minor level, the authors reported 3 categories of sport participation based on 0, <1, and ≥ 1 times per week. Would <1 not be the same as 0? As such, one of these categories would seem to be redundant.

Response: Thank you for identifying this area of potential ambiguity. The term “<1” refers to “less than once a week”, on average over the last 12 months. To avoid misunderstanding we will write the responses using text (not numbers and symbols), according to the original format in the questionnaire. The sentence has been changed on page 6:

“Four alternatives were given for describing the frequency of participation in each of the sport categories: never, less than once a week, once a week, several times a week”.

3 c. In a point that aligns with 3a above, the groupings of sports/activities also seem arbitrary and in fact rather inconsistent (see page 6, para 2). For example, ‘running’ is included as an endurance sport, but later ‘jogging’ is excluded from the analyses entirely. Isn’t jogging the same thing as running? Also, how was the categorization ‘technical sports’ decided upon? Aren’t sports that appear elsewhere (e.g., soccer) technical, and also why would sports like skiing and snowboarding be considered technical sports and not ‘extreme sports’? Finally, why eliminate ‘walking’ from your analyses? This seems strange, especially given that walking represents one of the most common types of physical activities. Would it not have made sense to include walking and make comparisons between walking and other sports/activities?

Response: As you point out, there are weaknesses due to the classification of this exposure variable. The questions for assessing various types of sport participation were developed for the HUNT-study, and our study group was not involved in decisions concerning the content and design of the questionnaire. Different types of sports were categorized into nine categories, with several alternatives within each category. Therefore, we did not have the opportunity to examine each individual sport as an exposure.

Jogging has been considered as an activity that differs from running (endurance sport), as it is not an organized activity. Also, jogging is often a part of warm-up routines in other sports. Jogging and walking were merged into one category in the questionnaire, and we were therefore unable to

distinguish between them. We agree that the sports activities could be categorised differently, and as you point out, snowboarding may be considered an extreme sport in some cases. However, football, volleyball and handball have been defined as "team sports" in previous studies (Evans et al., 2016; Eime et al., 2013; McMahon et al., 2016; Slutzky & Simpkins 2009).

3 d. In sum, the operationalization of the predictor variables (for PA as per point 2a above, and for sport participation as per point 3a) in this study appear problematic. When these assessment procedures are considered along with the cross-sectional design, with all the assessments conducted through self-report, and the absence of a priori hypotheses (and the absence of information about hypotheses and the assessment/operationalizations within the protocol document/appendix) leads me to suspect that many of the reported findings may well be less than robust and likely spurious.

Response: We want to make it clear that there has been no departure from the original HBSC assessment of PA in this study (ref. response 2). Moreover, we would like to emphasize that self-report as the assessment method of choice has some advantages and may be suitable for large-scale population surveys (ref. response 1). This work has been carried out based on predefined hypotheses, and we agree that the hypotheses should have been presented in the manuscript. We have now specified the a priori hypotheses in the Introduction, page 5:

“The aim of this study was… We hypothesised that a high level of PA and participation in sports would be associated with lower levels of psychological distress, higher self-esteem and greater life satisfaction, particularly among high school students and participants in team sports”.

4. I have several concerns about the operationalization of the criterion/dependent measures in the study. Specifically,

a. Why dichotomize measures of all of your mental health measures into high versus low rather than using the full range of score. It is widely recognized that categorization is unnecessary for statistical analysis and it has some serious drawbacks (see Altman, 2006, BMJ).

Response: We agree that dichotomizing these measures has drawbacks and might lead to a loss of information and variability in our data. The cut-off point at 2.0 for SCL-5 to distinguish between those with and without psychological distress, and the midpoint of the scale for the RSES to separate low and high self-esteem, have both been shown to be clinically relevant cut-points, with satisfactory sensitivity, specificity, positive predictive value, and negative predictive value (Strand et al 2003, Isomaa et al 2013). Further, when we investigated the structure of our data and tried to fit linear models to keep the outcomes as continuous variables, the fit was very poor. Moreover, the outcomes are categorical ordinal variables, thus a major assumption of linear regression would be violated. On the other hand, logistic regression fitted very well to our data. In addition, we would like to argue that odds ratios might be more intuitive to interpret compared to a unit change which would be estimated using linear regression.

We have emphasized limitations in Strengths and limitations, page 19:

"…Furthermore, dichotomization of the mental health outcomes makes them prone to misclassification. However, the cut-off values to distinguish those with a high vs low degree of psychological distress (SCL-5) and low self-esteem (RSES) have both been shown to be clinically relevant cut-points 21 24".

4 b. The authors appear to have used abbreviated/short versions of some of the instruments. For example, rather than use the 10-item Rosenberg Self-esteem scale, a 4-item measure was used. Please provide evidence of the reliability/validity of measures derived from this shortened instrument based on data from adolescents (as per the sample from this study). The same goes for the other abbreviated measures as well (e.g., short version of SCL-5).

Response: Thank you for highlighting the importance of discussing the use of shortened instruments. In the HUNT-study abbreviated instruments have been used for most mental health outcomes due to limited space in the questionnaire for use of the original instruments, as well as time concerns.

The Norwegian Mother and Child Cohort Study (MoBa), one of the world's largest health surveys, found that the four-item version of RSES correlates at 0.95 with the full scale and explains 0.90% of the full-scale variance. Cronbach alpha for the four-item version was 0.80. They argue that the precision of the four-item version remains sufficient for epidemiological purposes (Tambs & Røysamb, 2014). Symptoms of anxiety and depression were measured using a five-item shortened version of the Hopkins Symptom Checklist (SCL-25), which has shown high correlation with SCL-10 and SCL-25 (Tambs & Moum, 1993, Strand et al., 2003), as well as good internal consistency (Cronbach's alpha 0.87) (Strand et al., 2003). Specificity at 82% and sensitivity at 96% was found using the cut-off point of 2.0 to determine if adolescents had symptoms of anxiety and depression (Strand et al., 2003). Evidence of the reliability and validity of these shortened instruments has been added in Methods (Outcome variables), page 7:

"The five-item version (SCL-5) has shown high correlation with the 25-item SCL-25 (r = 0.92) 22 and good internal consistency (Cronbach's alpha 0.87) 21"….

"The four-item version of the RSES is found to correlate at 0.95 with the full scale and to explain 90% of the full-scale variance, and has good internal consistency (Cronbach's alpha 0.80) 25".

Further discussion has been added in the section Strengths and limitations, page 19:

"Measures of psychological distress (SCL-5) and self-esteem (RSES) were shortened versions of the original instruments, however, the measurement precision of these versions is found to be high and sufficient for use in population-based studies 21 22 25".

4 c. General well-being appears to have been assessed using a one-item measure with no known psychometric properties. What is the source of this one-item instrument? Please provide the relevant citation. If general well-being was a focal measure, why not use a multi-item questionnaire, with known reliability/validity evidence?

Response: Life satisfaction is a component of subjective well-being reflecting the cognitive evaluation of whether one is happy with one's life. A single-item life satisfaction measure has been used in many studies with large samples and has demonstrated a substantial degree of criterion validity compared to the multiple-item Satisfaction with Life Scale (SWLS) (Cheung & Lucas, 2014). The single-item life satisfaction measure is found to perform almost as well as the SWLS in adolescent samples (Jovanovic, 2016). Citations have been added in Methods (Outcome variables), page 8:

"A single-item life satisfaction measure is shown to perform almost as well as the multiple-item Satisfaction with Life Scale (SWLS) 26 27".

5. The authors used an alpha level of $p < .05$ for all of their analyses. Given the extensive number of analyses conducted in the study, was there any consideration to account for (i.e., minimize) familywise error through some correction to the p values used?

Response: Multiple testing is a challenge especially when working with large, population-based data. However, as our study was considered exploratory, corrections for multiple testing were not performed. In our interpretation we tried to emphasize the actual point estimates and their 95% confidence intervals, and the clinical relevance of the revealed association instead of just presenting p-values which we agree are not very informative. As you point out, our results should be interpreted with caution due to the large number of hypotheses being tested. We have therefore included these limitations in the discussion, Strengths and limitations, page 19:

"The results of this study should be interpreted with caution due to multiple testing, and replication of the results is warranted".

6. Why were no a priori hypotheses used? In light of the fact that previous research has provided links between sport participation and mental health outcomes (e.g., Chekroud et al, 2018, Lancet Psychiatry, along with other studies cited in this paper such as Sabiston et al., 2016, JSEP) I'd have thought that a sufficient basis would have existed to map out some well-considered hypotheses. The study is very exploratory in nature.

Response: As outlined in Response 3 d), a priori hypotheses were used, and we have included the hypotheses in the Introduction, page 5.

MINOR CONCERNS

1. In the introduction (page 4, para 2) the authors cite previous research linking PA and sports participation during adolescence to lifelong PA and well-being and use very causal language (see reference to 'influence'). However, the studies that the authors cite do not appear to have used experimental/causal designs and so the use of causal language is not well justified.

Response: Thank you for this remark. We agree that the causal language may not be appropriate, and have modified the sentence, page 4:

“Engaging in PA and sports during adolescence is associated with development of lifelong PA 6-8 and psychological well-being 1 9”.

2. It is unclear if the measure of interpersonal violence (page 8, para 2) is based on a published instrument with known psychometric properties or was developed for this study.

Response: Thank you for reminding us to include this information. The measures of interpersonal violence used in the brief Young-HUNT3 lifetime trauma screen have been derived from The University of California at Los Angeles Post-traumatic Stress Disorder Reaction Index (UCLA PTSD Reaction Index) and adapted to the Norwegian context in collaboration with the authors. Information about the measures of interpersonal violence has been added, page 8:

“Exposure to interpersonal violence was assessed with questions derived from The University of California at Los Angeles Post-traumatic Stress Disorder Reaction Index (UCLA PTSD Reaction Index) 29”.

3. The authors reported missing data (13% in girls, 15% in boys) for the PDS scores. Imputation is well justified if the data are missing at random (MAR) or missing completely at random (MCAR). It may be prudent to report the patterns of missing data (based on Little’s Chi square test), and not just the amount of missing data, as a means of justifying the imputation procedures that were used.

Response: Our data did not reveal any patterns concerning missingness and we assumed missing at random. Any model not based on the imputation method would only introduce more bias into our data, so we had to use model-based imputation. Linear regression models stratified by gender were fitted with age and BMI. The details are presented in the Methods (Statistical analyses), page 9.

Reviewer #3

Elaine McMahon

Institution and Country: National Suicide Research Foundation, School of Public Health, University College Cork, Ireland

This manuscript makes a worthwhile contribution to the literature on physical activity and mental health in adolescents. Its strengths include a large sample size across a broad age range as part of the Young-HUNT study in Norway and a number of relevant mental health and activity/sport participation indicators. The analytical approach is appropriate and the findings are well presented and discussed. There are a few issues which I feel should be addressed however.

Responses to Reviewer #3's comments

1. The authors correctly say in the Introduction that mental ill-health commonly has onset in adolescence, with prevalence of mental ill-health increasing with age in adolescence. However, in Table 1 we see that in two of the three mental health measures examined in this study, prevalence decreases or stays broadly the same with increasing age (low self-esteem and low life satisfaction). Only psychological distress is more prevalent in the older group. I think this reflects the fact that self-esteem and life satisfaction, although useful indicators in and of themselves, are not proxy measures for aspects of mental health such as depressive symptoms or anxiety levels which increase throughout adolescence. This should be discussed as a limitation. The authors correctly emphasise the important links between PA and self-esteem in the Discussion, it should just be noted that depression and anxiety were not assessed.

Response: Thank you for highlighting the importance of discussing this issue. The three outcome measures in this study have been used to investigate various aspects of mental health. Discussion have been added in Strengths and limitations, page 19:

“In contrast to psychological distress, low self-esteem and low life satisfaction were not more prevalent in the older age group, reflecting the measurement of different phenomena. Psychological distress is found to function as a proxy measure of anxiety and depression 21 22, while self-esteem and life satisfaction are more closely related to subjective well-being 26 35”.

- Moreover, we realize that the description of the aspects of mental health examined should be described more precisely in references to previous studies. Changes have been made accordingly, page 4: “Prevalence rates of psychological distress, such as anxiety and depression, increase with age, especially from the mid-teens (14-16 years) 11 17”….“Currently the evidence indicates that PA may have a positive impact on anxiety, depression and self-esteem among adolescents, although our knowledge is limited1”. Page 20: “PA may also be a helpful intervention for adolescents struggling with depressive symptoms”.

Furthermore, in the summary of results from the current study, the three different outcome measures have been described in more detail, page 18:

“Our results showed that higher levels of PA were favourably associated with self-esteem and life satisfaction throughout adolescence, as well as with reduced likelihood of psychological distress in senior high school students”.

2. The percentage of participants engaging in sports, and in particular team sports, is very high by international standards. This could be discussed, and reasons suggested, eg Norwegian school system prioritising the provision of sports, active community groups etc.

Response: Thank you for pointing this out. We have added information and suggested reasons for the high participation rate among Norwegian adolescents in the discussion, page 19:

“Norwegian society is rooted in egalitarian ideals, with "Sport for All" as a high priority and policy aim 36; this may be part of the reason why sport participation found in this study is high by international standards”.

3. As the authors mention, the benefits of high PA and of team sports are more striking in the older group. This requires further discussion as it is an interesting finding. An examination of the potential moderating role of pubertal stage in associations between PA and mental health measures would be very informative. Such a moderating role of pubertal stage may partly explain the stronger associations in the older group. The inclusion of the assessment of pubertal stage is a strength of the study but it hasn’t been used to examine more closely some of the effects which appear to be age-specific.

Response: Thank you for drawing our attention to the potential impact of development in the relationship between PA and psychological distress, self-esteem and life satisfaction. As you point out, biological maturation may play an important role, and we therefore adjusted for pubertal stage (PDS scale). Unfortunately, we were unable to assess moderation by pubertal stage in this study as participants in the Young-HUNT3 study were 13-19 years old; mainly pubertal or post-pubertal, with hardly anyone in the prepubertal group (particularly so for girls, with a mean age at menarche of 12.5 years). We did, however, have the chance to stratify junior versus senior high school students, as a proxy for adolescent development. As early adolescent stage versus mid-late adolescence is characterized by a major shift in psychosocial development tasks, where peer support and peer interaction play an increasingly important role (Christie & Viner, 2005), we chose to present stratified data in this study. Discussion about the differences in finding across age groups has been added, page 20-21:

"Explanations for why PA and sport participation may be of greater importance in reducing psychological distress among older adolescents could relate to how peer support and peer interaction play an increasingly important role during adolescence 50. Thus, social and physical activities with peers may be particularly beneficial for older adolescents, helping to distract them from depressive thoughts and to reduce the sense of isolation”.

4. Introduction and Methods: the authors should mention that the recommendation of 60 minutes per day and the item used to assess PA both refer to moderate-to-vigorous activity. This terminology should be used to clarify the intensity of activity being examined.

The term “mental health problem” is used throughout the manuscript. Some view this as an inappropriate phrase, preferring instead “mental ill-health”.

Response: Thank you for the suggestions. We have clarified the intensity of the PA assessment:

- Introduction, page 4: “Although few adolescents are satisfying the recommended 60 minutes of moderate to vigorous PA per day worldwide 12,13”.

- Methods, page 6: “The level of intensity during exercise where you breathe heavily and/or sweat refers to moderate to vigorous activity”.

Referring to response 1) we have changed the term “mental health problem” in sentences where a more precise description of the outcome is appropriate.

5. Table 1 includes some inaccurate labels, with Mean [SD] where n (%) should be for some variables.

Response: Thank you for this remark. Changes has been made accordingly in Table 1, page 12.

6. The terms “wellbeing” and “life satisfaction” are used interchangeably throughout the manuscript. I believe that the survey item used assessed life satisfaction which is distinct from wellbeing which generally reflects the absence of significant symptoms of mental ill-health.

Response: We apologize for this lack of clarity. Life satisfaction is a component of subjective well-being reflecting the cognitive evaluation of whether one is happy with one’s life. To avoid ambiguity, we have used the term “life satisfaction” throughout the revised manuscript. Changes are made on page 2 (abstract), page 8 (outcome variables) and page 12 (Table 1).

7. It would be worthwhile to note that the “High PA” group in fact are still falling short of the recommended levels of activity. They are more active than their peers, but daily activity is still very rare.

Response: We agree that this should be noted. However, activities within school hours were not covered by this question. In Norway, a minimum of 2 school hours per week is allocated for physical education for adolescents. We have added information in Strengths and limitations, page 19:

“It should be noted that a proportion of the “High PA” group in this study may not fulfil the recommended levels of daily PA according to the WHO guidelines 34, as this group includes all those who played sports or exercised ≥ 4 days/week”.

8. The Abstract should describe how PA was assessed.

Response: Thank you for this remark. A description of how PA was assessed has been included in the Abstract, page 2: “...according to self-reported PA level and type of sport participation…”

VERSION 2 – REVIEW

REVIEWER	Shuichi Suetani QCMHR
REVIEW RETURNED	07-Mar-2019

GENERAL COMMENTS	Authors have adequately addressed the issues previously discussed in my previous review.

REVIEWER	Mark Beauchamp University of British Columbia, Canada
REVIEW RETURNED	11-Mar-2019

GENERAL COMMENTS	I thought that the authors were very responsive to my previous comments/observations, and made a concerted effort to address those concerns. The manuscript is certainly improved. I have a couple of remaining questions/suggestions. These are as follows: 1. In my previous review I highlighted concerns with the absence of reported a priori hypotheses. In response the authors included the following hypothesis “We hypothesised that a high level of PA and participation in sports would be associated with lower levels of psychological distress, higher self-esteem and greater life satisfaction, particularly among high school students and participants in team sports”. It might be useful to provide an explanation/justification in the introduction for why stronger effects were expected for high school students and team sport participation. Currently the rationale provided is “based on our knowledge and the literature” – I suggest articulating that rationale in the introduction and cite the appropriate supporting literature. 2. It is still not evident why the authors did not include their measures of walking/jogging in their analyses. I recognize that a measure that combines walking with jogging is less than ideal, but some of the other physical activity and sport categorizations (and combinations subsumed within these) used in the HUNT study also appear less than ideal. On a substantive basis, walking represents one of the most popular means of physical activity participation, and so it would seem a little remiss to eliminate that measure from this study. Indeed, why not include that walking/jogging measure, while also recognizing the limitations of such a blended measure (especially as the authors state that this is/was an exploratory study)?

VERSION 2 – AUTHOR RESPONSE

Responses to Reviewer #2's comments

1. In my previous review I highlighted concerns with the absence of reported a priori hypotheses. In response the authors included the following hypothesis “We hypothesised that a high level of PA and participation in sports would be associated with lower levels of psychological distress, higher self-esteem and greater life satisfaction, particularly among high school students and participants in team sports”. It might be useful to provide an explanation/justification in the introduction for why stronger

effects were expected for high school students and team sport participation. Currently the rationale provided is “based on our knowledge and the literature” – I suggest articulating that rationale in the introduction and cite the appropriate supporting literature.

Response: Thank you for the constructive feedback and for the suggestions to improve the paper.

We expected stronger effects for high school students and team sport participation as peer support and peer interaction become more salient during adolescence (Brown & Larsoon 2009), and the transition from junior high to senior high school creates changes in social contexts and norms that may enhance the importance of peer relationships. During adolescence, at the same time as PA-levels and sport participation begin to decline (Dumith et al. 2011, Dalene et al. 2018, Sagatun et al. 2008, Baldurdottir et al. 2017), high school students may begin to experiment with alcohol, smoking or other risk behaviors, which can increase the burden on their health. Maintaining structure and social meeting places through participation in sports may therefore be particularly beneficial for adolescents in this age group.

We have provided an explanation for our hypothesis in the Introduction, page 5:

As early adolescent stage versus mid-late adolescence is characterized by a major shift in psychosocial development tasks, where peer relationships become more salient (Christie & Viner 2005, Brown & Larsson 2009), the social benefits of sports participation may be of greater importance with increasing age through adolescence. We therefore hypothesised that a high level of PA and participation in sports would be associated with lower levels of psychological distress, higher self-esteem, and greater life satisfaction, particularly among high school students and participants in team sports.

2. It is still not evident why the authors did not include their measures of walking/jogging in their analyses. I recognize that a measure that combines walking with jogging is less than ideal, but some of the other physical activity and sport categorizations (and combinations subsumed within these) used in the HUNT study also appear less than ideal. On a substantive basis, walking represents one of the most popular means of physical activity participation, and so it would seem a little remiss to eliminate that measure from this study. Indeed, why not include that walking/jogging measure, while also recognizing the limitations of such a blended measure (especially as the authors state that this is/was an exploratory study)?

Response: Thank you for pointing out this lack of clarity. Measures of walking/jogging were not excluded, but they overlap with the three exposure categories (no/infrequent sport or PA participation, individual sport participation and team sport participation).

We agree that the potential benefits of jogging/walking on mental health in adolescents would be an interesting topic to further investigate. However, as you point out, walking and jogging are popular means of physical activity participation. In our sample we found that 64% of the adolescents reported jogging/walking at least once a week, most often in combination with other sport activities. Jogging/walking at least once a week were reported among 61% of those participating in individual sports and 71% of those participating in team sports, and among 33% of those with no/infrequent sport participation or low PA level.

We want to apologize for this lack of information, and we realize the need to elaborate on the categorization of this variable. To avoid misunderstanding, the description of the sport participation exposure has been re-written in the Method section, page 7:

“Responses to “jogging/walking” were not defined as separate sport activities/participation, as they may also be performed in non-sport contexts. The activity “jogging/walking” was, however, included in all exposure categories; “jogging/walking” at least once a week was reported among 61% of those participating in individual sports, among 71% of those participating in team sports, and among 33% of those with no/infrequent sport participation or low PA level”.

In light of your comment we did considered creating a new exposure variable that additionally contains a category including those who report jogging/walking but no sport participation, as an attempt to assess the isolated effect of this type of activity. However, due to limited numbers in the category “only jogging/walking” (N = 257) we had insufficient statistical power to model the sport exposures into these four different categories.