**Author’s response to reviews**

**Title:** Trends in childhood and adolescent internalizing symptoms: results from Swedish population based twin cohorts

**Authors:**

Karolina Sorman (Karolina.Sorman@ki.se)

Natalie Durbeej (Natalie.Durbeej@ki.se)

Eva Norén-Selinus (Eva.Noren@ki.se)

Sebastian Lundström (sebastian.lundstrom@gnc.gu.se)

Paul Lichtenstein (Paul.Lichtenstein@ki.se)

Clara Hellner (Clara.Gumpert@ki.se)

Linda Halldner (Linda.Halldner@ki.se)

**Version:** 2  **Date:** 20 Jun 2019

**Author’s response to reviews:**

Author response to reviews of Manuscript PSYO-D-19-00002R1

Trends in childhood and adolescent internalizing symptoms: results from Swedish population based twin cohorts

Karolina Sorman, Ph.D.; Natalie Durbeej; Eva Norén-Selinus; Sebastian Lundström; Paul Lichtenstein; Clara Hellner; Linda Halldner

re-submitted to BMC Psychology

June 20th, 2019

Dear Dr. Harris

Thank you very much for taking the time to review our manuscript for publication in BMC Psychology, and for the opportunity to re-submit the manuscript.
We have strived towards addressing all points raised by the Reviewers by making appropriate changes in the revised manuscript and by providing responses in this cover letter. In the following letter, we address all comments and concerns point-by-point and declare how we revised the manuscript accordingly.

Reviewer #1

Tracey Wade (Reviewer 1): The authors have responded in some detail to all the reviewers' queries. Points for further consideration are listed below.

1. Along with Reviewer 4, I still consider the rationale for the manuscript to be weak.

RESPONSE: Thank you for pointing this out. The rationale was to investigate trends in internalizing symptoms, both anxiety and depression, in successive birth cohorts of youth. Exploring temporal trends of these symptoms is an important research endeavor to investigate whether the increase in anxiety and depression diagnoses and treatment during the past decades, corresponds to actual increasing levels of internalizing symptoms. This has been clarified under Methods page 4 (the headline “Aims of the study” has been revised to “Study rationale”). Along the same lines, in the first paragraph of the Discussion it is also stated:

“Epidemiological data from high-income countries has demonstrated increases in clinical parameters (i.e., diagnosis and treatment) of depression and anxiety disorders during the past decades (1). The current study examines whether such clinical trends correspond to changing levels of internalizing symptoms in population-based samples, using cross-cohort comparisons.”

**********************************************************************

2. The last sentence of the abstract is not particularly helpful - in essence "keep assessing everybody throughout early life" - this recommendation is unlikely to be acted on and I wonder if the authors can present a better take home message based on the evidence generated by their research.

RESPONSE: Thank you for this comment. We agree that the last sentence in the abstract had an unfortunate wording. We have now altered the conclusion in the abstract.

**********************************************************************

3. Just because Cronbach alpha is widely used does not make it acceptable - see Peter's paper in The European Health Psychologist on "The Alpha and the Omega of Scale Reliability and Validity" for advice on computing omega, or Daniel McNeish's paper in Psychological Methods which offer excel sheets to calculate the H coefficient.
RESPONSE: We have now calculated H coefficients as a measure of scale reliability and report these coefficients instead of Chronbach’s alpha values. We also refer to the article by Daniel McNeish (2018) which has been added as a reference. See sections Assessment at age 9 and Assessment at age 15.

**********************************************************************

Nora Trompeter (Reviewer 2): Thank you to the authors for their careful answers and revisions based on previous recommendations. All my queries have been addressed and I have no further comments.

RESPONSE: Thank you.

**********************************************************************

David Chinn (Reviewer 3): I am satisfied by the changes made in response to the comments by my fellow reviewers and I. I recommend that the article is accepted.

RESPONSE: Thank you.

**********************************************************************

Bryan Rodgers, Ph.D. (Reviewer 4): I will restrict my comments to areas that I raised in my original review and to the responses made by the authors in their Response to Reviewers section. I focus entirely on the original points 2 and 4, as the remaining issues are of little relevance unless these two important points are addressed appropriately.

2) What is meant by the term "internalizing symptoms" is fundamental to the present manuscript and the (revised) paper does not convey a clear picture. More specifically, the extent to which parent reports are valid indicators of "depression and anxiety" in children is an issue that has clouded inquiry in this field for decades. With the benefit of so much past experience, one would hope that contemporary research does not fall into the same traps as in the past. Analysis of trends over time in internalizing symptoms requires some thought about what this concept embraces (and what it does not include). For the present investigation, only parent reports are available for children aged 9 years whereas parent and self-reports are available for age 15 years, meaning that the choice of informant is crucial to the interpretation of findings. The authors have failed to grasp this fundamental point and appear very selective in the literature they choose to cite and not to cite. For example, the defence of the citation of Hariz et al. (2013) as an indication of the validity of SCARED parent reports in indicating children with clinical disorders ignores what is written in that paper, specifically that the high correlation between children self-reports and parent ratings found for the Arabic version was not found previously for the English version of the scared (Birmaher et al., 1999). I agree with the present authors that the literature on this topic is complicated but their own report glosses over such complications and, as such, presents what is a potentially misleading picture.
RESPONSE:

Our aim with this study was to examine if symptoms of anxiety and depression are in fact increasing amongst children and adolescents in the way media reports regularly claim. Data presented in popular media is often based on surveys using non-validated instruments. Furthermore, general mental health problems are mixed up with diagnosed mental disorders. In our study, we had access to data from three different scales (sMFQ, SCARED, and SDQ) across consecutive birth cohorts. We believe that these scales have been validated enough to claim that they measure symptoms comprised by the term internalizing symptoms.

We agree, however, on the importance of pointing out the conceptual debate/uncertainty around the concept “internalizing symptoms”. In this study, we did not aim to propose a conceptualization of the term “internalizing symptoms”. To avoid confusion, we have now in the background section defined in which way we use the term in our paper.

Regarding the issue of informants, we agree that the issue of validity of parent information is not fully solved, which we state in the discussion as a limitation of our study. However, there is some evidence that parent-information regarding symptoms of anxiety and depression in youth could be as valid as self-rated symptoms (see Table 4 in Jensen PS, Rubio-Stipec M, Canino G, Bird HR, Dulcan MK, Schwab-Stone ME, et al. Parent and Child Contributions to Diagnosis of Mental Disorder: Are Both Informants Always Necessary? Journal of the American Academy of Child & Adolescent Psychiatry. 1999;38(12):1569-79), and may even be more informative than self-rated information for children specifically (see Table 1 in Smith SR. Making Sense of Multiple Informants in Child and Adolescent Psychopathology: A Guide for Clinicians. Journal of Psychoeducational Assessment. 2007;25(2):139-49.)

We agree that scores on the measures used in this study are not directly linked to clinical symptoms (i.e., scores might be influenced by multiple factors including cultural factors.).

I am concerned that the authors state in their response "The aim of the study was to investigate any secular changes between cohorts, and therefore further validation of the scales used, were outside the scope of this paper." Given that the manuscript already uses citations to support the validity of the scales used in the study, what particular reason is there for NOT reporting the correlations between parent reports and self-reports at age 15?

RESPONSE: Thank you for this remark. We have now calculated the Pearson correlation between the SDQ self-report version and the SDQ parent-report version. We report this correlation (Pearson r = .41) in the method section. See section Assessment at age 15.
I have also realised that the Sharp et al. (2006) reference, now numbered 25, is inappropriate in that this study only collected self-report data using the sMFQ but it has been cited in support of the validity of "parent-rated scores". Apologies for not pointing this out in my original review.

RESPONSE: We thank the reviewer for noting this mistake. We have now corrected this by referring to Thapar and McGuffin 1998 instead as reference number 27 (previously number 25). This study contains data on parent rated scores (the results demonstrating that the scores were significantly higher for those in the depressed group compared to non-depressed group).

4) The authors reject my suggestion to report interaction terms between time and sex on the basis that their separate analyses of males and females were "sensitivity analyses". This is simply incorrect. They state in their response "The aim of the study was not to explore sex differences" but it would be difficult to find a more inaccurate phrase to describe this study. A quick glance at the Results and Conclusions of the Abstract undermines this statement. The reported separate analyses for boys and girls are NOT sensitivity analyses by any conventional use of that term. The authors appear not to understand that a single analysis including an interaction term can provide both a test of different slopes for males and females and also the results they already choose to report from separate analyses. First, this avoids the increased risk of Type 1 errors arising from multiple tests of significance. More importantly, it provides extra information not obtained by the separate analyses. For example, if a finding was obtained showing a greater slope in females compared with males this could help explain why earlier studies (possibly with less statistical power) have reported significant trends for girls but not for boys. The present paper concludes that "The results demonstrated some trends in internalizing symptoms across birth cohorts, for both boys and girls." To omit from this whether these trends differ (or not significantly so) between boys and girls is an incredible waste of information. To support this point the regression results reported in Table 2 show a beta of 0.054 for boys' SDQ self-reports and a value of 0.155 for girls, which is almost a three-fold difference.

RESPONSE: We are grateful for this comment by the reviewer and understand the value of adding interaction terms to the regression models. We apologize for using the term “sensitivity analyses”. It has now been deleted from the manuscript. We have computed new regression analyses with 1) birth cohort, 2) sex and 3) birth cohort*sex (interaction term) as independent variables and all scales (SCARED, SMFQ and SDQ (self-report and parent report)) as outcomes. However, we encountered a significant problem with these analyses.

The problem refers to multicollinearity between sex and the interaction term, as the correlation between these two variables was 1.0. Multicollinearity gives biased estimates and should be avoided when computing regression models. There was multicollinearity between sex and sex*birth cohort in all the regression analyses computed, resulting in that the variable sex was thrown out from all of the models. We know that there are statistical methods for dealing with
multicollinearity between continuous independent variables. However, we are not familiar with such methods when it comes to having categorical variables as we do. If we want to explore the relation between birth*sex and the outcome, we also need to include sex and cohort as stand-alone independent variables, thus, we cannot omit sex from the analyses. Overall, as sex was thrown out from the models, this suggests that our data is not suitable for regression analyses including interaction terms.

Secondly, although the reviewer is correct in that an interaction term would provide extra information, we do believe that our separate analyses provide sufficient information in order to fulfil the aim of the study. Certainly, we would have included analyses with interaction terms if we had not encountered the problem with multicollinearity.