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.

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

(This paper received three reviews from its previous journal but only two reviewers agreed to published their review.)

ARTICLE DETAILS

TITLE (PROVISIONAL)	Young age at school entry and attention deficit hyperactivity disorder related symptoms during primary school: Results of a prospective cohort study conducted at German Rudolf Steiner Schools
AUTHORS	Wendt, Janine; Schmidt, Martina F; König, Jochem; Patzlaff, Rainer; Huss, Michael; Urschitz , Michael S

VERSION 1 – REVIEW

REVIEWER	Bo Xiang Department of Psychiatry, Affiliated hospital of Southwest Medical University, China
REVIEW RETURNED	25-Dec-2017

GENERAL COMMENTS	1. The number of missing score of Hyperactivity-Inateention Subscale at baseline was 1383, the authors need to explain the problem in detail. 2.ADHD related symptoms were only evaluated using SDQ, and parents and teachers reported and without clinical diagnosis, the results were necessarily reliable. It's needed for more scale or clinical diagnosis to support. 3.The design of Table 2 was perplexing and imcompact, the primary outcome need to be explain more detailed.

REVIEWER	Linda Halldner,MD, PhD Department of Medical Epidemiology and Biostatsitics, Karolinska Institutet, Sweden
REVIEW RETURNED	25-Dec-2017

GENERAL COMMENTS	This prospective study of adhd-symptoms in relation to age at school start does add knowledge to the field of relative age and adhd, as it is prospecive and has information of adhd-symtoms at baseline. In the abstract: 1) it is not clear which children were selected to participate in the study (which age categories/date of births) 2) The conclusion is that "maturity testing" before school entrance may be not efficient. As I understand the results the data suggests that the maturity testing is not sufficient to avoid negative effects of

age at school start, but there is no comparison with schools that do not apply "maturity testing" before school start.

In the strengths and limitations section is stated "ADHD-related symptoms were assessed in a multi-setting approach"; I would rather call it a dual setting (school and home). In the same section, bias due to sole adjustment for parent-reported ADHD-related symptoms is mentioned; It should be clarifed that this adjustment refers to adjustment at baseline. Further, I miss comments on the response rate, and possible biasses due to systematic differences between participating schools and/or children.

The results are in general clearly presented; I however miss the data on correlations between parent- and teacher-rated symtoms in the results section (mentioned in the abstract). The intra-class correlations are reckonded moderate, but at 4th grade are below 0.5, wich is rather a poor intra class correlation

In the results section (p.13) it is stated that ASE is negatively associated with adhd-related symptoms regardless of source information, however this association is not significant for parent-related symtoms according to table 2a.

I miss a section on the valditiy of SDQ Hyperactivity-Inattention subscale in relation to ADHD diagnosis, especially in the light of that 25% of boys in 2nd grade were above cut-off. Is there a possibility that the SDQ measures not only adhd-related symptoms?

In the discussion (p.18) it is discussed that children with "adhd-related symptoms were obviously not appropriately identified and academically managed"; again is SDQ valid for adhd-symtoms only?, and would the results be the same using higher SDQ cut-offs?, and would it not be possible that the associations of ASE and adhd would be even higher if there was not a preceding "school maturity test"?

REVIEWER	Amelie Wuppermann
	Assistant Professor of Economics, University of Munich, Germany
REVIEW RETURNED	03-Jan-2018

GENERAL COMMENTS	Report on "Young age at school entry and attention-deficit hyperactivity disorder related symptoms during primary school: Results of a prospective cohort study"
	The study empirically investigates the causal impact of relative age for grade (or age at school entry) on symptoms of attention deficit hyperactivity disorder (ADHD) among primary school children in a special type of schools (Rudolf Steiner Schools) in Germany. By restricting the study sample to children who comply with the school entry cutoff dates the authors make use of variation in relative age for grade that is directly related to the timing of birth relative to the cutoff, which should not be related to risks of ADHD for other reasons. To additionally take possible confounders into account the authors adjust their analyses for gestational age at birth, family structure, socio-economic status and migrant background as well as ADHD symptoms before school entry. The results indicate that teacher reports of ADHD symptoms are associated with relative age for grade in 2nd and 4th grade in German Rudolf Steiner Schools. Results are much weaker and only significant for 2nd grade for symptoms as reported by parents.
	Overall, this is a very nice study that is clearly written and investigates a very important topic. More broadly speaking it

relates to the literature on determinants of Attention Deficit Hyperactivity Disorder, the most common mental health condition among children in the developed world. I believe this study clearly merits publication in BMJ Open but would like to see my comments and
concerns below addressed before publication.

The authors suggest that results in much of the previous literature on this question cannot be interpreted causally as they are based on cross-sectional study designs and thus cannot adequately take differences in ADHD symptoms prior to school entry into account. In addition, the authors state that adjustment for possible other confounders in the previous literature is insufficient to allow for causal interpretation. While it is certainly true that much of the previous literature is based on crosssectional study designs, the harsh statement that this precludes causal inference is not warranted and should at least be toned down: the previous literature uses timing of birth in narrow windows around school entry cutoff dates as a "natural" or "quasi" experiment generating variation in age for grade. The beauty of these type of natural experiments is that there is no reason to believe that children born right before a cutoff date and right after a cutoff date (let´s say in June or July in the case of many German states where the cutoff is June 30) vary in anything but the time and age when they enter school. In this sense, this natural experiment of timing of birth mimics a randomized trial, in which subjects are randomly assigned to a treatment and a control group. Hence, there is no reason to expect differences in ADHD symptoms before school entry across the two groups, making it unnecessary to adjust for these differences (or other potential confounders) as there should be none. (This again, is similar to randomized trials in which adjustment for confounders is not necessary.) But of course, I agree with the authors in that it is better to empirically investigate the existence of differences across treatment and control group than just to argue that differences should not exist and thus do not need to be adjusted for. Schwandt and Wuppermann (2016), for example, do briefly investigate differences in ADHD symptoms for children born before and after school entry cutoff dates before school entry (and find none). But with the cross-sectional design, the analysis does face the limitation that it cannot investigate this for the same children for whom also ADHD diagnoses are observed later in their school career. Overall, it seems warranted to tone down the harsh statement that causal inference is not possible based on the extant literature. Instead the authors could state the longitudinal study design of their study as contribution to the literature which allows for actually testing (or adjusting) for differences in ADHD symptoms prior to school entry. (However, the benefit of this analysis is not only limited as only parent reports of symptoms are available at baseline (as noted by the authors) but also because baseline measures are missing and consequently have to be imputed for more than half of the sample (Table 1)).

In the current version of the paper, the authors stress the importance of school entry policies for Rudolf Steiner Schools. However, there are likely additional differences when comparing conventional primary school education to Rudolf Steiner Schools, e.g. in teaching style and emphasis on achievement or marks. If the association between relative age for grade (or school entry age) and ADHD symptoms was different in Rudolf Steiner Schools compared to the general population studied in the earlier literature,

these differences could be results of either: differences in school entry policies or differences in teaching styles (or – as further discussed below – differences in sample composition as students in Rudolf Steiner Schools may vary from the general population). It thus does not seem warranted to only draw conclusions concerning Rudolf Steiner School entry policies.

The number of children with information on ADHD symptoms varies significantly between grade 2 (1965 children) and grade 4 (2617). Are there any important differences across these two groups of children (or participating schools)? The authors should investigate whether their results change when they only include children for whom information on ADHD symptoms is available in both waves in the analysis in order to shed light on the question whether differences in the results between 2nd and 4th grade could stem from the different groups of children on which the results are based.

On pages 8/9 the authors describe their sample selection. In order to rely only on variation in age at school entry resulting from timing of birth (relative to the cutoff date), the authors should only include children in the analysis who comply with the cutoff date. Whether this is actually done, does not become entirely clear from the description: "…we only included children who were participants in 2nd or 4th grade and were born between 30 June 2001, and the respective federal state-specific school entry cutoff date in 2002 (30 June 2002 to 31 December 2002)." First, depending on the federal state June 30, 2001 is not the appropriate start date. For example, children born between June 30, 2001 and December 31, 2001 in Berlin would have entered school already in 2007, not in 2008, if they complied with the cutoff. (To use the correct start date it should further be taken into account whether a change in the cutoff date appeared in 2008 compared to 2007. In Bavaria, for example, the cutoffs were September 30 in 2007 and October 31 in 2008. Thus in Bavaria all children born between October 1, 2001 and October 31, 2002 should have entered school in 2008 if they complied with the cutoff dates). Second, with the present wording it is not clear that the authors rely on the cutoff dates that applied to school entry in 2008.

A reference is missing on page 12.

Figure 3 suggests that roughly 31% of boys who enter school at the age of 6 have an ADHD indication in grade 2 as based on teacher
reports. On page 14, the number of 25% is stated, which – according to Figure 3 – seems to be the value for boys in 4th grade who entered school at the age of 6.

In general, the share of children with ADHD indication seems very high, given that in administrative data roughly 5% of boys in 2nd grade and 7.5% in 4th grade receive ADHD diagnoses (see Schwandt and Wuppermann, 2016). Similarly diagnosis prevalence of ADHD reported by parents in representative survey data for boys aged 7-10 in Germany is only 7.3% in the years 2009-2012 (Schlack et al. 2014, "Hat die Häufigkeit elternberichteter Diagnosen einer Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) in Deutschland zwischen 2003–2006 und 2009–2012 zugenommen?" Bundesgesundheitsblatt, 57:820–829). Are Rudolf Steiner students known to exhibit more

ADHD symptoms than the general population (e.g. as parents of "difficult" children are more prone to sending them to Rudolf Steiner Schools)? If so, that is something that may impede the generalizability of the findings and should be discussed.

The authors discuss possible causal pathways between age at school entry and ADHD symptoms on page 18, stressing channels that lead to an actual development of ADHD symptoms among the younger children. This discussion lacks the possibility that children who are young for their grade behave adequately for their age. Only when compared to their older classmates, their behaviour appears more hyperactive or impulsive and they thus receive higher ratings of ADHD symptoms – particularly when assessed by their teachers. As teachers are more prone to comparing children within a grade than parents, this interpretation is also supported by the different findings when teacher or parent reports are analyzed (see e.g. Elder 2010).

REVIEWER	Annie Garner
	Saint Louis University USA
REVIEW RETURNED	05-Jan-2018

| GENERAL COMMENTS | Thank you for the opportunity to review the manuscript entitled, "Young age at school entry and attention-deficit hyperactivity-disorder-related symptoms during primary school: Results of a prospective cohort study". The manuscript examines the relationship between age at school entry and symptoms of ADHD. This study has several strengths including its longitudinal design which allowed for control of baseline ADHD symptoms and an assessment of the relationship between age at school entry with ADHD symptoms at different grade levels. This topic has significant public health implications given that ADHD is associated with poor outcomes across multiple domains across the lifespan. Overall, I found the manuscript to be well-written and succinct. I see one potentially significant concern with statistical analyses as well as several comments that I think would strengthen the manuscript.

Major concern
• As far as I understand, the analyses account for the fact that each participant has two observations, a parent and a teacher. Do the analyses also account for the non-independence of the observations at the level of the classroom and school? Failing to account for the statistical dependency that occurs when children share the same educational context could lead to biased standard errors and incorrect conclusions about statistical significance. The authors are encouraged to re-run analyses accounting for this nesting.

Other comments (Organized by section)
Strengths and Limitations of this Study
• It is not clear which of the bullet points are strengths and which are weaknesses.
• The meaning of the last bullet point in this section is not clear without having read the article. I suggest that the authors reword this sentence so that it could stand alone and be understood without having read the full manuscript:
Ex. "Adjusting for parent-report of ADHD symptoms prior to school entry might not have been a sufficient control for analyses |

involving teacher report of ADHD symptoms at baseline and subsequent timepoints"

Introduction
• The introduction is very short. Several variables are controlled for in analyses but the research that supports the control of these variables is not discussed in the manuscript. The authors are encouraged to significantly expand upon their review of the literature.
• The Rudolf Steiner Schools are discussed in the introduction. It seems like the Setting and Study Design section would be a more appropriate place for this description.

Procedures and Instruments
• Please provide cronbach's alpha for the Hyperactivity-Inattention subscale of the SDQ for each assessment point.

Restrictions, Definitions, Statistical Analyses
• Please provide a rationale for why migrant background was coded as German born and German nationality versus non-German nationality or born outside of Germany. Is there research to support this distinction?

Results
• The authors report that the relationship between age at school entry and ADHD symptoms was stronger in the 2nd grade compared to the 4th grade. Was this tested statistically? Is a β of -.76 in the 2nd grade statistically different from -.57 in the 4th grade?

Discussion
• The authors are also encouraged to consider the impact of the Rudolf Steiner Schools' enrolment policy which relies on results of mandatory preschool examination rather than fixed cut-off dates. Doesn't this policy truncate the age range of the sample such that you are not likely to have many "very young" children in the sample? What does this truncation do to the analyses and how does it impact the interpretation of these results?
• Is it really a strength that the population is homogeneous? It may also be considered a limitation as these results might not generalize to other countries.

REVIEWER	Dr Megan Bell
	Telethon Kids Institute, Perth, Western Australia, Australia
REVIEW RETURNED	26-Mar-2018

| GENERAL COMMENTS | This is an interesting study on the association between age at school entry, and ADHD symptoms in later childhood. The authors find that young age at school entry is associated with higher levels of teacher-rated ADHD indications in later childhood, regardless of baseline level of ADHD symptoms. The analyses are appropriately constructed and clearly explained. The authors have included statistical controls for a number of confounders known to be associated with ADHD symptoms. The limitations are clearly and appropriately highlighted, and the recommendations for future research are relevant. The implications of the findings for school-entry assessments are well described. |

	One additional limitation which may be of relevance to include, is the high proportion (48%) of missing data on gestational age.

REVIEWER	Matej Stuhec, Ph.D., Pharm.D. 1. Faculty of Pharmacy, Department for Biopharmacy and Pharmacokinetics, Ljubljana, Slovenia. 2. Psychiatric Hospital Ormoz, Department for Clinical Pharmacy, Ormoz, Slovenia.
REVIEW RETURNED	01-Apr-2018

GENERAL COMMENTS	"Young age at school entry and attention-deficit hyperactivity-disorder-related symptoms during primary school: Results of a prospective cohort study" The authors have written about an understudied topic. There are some data on this topic although this topic is interesting. This paper is easy to read. Although this paper is an interesting one it has some important limitations (methodology, limitations) and questions, which should be adressed by authors. The Results are well presented. Language editing is also required (syntax, spelling). However, in the current form presented, I suggest a major revision before publication. General remarks: Positive: - Few data on this topic - New evidence Negative: - Important questions in methodology - Some important limitations Specific remarks: Title: Young age at school entry and attention-deficit hyperactivity-disorderrelated symptoms during primary school: Results of a prospective cohort study It would be better to change to: Young age at school entry and attention deficit hyperactivity disorder related symptoms during primary school: Results of a prospective cohort study Abstract: Conclusions are not completely in line with conclusion of the main text (Discussion). The authors wrote: »Current school enrolment procedures may be not effective in preventing ASE-related negative health outcomes«. In my opinion the authors didn't confirm it and therefore I suggest to remove it. STRENGTHS AND LIMITATIONS OF THIS STUDY »homogeneous population throughout Germany«. It is really homogeneous if the authors didn't include all German states? The authors stated that only Hesse region has approved study protocol. Introduction: First paragraph: I suggest that the author include also newest references about ADHD epidemiology in different countries worldwide. According to the last studies the prevalence is about 8%. Please use the following paper to cite the prevalence in different countries: Croat Med J. 2015 Apr;56(2):159-65 and Eur Psychiatry. 2017 May;42:129-133, where the edpidemiology of

ADHD in different countries has been cited. I suggest that the authors include these important references.

The last paragraph is unclear. Please re-modify this paragraph according to the aims of this study (project vs. paper).

Methods:
Unclear what means 2nd Grade and 4th Grade in Germany. Different countries have different grades. The author should explain it.

Were kids who were already treated excluded?
What the author mean under »linear mixed-effects regression analysis«? »linear regression«, »multiple regression«. If they include continuous variable as dependent variable then linear regression was used. On the other hand if they used non-continuous variable »logistic regression« was used.
Which method was used for missing data (e.g. LOCF)?

Results:
Correlation coefficient has been calculated, although was not specified in the methods.
First paragraph. Please add an appropriate reference.
The authors wrote: »Compared to girls, the Hyperactivity-Inattention Subscale score was higher among boys« although they didn't use statistics (e.g. Chi2).
Table 1: Please use point (not comma 49,8)
Discussion:
I think that the authors didn't confirm the following sentence: »The association remained after adjusting for potential confounders and prevalent symptoms at school entry and was stronger in the 2nd grade compared to the 4th grade«. If they didn't use ANOVA etc. between group analysis they cannot conclude that frequency in one group was bigger than another.
The authors wrote: »We observed moderate agreement«, although they didn't use appropriate methods (e.g. statistics) within methods (e.g. what means moderate?).
The authors wrote in the summary: »In summary, the youngest children within a school year may be at an increased risk …«. In my opinion the authors should include the main findings and not speculate »may be«.
I think that casual relationship has not been tested/confirmed by appropriate statistics. Please re-modify.
Limitation section should be improved. The authors should explain why they used adjusted method instead of multivariable regression analysis (easier to distinguish among different impacts).

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Bo Xiang

Comment: The number of missing score of Hyperactivity-Inattention Subscale at baseline was 1383, the authors need to explain the problem in detail.

Response: Because of the open cohort study design, not all children and schools that participated in 2010 or 2012 have already participated in 2008. Indeed, 1309 of 2671 children

analyzed had a parent assessment in 2008 and 1288 had completed at least 3 of 5 items of the Hyperactivity-Inattention Subscale. We have added an explanatory footnote to Table 1 on page 14.

Comment: ADHD related symptoms were only evaluated using SDQ, and parents and teachers reported and without clinical diagnosis, the results were necessarily reliable. It's needed for more scale or clinical diagnosis to support.

Response: Thank you for this comment. The Hyperactive-Inattention Subscale of the SDQ is a commonly used and validated instrument to measure the key components of ADHD-related behavior. We mentioned this in the methods (page 8, par 3) and discussion (page 20, para 2) and present a corresponding citation. Moreover, we now added a statement (and reference) that the SDQ was found to be a valid tool for discriminating cases with ADHD from those without ADHD, as well as from children experiencing other mental health diagnoses in the UK. In fact, the association between ASE and ADHD diagnoses and/or medication has been already investigated by others and was not the primary aim of the present study. This is now stated in the Discussion (page 20, para 2).

Comment: The design of Table 2 was perplexing and imcompact, the primary outcome need to be explain more detailed.

Response: Thank you for this comment. We have separated Table 2a&b into Table 2 and Table 3 (see page 15 & 16) and added a further line in the table explaining the depending
2
variable. Furthermore, we added that the association between ASE and parent-reported scores did not reach statistical significance (page 15, para 2). Descriptive statistics of the primary outcome is given in the first paragraph on page 13 and in Figure 2.

Reviewer: 2
Reviewer Name: Linda Halldner,MD, PhD

Comment: Abstract: it is not clear which children were selected to participate in the study (which age categories/date of births)

 Response: Thank you for this advice. We added in the abstract that the children were born between 1 July 2001 and 31 October 2002. Additionally, we added the age range (page 2).

 Comment: Abstract: The conclusion is that "maturity testing" before school entrance may be not efficient. As I understand the results the data suggests that the maturity testing is not sufficient to avoid negative effects of age at school start, but there is no comparison with schools that do not apply "maturity testing" before school start.

Response: We fully agree. Due to over-interpretation of our results we deleted this sentence (page 3).

Comment: strengths and limitations: It is stated "ADHD-related symptoms were assessed in a multi-setting approach"; I would rather call it a dual setting (school and home).

Response: We are grateful for this suggestion. We replaced "multi-setting approach" by "dual-setting approach" (see page 4 and 18).

Comment: strengths and limitations: bias due to sole adjustment for parent-reported ADHDrelated symptoms is mentioned; It should be clarifed that this adjustment refers to adjustment at baseline.

Response: Thank you for this advice. We fully agree that this important information was missing. We added "at baseline" in the last bullet point in the Strengths and Limitations (page 4).

Comment: I miss comments on the response rate, and possible biases due to systematic differences between participating schools and/or children.

Response: Response rates of schools are given in the methods sections (page 7, para 2). Furthermore, we stated that we were not able to perform a non-responder analysis due to missing information for the entire source population of pupils (page 21, para 2).

Comment: Results: I however miss the data on correlations between parent- and teacher-rated symptoms in the results section (mentioned in the abstract). The intra-class correlations are reckoned moderate, but at 4th grade are below 0.5, wich is rather a poor intra class correlation

Response: Thank you for this advice. We added the intra-class correlation coefficients in the results section (page 15, para 1). Furthermore, we changed the statement from "moderate" into "poor" agreement in the abstract (page 1) and the discussion section (page 19, para. 2 & page 20 para.1).

Comment: In the results section (p.13) it is stated that ASE is negatively associated with adhd-related symptoms regardless of source information, however this association is not significant for parent-related symptoms according to table 2a.

Response: We are grateful for this suggestion. We added the statement in the results section on page 15 (para. 2) that the associations between ASE and parent-reported symptoms are not significant.

Comment: I miss a section on the validity of SDQ Hyperactivity-Inattention subscale in relation to ADHD diagnosis, especially in the light of that 25% of boys in 2nd grade were above cut-off. Is there a possibility that the SDQ measures not only adhd-related symptoms?

Response: Thank you for this comment. The Hyperactive-Inattention Subscale of the SDQ is a commonly used and validated instrument to measure the key components of ADHD-related behaviour. We mentioned this in the methods (page 8, para. 3) and discussion (page 20, para. 2) and gave a corresponding reference. Moreover, we now added a statement (and reference) that the SDQ was found to be a valid tool for discriminating cases with ADHD from those without ADHD, as well as from children experiencing other mental health diagnoses in the UK. However, we fully agree that the HI Subscale of the SDQ may pick up other – not strictly ADHD-related – symptoms like stress or poor adaptation to school challenges. This is now mentioned in the Discussion (page 20, para. 2).

Comment: In the discussion (p.18) it is discussed that children with "adhd-related symptoms were obviously not appropriately identified and academically managed"; again is SDQ valid for adhd-symptoms only?, and would the results be the same using higher SDQ cut-offs?, and would it not be possible that the associations of ASE and adhd would be even higher if there was not a preceding "school maturity test"?

Response: Thank you for this comment. We used a score of <6 as "no indication of ADHD" and a score of ≥6 as "indication of ADHD" by applying German population reference values. We further combined borderline (score 6) and abnormal (score ≥7) hyperactivity/inattention symptoms to contrast children without and with indications of ADHD. We agree, that the

results would be probably different, if we use higher cut-offs. However, we refrained from performing these analyses because of the markedly lower frequency of ADHD-indications and, hence, poorer statistical power.

On the other hand, it is possible that the associations of ASE and ADHD-related symptoms would be even higher if there was not a preceding PSE with a focus on developmental aspects ("maturity testing"). This is now mentioned in the Discussion (page 21, para 2).

Reviewer: 3
Reviewer Name: Amelie Wuppermann

Comment: The authors suggest that results in much of the previous literature on this question cannot be interpreted causally as they are based on cross-sectional study designs and thus cannot adequately take differences in ADHD symptoms prior to school entry into account. In addition, the authors state that adjustment for possible other confounders in the previous literature is insufficient to allow for causal interpretation. While it is certainly true that much of the previous literature is based on cross-sectional study designs, the harsh statement that this precludes causal inference is not warranted and should at least be toned down: the previous literature uses timing of birth in narrow windows around school entry cutoff dates as a "natural" or "quasi" experiment generating variation in age for grade. The beauty of these type of natural experiments is that there is no reason to believe that children born right before a cutoff date and right after a cutoff date (let´s say in June or July in the case of many German states where the cutoff is June 30) vary in anything but the time and age when they enter school. In this sense, this natural experiment of timing of birth mimics a randomized trial, in which subjects are randomly assigned to a treatment and a control group. Hence, there is no reason to expect differences in ADHD symptoms before school entry across the two groups, making it unnecessary to adjust for these differences (or other potential confounders) as there should be none. (This again, is similar to randomized trials in which adjustment for confounders is not necessary.) But of course, I agree with the authors in that it is better to empirically investigate the existence of differences across treatment and control group than just to argue that differences should not exist and thus do not need to be adjusted for. Schwandt and Wuppermann (2016), for example, do briefly investigate differences in ADHD symptoms for children born before and after school entry cutoff dates before school entry (and find none). But with the cross-sectional design, the analysis does face the limitation that it cannot investigate this for the same children for whom also ADHD diagnoses are observed later in their school career. Overall, it seems warranted to tone down the harsh statement that causal inference is not possible based on the extant literature. Instead the authors could state the longitudinal study design of their study as contribution to the literature which allows for actually testing (or adjusting) for differences in ADHD symptoms prior to school entry. (However, the benefit of this analysis is not only limited as only parent reports of symptoms are available at baseline (as noted by the authors) but also because baseline measures are missing and consequently have to be imputed for more than half of the sample (Table 1)).

Response: Thank you very much for this detailed comment and advice. After reviewing the corresponding sentences, we agree that our statements on the causal inference of previous studies could be perceived as harsh. We, therefore, adjusted the corresponding phrases in the introduction section (page 5, para. 2). In addition, in the discussion section we mentioned that our study can be seen as a "contribution" to the pre-existing literature (page 19, para. 2). Moreover, we included some comments about "natural experiments" and the necessity to adjust for confounders and baseline symptoms as well as limitations of our analysis (page 19, para. 2).

Comment: In the current version of the paper, the authors stress the importance of school entry policies for Rudolf Steiner Schools. However, there are likely additional differences when comparing conventional primary school education to Rudolf Steiner Schools, e.g. in teaching style and emphasis

on achievement or marks. If the association between relative age for grade (or school entry age) and ADHD symptoms was different in Rudolf Steiner Schools compared to the general population studied in the earlier literature, these differences could be results of either: differences in school entry policies or differences in teaching styles (or – as further discussed below – differences in sample composition as students in Rudolf Steiner Schools may vary from the general population). It thus does not seem warranted to only draw conclusions concerning Rudolf Steiner School entry policies.

Response: We fully agree with the reviewer that there are several differences when comparing conventional primary school education to Rudolf Steiner Schools. However, we did not investigate whether the association between ASE and ADHD symptoms was different in Rudolf Steiner Schools compared to the general population studied in the earlier literature. Therefore, we refrained from adding a discussion on the various differences between public and Rudolf Steiner primary schools.

Comment: The number of children with information on ADHD symptoms varies significantly between grade 2 (1965 children) and grade 4 (2617). Are there any important differences across these two groups of children (or participating schools)? The authors should investigate whether their results change when they only include children for whom information on ADHD symptoms is available in both waves in the analysis in order to shed light on the question whether differences in the results between 2nd and 4th grade could stem 8 from the different groups of children on which the results are based.

Response: Thank you for this suggestion. We have shared this concern and have therefore chosen a statistical method including up to 4 assessments into a common model for correlated data. This models give unbiased estimates even if missingness depends on observed data (see Verbeke, G, Molenberghs, G, Linear Mixed Models for Longitudinal Data). In contrast, restricting analysis to complete cases may be generally prone to selection bias.

In order to further investigate this concern, we additionally checked whether the Hyperactivity-Inattention Subscale scores and ASE depend on the pattern of available data. We therefore, divided our analysis data set into three groups according to the observer (parents or teachers) and time of observation (2nd grade only, 4th grade only, or both grades). The results are as follows:

Information available at	Hyperactivity-Inattention Subscale score												Age at school entry		
	2nd grade						4th grade								
	Parents reports			Teachers reports			Parents reports			Teachers reports					
	N	Mean	SE	N	Mean	SE	N	Mean	SE	N	Mean	SE	N	Mean	SE
4th grade only	0	.	.	0	.	.	903	2.50	0.07	761	2.68	0.10	908	6.67	0.01
2nd grade only	292	2.43	0.12	231	2.88	0.20	0	.	.	0	.	.	294	6.62	0.02
2nd and 4th grade	1402	2.30	0.05	1317	2.60	0.08	1356	2.24	0.05	1361	2.57	0.07	1469	6.66	0.01
All	1694	2.32	0.05	1548	2.64	0.07	2259	2.34	0.04	2122	2.61	0.06	2671	6.66	0.01

Based on these results, we identified no meaningful differences and none were statistically significant (t test or one-way ANOVA). We, hence, concluded that data were missing at random and that the missingness pattern has likely not introduced relevant bias to the effect estimates.

In order to further investigate this concern, we have replicated some of the analyses. We have refitted model 2 of Table 2 restricting to 1277 children with teacher reports available at grade 2 **and** grade 4. However, results remained widely unchanged showing similar trends and effect sizes as in the primary analysis (see below). Due to these unchanged results, we decided to not present these results in the manuscript. However, upon request we can add these analyses to the supplement of the submission.

Association between age at school entry and Hyperactivity-Inattention Subscale score based on all available data (N=2671) or children with teacher reports available both for 2nd and 4th grade (N=1277) (multivariable linear regression for correlated outcome).

| | Second grade | | | | | | Fourth grade | | | | | |
| | Parent reports | | | Teacher reports | | | Parent reports | | | Teacher reports | | |
Data set	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value
All data	-0.24	0.14	0.0874	-0.68	0.19	0.0004	-0.10	0.13	0.4251	-0.58	0.17	0.0009
Teacher reports available at grade 2 and 4	-0.35	0.18	0.0459	-0.63	0.23	0.0055	-0.17	0.18	0.3255	-0.49	0.23	0.0297

The model contained the same variables for confounder adjustment as in Table 2, line 2, set 2: adjusted for gender, birth status, family form, CASMIN score, and migrant background.

Comment: On pages 8/9 the authors describe their sample selection. In order to rely only on variation in age at school entry resulting from timing of birth (relative to the cutoff date), the authors should only include children in the analysis who comply with the cutoff date. Whether this is actually done, does not become entirely clear from the description: "…we only included children who were participants in 2nd or 4th grade and were born between 30 June 2001, and the respective federal state-specific school entry cutoff date in 2002 (30 June 2002 to 31 December 2002)." First, depending on the federal state June 30, 2001 is not the appropriate start date. For example, children born between June 30, 2001 and December 31, 2001 in Berlin would have entered school already in 2007, not in 2008, if they complied with the cutoff. (To use the correct start date it should further be taken into account whether a change in the cutoff date appeared in 2008 compared to 2007. In Bavaria, for example, the cutoffs were September 30 in 2007 and October 31 in 2008. Thus in Bavaria all children

born between October 1, 2001 and October 31, 2002 should have entered school in 2008 if they complied with the cutoff dates). Second, with the present wording it is not clear that the authors rely on the cutoff dates that applied to school entry in 2008.

Response: Thank you very much for this really important point. We have now improved wording of the corresponding sentences in the Methods (page 9, para. 2). In addition, we have discussed the school entry policy with representatives of Rudolf Steiner Schools several times and found out that they do not fully comply with legal regulations. As now mentioned in the methods section (page 7, para. 3) the school enrolment policy is rather based on the results of a mandatory preschool examination (PSE) than on fixed cut-off dates. Moreover, representatives of Rudolf Steiner Schools reported that some schools have ignored the legal changes in school entry cut-offs starting in 2003 and continued their school entry policy by using June, 30th as their preferred cut-off date. Hence, we used a combination of Rudolf Steiner School-specific and federal state-specific cut-off dates as inclusion criterion for the primary analysis.

To assess whether our restriction criteria have introduced bias to the results, we additionally performed a sensitivity analysis, thereby restricting the sample to children, who fully comply with the federal state-specific cut-off dates. This analysis is now described in the methods section (page 11, para. 4), results are given in the Results (page 17) and in the supplementary Table 1. Furthermore, we discussed the findings in the discussion section (page 21, para. 2). Overall, the alternative restriction criterion did not change the results considerably.

Comment: A reference is missing on page 12.

Response: We are grateful for this hint and added the reference to Table 1 (para. 1. on page 13).

Comment: Figure 3 suggests that roughly 31% of boys who enter school at the age of 6 have an ADHD indication in grade 2 as based on teacher reports. On page 14, the number of 25% is stated, which – according to Figure 3 – seems to be the value for boys in 4th grade who entered school at the age of 6.

Response: Thank you for this comment, we have checked all numbers and Figures. In Figure 3, we present model-based predicted proportions of ADHD indication as function of age at school entry and stratified by gender and source of assessment. Indeed, the predicted proportion was as high as 31% for six years old boys at school entry assessed at 2nd grade by teachers. In the first paragraph on page 16 we just described the frequency of ADHD indications irrespective of age at school entry. The quoted number of 25% can be identified in Figure 3 at the age at school entry of 6.7 years.

Comment: In general, the share of children with ADHD indication seems very high, given that in administrative data roughly 5% of boys in 2nd grade and 7.5% in 4th grade receive ADHD diagnoses (see Schwandt and Wuppermann, 2016). Similarly diagnosis prevalence of ADHD reported by parents in representative survey data for boys aged 7-10 in Germany is only 7.3% in the years 2009-2012 (Schlack et al. 2014, "Hat die Häufigkeit elternberichteter Diagnosen einer Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) in Deutschland zwischen 2003–2006 und 2009–2012 zugenommen?" Bundesgesundheitsblatt, 57:820–829). Are Rudolf Steiner students

known to exhibit more ADHD symptoms than the general population (e.g. as parents of "difficult" children are more prone to sending them to Rudolf Steiner Schools)? If so, that is something that may impede the generalizability of the findings and should be discussed.

Response: Thank you for this comment. We have added a section on this limitation in the Discussion (page 18, para. 3).

Comment: The authors discuss possible causal pathways between age at school entry and ADHD symptoms on page 18, stressing channels that lead to an actual development of ADHD symptoms among the younger children. This discussion lacks the possibility that children who are young for their grade behave adequately for their age. Only when compared to their older classmates, their behavior appears more hyperactive or impulsive and they thus receive higher ratings of ADHD symptoms – particularly when assessed by their teachers. As teachers are more prone to comparing children within a grade than parents, this interpretation is also supported by the different findings when teacher or parent reports are analyzed (see e.g. Elder 2010).

Response: Thank you for this advice. We fully agree with your comment and added this statement in the discussion section (page 22, para. 2).

<u>Reviewer: 4</u>

<u>Reviewer Name: Annie Garner</u>

Comment: As far as I understand, the analyses account for the fact that each participant has two observations, a parent and a teacher. Do the analyses also account for the non-independence of the observations at the level of the classroom and school? Failing to account for the statistical dependency that occurs when children share the same educational context could lead to biased standard errors and incorrect conclusions about statistical significance. The authors are encouraged to re-run analyses accounting for this nesting.

Response: Thank you for this important comment. Indeed, we have used up to four observations per participant, as defined by the cross classification of the factors time of observation (2nd and 4th grade) and source of observation (parents and teachers). We have not accounted for dependencies resulting from the clustering at the level of classroom and school, because we do not expect inflation of standard errors. Note, that only 2.6% of the variance of our explanatory variable "age at school entry" is explained by the factor school. Hence, age at school entry is clearly a "within cluster" explanatory variable. Thus, we do not expect inflation of standard errors and omitted an adjustment for this nesting.

Comment: *Strengths and Limitations:* It is not clear which of the bullet points are strengths and which are weaknesses.

Response: Thank you for this advice. For the bullet points 2 and 3 we added that it was possible to assess ADHD-related symptoms in a dual-setting approach and that we were able to adjust for

baseline symptoms, respectively. Now it should be more clear that these two bullet points are strengths.

Comment: *Strengths and Limitations:* The meaning of the last bullet point in this section is not clear without having read the article. I suggest that the authors reword this sentence so that it could stand alone and be understood without having read the full manuscript: Ex. "Adjusting for parent-report of ADHD symptoms prior to school entry might not have been a sufficient control for analyses involving teacher report of ADHD symptoms at baseline and subsequent timepoints"

Response: Thank you for this advice. We fully agree with you that the meaning of the last bullet point was not clear. We therefore adapted his sentence.

Comment: *Introduction*: The introduction is very short. Several variables are controlled for in analyses but the research that supports the control of these variables is not discussed in the manuscript. The authors are encouraged to significantly expand upon their review of the literature.

Response: Thank you for this comment. We expanded the introduction with a sentence concerning known risk factors of ADHD, Moreover, we added a reference that describes the developmental trajectories of hyperactivity-impulsivity and inattention symptoms and prenatal, perinatal, and postnatal risk factors in the Methods (page 9, para. 3, ref. no. 7). Moreover, we discussed the approach of adjusting for potential confounders in the discussion section on several occasions. With regard to the word count, we further refrained from extensively discussing potential risk factors of ADHD, because none of them are correlated with ASE and confound the association between ASE and ADHD.

Comment: *Introduction*: The Rudolf Steiner Schools are discussed in the introduction. It seems like the Setting and Study Design section would be a more appropriate place for this description.

Response: We agree and moved this paragraph to the methods section (page 7, para. 3).

Comment: *Procedures and Instruments*: Please provide Cronbach's alpha for the Hyperactivity-Inattention subscale of the SDQ for each assessment point.

Response: Cronbach's alpha is now provided for the Hyperactivity-Inattention Subscale of the SDQ (page 14, para 1) and mentioned in the Methods.

Comment*: Restrictions, Definitions, Statistical Analyses:* Please provide a rationale for why migrant background was coded as German born and German nationality versus non-German nationality or born outside of Germany. Is there research to support this distinction?

Response: Thank you for this comment. The coding of a migrant background is based on the definition of the German Federal Ministry of Justice and Consumer Protection. We added this information (page 10, para. 1).

Comment: *Results*: The authors report that the relationship between age at school entry and ADHD symptoms was stronger in the 2nd grade compared to the 4th grade. Was this tested statistically? Is a β of -.76 in the 2nd grade statistically different from -.57 in the 4th grade?

Response: Thank you for this comment. Differences in effect sizes were not tested statistically and were reported only on a descriptive basis. This is now clarified in the Results on two occasions.

Comment: Discussion: The authors are also encouraged to consider the impact of the Rudolf Steiner Schools' enrolment policy which relies on results of mandatory preschool examination rather than fixed cut-off dates. Doesn't this policy truncate the age range of the sample such that you are not likely to have many "very young" children in the sample? What does this truncation do to the analyses and how does it impact the interpretation of these results?

Response: Thank you for this comment. As mentioned in the methods section, Rudolf Steiner Schools have lower proportions of early (2%) and higher proportions of delayed school entries (13%) compared to public schools in Germany (6% and 5%, respectively). This policy truncates the age range and reduced the fraction of "very young" children in the source population of our sample. In contrast, our restriction aims at reducing the fraction of "very old" children (following school entry deferral the year bevor), because these children have serious medical and/or educational reasons for the deferral and would have introduced bias if not excluded. Based on these particularities, we performed a sensitivity analysis with a more rigorous restriction, now excluding more children at the edges of the age distribution. However, this did not change the results of the study. This discussion is now presented (page 21, para. 2).

Comment: Is it really a strength that the population is homogeneous? It may also be considered a limitation as these results might not generalize to other countries.

Response: Thank you for this comment. A homogeneous sample usually increases internal and decreases external validity (generalizability) at the same time. We, however, agree that our results should not be generalized to other school settings or countries. This is now stated in the Discussion (page 18, para 1).

Reviewer: 5

Reviewer Name: Dr Megan Bell

Comment: One additional limitation which may be of relevance to include, is the high proportion (48%) of missing data on gestational age.

Response: Thank you for this comment. Because gestational age was not associated with age at school entry, the high proportion of missing values should not have introduced bias. Hence, we think that the proportion of missing values for gestational age is not important enough to discuss it in the Limitations.

Reviewer: 6

Reviewer Name: Matej Stuhec, Ph.D., Pharm.D.

Comment: *TITLE*: Young age at school entry and attention-deficit hyperactivity-disorder related symptoms during primary school: Results of a prospective cohort study It would be better to change to: Young age at school entry and attention deficit hyperactivity disorder related symptoms during primary school: Results of a prospective cohort study

Response: Thank you for this advice. We deleted the hyphens, accordingly.

Comment: *Abstract*: Conclusions are not completely in line with conclusion of the main text (Discussion).

The authors wrote: »Current school enrolment procedures may be not effective in preventing ASE-related negative health outcomes«. In my opinion the authors didn't confirm it and therefore I suggest to remove it.

Response: We fully agree with this comment. Due to over-interpretation we deleted this sentence in the Abstract.

Comment: *STRENGTHS AND LIMITATIONS OF THIS STUDY:* homogeneous population throughout Germany«. It is really homogeneous if the authors didn't include all German states? The authors stated that only Hesse region has approved study protocol.

Response: Thank you for this comment. We are not sure, if the reviewer meant "representative" instead of "homogeneous"? The sample is quite homogeneous, because all pupils attended the same educational setting (i.e. a Rudolf Steiner School) across Germany. In contrast to Rudolf Steiner Schools, public schools are very heterogeneous across Germany, because education is in the responsibility of the German federal states.

The principle investigator of the study was located in Hesse, hence, the study protocol was approved by the ethics committee of the federal physician chamber in Frankfurt/Main, Hesse. However, the approval was valid for all German federal states.

Comment: *Introduction*: First paragraph: I suggest that the author include also newest references about ADHD epidemiology in different countries worldwide. According to the last studies the prevalence is about 8%. Please use the following paper to cite the prevalence in different countries: Croat Med J. 2015 Apr;56(2):159-65 and Eur Psychiatry. 2017 May;42:129-133, where the edpidemiology of ADHD in different countries has been cited. I suggest that the authors include these important references

Response: Thank you very much for this important advice. By looking at the reference of Eur Psychiatry. 2017 May;42:129-133 we found out that the authors cited the same references as we did. In the publication Croat Med J. 2015 Apr;56(2):159-65 only older scientific publications were cited. Because of the fact that we investigated school-aged children, we added further prevalence estimates, which are given in "Polanczyk GV, Willcutt EG, Salum GA, Kieling C, Rohde LA. ADHD

prevalence estimates across three decades: An updated systematic review and meta-regression analysis. International Journal of Epidemiology 2014;43: 434–42" and "Willcutt EG. The prevalence of DSM-IV attention-deficit/hyperactivity disorder: A meta-analytic review. Neurotherapeutics: the Journal of the American Society for Experimental Neurotherapeutics 2012;9: 490–9" and "Polanczyk G. The Worldwide Prevalence of ADHD: A Systematic Review and Metaregression Analysis. The American journal of psychiatry 2007;164: 942." These publications are now cited in the Introduction (page 5, para 1).

Comment: *Introduction*: The last paragraph is unclear. Please re-modify this paragraph according to the aims of this study (project vs. paper).

Response: Thank you for this advice. To make it more clear we re-modified this paragraph (page 6, para. 1).

Comment: *Methods*: Unclear what means 2nd Grade and 4th Grade in Germany. Different countries have different grades. The author should explain it.

Response: Thank you for this comment. In Germany, school entry, grade 2, and grade 4 result in a current age of 6-7 years, 7-8 years and 9-10 years. This is now stated in the Methods (page 7, para. 2).

Comment: *Methods*: Were kids who were already treated excluded?

Response: Thank you for this comment. Unfortunately, we did not assess pre-existing or prevalent ADHD diagnoses or medication use for ADHD in this study. Hence, it is possible that children with a diagnosis of ADHD and an effective ameliorating treatment were misclassified as disease-free by the SDQ. However, this misclassification was non-differential (i.e. misclassification affected children of all ages in the same way) and would not explain our findings. In fact, the association between ASE and ADHD diagnoses and/or medication has been already investigated by others and was not the primary aim of the present study. This is now stated in the Discussion (page 20, para. 2).

Comment: *Methods*: What the author mean under »linear mixed-effects regression analysis«? »linear regression«, »multiple regression«. If they include continuous variable as dependent variable then linear regression was used. On the other hand if they used non-continuous variable »logistic regression« was used.

Response: We apologize for having caused confusion. We have used linear models for correlated data for our continuous primary outcome variable 'ADHD symptom score'. We have used the term linear mixed-effects model, which in the strict sense denotes a linear model that contains both fixed and random effects but is often used for the wider type of linear models that allow for more than one observation per subject. Having observed the ADHD symptom score by parent and teacher ratings at 2nd and 4th grade, we have up to four observations per subject. We have chosen this comprehensive and somewhat complex model, because it is less prone to selection bias due to missing values (see Verbeke, G, Molenberghs, G, Linear Mixed Models for Longitudinal Data) as compared to model each score separately in a different data set. In the method section we have changed the wording

now avoiding the word 'mixed' and using 'multivariable linear regression for correlated data' instead (page 10, para. 2).

Comment: *Methods*: Which method was used for missing data (e.g. LOCF)?

Response: The methods used to cope with missing data are described in the statistical methods section. See also our responses to a comment of Reviewer 3, p7 bottom, and to your previous comment. Briefly, we fitted a common model for correlated outcomes for Row 1 and 2 in Table 2 and 3. We additionally used multiple imputation when adjusting for the baseline ADHD symptom score (Row 3 in Table 2 and 3). In answer to Wuppermann's comments we have added a complete case analysis to this response letter, as a complementary sensitivity analysis. We observed no shifts in all observed trends.

Comment: *Results*: Correlation coefficient has been calculated, although was not specified in the methods.

Response: Thank you for this advice. We have calculated intra-class correlation coefficients, which is now stated in the methods section (page 10, para 2).

Comment: *Results*: First paragraph. Please add an appropriate reference.

Response: We are grateful for this hint and added the reference to Table 1 (para. 1. on page 13).

Comment: *Results*: The authors wrote: »Compared to girls, the Hyperactivity-Inattention Subscale score was higher among boys« although they didn't use statistics (e.g. Chi2).

Response: These statements describe the obvious patterns in Figure 2. All four comparisons are statistically highly significant (p<0.0001, Satterthwaite t-test). Nevertheless, we are tending not to report these p-values, because the gender effect was not the focus of our paper and has been repeatedly reported by other authors. Upon request by this reviewer, we can add these p-values to the Results.

Comment: *Results*: Table 1: Please use point (not comma 49,8).

Response: Thank you for this hint. We changed the comma into a point in Table 1.

Comment: *Discussion*: I think that the authors didn't confirm the following sentence: »The association remained after adjusting for potential confounders and prevalent symptoms at school entry and was stronger in the 2nd grade compared to the 4th grade«. If they didn't use ANOVA etc. between group analysis they cannot conclude that frequency in one group was bigger than another.

Response: The sentences intended to sum up the essential results of Table 2 and 3 on a descriptive basis. This is now stated in the Results on several occasions. Each table presents associations between ASE and ADHD-related symptoms, adjusted for gender in the first line, and for further variables in lines 2 and 3. The explanatory variable ASE was used as a continuous variable. Effects are described as parameter estimates and odds ratios with 95% confidence intervals. In fact, we have

not compared frequencies in one group to frequencies in another group. The reported odds ratios are unit odds ratio and estimate the relative change in odds of ADHD-indications between children that were one year apart in age at school entry. We have added some additional words to clarify this point of concern (Table 3).

Comment: *Discussion*: The authors wrote: »We observed moderate agreement«, although they didn't use appropriate methods (e.g. statistics) within methods (e.g. what means moderate?).

Response: Thank you for this comment. We used descriptive statistics (Figure 2) and the intra-class correlation coefficient between parent and teacher reports to assess agreement. This is now stated in the Methods and Results. Accordingly, we changed the word "moderate" to "poor", because the correlation was mostly <0.5.

Comment: *Discussion*: The authors wrote in the summary: »In summary, the youngest children within a school year may be at an increased risk …«. In my opinion the authors should include the main findings and not speculate »may be«.

Response: Thank you for this comment. We re-modified this sentence (page 23, para. 1).

Comment: *Discussion*: I think that causal relationship has not been tested/confirmed by appropriate statistics. Please re-modify.

Response: Thank you very much for this important comment. Due to the observational study design, causal inference cannot be warranted. We, therefore, expanded the discussion on causal inference and the methods used in the present study (page 19, para. 2). Indeed, we had no control over the explanatory variable 'age at school entry' like in a randomized trial. As common practice for cohort studies, we used different adjustment models and obtained similar effect estimates, which renders confounding an unlikely explanation for the observed effects.

We further have restricted our analysis to children complying strictly to school entry rules, which makes selection bias also an implausible explanation. Finally, as argued by other reviewers (e.g. Wuppermann) timing of school entry has the character of a "natural experiment" because birth dates are uniformly distributed over the year and can be expected to be uncorrelated to bio-psycho-social characteristics of a child. We have modified the discussion somewhat and do not claim now to have proven a causal relationship between ASE and ADHD symptoms.

Comment: *Discussion*: Limitation section should be improved. The authors should explain why they used adjusted method instead of multivariable regression analysis (easier to distinguish among different impacts).

Response: Thank you for this advice. Obviously, there is misunderstanding: The only adjustment method applied was multivariable regression analysis. The regression models are extensively described in the statistical methods section. In Tables 2 and 3, we report effects for the interesting variable 'age at school entry' and not for other variables that have been included for the purpose of

confounder adjustment. Upon request we can add in footnotes to Table 2 and 3 that the effects of adjustment variables are not reported.

VERSION 2 – REVIEW

REVIEWER	Bo Xiang Affiliated Hospital of Southwest Medical University, China
REVIEW RETURNED	21-Jun-2018

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.

REVIEWER	Linda Halldner Karolinska Institutet and Umeå Universitet, Sweden
REVIEW RETURNED	29-Jun-2018

GENERAL COMMENTS	General comment: I believe that the authors have adequately answered my (and other reviewers´) comments and have made changes accordingly when appropriate. In my opinion the paper has much improved after revision. This work contributes with additional knowledge on the topic of relative age effects and should thus be published. Specific comment: In the introduction, line 5, p. 5, it says "ADHS" instead of ADHD.

REVIEWER	Amelie Wuppermann LMU Munich, Germany
REVIEW RETURNED	29-Jun-2018

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.

REVIEWER	Matej Stuhec, Pharm.D., Ph.D. Faculty of Pharmacy Ljubljana, Slovenia
REVIEW RETURNED	14-Jun-2018

GENERAL COMMENTS	The authors have answered on all my questions/remarks and therefore I suggest to accept this paper.