PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A nationwide, population-based study of school grades in practical and aesthetic subjects of children treated for a brain tumour
AUTHORS	Lönnerblad, Malin; van't Hooft, Ingrid; Blomgren, Klas; Berglund, Eva

VERSION 1 – REVIEW

REVIEWER	Reviewer name: Peter Flom Institution and Country: Peter Flom Consulting USA Competing interests: None
REVIEW RETURNED	13-Dec-2019

GENERAL COMMENTS	I confine my remarks to statistical aspects of this paper. I have a few issues to resolve before I can recommend publication. The biggest one is how average grade was computed. On p. 9, it says that the grades were fail, pass, etc. These cannot be averaged as is. Were they converted to numbers? If so, how and why? If not, then what was done? This will affect the choice of method as linear regression (named higher on p. 9) is only appropriate for a continuous dependent variable. Other issues: p. 6 line 36 P values and stat sig are not relevant here. Look at effect size. Table 1 Delete the p value column p 9 Do not categorize continuous independent variables. In *Regression Modeling Strategies* Frank Harrell lists 11 problems with this and concludes "nothing could be more disastrous". I wrote a blog post showing, graphically, what can go wrong: https://medium.com/@peterflom/what-happens-when-we-categorize-an-independent-variable-in-regression-77d4c5862b6c Table 2 a) Show all the odds ratios b) This is not the standard way to report on logistic regression. None of what is reported is wrong, and if it suits the editors, OK, but the usual method would make this 5 tables and would show the parameter estimate, the ORs and CIs for each effect. So there would be lines for PBTS, Male and interaction in the table for each DV (this could be formatted as one table or five tables) Table 3 - similar to my b) comment above. This is not the usual format. Show lines for PBTS, male and interaction, with each line having beta, SE, p and CI

REVIEWER	Reviewer name: Defne Saatci Institution and Country: Institute of Child Health University College London 30 Guilford Street London WC1N 1EH Competing interests: None
REVIEW RETURNED	18-Dec-2019

GENERAL COMMENTS	The manuscript is an interesting paper on assessing educational attainment in practical and aesthetic subjects (PRAEST) in paediatric brain tumour survivors. Despite their potential to significantly improve the quality of life of survivors, these subjects have not been sufficiently explored in literature to date. This manuscript provides novel opportunity to explore these subjects but has limitations that needs to be addressed. It is uncertain if the identified cohort of paediatric brain tumour survivors is representative of the true population - therefore making results difficult to interpret. Only 65% of (475 out of 724) the identified paediatric brain tumour survivors are included in the study. Demographics of the exclusions (35%) were not provided and therefore it is impossible to say if these exclusions significantly biased results. For example, did more males get excluded due to death/missing data than females and if so could this bias the overall results? This limitation should be addressed in the discussion. It is difficult to justify including survivors with relapses and children potentially still undergoing treatment (as authors describe inclusion as children one-year post diagnosis but not remission) in this study. These children would not normally be considered survivors. This group of children can have other reasons for poor educational performance such as missing school rather than differences in their abilities. This could particularly bias the primary outcome “failing of PRAEST subjects” as the authors define failing “as either absence from school or not obtaining expected goals”. Furthermore, as passing the PRAEST subjects are not required to pass on to upper secondary school in Sweden, survivors may choose to preferentially drop these subjects if they cannot attend school often (rather than failing due to their ability). Furthermore, it is not clear what proportion of this group is female or male, once again making it difficult to assess bias on the overall results. The methodology described in this manuscript is unclear. Were identified survivors from the Swedish Cancer Registry linked to Statistics Sweden? If so, what was the linkage methodology? How was this achieved without survivor consent – was ethics approval gained? The matching process also requires further clarification; what do authors mean by matched by age? As authors are assessing grades at year 9, most children will be the same age (15-16) so details of how age was matched would be useful. The statistical methods described by the authors are well-defined. It is unclear however why gender is individually chosen as a covariate in the regression models where it is already accounted for in the matching process? It would be more useful to know how gender and age-at-diagnosis or tumour type interact to influence outcomes? It is unclear if this study has sufficient power to detect a difference in male survivors’ performances. Control males have much lower average grades compared to control females.

	Thus, the differences in grades between control males and survivor males is much smaller and this is more likely to need a larger sample size to detect a true difference. It is important to consider this in the discussion as the recommendation from this manuscript focuses on supporting females only and does not highlight the need for further and larger studies to evaluate male outcomes. Specific comments: • Introduction: o Page 5, line 5 – could each subject be explained in detail? Not all countries have crafts as a subject and it would be useful to know what this means in Sweden. o Page 5, line 10 – full stop missing • Methods: o Page 6, lines 55-60 and page 7 lines 3-4. The description of the age groups needs further clarification. This is much more clearly explained in your previous publication (Lönnerblad, 2019) • Results: o Page 10 line 18 – authors describe results as 2-3 times higher odds in failing subjects. Accurate statement of the ORs would be more useful. o Page 10, line 27 – cannot find the Crafts p-value described in the text in the corresponding table 2. • Discussion: o Page 14, line 29 – authors state “Treatment for a high- grade or a low-grade tumour did not have any impact in any subject on average grade or failing.” Cannot infer results about treatment from tumour type.

VERSION 1 – AUTHOR RESPONSE

 # Reviewer: 1 Peter Flom
Institution and Country: Peter Flom Consulting USA
Comments to the Author:

I confine my remarks to statistical aspects of this paper.

I have a few issues to resolve before I can recommend publication.
The biggest one is how average grade was computed. On p. 9, it says that the grades were fail, pass, etc. These cannot be averaged as is. Were they converted to numbers? If so, how and why? If not, then what was done? This will affect the choice of method as linear regression (named higher on p. 9) is only appropriate for a continuous dependent variable.

Answer: We agree that the decision to approximate the numerals that in principle constitute an ordinal scale to interval-scale methods might be controversial, yet it is the official way to enable calculation of average grades in Sweden. We have now added information about this in the method section (page 6 line 39-48 in the first manuscript) and added the reference below:

https://www.skolverket.se/download/18.47fb451e167211613ef398/1542791697007/swedishgrades_bilaga.pdf

For examples by the National Agency for Education (Skolverket), see below:

https://siris.skolverket.se/reports/rwservlet?cmdkey=common&geo=1&report=gr_betyg_amne&p_flik=G&p_ar=2019&p_lankod=&p_kommunkod=&p_skolkod=&p_hmantyp=&p_hmankod=&p_flik=G

https://www.skolverket.se/getFile?file=3496 page 41

Other issues:

p. 6 line 36 P values and stat sig are not relevant here. Look at effect size.

Answer: We agree with the reviewer that effect sizes are often more relevant than p-values. However, the purpose of Table 1 is to show that the differences between PBTS and controls are not statistically significant, and that they, hence, are comparable.

Table 1 Delete the p value column

Answer: The p-values in Table 2 (previously Table 1) only demonstrate that the background variables of the PBTS subjects and their controls were sufficiently similar, and the p-values are reported to verify this. Therefore, we consider the column informative, or even necessary.

p 9 Do not categorize continuous independent variables. In *Regression Modeling Strategies* Frank Harrell lists 11 problems with this and concludes "nothing could be more disastrous". I wrote a blog post showing, graphically, what can go wrong: https://medium.com/@peterflom/what-happens-when-we-categorize-an-independent-variable-in-regression-77d4c5862b6c

Answer: The independent variables used in the logistical regressions were sex, diagnosis (or not), and age group. We have earlier demonstrated that age at diagnosis is of significant importance for the school results (Lönnerblad et al., 2019). Age at diagnosis is, as noted, a continuous variable, and we are aware of the problems inherent in categorizing a continuous variable, e.g. overly simplified assumptions, loss of power. However, the categorization of age groups follows the Swedish school system, with pre-school, early years (when children learn basic skills, particularly reading, writing and math) and middle/upper primary school, as described in the first manuscript (page 6 line 53 ff). Thus, we find the applied categorization relevant to interpretations of the school results, assuming that contracting a disease and being subjected to the treatment will have different consequences depending on the school stage attended at the time.

Table 2 a) Show all the odds ratios b) This is not the standard way to report on logistic regression. None of what is reported is wrong, and if it suits the editors, OK, but the usual method would make this 5 tables and would show the parameter estimate, the ORs and CIs for each effect. So there would be lines for PBTS, Male and interaction in the table for each DV (this could be formatted as one table or five tables)

Answer: a) We have now added all odds ratios. **b)** To save space, and to make the table more readable, the parameter estimates were excluded, but can easily be calculated by ln(OR). For the same reason, estimated standard errors were also excluded. However, we have added a line with the p-values for each interaction sex*group, where group indicates PBTS or control, to Table 2 (now Table 3). We see the reviewer's point about including parameter estimates and SEs, but find the tables to be more readable without them.

Table 3 - similar to my b) comment above. This is not the usual format. Show lines for PBTS, male and interaction, with each line having beta, SE, p and CI

Answer: We have now added a line with the p-values for each interaction sex*group, and added the parameter estimates and 95% CIs of the difference in average grade between PBTS and controls in Table 3 (now Table 4).

 # Reviewer: 2 Defne Saatci

Institution and Country: Institute of Child Health University College London
30 Guilford Street
London WC1N 1EH

Comments to the Author:

The manuscript is an interesting paper on assessing educational attainment in practical and aesthetic subjects (PRAEST) in paediatric brain tumour survivors. Despite their potential to significantly improve the quality of life of survivors, these subjects have not been sufficiently explored in literature to date. This manuscript provides novel opportunity to explore these subjects but has limitations that needs to be addressed.

It is uncertain if the identified cohort of paediatric brain tumour survivors is representative of the true population - therefore making results difficult to interpret. Only 65% of (475 out of 724) the identified paediatric brain tumour survivors are included in the study. Demographics of the exclusions (35%) were not provided and therefore it is impossible to say if these exclusions significantly biased results. For example, did more males get excluded due to death/missing data than females and if so could this bias the overall results? This limitation should be addressed in the discussion.

Answer: Thank you for this remark. We have now added more information about the demographics of the exclusions in Figure 1. We find that the added information demonstrates that no apparent bias was introduced by uneven exclusion rates.

It is difficult to justify including survivors with relapses and children potentially still undergoing treatment (as authors describe inclusion as children one-year post diagnosis but not remission) in this study. These children would not normally be considered survivors. This group of children can have other reasons for poor educational performance such as missing school rather than differences in their abilities. This could particularly bias the primary outcome “failing of PRAEST subjects” as define failing “as either absence from school or not obtaining expected goals”.

Answer:

We agree that the term ”survivors” could be considered controversial in the for the reasons indicated by the reviewer. We therefore performed a sensitivity analysis in which we excluded children less than two years post diagnosis (n=45; 10%), since no treatment protocol is as long as two years and all children have completed their treatment by then. The sensitivity analysis is presented in the table below and shows that there were virtually no differences when children less than two years post diagnosis were excluded, so it is unlikely that this procedure introduced an error. We added sentence about this in the Discussion (page 19, Strengths and Weaknesses).

Table with only PBTS more than 2 years post diagnosis included

Number, percentage and odds ratios (OR) with 95% confidence intervals (CI) and p-values for failing the subjects for PBTS **(n=430) at least 2 years post diagnosis** vs. controls (n=2197).

	N (%) PBTS	**N (%) Controls**	**OR (95% CI)**	**p-value PBTS vs. controls**
Home and consumer studies				
All	29 (6.7%)	112 (5.1%)	1.35 (0.88-2.05)	0.162 n.s.
Females	16 (7.8%)	38 (3.6%)	2.26 (1.24-4.14)	0.013
Males	13 (5.8%)	74 (6.5%)	0.886 (0.482-1.63)	0.767 n.s.
Physical education and health				
All	44 (10.2%)	153 (7.0%)	1.52 (1.07-2.67)	0.021
Females	28 (13.6%)	74 (7.0%)	2.09 (1.32-3.33)	0.003
Males	16 (7.1%)	79 (6.9%)	1.03 (0.59-1.80)	0.886 n.s.
Art				
All	33 (7.7%)	109 (5.0%)	1.59 (1.06-2.38)	0.027
Females	20 (9.7 %)	39 (3.7%)	2.81 (1.60-4.93)	0.001
Males	13 (5.8 %)	70 (6.2%)	0.94 (0.51-1.73)	1.000 n.s.
Crafts				
All	26 (6.0%)	91 (4.1%)	1.49 (0.95-2.33)	0.095 n.s.
Females	16 (7.8%)	32 (3.0%)	2.70 (1.45-5.02)	0.003
Males	10 (4.5%)	59 (5.2%)	0.855 (0.43-1.70)	0.741 n.s.

Music				
All	45 (10.5%)	129 (5.9%)	1.87 (1.31-2.68)	0.001
Females	26 (12.6%)	44 (4.2 %)	3.33 (2.00-5.55)	< 0.001
Males	19 (8.5%)	85 (7.5%)	1.15 (0.68-1.93)	0.583 n.s

Furthermore, as passing the PRAEST subjects are not required to pass on to upper secondary school in Sweden, survivors may choose to preferentially drop these subjects if they cannot attend school often (rather than failing due to their ability).

Answer: We agree, and we have now added this to the discussion in the section Strengths and weaknesses of the study. However, it is not possible for pupils to decide to drop subjects of their choice unless the school gives permission to do so.

Furthermore, it is not clear what proportion of this group is female or male, once again making it difficult to assess bias on the overall results.

Answer: This is a good point, and **w**e have now added this information to Figure 1. Also, please see our reply to the question above.

The methodology described in this manuscript is unclear. Were identified survivors from the Swedish Cancer Registry linked to Statistics Sweden? If so, what was the linkage methodology? How was this achieved without survivor consent – was ethics approval gained? The matching process also requires further clarification; what do authors mean by matched by age? As authors are assessing grades at year 9, most children will be the same age (15-16) so details of how age was matched would be useful.

Answer: We have now tried to describe the linkage more clearly in the Method section and added a section about consent. We have also changed “matched by age, /…/” to “matched by birth-year, /…/” in figure 1.

Ethical approval for this study was given by the Regional Ethical Review Board in Stockholm (no. 2017/995-31/5). This information was included in the submission but we see now that it is not shown in the manuscript; we have now added this information.

The statistical methods described by the authors are well-defined. It is unclear however why gender is individually chosen as a covariate in the regression models where it is already accounted for in the matching process? It would be more useful to know how gender and age-at-diagnosis or tumour type interact to influence outcomes?

Answer: The reviewer is correct in stating that no adjustment for gender is needed when comparing PBTS and controls, due to the matching process. However, gender was chosen as a variable of interest, because grades differ between the sexes. No interaction between gender and age at diagnosis was found, in contrast to our previous study of theoretical subjects (Lönnerblad et al., 2019). We had no pre-specified hypothesis of tumour grade (high or low) affecting final school grades differently in males compared to females, and therefore we did not check for such interactions. The same goes for interactions between age at diagnosis tumour grade.

It is unclear if this study has sufficient power to detect a difference in male survivors' performances. Control males have much lower average grades compared to control females. Thus, the differences in grades between control males and survivor males is much smaller and this is more likely to need a larger sample size to detect a true difference. It is important to consider this in the discussion as the recommendation from this manuscript focuses on supporting females only and does not highlight the need for further and larger studies to evaluate male outcomes.

Answer: This is a very good point, and we have now added this to the discussion in the section Unanswered questions and future research. It is possible that a larger population size would detect differences also between male PBTS and controls. Concerning power, the study clearly has sufficient power to detect the large differences between PBTS girls and control girls, while the differences between PBTS boys and their controls seem to be smaller, and are therefore harder to detect. Had the differences between PTBS boys and non-PBTS boys been as large as for the girls, we would most likely have gotten statistically significant results also for the boys.

Specific comments:

- Introduction:

o Page 5, line 5 – could each subject be explained in detail? Not all countries have crafts as a subject and it would be useful to know what this means in Sweden.

Answer: Thank you for the suggestion. We have now created a new table with the aims of the different subjects as they are explained in the Swedish curriculum for current years (2003-2012).

o Page 5, line 10 – full stop missing

Answer: We have added a full stop on this side and line.

- Methods:

o Page 6, lines 55-60 and page 7 lines 3-4. The description of the age groups needs further clarification. This is much more clearly explained in your previous publication (Lönnerblad, 2019)

Answer: We have now added information to these sentences and hope the explanation is better.

- Results:

o Page 10 line 18 – authors describe results as 2-3 times higher odds in failing subjects. Accurate statement of the ORs would be more useful.

Answer: Accurate statements of the ORs are now added to the results section.

o Page 10, line 27 – cannot find the Crafts p-value described in the text in the corresponding table 2.

Answer: The p-values for the interaction effects between sex and controls or PBTS were only presented in the text, not in Table 2, but we have now added them (now in Table 3).

- Discussion:

o Page 14, line 29 – authors state "Treatment for a high- grade or a

low-grade tumour did not have any impact in any subject on average grade or failing.” Cannot infer results about treatment from tumour type.

Answer: It is true that we do not have information about the treatments for each patient, but high-grade tumours are almost invariably treated with cranial radiotherapy and more intense chemotherapy, and these patients are hence more prone to develop more pronounced cognitive deficits. Radiotherapy is particularly known to cause debilitating cognitive and other side effects.