Author's response to reviews

**Title:** Influence of family size and birth order on risk of cancer: a population-based study

**Authors:**

Melanie Bevier (m.bevier@dkfz-heidelberg.de)
Marianne Weires (m.weires@dkfz-heidelberg.de)
Hauke Thomsen (h.thomsen@dkfz-heidelberg.de)
Jan Sundquist (jan.sundquist@med.lu.se)
Kari Hemminki (k.hemminki@dkfz-heidelberg.de)

**Version:** 5 **Date:** 19 November 2010

**Author's response to reviews:** see over
Dear Editor,

With this letter we submit the revised manuscript “Influence of family size and birth order on risk of cancer: a population-based study” for publication in BMC Cancer.

Please find below the point-by-point response to the reviewer's comments:

REVIEWER 1

An impressing sample size, providing an outstanding chance for an investigation on family size and birth order.

Major Compulsory Revisions:

1. You cannot do your investigating with persons alive. Neither the healthy persons nor the cancer patients. In order to be sure that a given person in you material has cancer or not, the person must have been observed until death. And to be sure that a patient with one type of cancer later develop yet another type of cancer, the patients will have to be observed as long as she or he is alive.

   For a new run, selected only persons in your Cancer Registry, who are no longer alive. Drop the parameters "Follow up stated at year of immigration, birth year, or start year of cancer registry (1961)" as stated in Methods, and use only "age at onset of cancer", even in cases with onset of cancer before 1961, when your Cancer Registry started. You will have to trace them all to complete your sample, even whose who had onset of cancer before 1961.

   - In the analysis we focused on primary cancers. We stopped the follow-up period at the time where the first cancer occurred. So it is irrelevant if the person is still alive or not or was affected with other cancers later. First registered cancer was the event of interest.

   We wanted to analyze the time interval between 1961 and 2006 to see which persons had cancer during this time. We adjusted for several covariates to calculate expected number of cancer cases in a reference group (for example first born children) and compare these numbers with the observed number of cases in the non-reference group. This gives us a relative risk estimate for the observed cancer site.

2. You cannot excluded cancer parents the way you do it with a remark in Material "that family history should not be taken into account". First of all, the removal of affected parents does not at all secure that the family history is excluded, viz. that the inborn burden of cancer susceptibility in your sample is excluded. The burden will be reduced but not minimized, but worse: you violate your data severely creating a selective loss. The impact of environmental heterogeneity on family size (I guess that you mean size of the sibship) and on the birth order
should be recorded in your maximum likelihood test where you have relevant parameters like socioeconomic stage (where you forgot a description so that the reader cannot see which stage) and area (rural, urban etc. I guess, because neither here you provide any description) nicely ready for a comparison with birth order and size of the family/sibship. It is from the maximum likelihood that you should land your findings, and not from a dangerous selection among input data.

- I corrected the sentence that only parental cancer history should not be taken into account. Even if I include them into my analyses, the results will not change significantly. The explanations of the covariates I used in my analyses, I included now in the manuscript so it is more clear which information I had.

3. You cannot do your investigation with childhood cancer. It makes no sense to ask for childhood variables like birth order and family size and their influence on later cancer among adults, when the child has a cancer already. Your set-up is unable to embrace childhood cancer.

You'll have to excluded childhood ALL, brain- and kidney tumours in children etc. and if you can't (because adult and childhood cancer might be mixed in the ICD7 diagnoses), then you will have to exclude the whole ICD7 compound for the diagnoses in question. Hence, your analysis will not cover any type of childhood cancer. If you wish to include childhood cancer, you will need quite a different set up. Be very careful with the rather large proportion of your patients who have more than one cancer diagnosis in adult age and have a strategy for twins.

- As I studied only primary cancers, the persons that had childhood cancer during the period of follow-up were stopped at the year of diagnosis. That means that for childhood cancers I stopped the follow-up after the diagnosis and did not take any later cancer into account - if there was any.

4. As far as I can see, you do not take all patients through the data processing? I don't know whether I am right or not, but you have a total of 134,896 cancer patients (Table 1) and you have 59,053 patients in Table 2 (birth order) and 109,687 patients in Table 3 (family size).

One would like to know why there is this difference. Hopefully not a selective loss, but you provide absolutely no text to help the reader.

Why are the number of persons investigated in Table 2 and Table 3 not equal to the total number of patients included?
I took all patients through data processing. The cancer cases in Table 2 and 3 are divided in the cancer cases that occur in the reference group (the cancer cases that were diagnosed at first borns or in one-child families) and the non-reference group (at least second born children or children in families with at least two children). I split this number up to make clear how the cancer cases were distributed into the reference and the observed group.

5. Your description of the data-processing is insufficient and one cannot seen the details and what has been done. A Poisson Regression combined with a Maximum Likelihood Test sounds quite right. That part of Material and Methods is free from any information, an exp(beta) makes no sense to me. A maximum likelihood test is a fascinating and sophisticated operation which deserves all the best and a proper description of censors, approximators and trends. Here, you provide nothing. In your next version of the manuscript, please describe the data programs used, and the data-processing done, in such a way that the reader can see what you have asked your computer to do.

Tables 1-3 are impressive with all RRs smaller than 2 ! You do not clearly state the level at which you accept significance ? Do so in the next version, please, and take out all significantly high or low RR's for two new tables (Table 4 birth order and Table 5 family size) and give us all details on influence of age at onset, sex and age of the patients, sex and age of older and younger sibs, age of parents at birth of the child, socioeconomic stage, region where the patienten lived etc. related to ICD7 and sib size, and birth order. As it is now you don't report a word about the power of there variables. I am not comfortable with an age estimation either over or under 50 yr, and I am sure that you can show up a much better description (e.g.: age < 40, age 41-50, 51-60, 61-70, 71-80, > 80 yrs). I am not sure that I really see the need for your reference cases and there is no help for me in your text.

- PROC GENMOD in SAS is used to perform a Poisson regression analysis with a log link function. A Poisson distribution is used to model the distribution of cell counts, where the cells are derived through the classification of the included covariates. As an offset-variable the logarithm of the calculated person-years is used. The cancer status (affected or not affected) is specified as a response variable, whereas the included covariates are explanatory variables.

Significance is accepted at a 5% confidence level.

We split up the analyses at age of 50 years because we wanted to see whether the effect of birth order and family size change from early to late adulthood. In these analyses we adjusted for age still using 5-year-bands.
Minor Essential Revisions

Please consider English revision of the language when writing you new version. You probably do not mean it, but indeed: you wrote "Some subtypes of leukemia as acute lymphoblastic leukemia come from viral or bacterial infections" (your text in the middle of Discussion). There are unfortunately other places like that in the manuscript. And please, in the next version of Discussion, stick to your findings and discuss the implications and restrictions. Never mind the figures, I think you can do without

- I revised the language in the new version. I also changed the discussion part to discuss clearly the implications and restrictions.

REVIEWER 2

Title: Influence of family size and birth order on risk of cancer: a population-based study

Version: 2 Date: 13 July 2010

Reviewer: Yani Lu

Reviewer's report:

Summary:

Bevier et al. examined the association between family size, birth order and risk of cancer using the Swedish Family-Cancer Database. The authors proposed to check if the effects of family size and birth order on the risk of cancer were carried on from childhood to adulthood. They divided the age at diagnosis in two groups, below and over age 50 years. The authors observed that increasing birth order decrease the risk of endometrial, testicular, skin, thyroid and connective tissue cancers and melanoma, but increase the risk of lung, male and female genital cancers; moreover, family size was also associated with several types of cancers. The authors concluded that the effect of birth order decreases from childhood to adulthood for lung and endometrial cancer.

Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached)

1. The authors proposed to check whether the effects of family size and birth order on the risk of cancer are carried on from childhood to adulthood. However, the authors divided the age at diagnosis into groups of below and over age 50.
Both groups of cancer were adult-onset. If all the cancers before age 50 in this dataset were diagnosed after age 18, then, in fact, the authors checked the effects of family size and birth order on cancers of early age adult-onset (before age 50) and late age adult-onset (after age 50). The authors did not provide the basic information of age distribution; it is hard for readers to judge what age group the data can address.

- To analyze an effect from childhood to adulthood I analyzed all individuals and separated for age at diagnosis less than 20 years and at least 20 years. There were no significant associations found for any cancer site. I could only check whether the influence of birth order or family size changed during life - from early to late adulthood (where the cut-off is at age of diagnosis 50 years). I also provided an overview of the age distribution of the individuals included in the study.

2. The authors did not state any limitation of this study. Obviously, several limitations exist for this study. For example, the authors reported the association between birth order and lung cancer. However, smoking history was not taken into account in the analysis. Besides the adverse effect of active smoking, passive smoking has been found to increase the risk of lung cancer. The increased risk for increasing birth order may due to higher level of passive smoking during childhood (older brother or sister may smoke) and active smoking later in life.

- I included a paragraph in the manuscript where I stated the limitations of the study. It could not be taken into account the active nor the passive smoking as the information was missing in our data. We also had no information on obesity to explain some association on the risk of endometrial cancer.

3. Hemminki et al examined the same topic using the Family-Cancer Database 10 years ago. In that study, the authors found an increased risk for breast cancer by birth order and a decreased risk for melanoma by birth order. In this recent study, the authors did not find any association for breast cancer. The authors need to address the potential reasons for the inconsistent results.

- In the previous study the analyzed population was around 10 years younger. In our study we used a larger database where we have older individuals registered which leads to more representative results.
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. The authors grouped subjects into two groups: age at diagnosis below and over age 50 years old. Where are those subjects whose ages at diagnosis are equal to 50 years old?

- The individuals that were diagnosed at the age of 50 were included in the group of individuals diagnosed over age 50. I corrected that in the manuscript.

2. In the first paragraph of introduction, the authors mentioned that higher parental age at conception has not been reported to be a risk factor. In fact, higher paternal age has been linked to prostate cancer, non-Hodgkin lymphoma and breast cancer; higher maternal age also has been linked to breast cancer.

- I included the references where parental age at conception was reported to be a risk factor.

3. In the methods part, the authors mentioned that individuals born after 1932 have been registered. The cancer registry covers all cancers from 1961 to 2006. What about subjects who were diagnosed between years 1932 to 1961. How did the authors treat these people? Were they excluded from the study?

- I only had a look at the time period after 1961 as we had the cancer data from then onwards. I analyzed the individuals whether they were affected with cancer during this period of time.

4. In the methods part, “The family size (grouped 1, 2, 3-4, 5-17) is defined as the number of children per mother. What about those who had divorce?

- As it is more common that the children live with their mother (even after a divorce) we defined the variable of family size as the number of children per mother.

5. In the methods part, “Individuals were categorized according to their age at diagnosis of cancer below age 50 years and above age 50 to distinguish between the effect of birth order and family size in childhood and adulthood.” The cut at age 50 is too high to address the
different effect of exposures in childhood and adulthood. It may be more appropriate to check their effect on early age adult-onset cancer and late age adult-onset cancer.

- I additionally analyzed cancers with a stratification for age at diagnosis of 20 years. In this analysis no significant associations with birth order or family size were found for any cancer site. The analysis with a cut at age of 50 has to be interpreted as varying effect through early and later adulthood.

6. In the methods part of the covariates: please list the detail categories in these variables. For example, “age” means “age at diagnosis” or “age at registry”? What is the age range? What means “period”? Calendar period? What are these 5 groups? What are the 4 groups of region? How socioeconomic status was defined? What are those 6 groups? Is there any subject with unknown values of these categories?

- I included a detailed explanation of the covariates in the methods part. If there were any unknown values they were coded as 'other'.

7. In the results part for Table 3: This paragraph is confusing regarding to results for which group of age at diagnosis. Please clarify the age group after describing the relative risk. The authors mentioned that “In a separate analysis for the age at diagnosis, the risk was increased for stomach cancer……” It seems like that the authors were presenting the results in Table 3. It this is true, it is not appropriate to describe it as “a separate analysis”.

- I described the analyses for age less than 50 years and above 50 years as a "stratified analysis" as there are the results presented in the table for the stratified analysis for age at diagnosis.

8. In the results part: “The risk for cancer of the thyroid gland was marginally significantly decreased for birth order……” The authors did not report the results for specific birth order on the risk of other types of cancer. Does that mean the results for other types of cancer are null? Please clarify. Same to the next paragraph on family size, only testicular cancer was mentioned. What about other types of cancer?

- I just mentioned the results for specific birth order for thyroid cancer and no other cancer sites as they showed significant associations which was not observed for the other cancer sites.
9. In the discussion part: the authors drew the conclusion from the results of lung cancer, “The effect of birth order and family size decreased from childhood to adulthood.” However, this conclusion is not appropriate since we don’t know what their effect on cancer in childhood.

- I changed the conclusion in an appropriate way as we could not identify an effect on cancer in childhood. The effect of birth order and family size decreased from early to later adulthood for lung cancer.

10. The results for family size and endometrial cancer (age at diagnosis below 50) is controversy between Table 3 and Figure 2. In figure 2, using family size of 5-17 as reference group, the relative risks for family size of 1, 2, 3-4 are 0.65, 0.64, and 0.72 separately for age at diagnosis below 50 years. In Table 3, using family size of 1 as the reference group, the relative risk was 0.72 (95% CI=0.64-0.81) for age at diagnosis below 50 years. Please check the analysis and corresponding results.

- I did the analysis again and changed the figures for family size and endometrial cancer.

11. Table 3 presented the results for family size; the reference group should “first born child” or “family size of 1”?

- I changed the reference group for the analyses in Table 3 (revised manuscript: Table 4) in "one-child families".

Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore):

1. In the methods part for cancer sites grouping: “Cancer sites were grouped by …..” should be edited to emphasize that these are special cancer sites instead of all cancer sites. “Liver” should be “liver and gall bladder”?

- I changed the explanation of cancer site grouping ICD7-codes 155.0-156.9 to "liver and gallbladder".
2. In Table 3, the RR for Nervous system (age at diagnosis <50 years) is statistically significant (RR=1.16, 95% CI=1.05-1.29), but was not bolded.

- I bolded that RR for Nervous system cancer in Table 3 (revised manuscript: Table 4).

3. The first word of the notes for Table 2 and Table 3 should be capitalized.

- I capitalized the first words of the notes in the tables.

REVIERE 3

Title: Influence of family size and birth order on risk of cancer: a population-based study

Version: 2 Date: 8 August 2010

Reviewer: Jacqueline Clavel

Reviewer's report:

General comments

This paper reports an analysis of the influence of birth order and sibship size on the risk of adult cancer. It is based on the large and unique Swedish family cancer database. The question is well defined and interesting. The paper tends to be too fast on many points and should be improved by a more precise and rigorous presentation.

Minor Essential Revisions

There are some misleading phrasings:

1- Introduction:

a. the sentence line 2 suggests the relationship between birth order and birth weight is well established. Is it really so? If yes, please explain and develop this statement, else rephrase it so that local observations do not appear as generally admitted results.
- I changed the formulation in the background paragraph that the relationship between birth order and birth weight as these were only local observations.

b. Similarly, observing a negative relationship (as reported between testicular cancer and high birth order) should not be described as detecting a protective effect unless the relationship is causal. Therefore, the sentence should be either developed, in case evidences for a causal association are available, or moderated.

- I reformulated the negative relationship between testicular cancer and high birth order: "Negative association has also been detected in testicular cancer for children of higher birth order."

c. Again, family size and birth order influence the early environment and lifestyle, but they are not actual environmental factors.

- I corrected the misleading sentence of birth order and family size as environmental factors and wrote that family size and birth order influence childhood environment and lifestyle.

d. Lastly, some information given at the top of page 4 should rather take place in the Methods section.

- I shortened the last part of background section and put the detailed information about the study population in the methods section.

2- Methods

a. Clinical information on tumors is available for most cases. However, only CIM7 is used. What information is available and for how many cases and for which cancers?

- The clinical information on tumors was used to translate the ICD-codes back to ICD-7 that all information on cancer sites were comparable (with the registered cancers from the beginning of the registry where ICD-7 coding was used).
b. The reason why individuals with a family history of cancer are excluded should be explained more clearly. If the motivation by the risk of confounding, it would have been more interesting, in my opinion, to analyze the data with and without these individuals and to elicit this confounding, since this database offers this rare opportunity. If it is a question of selection bias due to the fact that parental cancer could have censored the size of the family, it should be explained (may be the authors have a possibility to evaluate such a bias, which could be interesting?). May be the motivation is elsewhere.

- The aim was to look at the effect of birth order and family size. To avoid any confounding effect by parental cancer cases we excluded the individuals having affected parents. However, I did another analysis were I included all offspring (having or not having affected parents) but the results did not change significantly.

c. In an attempt to alleviate the issue of multiple comparisons, the authors used 1% confidence intervals. However, given the number of tests, it is far to address the issue and finally, does not really weaken it either. I think that 5% confidence intervals are sufficient.

- We wanted to be even more confident that the confidence interval includes the true population value so we additionally analyzed the data using 1% confidence intervals (instead of only showing 5% confidence intervals).

d. Please, give the power of the study for different cancers before and after 50 years old.

- As we analyze a very large study population, we get a power of 0.999 for rare cancers as nose as well as for common cancers as breast cancer or melanoma. The power is the same for before and after 50 years of age.

3- Results:

a. Page 6: The comments on the methods used in table 2 should take place in the Method section and in footnotes, rather than in the text of the Results.

- I shortened the part on the methods used in the results part.
b. page 7 lines 4: “Smaller family size had a protective effect on”. Something like “smaller family size was negatively associated with” would be more adequate, given that causality is not established. Also Page 6: “the effect of family size.” would be replaced by something like “the association between family size…”.

- I wrote "Smaller family size was negatively associated with stomach cancer..." instead of "smaller family size had a protective effect on..." as the causality is not established.

c. Page 7 lines 6-8: “as most… non-significant results” is not exact. It means that the results would have been significant if the numbers had been sufficient. This is only an assumption, the facts are that numbers are small, power is too low, and results are not significant.

- I corrected this in the following way: "Most of the stomach cancer cases occur with an age of diagnosis of at least 50 years and the separate analysis for the younger age group shows no significant associations."

d. “Risk” is sometimes used instead of “Relative Risk”.

- I wrote "relative risk" instead of "risk".

e. Legends of the tables:

  i. poisson should begin with a capital letter

- I changed "Poisson" (written with a capital letter).

  ii. the legend of table 3 on family size indicates “first-born child” as reference. Is it really the case?

- One-child families are used as a reference in the analyses presented in Table 3 (revised manuscript: Table 4). I changed it in the legend.
Major Compulsory Revisions

4- Results

a. Is there any possibility to distinguish subtypes given the available clinical information? In particular, could chronic lymphoid leukemia, chronic myeloid leukemia and acute leukemia be separated? papillary and medullary thyroid cancers?

- I did additional analysis regarding subtypes of leukemia and thyroid cancers. The results for papillary, medullary, follicular and other thyroid cancer, I included as a table in the manuscript. For lymphoid leukemia there was a relative risk of 1.30 (95% CI: 1.04-1.61) found for increasing family size. The other analyses showed no significant associations.

b. What are the relationships between the variables of interest, birth order and size of family, and between them and socioeconomic status? Are these relationships different before and after 50 years old?

- Larger family size is associated with lower socioeconomic status. To our knowledge there is no different relationship before and after 50 years of age.

c. Is obesity available in the database? If yes, the relationship between endometrial cancer and birth order.

- No, we do not have any information on obesity in the database.

d. Figures: I would have preferred tables rather than figures, all the more that figure 1 puts together figures on lung cancer with birth order and figures on stomach cancer and family size for stomach cancer, without confidence interval.

- In figure 1 I pointed out the relative risk for stomach cancer and how this risk is changing for family size stratified for birth order. Vice versa, I showed the relative risks for lung cancer. Adding confidence intervals in the graph would have been too confusing, therefore I added the tables with the relative risks and confidence intervals.
Tables 2 and 3 are the most important and should be included in the paper rather than additional. The figures should be put in a table. Table 1 could be described in the text and given as an additional file.

- I included Tables 2 and 3 (revised manuscript: Tables 3 and 4) now in the manuscript.

5- Discussion

a. The discussion section does not discuss the validity of the results, and often takes the observations as if they were causal. The discussion of the relationships between endometrial cancer, socioeconomic status and obesity is too far from the facts and the observations and should be improved.

- I changed the discussion part, especially the discussion of the relationships between endometrial cancer, socioeconomic status and obesity.

In conclusion, the paper could be good because the data are very interesting but it has not been written thoroughly enough and should be improved

Sincerely yours,

Melanie Bevier (on behalf of the co-authors)

Division of Molecular Genetic Epidemiology C050
German Cancer Research Center (DKFZ)
Im Neuenheimer Feld 580, 69120 Heidelberg, Germany
Tel: +49-6221-42 18 05
Fax: +49-6221-42 18 10
E-mail: m.bevier@dkfz-heidelberg.de