PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Prospective Investigation of Folic Acid Supplements Before and During Early Pregnancy and Pediatric and Adult Cancers in the Chinese Children and Families Cohort: A Pilot Study in a Sample of Rural and Urban Families
AUTHORS	Linet, Martha; Wang, Linhong; Wang, Ning; Berry, Robert; Chao, Ann; Hao, Ling; Li, Zhu; Fang, Liwen; Yin, Peng; Potischman, Nancy; Sun, Xin; Meng, Fanweng; Yang, Ruilan; Cong, Shu; Fan, Jing; Kitahara, Cari; Liang, Xiaofeng; Liu, Fang; Lu, Xiaojun; Lv, Fan; Mu, Chunhua; Sampson, Joshua; Tang, Yongmin; Wan, Weiqing; Wang, Baohua; Wang, Hongsheng; Zhang, Leping; Wang, Yu

VERSION 1 – REVIEW

REVIEWER	Lauren C. Houghton Columbia University, New York, USA
REVIEW RETURNED	09-Mar-2018

GENERAL COMMENTS	This manuscript describes, in great detail, a pilot study that has the potential to be resourceful cohort study. However, the inability to link the cohort with hospital records seems to be a substantial limiting factor for future scaling of the project. In general, the conclusion of the abstract should be more specific and the discussion of the paper could be strengthened by including more detail about the cultural context of this study. Given that the pilot generated distribution of the demographics and risk factors, are there any potential exposures that may be particularly hypothesis generating in this context? Page 7, line 12-17: The second sentence of the introduction is unclear, please provide more detail for a reader that is not familiar with the biological evidence. Page 9, line 22: Were there many infants without known sex? If so, could this be a source of bias? Page 11 line 54- Page 12, line 6: The inability to link the cohort participants with the registries seems a limiting factor for scaling the project, more detail as to why this was not possible is warranted. Page 12, lines 38-48. Subject involvement section seems unnecessary especially since it does not provide references. Was there any consultation with the subjects themselves regarding study design? Such qualitative data may help in future scaling of the cohort.

	Page 13, line 17: What exclusions are you referring to? Page 18, line 34-35: There seems to be a typo here. Page 20, line 30: 23% as this is not a small percentage. Please remove the qualifier “only” Page 25, lines 5-22. The proposed data sharing seems to be a conservative approach, especially since much of externally funded NIH research must have more open access data sharing policies than what is proposed here.

REVIEWER	Denhard de Smit Clinical Genetics and APH Research Insitute, VU University Medical Center, Amsterdam, The Netherlands
REVIEW RETURNED	18-Mar-2018

GENERAL COMMENTS	Dear authors, Have you checked whether the use of Chi-square is applicable in all (sub)tables? The subtables 3 -Alcohol use and 4 - Living arrangements/ Current activity/ Smoking exposure/ Alchohol use, alle have one or more cells with an expected observation under H0 of less than 5 which makes the Chi-square unreliable. If you corrected for this, please mention it explicitly. Furhermore you have included in the analysis rows for 'Unknown'. These should in my opinion be left out of the Chi-square test. This not a category of the variable that you are testing. Unknown should be treated as missing data.

VERSION 1 – AUTHOR RESPONSE

REVIEWERS’ COMMENTS TO AUTHOR:

REVIEWER: 1

Reviewer Name: Lauren C. Houghton

Institution and Country: Columbia University, New York, USA

Please state any competing interests or state ‘None declared’: I am a former fellow of NCI where some of the authors of this manuscript currently work

Please leave your comments for the authors below This manuscript describes, in great detail, a pilot study that has the potential to be resourceful cohort study. However, the inability to link the cohort with hospital records seems to be a substantial limiting factor for future scaling of the project.

***Reviewer’s comment*: In general, the conclusion of the abstract should be more specific and the discussion of the paper could be strengthened by including more detail about the cultural context of this study. Given that the pilot generated distribution of the demographics and risk factors, are there any potential exposures that may be particularly hypothesis generating in this context?**

Authors’ response: The conclusion of the Abstract has been revised to be more specific (see below):

> “Overall, 20 years after the original Community Intervention Program the pilot study achieved high levels of follow-up and family member interview participation, and identified substantial numbers of pediatric malignancies during 1994-2013 in catchment area hospitals. Next steps and strategies for overcoming limitations are described.”

We do not understand what the reviewer means by cultural context of the study, and thus we are unable to respond to this point. If the full-scale prospective follow-up study is funded, there are many potentially interesting 'exposures' that could be explored in relation to pediatric malignancy risk based on the substantial maternal sociodemographic, lifestyle, and reproductive characteristics, medical conditions and treatments, occupational and other factors during the periconceptional and prenatal periods, and on offspring characteristics during the early neonatal period as obtained from maternal interviews and medical records during the original 1993-1995 Community Intervention Program investigation (see Introduction, para 2, last sentence). A description of the extensive data collected is beyond the scope of the current paper, but the data collection instruments, field and coding manuals are available from the corresponding author.

Reviewer's comment: Page 7, line 12-17: The second sentence of the introduction is unclear, please provide more detail for a reader that is not familiar with the biological evidence.
Authors' response: Support for the *in-utero* origin of pediatric leukemia, mostly acute lymphoblastic leukemia, has been provided by observations on twins showing that the chromosomal rearrangements seen in leukemia cases could occur *in-utero*, coupled with work demonstrating that chromosomal translocations such as *ETV6-RUNX1* are present in the blood spot cards of children who later developed *ETV6-RUNX1* positive ALL (reviewed in Greaves and Wiemels Nat Rev Cancer 2003;3:639-49; see Alpar D et al Leukemia 2015;29:839-46 and Bateman CM et al Leukemia 2015;29:58-65). We are reluctant to expand beyond one sentence since this could mislead readers into thinking that the current paper will focus on this topic. To clarify, we have revised the sentence as follows:

> "Support for the *in-utero* origin of pediatric leukemia, mostly acute lymphoblastic leukemia, has been provided by observations on twins showing that the chromosomal rearrangements seen in leukemia cases could occur *in-utero*, coupled with work demonstrating that certain chromosomal translocations are present in the blood spot cards of children who later developed childhood acute lymphoblastic leukemia with the same translocations."

Reviewer's comment: Page 9, line 22: Were there many infants without known sex? If so, could this be a source of bias?
Authors' response: In the original full-scale Chinese Community Intervention Program population the sex of 570 infants was unknown and 70 were ambiguous; thus 640/247,831 or 0.26% were of unknown sex. Given this very low percent, unknown sex is unlikely to be a source of bias. In the pilot study, sex was known for all offspring.

Reviewer's comment: Page 11 line 54- Page 12, line 6: The inability to link the cohort participants with the registries seems a limiting factor for scaling the project, more detail as to why this was not possible is warranted.
Authors' response: We agree and have underscored this limitation more clearly in the Limitations section of the Discussion:

> "The ideal approach for identifying incident pediatric cancers in the Chinese Children and Families Cohort Study would have been linkage with population-based cancer registries, but there have been no long-standing population-based cancer registries of high quality in the geographic regions where the original Chinese Community Intervention Program was carried out."

Reviewer's comment: Page 12, lines 38-48. Subject involvement section seems unnecessary especially since it does not provide references. Was there any consultation with the subjects themselves regarding study design? Such qualitative data may help in future scaling of the cohort.
Authors' response: We prefer to leave as is the description of the information provided back to subjects since this was how subjects were involved in the pilot study. The provision of some results was very well received by the pilot study family members. We did not consult with the subjects about the study design.

Reviewer's comment: Page 13, line 17: What exclusions are you referring to?
Authors' response: We have clarified the description of 'exclusions' and modified the sentence as follows:

"Among the 469 families (84% of the 560 selected) targeted for interviews (after excluding those who could not be found, had moved away, or had refused to be contacted) the interview participation was very high, e.g., 98% of mothers, 95% of fathers, and 99% of offspring."

Reviewer's comment: Page 18, line 34-35: There seems to be a typo here.
Authors' response: Thank you for pointing out the typo. We have modified the sentence as follows:
"The two groups were similar in birth year, age at interview, and sex distribution, but somewhat more urban than rural offspring reported living with their parents."

Reviewer's comment: Page 20, line 30: 23% as this is not a small percentage. Please remove the qualifier "only"
Authors' response: We have deleted the modifier and changed the wording to:
"A broad range of incident pediatric malignancies and related disorders were identified in the 8 hospitals; of the total cancers, 77% were designated by type and 23% were unspecified."

Reviewer's comment: Page 25, lines 5-22. The proposed data sharing seems to be a conservative approach, especially since much of externally funded NIH research must have more open access data sharing policies than what is proposed here.
Authors' response: We note that qualified researchers can contact the Steering Committee of the Chinese Children and Families Cohort Study (to which the corresponding author will direct any requests) to seek collaboration with study investigators in research projects that would use the study data. This data sharing policy was developed after extensive and lengthy discussions with the three collaborating organizations, namely the Chinese Center for Disease Control, the U.S. Center for Disease Control, and the U.S. National Cancer Institute. The study protocol, data collection instruments, field and coding manuals are all available from the corresponding author. As indicated in the description of the funding, the study was funded by the three organizations, not just the NIH.

REVIEWER: 2
Reviewer Name: Denhard de Smit

Institution and Country: Clinical Genetics and APH Research Institute, VU University Medical Center , Amsterdam, The Netherlands

Please state any competing interests or state 'None declared': None

Please leave your comments for the authors below Dear authors,

Reviewer's comment: Have you checked whether the use of Chi-square is applicable in all (sub)tables? The subtables 3 -Alcohol use and 4 - Living arrangements/ Current activity/ Smoking exposure/ Alcohol use, all have one or more cells with an expected observation under H0 of less than 5 which makes the Chi-square unreliable. If you corrected for this, please mention it explicitly. Furthermore, you have included in the analysis rows for 'Unknown'. These should in my opinion be left out of the Chi-square test. This not a category of the variable that you are testing. Unknown should be treated as missing data.
Authors' response: To address the reviewer's concern about the use of the Chi-square given the small numbers for some variables and the relatively small size of our pilot study, we have re-calculated urban versus rural geographic differences using Fisher's Exact test and provide p-values for Fisher's Exact in all tables. We did not include the unknowns in calculating the original Chi square nor did we include the unknowns in calculating Fisher's Exact p-values; we include the unknown numbers in the tables for completeness and treated the unknowns as missing data.

VERSION 2 – REVIEW

REVIEWER	Lauren Houghton Columbia University Mailman School of Public Health
REVIEW RETURNED	23-May-2018

GENERAL COMMENTS	Thank you for revising and clarifying the manuscript. I understand that community-based participatory research was not part of the pilot

	study, but it seems that if the study is to be scaled up, incorporating participant input and feedback could strengthen the study both its design and the results--a recommendation for the future.

REVIEWER	D.J. de Smit Section for Community Genetics, Department of Clinical Genetics APH research institute, VU medical centre Amsterdam, NL
REVIEW RETURNED	30-May-2018

GENERAL COMMENTS	The revisions are adequate. This study is a good preparation for the attempt to learn more about the possible late onset effects of periconceptional folic acid supplementation. I wish you good luck with the next steps.