PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Preexisting type 2 diabetes and risk of lung cancer: a report from two prospective cohort studies of 133,024 Chinese adults in urban Shanghai
AUTHORS	Yang, Wan-Shui; Yang, Yang; Yang, Gong; Chow, Wong-Ho; Li, Honglan; Gao, Yu-Tang; Ji, Butian; Rothman, Nat; Shu, Xiao-Ou; Zheng, Wei; Xiang, Yong-bing

VERSION 1 - REVIEW

REVIEWER	Tim Pickles South East Wales Trials Unit, Cardiff University, Wales
REVIEW RETURNED	28-Feb-2014

GENERAL COMMENTS	Quick summary: This paper is very well written and is also very sound from a statistical viewpoint, to the extent that I have just a small number of comments, corrections and questions. I will highlight these section by section. Abstract: • This is not meta-analysis, so please remove any mentions of this • Could you add when the recruitment started? • Later reading shows that 2010 is the last year in which data was collected, but this fact is unclear here Methods: • What are these participation rates of 92.7% and 74.1%? What population is the denominator? • If the first follow up for the males was in 2004-2008, this implies that there was a second follow up. If so, when was it? • Cannot use numerical operators ≤ or > on text. Please find another way to describe these categories • The numbers describing income are in the footnotes of Table 1, so put a reference to this • Is physical activity calculated using IPAQ or similar? If so, please reference Results: • 'After median follow-up of 6.3 years for SMHS and 12.2 years for SWHS' should be in methods • (Assuming that higher HR is more likely incidence of lung cancer)Whilst not significant, the direction of these results

	suggests that those with diabetes are less likely to develop lung cancer, which is the opposite argument to that made in the introduction. Is this worth mentioning? • This is not meta-analysis, so please remove any mentions of this Table 1: • Cannot use numerical operators ≤ or > on text. Please find another way to describe these categories • Use mean (SD) rather than mean±SD • Remove all p-values – there is no hypothesis test here • Put 'Mean' next too all things you are calculating means for, else remove it from 'Mean age at baseline' • Add extra row above '<18.5 (%)' called 'BMI (kg/m^2) (%)', then remove '(%)' from each relevant row • Ambiguous 3000 and 500 Yuan boundaries for income Table 3 • Ambiguous 0.924 boundary for WHR in men • Ambiguous 0.831 boundary for WHR in women

REVIEWER	Rowan Chlebowski LABIOMED, USA
REVIEW RETURNED	30-Apr-2014

GENERAL COMMENTS	The investigators conducted two perspective population based cohort studies in China and examined the association between self-reported diabetes and lung cancer incidence. The lung cancer cases were identified through Cancer Registry linkages, the authors concluded that "there is little evidence that preexisting T2D may influence the incidence of lung cancer." The information provided is of interest as the relationship between diabetes and a number of cancers is of increasing concern. However, there are limitations to this analysis not identified by the authors. The diabetes assessment procedure is described in an abbreviated fashion making it not possible to determine exactly how it was performed. It seems to be based completely on self-report. However, the subject is said to have had to meet at least one of the following self-reported items, blood concentrations > 7 mm/L on two separate occasions, plasma glucose concentration > 11.1 at 2 hours or 75 g tolerance desk and use of insulin or other glycemic agent. It seems the first two would be unlikely to be recalled with the medical cut points by many lay cohort participants. It is unclear whether reports of use of hypoglycemic agents was confirmed by prescription review or other methods. The authors must provide more details how individuals with diabetes were identified. If it was only through self-report and the authors can provide no subset or prior publication of the validity of this approach the findings become suspect. In any event, part of the interest in diabetes and cancer relates to, as the authors point out, the potential role of medications used to treat diabetes which either potentially reduce the risk (metformin) or other

	agents which may increase risk. So analysis based on self-report of diabetes without medical records information provides information which is of less use to advancing questions of current scientific importance. Perhaps a topic which should be addressed by the authors is that in their population diabetes is of relatively low frequency so they are examining quite a different population than available for most Western databases. This point should be made. On the hand, this raises a question since diabetes was assessed at baseline yet the authors indicate that diabetes incidence is increasing and rapidly changing in their population. Thus, many individuals could have switched from no diabetes to diabetes state during the course of follow-up, which would again reduce their chances of showing a difference. Comments to the Editor: The authors provide information on the relationship between diabetes and lung cancer in an understudy population which is of some interest. They need to clarify how diabetes status was determined. If it is only self-report and no prior study or subgroup analysis in their current cohorts can support the reliability of self-reported diabetes by first hand medical record review, the findings have much more suspect. Other limitations have been pointed out to the authors. The report is of some value but does less to address questions of major importance regarding this issue as no medication use is provided and no analytic capacity to allow individual to switch from a no diabetes stage to diabetes stage during course of follow-up represent limitations. Consideration for acceptance could be made with modest enthusiasm.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1:
Reviewer Name Tim Pickles
Institution and Country South East Wales Trials Unit, Cardiff University, Wales
Please state any competing interests or state 'None declared': None declared
Quick summary:
This paper is very well written and is also very sound from a statistical viewpoint, to the extent that I have just a small number of comments, corrections and questions. I will highlight these section by section.
Abstract:
• This is not meta-analysis, so please remove any mentions of this
We apologize for the error and revised accordingly. Thank you.
• Could you add when the recruitment started?
For the SWHS, the recruitment for female residents of Shanghai aged 40-70 years old started in 1996 and was completed in 2000, with an overall participation rate of 92% (74941/81170). For the SMHS, the recruitment for men aged 40-74 years old with no history of cancer in Shanghai started in April 2002 and was completed in June 2006, with an overall participation rate of 74% (61491/83125). We clarified this point in the sections of Abstract and Methods.
• Later reading shows that 2010 is the last year in which data was collected, but this fact is unclear here

We agree and clarified this point in the section of Methods.

Methods:
• What are these participation rates of 92.7% and 74.1%? What population is the denominator?
For the SWHS, of the 81,170 eligible women (i.e. aged 40-70y) who lived in the study communities in Shanghai, during the time period of the baseline survey (1996-2000), 75,221 participated in the study, with an overall participation rate of 92.7% (75,221/81,170). Of the 75,221 study participants, 280 women who did not meet the age eligibility requirements for the study were excluded, resulting in a cohort of 74,941 subjects.
For the SMHS, of the 83,125 eligible men (i.e. aged 40-74y) who lived in the study communities in Shanghai, during the time period of the baseline survey (2002-2006), 61,582 participated in the study, with an overall participation rate of 74.1% (61,582/83,125). Of the 61,582 study participants, 91 men who were found as cancer cases before baseline survey during the follow-up were excluded, resulting in a cohort of 61,491 subjects.
• If the first follow up for the males was in 2004-2008, this implies that there was a second follow up. If so, when was it?
Yes, for the SMHS, the 2nd follow up started in June, 2008 and was completed in May, 2011. But the relevant data is not available to date (the date for the present analysis). Likewise, for the SWHS, although the 4th follow up was launched in April 2007 and completed in October 2010, the pertinent data is unavailable. We therefore used the updated information of 1st follow up for the SMHS, and updated information of 1st, 2nd, and 3rd follow ups for the SWHS.
• Cannot use numerical operators ≤ or > on text. Please find another way to describe these categories
We agree and revised accordingly. Thanks.
• The numbers describing income are in the footnotes of Table 1, so put a reference to this
We agree and revised accordingly. Thanks.
• Is physical activity calculated using IPAQ or similar? If so, please reference
The physical activity was calculated according to Compendium of Physical Activities by Ainsworth et al. (Med Sci Sports Exerc, 2000,32:S498-S504; Med Sci Sports Exerc, 1993,25:71-80.). We clarified this point in the section of Methods.
Results:
• 'After median follow-up of 6.3 years for SMHS and 12.2 years for SWHS' should be in methods
We agree and revised accordingly. Thanks.
• (Assuming that higher HR is more likely incidence of lung cancer)Whilst not significant, the direction of these results suggests that those with diabetes are less likely to develop lung cancer, which is the opposite argument to that made in the introduction. Is this worth mentioning?
• This is not meta-analysis, so please remove any mentions of this
We apologize for the error and revised accordingly. Thanks.
Table 1:
• Cannot use numerical operators ≤ or > on text. Please find another way to describe these categories
We agree and revised accordingly. Thanks.
• Use mean (SD) rather than mean±SD
We agree and revised accordingly. Thanks.
• Remove all p-values – there is no hypothesis test here
We agree and revised accordingly. Thanks.
• Put 'Mean' next too all things you are calculating means for, else remove it from 'Mean age at baseline'
We agree and revised accordingly. Thanks.
• Add extra row above '<18.5 (%)' called 'BMI (kg/m^2) (%)', then remove '(%)' from each relevant row
We agree and revised accordingly. Thanks.
• Ambiguous 3000 and 5000 Yuan boundaries for income
In the baseline questionnaires, participants were asked to evaluate 4 different levels of family income per year for herself, i.e. 1) less than 10,000 yuan; 2)10,000-20,000 yuan; 3) 20,000-30,000 yuan and 4) more than 30,000 yuan. Thus we used 3 cut-points (i.e. 10,000, 20,000, 30,000) to split income data into 4 groups for women (see footnote in Table 1).
In the baseline questionnaires in the SMHS, participants were asked to evaluate 7 different levels of family income per month for himself, i.e. 1) less than 500 yuan; 2)500-1000 yuan; 3)1000-2000 yuan; 4)2000-3000 yuan; 5)3000-4000 yuan; 5)4000-5000 yuan; and 7) more than 5000 yuan. Due to limited number of subjects among 1st, 2nd, 3rd income levels, we used 3 cut-points (i.e. 1000, 3000,

5000) to split income data into 4 groups for men (see footnote in Table 1).
Table 3
• Ambiguous 0.924 boundary for WHR in men; Ambiguous 0.831 boundary for WHR in women
We used two tertiles to split WHR data into three groups. We clarified this point in the footnote (Table 3).

Reviewer #2:
Reviewer Name Rowan Chlebowski
Institution and Country LABIOMED, USA
Please state any competing interests or state 'None declared': None declared
The investigators conducted two perspective population based cohort studies in China and examined the association between self-reported diabetes and lung cancer incidence. The lung cancer cases were identified through Cancer Registry linkages, the authors concluded that "there is little evidence that preexisting T2D may influence the incidence of lung cancer."
The information provided is of interest as the relationship between diabetes and a number of cancers is of increasing concern. However, there are limitations to this analysis not identified by the authors.
The diabetes assessment procedure is described in an abbreviated fashion making it not possible to determine exactly how it was performed. It seems to be based completely on self-report.
In our analysis, diabetes cases were identified based completely on the self-reported data. Self-reported DM was recorded on the baseline questionnaires (2002–2006 for the SMHS and 1996–2000 for the SWHS), and updated in each of the subsequent follow-up questionnaires (2004–2008 for the SMHS, and 2000–2002, 2002–2004 and 2004–2007 for the SWHS). Participants were asked whether they had ever been diagnosed with DM by a physician (yes/no) and if yes, the age at diagnosis was recorded. From the beginning with the 2004–2008 follow-up questionnaires for men and 2000–2002 follow-up questionnaires for women, and for all subsequent surveys, the question was modified, and participants were additionally asked in what year and month and in which hospital their DM had been diagnosed since the most recent survey.
In present study, a case of T2D was considered to be confirmed if the participant reported having been diagnosed with type 2 diabetes and met at least one of the following criteria: (i) fasting plasma glucose concentration >7 mmol/l on two separate occasions, (ii) plasma glucose concentration >11.1 mmol/l at 2 h for a 75 g oral glucose tolerance test and (iii) the use of insulin or other hypoglycemic agents.
We added these points in the section of Method for clarity.
However, the subject is said to have had to meet at least one of the following self-reported items, blood concentrations > 7 mm/L on two separate occasions, plasma glucose concentration > 11.1 at 2 hours or 75 g tolerance desk and use of insulin or other glycemic agent. It seems the first two would be unlikely to be recalled with the medical cut points by many lay cohort participants.
We agree that the first two items would be unlikely to be recalled with the medical cut points by many subjects in our cohorts. However, these two items is not essential condition for diabetes ascertainments in our analysis. In addition, we conducted a sensitivity analysis and found that the results were similar when diabetes cases were indentified only relying on the question that whether the participants had ever been diagnosed with diabetes by their physicians (data not shown).
It is unclear whether reports of use of hypoglycemic agents was confirmed by prescription review or other methods. The authors must provide more details how individuals with diabetes were identified. If it was only through self-report and the authors can provide no subset or prior publication of the validity of this approach the findings become suspect.
The question of validity of self-reported exposure data obtained through questionnaires is indeed very important. We realized that one of major limitations to this study is that the data on type 2 diabetes as well as the history of hypoglycemic agents use is self-reported and its validation is consequently doubted. However, previous validation studies (Eur J Public Health 2007;17:199-205, Public Health 2000:114:137-142, Am J Prev Med 2000;18:215-218, Clin Trials 2008;5:240-247) showed that a self-reported history of diabetes may be reasonably accurate and could provide a useful estimate for broad measures of diabetes in the large-scale observational study.
In addition, in 2009, we selected an area from one of our cohort study communities in Shanghai, in which 2000 participants were asked again to donate fasting blood sample with aims of studying some biomarkers in blood. The data was now using to examine the validity of exposure data of type 2 diabetes in our cohorts. Self-report diabetes was confirmed by measuring fasting plasma glucose concentration, or by reviewing their medical records if they did not provide the blood sample. The self-reported diabetes was in good agreement with our laboratory data (WHO criteria) and data extracted

from medical records, with an overall agreement rate of 93.3%. Among 432 subjects of 2000 cohort participants, 45 patients told us they are type 2 diabetes (self-reports of diabetes) from the questionnaires in this study. In which, 31 provided the fasting blood sample and were all confirmed by our experimental data; 11 were confirmed by review of their medical or treatment records; the remaining 3 subjects can not provide available information for our review; so the agreement is 93.3%. The data from all 2000 participants was still processing. We will publish the results for the diagnosis validation of type 2 diabetes soon.

We clarified these points in the section of Methods and Discussion. So we add a sentence “a validation study showed that the self-reported diabetes was in good agreement with the measurement of fasting plasma glucose concentration and medical treatment records in our cohorts (data was not shown).” in the part of diabetes diagnosis of Methods and Discussion.

In any event, part of the interest in diabetes and cancer relates to, as the authors point out, the potential role of medications used to treat diabetes which either potentially reduce the risk (metformin) or other agents which may increase risk. So analysis based on self-report of diabetes without medical records information provides information which is of less use to advancing questions of current scientific importance.

We agree. In order to address the issue raised by the reviewer, a separate analysis that excluded treated diabetes was conducted and yielded similar results with overall analysis. However, the information for hypoglycemic drug use is also derived from self-reported data, thus the misclassification bias cannot be ruled out. We added the relevant contents in the sections of Methods, Results, and Discussion.

Perhaps a topic which should be addressed by the authors is that in their population diabetes is of relatively low frequency so they are examining quite a different population than available for most Western databases. This point should be made.

The prevalence of diabetes has increased substantially in China, with the age-standardized rates from 2.4% in 1994 (Diabetes Care 1997;20:1664-9) to 9.7% in 2007 to 2008 (N Engl J Med 2010;362:1090-101), which may parallel a marked lifestyle transition (Diabetes Care 2011; 34: 1249–57). Unlike the steady transition in most Western developed countries, these changes have occurred within a very short period in China. However, no prospective study, to date, has evaluated the effect of diabetes on the lung cancer risk to our best knowledge. We clarified this point in the section of Introduction.

On the hand, this raises a question since diabetes was assessed at baseline yet the authors indicate that diabetes incidence is increasing and rapidly changing in their population. Thus, many individuals could have switched from no diabetes to diabetes state during the course of follow-up, which would again reduce their chances of showing a difference.

The reviewer is right that only using a single measurement of diabetes at baseline may have resulted in some underestimation of the true associations since the diabetes newly identified during the follow-up periods were neglected.

However, in our analysis (including the previous version of our manuscript), type 2 diabetes (yes/no) was modeled as a time-varying exposure through the use of repeated measures of diabetes status, meaning that information on type 2 diabetes reported in questionnaire n, was used to prospectively categorize participants for the periods between completion of questionnaires n and n + 1, and the risk person-years was allocated to the corresponding groups. In the SWHS, for example, a participant who first reported having T2D in 2001 would contribute person-time to the non-diabetes group from 1996 to 2000, whereas from 2001 onward, this participant would contribute person-time to the diabetes group.

Therefore, the diabetes newly identified during the follow-up periods has been taken into account in our study.