PEER REVIEW HISTORY

This paper was submitted to the JECH 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.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Fruit and vegetable intake and risk of type 2 diabetes mellitus: meta-analysis of prospective cohort studies
AUTHORS	Tang, Zhenyu; Li, Min; Fan, Yingli; Zhang, Xiaowei; Hou, Wenshang

VERSION 1 - REVIEW

REVIEWER	Patrice Carter Leicester Diabetes Centre, The University of Leicester, Leicester, UK
REVIEW RETURNED	21-May-2014

GENERAL COMMENTS	This meta-analysis of prospective cohort studies provides an update to previous reviews conducted on the same topic (Carter et al 2010, Cooper et al 2012). The examination of potential dose response between intake and risk of T2DM provides interesting data. However, the findings do not show anything new; to provide more novel and interesting data would it be possible to examine categories of different fruit and vegetables, as done with green leafy vegetables? Major points to consider: Methods. The paper appears to include the same population, specifically The Nurses' Health Study on multiple occasions (Colditz, Bazzanno and Muraki)? Please clarify if this is the case or not. A previous systematic review (Carter et al 2010) does not appear to have been referenced adequately in certain parts of this review. For example in the Validity Assessment, the tool used to score papers is clearly the same as used by Carter et al, which the authors themselves created, this needs acknowledging. Minor points to consider: Introduction. On page 3, line 30, it states that previous reviews were restricted by language, this is not the case, please amend. Discussion. The discussion does not provide any new insights to

	mechanisms or interpretation than that discussed in previous reviews. I would highly recommend a statistician reviews the methods section.

REVIEWER	Peilin Shi Harvard School of Public Health, USA
REVIEW RETURNED	21-Jul-2014

GENERAL COMMENTS	Major Comments: 1. This article involved the Nurses' Health Studies (NHS) established in 1976 with funding from National Institutes of Health. The authors in their study included three articles involved NHS, Colditz et al 1992, Bazzano et al 2008, and Muraki et al 2013. It is not clear if same participants counted more than once. Authors should read the original articles to check them carefully. 2. The PRISMA Guideline mainly focused on randomized trials, although PRISMA can also be used as a basis for reporting systematic reviews of other types of research, particularly evaluations of interventions. The MOOSE (Meta-analysis of Observational Studies in Epidemiology) statement is intended for the reporting of meta-analyses of observational studies. Thus the authors should use The MOOSE Guideline Check List other than The PRISMA Guideline. Minor Comments: 1. Last line on page 2: GLV should be spelled out the first time it is used; 2. The authors should not use same tiles for the different figures, for example, Fig 2 vs. Supplemental fig A, Fig 4 vs. Supplemental fig B; 3. If random effects models were used in all analyses (irrespective of I square or P-value) then it should be specified why. 4. The title and legend in some figure are not clear to indicate the study. For example, in Figure 2, Cooper (2012) and Cooper (201212) should be, something like, Cooper (study a, 2012), Cooper (study b, 2012) ; Muraki (2013), Muraki (2013), Muraki (2013), should be Muraki (Cohort a: 2013), Muraki (Cohort b: 2013), Muraki (Cohort c: 2013),…; 5. The title of Figs 2, 4, 5 should include the information "for highest versus lowest intake of …"; 6. As compared with the fixed effect model, the weights assigned under random effects are more balanced. Large studies are less likely to dominate the analysis and small studies are less likely to be trivialized. "Weight %" should be added in all forest plots for the random effect models; 7. The line 5 from bottom of page 5, add "Cooper et al (study a:2012) and Cooper et al (study b:2012) ";

	8. Put "The analysis included 15 cohorts among the eleven articles, where Ford et al and Kurotani et al study examined male and female separately, Cooper et al have two studies in 2003 and Muraki et al report included data from three independent cohorts." in footnote of Table A; 9. Page 5, authors indicated that "We considered low, moderate, and high degrees of heterogeneity to be I2 values of 25%, 50%, and 75%, respectively." It should be specified why. 10. The authors said, on page 9, that "Compared with the previous meta-analyses, our study has several strengths. The present meta-analysis included 2.6-times more participants and 2.8-times more T2D cases, which significantly increased the statistical power to detect potential associations." I am not sure if the participants and T2D cases repeatedly counted because several studies from same cohort (NHS). 11. The authors should use funnel plot to check for the existence of publication bias in this systematic review and meta-analyse. I would suggest 'major/substantial revision'. It might take at least one round of major revision before it might be acceptable for publication.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1 (Reviewer Name Patrice Carter):

Q: This meta-analysis of prospective cohort studies provides an update to previous reviews conducted on the same topic (Carter et al 2010, Cooper et al 2012). The examination of potential dose response between intake and risk of T2DM provides interesting data. However, the findings do not show anything new; to provide more novel and interesting data.

R: Thank you for your careful review and very constructive suggestion. In fact, compared with the previous meta-analyses (Carter et al 2010, Cooper et al 2012), our study has several strengths. We included some new prospective studies (Kurotani et al 2013, Cooper et al 2012, and Muraki et al 2013). Therefore, this paper included 2.6-times more participants and 2.8-times more T2DM cases, which significantly increased the statistical power to detect potential associations. In addition, we also investigated a dose-response relation between fruit, vegetables, fruit and vegetables combined consumption and risk of T2DM. Therefore, the results should be more novel and interesting.

Q: Would it be possible to examine categories of different fruit and vegetables, as done with green leafy vegetables?

R: It is possible that different types of fruit and vegetables might have different effects on T2DM risk. In fact, we tried to further investigate the association between categories of different fruit and vegetables and risk of T2DM. However, due to the limited number of literature, we failed it. Another possible explanation for the differences between the studies might be the classification of food groups. For example, GLVs'criteria was inconsistent: some studies included spinach and lettuce; some included spinach and greens; others did not provide specific description. If they were included with an uniform definition of each groups, the effects on T2DM risk might differ. Future studies are needed to be more specific about the role of different categories of fruit and vegetables and risk of

T2DM.

Major points
Q: Methods. The paper appears to include the same population, specifically The Nurses' Health Study on multiple occasions (Colditz, Bazzano and Muraki)? Please clarify if this is the case or not.
R: We are very sorry that we did not explain this critical point clearly. Among The Nurses' Health Studies, the same population was reported for the three studies (Colditz et al 1992, Bazzano et al 2008, Fung et al, 2004 (one of three prospective longitudinal cohort studies including Muraki et al's paper). As described in the Study selection. If data were reported more than once, we included the study with the longest follow-up time. Therefore, we included the study by Bazzano et al, which followed for 18 years. Accordingly, we have made corrections in the revised manuscript.
Q: A previous systematic review (Carter et al 2010) does not appear to have been referenced adequately in certain parts of this review. For example in the Validity Assessment, the tool used to score papers is clearly the same as used by Carter et al, which the authors themselves created, this needs acknowledging.
R: We are very sorry that we did not discriminate this original reference clearly. In the Validity assessment, the statement has been edited as you suggested.

Minor points
Q: Introduction. On page 3, line 30, it states that previous reviews were restricted by language, this is not the case, please amend.
R: We are very sorry that we did not explain the difference accurately. The sentence has been revised as you suggested.

Q: Discussion. The discussion does not provide any new insights to mechanisms or interpretation than that discussed in previous reviews.
R: Thank you very much for your suggestion. In the Discussion, the paragraph has been revised you suggested. We hope it meets your requirement.

Q: I would highly recommend a statistician reviews the methods section.
R: Thank you very much for your suggestion. Accordingly, we have invited an expert statistical scientist to review the manuscript. We believe that the manuscript has been significantly improved and hope it meets your requirement.

Reviewer #2 (Reviewer Name Peilin Shi):
Major points
Q1: This article involved the Nurses' Health Studies (NHS) established in 1976 with funding from National Institutes of Health. The authors in their study included three articles involved NHS, Colditz et al 1992, Bazzano et al 2008, and Muraki et al 2013. It is not clear if same participants counted more than once. Authors should read the original articles to check them carefully.
R1: We are very sorry that we did not explain this critical point clearly. Among The Nurses' Health Studies, the same population was reported for the three studies (Colditz et al 1992, Bazzano et al 2008, Fung et al, 2004 (one of three prospective longitudinal cohort studies including Muraki et al's paper). As described in the Study selection. If data were reported more than once, we included the study with the longest follow-up time. Therefore, we included the study by Bazzano et al, which

followed for 18 years. Accordingly, we have made corrections in the revised manuscript.

Q2: The PRISMA Guideline mainly focused on randomized trials, although PRISMA can also be used as a basis for reporting systematic reviews of other types of research, particularly evaluations of interventions. The MOOSE (Meta-analysis of Observational Studies in Epidemiology) statement is intended for the reporting of meta-analyses of observational studies. Thus the authors should use The MOOSE Guideline Check List other than The PRISMA Guideline.
R2: Thank you very much for your suggestion. In the Supplementary file, the MOOSE Guideline Check List has been edited you suggested.

Minor points
Q1: Last line on page 2: GLV should be spelled out the first time it is used;
R1: Thank you for your suggestion. We have made correction accordingly in page 2, last line, and the name “GLV” has corrected as “green leafy vegetables”.

Q2: The authors should not use same tiles for the different figures, for example, Fig 2 vs. Supplemental fig A, Fig 4 vs. Supplemental fig B;
R2: In Figs and Supplemental figs, the titles have been revised as you suggested.

Q3: If random effects models were used in all analyses (irrespective of I square or P-value) then it should be specified why.
R3: Taking into account the within-study and between-study variability, we pooled relative risks and their standard errors with random effects model (DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-88.).

Q4: The title and legend in some figure are not clear to indicate the study. For example, in Figure 2, Cooper (2012) and Cooper (201212) should be, something like, Cooper (study a, 2012), Cooper (study b, 2012) ; Muraki (2013), Muraki (2013), Muraki (2013), should be Muraki (Cohort a: 2013), Muraki (Cohort b: 2013), Muraki (Cohort c: 2013),…;
R4: In Figures, the title and legend have been revised as you suggested.

Q5: The title of Figs 2, 4, 5 should include the information “for highest versus lowest intake of …”;
R5: In Figs 2, 4, 5, the titles have been revised as you suggested.

Q6: As compared with the fixed effect model, the weights assigned under random effects are more balanced. Large studies are less likely to dominate the analysis and small studies are less likely to be trivialized. “Weight %” should be added in all forest plots for the random effect models;
R6: In all forest plots, the “Weight%” has been added as you suggested.

Q7: The line 5 from bottom of page 5, add “Cooper et al (study a:2012) and Cooper et al (study b:2012)”;
R7: Thank you for your suggestion. The statement has been added as you suggested. Additionally, we have made corrections accordingly in the main text, table and figures, and the name “Copper et al” has corrected as “Cooper et al (study a:2012)” and “Cooper et al (study b:2012)”, respectively.

Q8: Put “The analysis included 15 cohorts among the eleven articles, where Ford et al and Kurotani et al study examined male and female separately, Cooper et al have two studies in 2003 and Muraki et al report included data from three independent cohorts.” in footnote of Table A;
R8: Thank you for your careful review and very constructive suggestion. In Table A, the sentence in footnote has been added as you suggested.

Q9: Page 5, authors indicated that “We considered low, moderate, and high degrees of heterogeneity to be I2 values of 25%, 50%, and 75%, respectively.” It should be specified why.
R9: Thank you for your careful review. A naïve categorisation of values for I2 would not be appropriate for all circumstances, although we would tentatively assign adjectives of low, moderate, and high to I2 values of 25%, 50%, and 75%. Quantification of heterogeneity is only one component of a wider investigation of variability across studies, the most important being diversity in clinical and methodological aspects. Meta-analysts must also consider the clinical implications of the observed degree of inconsistency across studies. For example, interpretation of a given degree of heterogeneity (25%, 50%, and 75%) across several studies will differ according to whether the estimates show the same direction of effect. (Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60; Higgins JP. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. Int J Epidemiol. 2008 37:1158-60.)

Q10: The authors said, on page 9, that “Compared with the previous meta-analyses, our study has several strengths. The present meta-analysis included 2.6-times more participants and 2.8-times more T2D cases, which significantly increased the statistical power to detect potential associations.” I am not sure if the participants and T2D cases repeatedly counted because several studies from same cohort (NHS).
R10: We are very sorry that we did not explain this critical point clearly. Among The Nurses’ Health Studies, the same population was reported for the three studies (Colditz et al 1992, Bazzano et al 2008, Fung et al, 2004 (one of three prospective longitudinal cohort studies including Muraki et al’s paper). As described in the Study selection. If data were reported more than once, we included the study with the longest follow-up time. Therefore, we included the study by Bazzano et al, which followed for 18 years. Accordingly, all amendments have been highlighted in red in the page 9.

Q11: The authors should use funnel plot to check for the existence of publication bias in this systematic review and meta-analyses.
R11: Thank you for your careful review and very constructive suggestion. In fact, the possibility of publication bias was also evaluated using visual inspection of a funnel plot except the Begg rank correlation test and Egger’s regression test (figures not shown).