PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Collaborative care for co-morbid depression and diabetes: a systematic review and meta-analysis
AUTHORS	Atlantis, Evan; Fahey, Paul; Foster, Jann

VERSION 1 - REVIEW

REVIEWER	Aimin Guo School of General Practice and Continuing Education, Capital Medical University, Beijing, China
REVIEW RETURNED	29-Jan-2014

GENERAL COMMENTS	This systematic review and meta-analysis covers an important topic. It evaluates the effectiveness of collaborative care for patients suffering from both depression and diabetes in primary care setting. The study concludes that collaborative care significantly improves both depression and glycaemia outcomes. The paper is well organized and well written. Only minor revisions are required. Minor revisions: 1. Statistical methods section (page 9 line 11-12): The authors states “P-value < 0.05 was considered statistically significant…”. However, the authors did not provide P value for the effect size of all the main results. Consider for example, for the effect of collaborative care on HbA1c, the authors only provides pooled WMD and 95% CI (“-0.33% [95% CI: -0.66% to -0.00%]”) (page 12 line 3). It would be better if P value is also provided. (The upper level of 95% CI is very close to zero [-0.00%], which indicates the P value is very close to 0.05.) 2. Quantitative data synthesis section (page 12 line 7-8): Sensitivity analyses show insignificant results for collaborative care on HbA1c. The authors may want to discuss more about it in the discussion. 3. The inclusion and exclusion criteria should be made clearer in the methods section. Two studies included patients without diabetes. In Morgan’s study (doi:10.1136/bmjopen-2012-002171), 45.3% patients in the intervention group are not diabetes patients. In Katon’s study (DOI: 10.1056/NEJMoa1003955), 11% patients in the intervention group are not diabetes patients. It would be better if the authors give a detailed inclusion criteria for patients with co-morbid depression and diabetes (for example, studies are included if more than 80% patients in the study have co-mobid depression and diabetes…). 4. Flowchart section (Page 30): The authors searched seven databases such like PubMed, Scopus, Cochrane Library, etc. The authors also searched reference lists of retrieved articles. They may

	want to list how many citations were searched from each of them, rather than just give a total number. 5. Primary outcomes section (Page 8 line 6): The authors may want to provide a reference for the Glass's Delta method.

REVIEWER	Prof James Dunbar Greater Green Triangle University Department of Rural Health Flinders and Deakin Universities Australia Chief Investigator, TrueBlue project mentioned as reference 25
REVIEW RETURNED	30-Jan-2014

GENERAL COMMENTS	This is a useful contribution to the literature on comorbid depression and diabetes. It is the first review looking at the effect of collaborative care on improving measures of depression and glycaemia. Both diabetes and depression are major contributors to the burden of disease. Together they are an important health problem. Abstract I think the second bullet-point might be better to say that the collaborative care model fits better within some health systems than others and that the extent of benefits in randomised trials are proportionate to the extent which the measures of HbA1c and depression are out of the target range in the guidelines. Results It is interesting that the baseline HbA1c's range from 6.9 which is within the guideline target to 9.1 which is well outside. The higher the 'out of range' starting point, the more likelihood of improvement. This should be mentioned in the discussion. Perhaps a bit more could be made of the result of meta-regression model which didn't show depression scores predicting HbA1c values. It seems unlikely that managing depression alone results in the improvement of HbA1c's. TeamCare published by Katon and TrueBlue which is the sister study that I ran both decided that the reason diabetes measures hadn't improved with improvement in depression measures in trials prior to 2007 was that there hadn't been a concentration on the diabetes outcomes. We also agreed to look at managing depression, CVD and diabetes together because the overlap between CVD and diabetes is so large, both in risk factors and overt disease. In the discussion it says that properly integrated lifestyle intervention wasn't any of the RCTs offered in these two studies. This is untrue of TrueBlue and TeamCare. Indeed in your own table you describe how it was conducted. It is interesting that you refer to the prevention of type 2 diabetes. My group ran the Council of Australian Government's National Demonstrator Greater Green Triangle Diabetes Prevention Program 2004-2006 so a great deal of that methodology was incorporated into TrueBlue. Nevertheless you need to be realistic about what can be expected by way of this kind of intervention in day-to-day general practice. I think in your emphasis on diabetes care and lifestyle intervention, you should refer to TeamCare and TrueBlue at least. In limitations you are suggesting that because most of the trials have been done in the US, further research is required in other countries. I disagree having set up in Australia. I also have long experience of

	the NHS in UK and it is not the need for further research. It is the need for health system change to be able to undertake the work. After all, UK has many experts in this area who have had difficulty making it work within their system. In limitations, I think the biggest limitation is that you have only looked at HbA1c. It is CVD which kills people with diabetes and that is the reason that TeamCare and TrueBlue both incorporated it into their collaborative care. If you look up UK PDS and the more recent ACCORD studies, you will see that attention to blood pressure, cholesterol and smoking are at least important as attention to HbA1c. In fact, much of the evidence would point to blood pressure being the most important of all measures. Therefore I think that your conclusion should be that not only should collaborative care focus on the management of depression but it needs to concentrate on better management of diabetes, cardiovascular risk and comorbid vascular disease.

REVIEWER	Marco Menchetti Department of Medical and Surgical Sciences University of Bologna
REVIEW RETURNED	09-Feb-2014

GENERAL COMMENTS	The review does not add much to the existing literature, including 2 meta-analyses with the same studies This is a well written and well conducted systematic review and meta-analysis. The major consideration for the authors is for them to distinguish their review from previous recent meta-analyses, and in particular those cited in their introduction (Van der Feltz-Cornelis et al. 2010; Huang et al., 2013). In addition, there is the recently published Cochrane review of collaborative care that they did not mention in their introduction (Archer et al., 2012). The authors stated that there are newly published RCTs (Bogner et al., 2012; Katon et al., 2010) but these studies were also included in previous reviews. The authors should explain more in depth this overlap and what does their paper add to the literature. Possible inconsistencies among findings found in previous reviews and findings of the present work sholud be addressed.

REVIEWER	Andrew Hinde Department of Social Statistics and Demography University of Southampton SOUTHAMPTON
REVIEW RETURNED	20-Feb-2014

GENERAL COMMENTS	My main concern is about the description of the statistical methods used. I feel they are described to briefly and ambiguously to allow a reader to replicate the study. The specific issues I am concerned about are listed below. (1) On p. 7, ll. 22-23 you say that '[s]tudies were considered "better quality" if they received a score higher than 4, since that meant that they had most of our quality items'. First one might dispute the implication that the six criteria you use are being given equal weight (arguably randomisation is more critical than, say, the balancing of

	drop-out rates). Second, you seem to have included one study with a score of 3.5 (see table on p. 37). (2) On p. 8, l. 6 you state that you computed standardised mean differences using Glass's Delta method. Could you be more specific as to whether you used the standard deviation of the control group (which is what Glass proposed) or a 'Delta-family' method where you might use some combination of the standard deviations of the control and treatment groups. (3) On p. 8, l. 14 you refer to the 'inverse variance weighted mean difference'. What variances were used for the weights? Later you say that you pooled the standardised mean differences (SMDs) from each randomised control trial 'to produce an overall estimate of effect' (ll. 16-18). Again, did you weight these SMDs and, if so, how? I could not follow exactly what you did from the rather terse description given in this paragraph.

VERSION 1 – AUTHOR RESPONSE

Reviewer Name Aimin Guo
Institution and Country School of General Practice and Continuing Education, Capital Medical University, Beijing, China
Please state any competing interests or state ‘None declared’: None declared

This systematic review and meta-analysis covers an important topic. It evaluates the effectiveness of collaborative care for patients suffering from both depression and diabetes in primary care setting. The study concludes that collaborative care significantly improves both depression and glycaemia outcomes. The paper is well organized and well written. Only minor revisions are required.

Minor revisions:
1. Statistical methods section (page 9 line 11-12): The authors states “P-value < 0.05 was considered statistically significant…”. However, the authors did not provide P value for the effect size of all the main results.
>>> We have made the following revision to avoid confusion (page 9): “Effects were considered statistical significant when the associated 95% confidence intervals did not include zero and heterogeneity was considered statistically significant where the associated P-value was less than 0.05.”

Consider for example, for the effect of collaborative care on HbA1c, the authors only provides pooled WMD and 95% CI (“-0.33% [95% CI: -0.66% to -0.00%]”) (page 12 line 3). It would be better if P value is also provided. (The upper level of 95% CI is very close to zero [-0.00%], which indicates the P value is very close to 0.05.)
>>> This comment confirms that the confidence interval conveys the same information as the P-value. (Indeed the –ve sign on the -0.00% confirms that the p-value will be marginally less than 0.05 - not just very close to).

2. Quantitative data synthesis section (page 12 line 7-8): Sensitivity analyses show insignificant results for collaborative care on HbA1c. The authors may want to discuss more about it in the discussion.
>>> The smaller the sample size, the lower the statistical power. As studies are excluded, results will tend to be less significance. The more important aspect of the sensitivity analysis is whether effect size changes with the tightening of inclusion criteria (not whether p-values change).

3. The inclusion and exclusion criteria should be made clearer in the methods section. Two studies included patients without diabetes. In Morgan's study (doi:10.1136/bmjopen-2012-002171), 45.3% patients in the intervention group are not diabetes patients. In Katon's study (DOI: 10.1056/NEJMoa1003955), 11% patients in the intervention group are not diabetes patients. It would be better if the authors give a detailed inclusion criteria for patients with co-morbid depression and diabetes (for example, studies are included if more than 80% patients in the study have co-mobid depression and diabetes…).
>>> We have made the following revision: "in adults, most of who had to have had co-morbid diabetes, were eligible."

4. Flowchart section (Page 30): The authors searched seven databases such like PubMed, Scopus, Cochrane Library, etc. The authors also searched reference lists of retrieved articles. They may want to list how many citations were searched from each of them, rather than just give a total number.
>>> As stated on page 6, "Reference lists of potentially eligible articles were searched by hand to identify additional studies missed by our search strategy." This equalled 18 articles. We have changed the wording on the flowchart to "potentially eligible citations" for greater clarity.

5. Primary outcomes section (Page 8 line 6): The authors may want to provide a reference for the Glass's Delta method.
>>> This is the original article: Glass, Gene V. "Primary, secondary, and meta-analysis of research." Educational researcher (1976): 3-8.
>>> We would be happy to cite this article if required by the Editor.

Reviewer Name Prof James Dunbar
Institution and Country Greater Green Triangle University Department of Rural Health
Flinders and Deakin Universities
Australia
Please state any competing interests or state 'None declared': Chief Investigator, TrueBlue project mentioned as reference 25

This is a useful contribution to the literature on comorbid depression and diabetes. It is the first review looking at the effect of collaborative care on improving measures of depression and glycaemia. Both diabetes and depression are major contributors to the burden of disease. Together they are an important health problem.
Abstract
I think the second bullet-point might be better to say that the collaborative care model fits better within some health systems than others and that the extent of benefits in randomised trials are proportionate to the extent which the measures of HbA1c and depression are out of the target range in the guidelines.
>>> Feasibility and appropriateness of collaborative care in specific health care settings are excellent questions, but were outside the focus of our systematic review.

Results
It is interesting that the baseline HbA1c's range from 6.9 which is within the guideline target to 9.1 which is well outside. The higher the 'out of range' starting point, the more likelihood of improvement. This should be mentioned in the discussion. Perhaps a bit more could be made of the result of meta-regression model which didn't show depression scores predicting HbA1c values.
>>> This is an excellent point, and worth including in the limitations section, as follows (page 15): "Secondly, baseline mean HbA1c level was close to the upper limit of the normal range in several studies, which would have underestimated the effect size for, and therapeutic benefit of, collaborative

care for glycaemic control."

It seems unlikely that managing depression alone results in the improvement of HbA1c's. TeamCare published by Katon and TrueBlue which is the sister study that I ran both decided that the reason diabetes measures hadn't improved with improvement in depression measures in trials prior to 2007 was that there hadn't been a concentration on the diabetes outcomes. We also agreed to look at managing depression, CVD and diabetes together because the overlap between CVD and diabetes is so large, both in risk factors and overt disease.
In the discussion it says that properly integrated lifestyle intervention wasn't any of the RCTs offered in these two studies. This is untrue of TrueBlue and TeamCare. Indeed in your own table you describe how it was conducted. It is interesting that you refer to the prevention of type 2 diabetes. My group ran the Council of Australian Government's National Demonstrator Greater Green Triangle Diabetes Prevention Program 2004-2006 so a great deal of that methodology was incorporated into TrueBlue. Nevertheless you need to be realistic about what can be expected by way of this kind of intervention in day-to-day general practice. I think in your emphasis on diabetes care and lifestyle intervention, you should refer to TeamCare and TrueBlue at least.
>>> We were referring to lifestyle interventions not being as per the IDF Global Guideline for effective management of type 2 diabetes. We have revised this section accordingly (page 14): "as per the current global guideline for effective management of type 2 diabetes,[42]"

In limitations you are suggesting that because most of the trials have been done in the US, further research is required in other countries. I disagree having set up in Australia. I also have long experience of the NHS in UK and it is not the need for further research. It is the need for health system change to be able to undertake the work. After all, UK has many experts in this area who have had difficulty making it work within their system.
>>> Global health care needs, as identified by clinicians or patients/consumers, are addressed through the generation of research evidence that is not only effective, but also feasible, appropriate and meaningful to specific populations, cultures and settings. Since there is an absence of a strong body of such evidence, we believe that our statement for generating more research in other health care settings/countries is accurate.

In limitations, I think the biggest limitation is that you have only looked at HbA1c. It is CVD which kills people with diabetes and that is the reason that TeamCare and TrueBlue both incorporated it into their collaborative care. If you look up UK PDS and the more recent ACCORD studies, you will see that attention to blood pressure, cholesterol and smoking are at least important as attention to HbA1c. In fact, much of the evidence would point to blood pressure being the most important of all measures. Therefore I think that your conclusion should be that not only should collaborative care focus on the management of depression but it needs to concentrate on better management of diabetes, cardiovascular risk and comorbid vascular disease.
>>> We agree that these are interesting and important questions, requiring further research, and have revised our conclusion accordingly (page 15): ", and other co-morbid cardiovascular risk conditions,"

Reviewer Name Marco Menchetti
Institution and Country Department of Medical and Surgical Sciences
University of Bologna
Please state any competing interests or state 'None declared': None declared

The review does not add much to the existing literature, including 2 meta-analyses with the same studies
>>> We have now highlighted the novelty of our research compared with previous systematic reviews

and meta-analyses.

This is a well written and well conducted systematic review and meta-analysis.
The major consideration for the authors is for them to distinguish their review from previous recent meta-analyses, and in particular those cited in their introduction (Van der Feltz-Cornelis et al. 2010; Huang et al., 2013).
>>> We have revised our discussion to explicitly describe the novelty of our research compared with previous systematic reviews, as follows (page 13): “Our results for better glycaemic control are novel and more comprehensive than those published from previous meta-analyses because we sought and obtained raw unpublished data from the authors of three studies [27-29].”

In addition, there is the recently published Cochrane review of collaborative care that they did not mention in their introduction (Archer et al., 2012).
>>> The systematic review by Archer et al., 2012 is not relevant for co-morbid depression and diabetes, since it aimed to: “.. assess the effectiveness of collaborative care for patients with depression or anxiety.”

The authors stated that there are newly published RCTs (Bogner et al., 2012; Katon et al., 2010) but these studies were also included in previous reviews. The authors should explain more in depth this overlap and what does their paper add to the literature. Possible inconsistencies among findings found in previous reviews and findings of the present work should be addressed.
>>> We have revised our discussion to explicitly describe the novelty of our research compared with previous systematic reviews, as follows (page 13): “Our results for better glycaemic control are novel and more comprehensive than those published from previous meta-analyses because we sought and obtained raw unpublished data from the authors of three studies [27-29].”

Reviewer Name Andrew Hinde
Institution and Country Department of Social Statistics and Demography
University of Southampton
SOUTHAMPTON
SO17 1BJ
United Kingdom
Please state any competing interests or state ‘None declared’: None declared.

My main concern is about the description of the statistical methods used. I feel they are described to briefly and ambiguously to allow a reader to replicate the study. The specific issues I am concerned about are listed below.

(1) On p. 7, ll. 22-23 you say that '[s]tudies were considered "better quality" if they received a score higher than 4, since that meant that they had most of our quality items'. First one might dispute the implication that the six criteria you use are being given equal weight (arguably randomisation is more critical than, say, the balancing of drop-out rates).
>>> We conducted five sensitivity analyses on each meta-analysis. Each of these analyses addressed the question “Would the observed effect size be changed if we tightened our inclusion criteria by excluding studies which ...” One of these five criteria was quality score. There is always a degree of subjectiveness in the choice of inclusion criteria and quality scores. While we acknowledge that using overall quality scores is problematic, we decided ‘a priori’ that they may be more useful to assess overall study quality as one of several potential sources of heterogeneity in one subgroup analysis, than to assess six individual quality items in separate subgroup analyses for each outcome. Furthermore, there is evidence supporting the use of quality scores in the assessment of heterogeneity and risk of bias in meta-analysis (Moher D1, Pham B, Jones A, Cook DJ, Jadad AR,

Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998 Aug 22;352(9128):609-13.)

Second, you seem to have included one study with a score of 3.5 (see table on p. 37).
>>> The table shows quality scores for all papers included in the meta-analysis. Sensitivity analyses are conducted on sub-groups of these papers. One of the five sensitivity analyses summarised in Tables 2 and 3 excludes all three studies with quality scores ≤4.0.

(2) On p. 8, l. 6 you state that you computed standardised mean differences using Glass's Delta method. Could you be more specific as to whether you used the standard deviation of the control group (which is what Glass proposed) or a 'Delta-family' method where you might use some combination of the standard deviations of the control and treatment groups.
>>> The sentence immediately prior confirms this method: “control group standard deviations carried forward from the baseline values”

(3) On p. 8, l. 14 you refer to the 'inverse variance weighted mean difference'. What variances were used for the weights? Later you say that you pooled the standardised mean differences (SMDs) from each randomised control trial 'to produce an overall estimate of effect' (ll. 16-18). Again, did you weight these SMDs and, if so, how? I could not follow exactly what you did from the rather terse description given in this paragraph.
>>> Meta-analysis combines the estimated effect sizes reported by different studies into a single estimate. The greater the uncertainty in the estimated effect size, the less weighting given to that study in the combined estimate. By “inverse variance weighted” we mean the estimated effect size of each study was weighted to take into account the level of certainty of that estimate. The variance is estimated variance of the effect size and, in formula terms, the weight is the inverse of the standard error of the effect size squared.