Article details: 2013-0030	
Title	Screening for depression: a systematic review and meta-analysis
Authors	Homa Keshavarz PhD, Donna Fitzpatrick-Lewis MSW, David L. Streiner PhD, Maureen Rice, Muhammad Usman Ali MD MS, Harry S. Shannon PhD, Parminder Raina PhD
Reviewer 1	Yann Le Strat
Institution	Centre for Addiction and Mental Health, Translational Addiction Research Laboratory, Toronto, Ont.
General comments	This paper is a meta-analysis. The main objective of the study is to assess the benefits of screening for depression in a community-based sample. Five controlled studies were included, all recruiting rural Japanese adults. Therefore, the major limitation of this study is the lack of generalizability of the result, as acknowledge by the authors. The paper is well written, the methodology used is relevant, and the findings of the study will be of high interest for the readers of the Journal. Major Revisions 1. P8L13: The aim of the study is to assess the benefits of a depression screening. However, it is not clear to us whether this screening is the only difference between the intervention group and the control group. It seems that the intervention group received (i) screening for depression, (ii) psychoeducation and (iii) psychiatric treatment if screening was positive, are we wrong? Which of these interventions were received by the control group is not mentioned in the text. Were the participants and control randomized? A description of the control group could be added in the results section. 2. How information on suicide was collected is not clear? Is it through a national database? Minor revision 1. P8L49: Please add the number of participants next to the person-years and suicide victim's count. 2. P8L20: A short description of the SDS should be added, as well as the cut-off used for the study, if any. 3. P8L56: Were the age-group analysis decided post-hoc? If yes, please add a mention in the text. 4. The paper deals with the impact of a depression screening rather than with the psychometric properties of the screening itself. In order not to be misleading for the reader, the title could be modified into "Impact of depression screening" for example. 5. P12L13: Please add a reference for suicide prevalence in Japan and Canada.
Reviewer 2	Laura Manea
Institution	Department of Health Sciences, University of York, UK
General comments	This is a clearly and concisely written paper that explores an important clinical question. To date, the benefits and harms of screening for depression in asymptomatic adults or higher risk groups in primary care or outpatients settings are still to be clearly summarized. The authors have carried out a systematic review in order to answer this question, however very limited evidence to inform this review was found. My main suggestions relate to a few details which I believe are missing from the paper. Methods The search strategy should be included in an appendix. Is there any particular reason why the databases were searched from 1994? This is not a critique, but simply wanting to know if there is a specific rationale. Study selection P. 5 lines 40-42 it is stated: 'The study settings were primary care or, of high risk groups, specialty clinics.' Could the authors please clarify which groups were considered high risk. Quality assessment Please explain in one sentence or two why the GRADE system was chosen to assess the quality of the evidence. I believe it would be useful for readers who might not be familiar with various formal systems for grading evidence. It might also help to have the actual ratings of the five criteria in an appendix. Statistical analysis It might be useful to briefly explain the measures employed to quantify the heterogeneity. Results Paragraph relating to high risk population P.11: which are the 5 high risk groups selected in the key questions and what do the authors mean by 'any high risk group'? Conclusions Given that their search found so limited evidence, the authors may wish to make some recommendations for future research.
Reviewer 3	Genevieve Gariepy

Institution	Douglas Mental Health University Institute, Montréal, Que.
General comments	This study is a systematic review and meta-analysis of the benefit and harm of depression screening programs in the population and sub-group of the population. The review question is: what are the benefits and harms of screening for depression in asymptomatic adults from the general population and in adults with high risk for depression in outpatient or primary care setting? The authors found 5 studies to include in their review. All studies were based on samples of elderly from rural Japan. Overall, these studies suggest that depression screening reduced the rate of suicide in this population. The review was overall well-written and well-structured. The review question is timely. Depression poses a significant health and economic burden on the population. Depression screening has been proposed by a number of researchers as a way to reduce this burden. The authors searched a number of relevant databases. Reviews were done in pair which helps to increase the confidence in results. The authors used the GRADE system to assess the quality of evidence which has been recommended by many organizations. Specific comments Introduction 1. The introduction explains that the paper is based on the report "Screening for Depression in Primary Care: Updated Recommendations from the Canadian Task Force on Preventive Health Care" (CTFPHC, 2004). This report was an update from a previous systematic review for the U.S. Preventive Services Task Force (USPSTF) to determine whether routine screening for depression improved detection, treatment and outcome (Pignone et al., 2002). Together, these 2 systematic reviews covered the period from 1994 to 2002. They found 14 relevant randomized trials that examined the effect of routine screening of adult patients for depression in primary care settings. In addition, a review by Gilbody et al. (Gilbody et al., 2008) also investigated the question to determine the specific clinical effectiveness of screening and case-finding instruments without additional enhancement of care in improving the recognition, management and outcome of depression. They found 16 randomized studies. It would be useful and important for these reviews to be included and discussed in the introduction. Given these previous reviews, how does the current review differ from previous ones? is it to update the previous reviews? does it have a different focus? Please explain for readers. The fact that none of the previously identified randomized trials are discussed or included in the present review is confusing. 2. In the Abstract, the Background suggests that the aim of the review is to provide information for the Canadian Task Force on Preventative Health Care on depression screening. This is not clearly stated in the introduction of the manuscript and should be added. In addition, the Canadian Task Force on Preventive Health Care already reported results from this review and guidelines in CMAJ last month ("Recommendations on screening for depression in adults", CMAJ, June 2013). This should be included in the introduction. Why is it important to have detailed results published? Please explain to readers. 3. It is not explicitly clear how a recovery rate of 50% in 3 months is an argument against depression screening. For example, some may argue that depression is often a recurrent and chronic mental disorder, and detection is important for monitoring. Others might suggest that 3 months of suffering without a diagnosis or guidance is a long time for one to bear. Please add a line to explain (e.g., Early depression detection and treatment would thereby use healthcare resources that may be costly and unnecessary for recovery). Methods SEARCH STRATEGY 4. I would suggest the authors update their search to July 2013. This reviewer ran the following search terms under "topic" in ISI Web of Science: "("screening for depression" OR "depression screening"). The search found several studies that were potential relevant to the review, including 1 study by Romera et al. (Systematic depression screening in high-risk patients attending primary care: a pragmatic cluster-randomized trial, BMC Psychiatry, March 2013), 1 unpublished study (poster) by Buckingham et al. (An evaluation of routine screening for depression in a diabetes centre, Diabetic Medicine, March 2013). 5. Please include information on the search terms.

	6. Please provide an example of a search strategy for one database in the appendix. STUDY SELECTION 7. One inclusion criterion is for the study to have used a "comparative study design", but it is not clear what this means specifically. Does this type of design exclude randomized trials that do not have pre-implementation information? If this is the case, please specify and justify. 8. It might be easier for readers if inclusion criteria were gathered in the same paragraph. For example, the criterion on study setting (p.5, line 43) could be moved with the other inclusion criteria (p.5, line 22-27). 9. Please indicate who conducted study selection using reviewers' initials. DATA ABSTRACTION AND QUALITY ASSESSMENT 10. The GRADE system rates the overall quality of evidence. One of the criteria of the GRADE system is Risk of bias (or "Study Limitations", according to reference 10 of the manuscript), which includes both study design and study quality. However, methodological quality of the studies was not reported in this review. Please add information on the methodological quality of the studies or justify why this was not assessed. Further information on the GRADE rating can be found at http://cebgrade.mcmaster.ca/ 11. Please indicate who conducted quality assessment using reviewers' initials. STATISTICAL ANALYSIS 12. P.6, line 35 refers to "suicide", which is in reference specifically to the studies that were selected in the review. Please change this section to keep the description of statistical methods general and not specific to the studies that were found. Results 13. Figures 3 to 6 can be dropped and results reported in the text to save on space. 14. GRADE Rating: The authors touch on 2 of the 5 GRADE criteria, but it is not clear how the evidence fare for the other 3 criteria. I think it would be helpful and interesting for readers to explicitly address each of the 5 GRADE criteria. 1) Risk of bias: a. study design: Discussed b. study quality: Not discussed 2) Consistency: Not discussed. 3) Directness: Discussed 4) Precision: Not Discussed 5) Reporting bias: Could not be performed – Could this tell us anything about quality of evidence? Authors are referred to the GRADE website for more information on each of the specific GRADE component: http://www.gradeworkinggroup.org/publications/JCE_series.htm Discussion 15. The discussion concludes that the "ultimate goal of screening for depression is to decrease morbidity and mortality related to this disease" (p.12, lines 52). However, the introduction informs readers that there are many benefits to screening, such as improved quality of life, which some may argue is as important as morbidity. The abstract is more restrictive and suggests that "the ultimate goal of screening is to decrease incidence of and mortality from this disease". I would suggest that the goal of screening is not to decrease incidence but to increase early detection. The ultimate goal of screening for many diseases, such as cancer, is indeed to decrease mortality, but this may be more nuanced in the case of depression. For depression, decreased quality of life, disability, and morbidity are some of the more devastating consequences of the disease. I would suggest nuancing these sentences and perhaps broadening the "goals" of depression screening to include more "soft" outcomes, such as early treatment/improved quality of life/decreased disability. 16. Since the authors found little and low quality evidence for their research question, it would be pertinent to provide recommendations for future studies for researchers.

Author response	

Reviewer Comments	**Author response**
Reviewer #1	
1. P8 L13: The aim of the study is to assess the benefits of a depression screening. However, it is not clear to us whether this screening is the only difference between the intervention group and the control group. It seems that the intervention group received I. screening for depression, II. psychoeducation and III. psychiatric treatment if screening was positive Are we wrong? Which of these interventions were received by the control group is not mentioned in the text. Were the participants and control randomized? A description of the control group could be added in the results section.	p. 10: added: *The control communities were similar demographically and were in the same geographical region as the intervention communities but they received no components of the program.*
2. How information on suicide was collected is not clear. Is it through a national database?	p. 11 added: *The outcome of interest was completed suicides determined from registrations of suicides at local Public Health Centres.*
P8 L49: Please add the number of participants next to the person-years and suicide victim's count.	p. 12 added: *Based on the information provided in the papers the estimated invention sample was 18,311 and control was 19,736.*
P8 L20: A short description of the SDS should be added, as well as the cut-off used for the study, if any.	Added: p. 11 *Zung Self-rating Depression Scale a 20 item scale that measures affective, psychological and somatic symptoms associated with depression (SDS)* No cut off information was provided by the study authors.
P8 L56: Were the age-group analysis decided post-hoc? If yes, please add a mention in the text.	This analysis has been removed following consideration of the editor's comments (see Editor comment 4).
The paper deals with the impact of a depression screening rather than with the psychometric properties of the screening itself. In order not to be misleading for the reader, the title could be modified into "Impact of depression screening" for example.	Thank you for this feedback, however we want to keep the title consistent with the CTFPHC guideline on screening for depression and what has already been made available on line in the CTFPHC website.
P12 L13: Please add a reference for suicide prevalence in Japan and Canada.	Reference #26 (Shaw 2010)
Reviewer #2	
Methods 1. The search strategy should be included in an appendix.	Added **Appendix A** (See Editor comment #4)
2. Is there any particular reason why the databases were searched from 1994? This is not a critique, but simply wanting to know if there is a specific rationale.	p. 16 *We chose 1994 as our beginning date for the search as that was when the DSM-IV was published and the definition of major depression changed.* (see Editor's comment #1)
Study selection 3. P. 5 lines 40-42 it is stated: 'The study settings were primary care or, of high risk groups, specialty clinics.' Could the authors please clarify which groups were considered high risk.	We added **Appendix B** that lists high risk factors.
Quality assessment 4. Please explain in one sentence or two why the GRADE system was chosen to assess the quality of the evidence. I believe it would be useful for readers who might not be familiar with various formal systems for grading evidence. It might also help to have the actual ratings of the five criteria in an appendix.	GRADE table added as **Table 1** p. 8 added: *GRADE within a systematic review assesses the overall confidence that additional research will or will not impact the direction of the effect.* p. 8 added: *This system of grading evidence has been widely used and has been endorsed by over 40 major organizations including the World Health Organization, Centers for Disease*

	Control and Prevention, and the Agency for Healthcare Research and Quality.
Statistical analysis 5. It might be useful to briefly explain the measures employed to quantify the heterogeneity.	See p.9 in the statistical analysis section " The Cochrane's Q (α=0.10) and I^2 statistic were employed to quantify the statistical heterogeneity between studies, where p<0.10 indicates a high level of statistical heterogenity between studies"
Results 6. Paragraph relating to high risk population P.11: which are the 5 high risk groups selected in the key questions and what do the authors mean by 'any high risk group'?	This paragraph has been edited and moved to the methods section and a table of high risk factors has been added as **Appendix B.**
Conclusions 7. Given that their search found so limited evidence, the authors may wish to make some recommendations for future research.	p. 17: research implications section added: *This review found a) limited evidence to estimate the effectiveness of screening for depression in primary care with individuals at average risk for depression, b) no evidence for screening in high risk populations and c) no evidence of the harms of screening. RCT research comparing screening and no screening should help to clarify that issue. Future research must have a broader scope demographically, geographically and culturally. Trials on effectiveness of screening in people who are at increased risk of major depressive disorder are also needed in order to help in the early diagnosis and treatment of people most likely to be affected by depression. More evidence is needed on the harms of screening for depression (e.g., false positive rates) and the related potential for unnecessary, and possibly harmful, diagnostic and treatment procedures.*
Reviewer #3	
Introduction 1. The introduction explains that the paper is based on the report "Screening for Depression in Primary Care: Updated Recommendations from the Canadian Task Force on Preventive Health Care" (CTFPHC, 2004). This report was an update from a previous systematic review for the U.S. Preventive Services Task Force (USPSTF) to determine whether routine screening for depression improved detection, treatment and outcome (Pignone et al., 2002). Together, these 2 systematic reviews covered the period from 1994 to 2002. They found 14 relevant randomized trials that examined the effect of routine screening of adult patients for depression in primary care settings. In addition, a review by Gilbody et al. (Gilbody et al., 2008) also investigated the question to determine the specific clinical effectiveness of screening and case-finding instruments without additional enhancement of care in improving the recognition, management and outcome of depression. They found 16 randomized studies. It would be useful and important for these reviews to be included and discussed in the introduction. Given these previous reviews, how does the current review differ from previous ones? Is it to update the previous reviews? Does it have a different focus? Please explain for readers. The fact that none of the previously identified randomized trials are discussed or included in the present review is confusing.	p. 6 added: *A decision was made to undertake a de novo review given the focus for the CTFPHC is the screening for depression in people with no apparent depression symptoms versus no screening. Our review this differed from the reviews by Pignone and colleagues (2002) and that by O'Connor et al., (2009) which served as the evidentiary base for the 2009 USPSTF screening recommendations for adults {501397}. Those reviews included papers where the populations were all screened and the comparison was treatment/no treatment or feedback/no feedback. Likewise, the review by Gilbody et al. (2005) was also outside the scope of our review.*
2. In the Abstract, the Background suggests that the aim of the review is to provide information for the Canadian Task Force on Preventative Health Care on depression screening. This is not clearly stated in the introduction of the manuscript and should be added. In addition, the Canadian Task Force on Preventive Health Care already reported results from this review and guidelines in CMAJ last month ("Recommendations on screening for depression in adults", CMAJ, June 2013). This should be included in the introduction. Why is it important to have detailed results published? Please explain to readers.	See p. 5 where we added the reference to the guideline paper that was published in the CMAJ 2013. Also added: *This review provides the full evidence taken into consideration by the CTFPHC when making their guideline.*

3. It is not explicitly clear how a recovery rate of 50% in 3 months is an argument against depression screening. For example, some may argue that depression is often a recurrent and chronic mental disorder, and detection is important for monitoring. Others might suggest that 3 months of suffering without a diagnosis or guidance is a long time for one to bear. Please add a line to explain (e.g., Early depression detection and treatment would thereby use healthcare resources that may be costly and unnecessary for recovery).	p. 5: This section has been edited removing the statement about the 50% recovery rate.
Methods SEARCH STRATEGY 4. I would suggest the authors update their search to July 2013. This reviewer ran the following search terms under "topic" in ISI Web of Science: "("screening for depression" OR "depression screening"). The search found several studies that were potential relevant to the review, including 1 study by Romera et al. (Systematic depression screening in high-risk patients attending primary care: a pragmatic cluster-randomized trial, BMC Psychiatry, March 2013), 1 unpublished study (poster) by Buckingham et al. (An evaluation of routine screening for depression in a diabetes centre, Diabetic Medicine, March 2013).	p. 6 added: *We also conducted a Medline search for potentially relevant RCTs six weeks prior to the guideline publication (April 2013).* Added in the results: *Our targeted search in Medline in April 2013 located 838 citations. Those citations were also reviewed for inclusion however none met the inclusion criteria of this review.*
5. Please include information on the search terms.	See **Appendix A**
6. Please provide an example of a search strategy for one database in the appendix.	See **Appendix A** which is the MEDLINE search
STUDY SELECTION 7. One inclusion criterion is for the study to have used a "comparative study design", but it is not clear what this means specifically. Does this type of design exclude randomized trials that do not have pre-implementation information? If this is the case, please specify and justify.	All on topic RCTs were included and we did not consider pre-implementation information as an inclusion/exclusion criterion.
8. It might be easier for readers if inclusion criteria were gathered in the same paragraph. For example, the criterion on study setting (p.5, line 43) could be moved with the other inclusion criteria (p.5, line 22-27).	We have changed this section by moving line 43 to follow line 27.
9. Please indicate who conducted study selection using reviewers' initials.	Reviewer initials have been added on p. 7
DATA ABSTRACTION AND QUALITY ASSESSMENT 10. The GRADE system rates the overall quality of evidence. One of the criteria of the GRADE system is Risk of bias (or "Study Limitations", according to reference 10 of the manuscript), which includes both study design and study quality. However, methodological quality of the studies was not reported in this review. Please add information on the methodological quality of the studies or justify why this was not assessed. Further information on the GRADE rating can be found at http://cebgrade.mcmaster.ca/	GRADE table added as **Table 1** The risk of bias on the primary studies was done with Newcastle Ottawa scale
11. Please indicate who conducted quality assessment using reviewers' initials.	Reviewer initials have been added on p. 7.
STATISTICAL ANALYSIS 12. P.6, line 35 refers to "suicide", which is in reference specifically to the studies that were selected in the review. Please change this section to keep the description of statistical methods general and not specific to the studies that were found.	This section has been edited to make it more generic.
Results 13. Figures 3 to 6 can be dropped and results reported in the text to save on space.	These figures have been deleted. The text has also been deleted (See Editor

	comment # 4).
14. GRADE Rating: The authors touch on 2 of the 5 GRADE criteria, but it is not clear how the evidence fare for the other 3 criteria. I think it would be helpful and interesting for readers to explicitly address each of the 5 GRADE criteria. 1) Risk of bias: a. study design: Discussed b. study quality: Not discussed 2) Consistency: Not discussed. 3) Directness: Discussed 4) Precision: Not Discussed 5) Reporting bias: Could not be performed – Could this tell us anything about quality of evidence? Authors are referred to the GRADE website for more information on each of the specific GRADE component: http://www.gradeworkinggroup.org/publications/JCE_series.htm	Summarized GRADE table added as **Table 1**
Discussion 15. The discussion concludes that the "ultimate goal of screening for depression is to decrease morbidity and mortality related to this disease" (p.12, lines 52). However, the introduction informs readers that there are many benefits to screening, such as improved quality of life, which some may argue is as important as morbidity. The abstract is more restrictive and suggests that "the ultimate goal of screening is to decrease incidence of and mortality from this disease". I would suggest that the goal of screening is not to decrease incidence but to increase early detection. The ultimate goal of screening for many diseases, such as cancer, is indeed to decrease mortality, but this may be more nuanced in the case of depression. For depression, decreased quality of life, disability, and morbidity are some of the more devastating consequences of the disease. I would suggest nuancing these sentences and perhaps broadening the "goals" of depression screening to include more "soft" outcomes, such as early treatment/improved quality of life/decreased disability.	Abstract – we removed incidence and added morbidity
16. Since the authors found little and low quality evidence for their research question, it would be pertinent to provide recommendations for future studies for researchers.	Research implication section has been added (see reviewer 2, comment # 7): *This review found a) limited evidence to estimate the effectiveness of screening for depression in primary care with individuals at average risk for depression, b) no evidence for screening in high risk populations and c) no evidence of the harms of screening. RCT research comparing screening and no screening should help to clarify that issue. Future research must have a broader scope demographically, geographically and culturally. Trials on effectiveness of screening in people who are at increased risk of major depressive disorder is also needed in order to help in the early diagnosis and treatment of people most likely to be affected by depression. More evidence is needed on the harms of screening for depression (e.g., false positive rates) and the related potential for unnecessary, and possibly harmful, diagnostic and treatment procedures.*