Author’s response to reviews

Title: Screening for depression in women during pregnancy or the first-year postpartum and in the general adult population: a protocol for two systematic reviews to update a guideline of the Canadian Task Force on Preventive Health Care

Authors:

Candyce Hamel (cahamel@ohri.ca)
Eddy Lang (Eddy.Lang@albertahealthservices.ca)
Kate Morissette (kate.morissette@canada.ca)
Andrew Beck (abeck@ohri.ca)
Adrienne Stevens (adstevens@ohri.ca)
Becky Skidmore (bskidmore@rogers.com)
Heather Colquhoun (heather.colquhoun@utoronto.ca)
John LeBlanc (John.LeBlanc@Dal.Ca)
Ainsley Moore (ainsleyemoore@gmail.com)
John Riva (rivaj@mcmaster.ca)
Brett Thombs (brett.thombs@mcgill.ca)
Ian Colman (icolman@uottawa.ca)
Sophie Grigoriadis (sophie.grigoriadis@sunnybrook.ca)
Stuart Nicholls (snicholls@ohri.ca)
Beth Potter (Beth.Potter@uottawa.ca)
Kerri Ritchie (kritchie@ottawahospital.on.ca)
Julia Robert (jrobert72@gmail.com)
Priya Vasa (VasaP@smh.ca)
Bianca Horner (hornerb@dal.ca)
1. The authors are undertaking two complementary reviews (general population and specific sub-population). While the methods are often similar, they are often different due to differences in key questions, search strategies, inclusion criteria, etc. While this falls in line with the needs of the guideline developers, it may be difficult for the reader to follow the logic and easily note where the differences should be noted. This leads to two options, the first being to clearly separate make distinctions where the review methods are dissimilar (e.g. years of searching) and the other is to split the protocol and subsequent reviews into two separate documents. Additionally, if it’s not clear if the identification of the evidence will be conducted for both reviews simultaneously (e.g. a citation will be screened once and determine if it may be included in one review, both reviews, or neither review) or if there will be separate independent processes conducted for each review.

Response: These are two separate reviews that will be performed, however, there is a lot of overlap in the methods. We have reformatted the document in a more intuitive way, with differences between reviews clearly stated.
2. For the introduction, it is suggested to streamline it further for the reader by beginning with the sections on depression in the general population, then on the subpopulation of women during pregnancy and up to one-year postpartum and finalizing it with the objective of this protocol to outline the methodological process for synthesizing systematic reviews for the CTFPHC guidelines.

Response: We have restructured the introduction. Repetitive information has been removed.

3. Additionally, the use of asterisk should rarely occur except in tables. Rather, please elaborate in writing the clarifications you want to make.

Response: This has been edited so there is no asterisk and the associated text is found within criteria ii.

“(ii) patients who are known to have a current episode of depression or are already being treated for depression close to the time of eligibility assessment are excluded, as screening is intended to identify undetected cases and those who are known to have depression would not be screened in actual clinical practice;”

Comments requiring consideration

- The authors have noted that 'based on the evidence from both key questions, if screening is found to be effective and the CTFPHC is interested in a key question on patient values and preferences. The potential key questions are...' While this is practical from a guideline perspective, it can be considered to be highly controversial and may be biased as the decision to review these key question will data-driven. These decisions should be made a priori and not subject to post-hoc decisions.

Response: This systematic review is being conducted to inform a guideline on screening for depression. This protocol describes the methods for the first systematic review, which will address KQ1 and KQ1a (effectiveness of screening). We are signaling our intention to conduct a separate systematic review on KQ2 and KQ2a on patient values and preferences should the working group decide it is needed to inform the guideline. This will be decided after reviewing
the evidence from the outcomes. If the Task Force working group believes that systematic review information on patient values and preferences would potentially change recommendations, beyond what is known from the KT Team’s focus groups, then we will move forward to do the review. We will follow GRADE methods and weigh all a priori identified critical outcomes for both benefit and harm. If we do pursue a systematic review on KQ2 and KQ2a (patient values and preferences), a separate protocol will be developed (including the relevant PICO criteria). Text has been added to this section in the protocol to describe the reason for including this piece, and the process of how it will be addressed.

- The authors noted that 'We will invite patients to partner with the team to gain from their perspectives and learn from their knowledge regarding the prioritization of the outcomes... The list of outcomes will be finalized after the input from patients.' If the list of outcomes are already declared then what is the role of including patient partners. Patients partners should be interviewed using structured questions in a 'safe environment' long before the outcomes are decided and finalized or a protocol is completed. Including them after the completion of the protocol undermined the true value of their insights and decision-making abilities. Please clarify or remove this statement.

Response: Since the time the protocol was initially developed, the patients’ rating of outcomes has been completed and results have been incorporated into the protocol. The patients rated the outcome ‘labeling/stigma’ as an important outcome, therefore, the Working Group has added it as an outcome in this review.

- Inclusion criteria: Please define the patient age range for the review on women during pregnancy and up to one-year postpartum. Please clarify if family history of depression will be used in the inclusion or exclusion process.

Response: As stated in the PICOs table, we will include pregnant and postpartum women of any age. We have also included examples within the PICOs table, which states that family history of depression could be a characteristic that may suggest elevated risk of depression. “*characteristics as defined in primary studies (e.g., trauma early in life, a family history of depression)”
- Exclusion criteria: The authors note that 'if >80% of women have a recent history of depression, have a current diagnosis or are receiving treatment for depression or other mental disorders, unless results are provided separately from the population of interest'. Maybe this is a typo, but for example if 75% of participants have a history of depression then the contamination rate would be huge and makes this criteria invalid. I'm assuming that the authors meant >20% (e.g. the 80:20 rule).

Response: Thank you for picking this up, that is exactly what we meant. 80% has been changed to 20% in both PICOs tables.

- Search strategy: The authors note that the final search will be peer-reviewed using the PRESS 2015 guideline. This should be done prior to finalizing the review and submitting it for publication as peer-reviewing of the search strategy may lead to major changes in the search terms. Please conclude this prior to re-submission.

Response: Thank you for this feedback. The search strategy has since undergone peer-review using the PRESS 2015 guideline.

- For grey literature, when you mention searching the websites, are you also referring to the related conference proceedings by these organizations. If so, what is the range of years that you will hand-search (e.g last 3 years, last 5 years, etc)? Additionally, the argument for one week of searching by one person for practical purposes is understood, but not clear what this will cover. Some items should be clearly prioritized to the reader as what will be covered first and what will be covered if time allows.

Response: No, as these have been removed from the main search strategy (database dependent), we will not be considering conference proceeding as part of the grey literature search. As far as prioritizing what will be reviewed first, we will focus on any reports of trials, as this is the only study design of interest for these reviews.

- For title/abstract screening, I would disagree that this would be considered 'independent'. The second reviewer is screening citations that were excluded by the first reviewer. As such they already know the decision of the first reviewer and as such they are neither blinded, nor
independent of the original decision. Please update the statement to reflect this. Further, in most cases (and therefore should be the rule) co-publications or multiple reports of the same study should not be confirmed to this stage and require reviewing the full-text publications. In cases, where all the citations from a trial are excluded then it makes no sense to categorize them at this stage anyways.

Response: We have removed the word ‘independent’. The next sentence does describe the process in further detail. “As these are done concurrently and randomly, each reviewer will not necessarily know if the reference has already been considered irrelevant by the other reviewer.” The second reviewer does not wait until the first reviewer has assess the record, and therefore reviewers do not necessarily know if the record has already been assessed as excluded by the other reviewer.

As far as identifying co-publications, this is done at full-text, not at title and abstract screening. We have added additional wording to make this clearer: “Reports that are co-publications or multiple reports of the same study will be identified at full-text review and labelled as such.”

- The statement "conference abstracts have been removed from the search results in Embase and Cochrane, a feature only available in these two databases" is somewhat (in)correct. While Medline does not contain many conference abstracts, it does have a 'clinical-conference' publication type (clinical-conference.pt.) that can be used to remove conference abstracts.

Response: Thank you for this suggestion. We had our medical information specialist look into this and here is her response: “I looked at a subset of these records as it was clear that many records were more than one page long (as is generally the case with abstracts). Many share the “clinical conference” publication type with the “Journal article” publication type, so it would be incorrect to blankly remove records tagged with the ‘clinical conference’ publication type.”

- For risk of bias assessments, it is recommended to use Risk of Bias 2.0 (https://sites.google.com/site/riskofbiastool//welcome/rob-2-0-tool) instead of the original version available in the Cochrane Handbook.

Response: Since it is possible that changes will be made in 2016 version, as it is still in a draft version, we will use the 2011 version until the 2016 version is finalized. A search of recently published Cochrane reviews show that they are still using the 2011 ROB version.
- For the decision to pool or not based on the quantification of the amount of heterogeneity, the reviewers have noted that 'should considerable statistical heterogeneity exist, we will present all studies in a forest plot, but will not provide the pooled estimate. Please clarify from the categorical breakdown you have provided (e.g. moderate to substantial heterogeneity) what you will use to make this decision. Please also note that I-squared alone should not be used to determine if pooling should be conducted or not (Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. Res Synth Methods. 2017 Mar;8(1):5-18).

Response: As stated in the protocol, we will be using both the Cochran’s Q and the I2 to determine statistical heterogeneity.

Reviewer 2

Perhaps the major issue for this reviewer was the decision to conduct the two parallel systemic reviews. It is acknowledged that both are related to the same clinical disorder of Major Depression, and the authors in many respects try to make the reviews parallel.

- For example, they propose parallel computation of using risk ratios and risk differences to calculate the benefits and risks of white screening, and generally comparable criteria for inclusion of studies (with the obvious difference that the peripartum analysis will be related only to women in the pregnancy and immediate postpartum.). This attempt at parallel analyses, however, has some severe limitations. For example, one of the common screening tools for depression in peripartum women is the Edinburgh Perinatal/Postpartum Depression Scale, which is not appropriate or used in the general population. Thus, the results of the screening analyses cannot be identical because of this methodological difference.

Response: I believe this is a misunderstanding. We will not be performing any parallel computation. The results from the pregnancy and postpartum population will be used to inform the guidelines for this population. The general adult population, which may include pregnant and partparum women, is a separate analysis, and results from this population will be used to inform the guideline on the general adult population.
- As a second example, the outcomes associated with the screening process will differ between the two data sets. In the general population, the listed outcomes include:

  o Symptoms of depression (continuous or dichotomous) or diagnosis of MDD (using a validated diagnostic interview)
  o Health-related Quality of life
  o Day-to-day functionality
  o Lost time at work/school
  o Impact on lifestyle behaviour (alcohol abuse, smoking, drugs, gambling, etc.)
  o Suicidality (suicide ideation, attempt or completion)
  o False positive result (positive screen in absence of depressive disorder), overdiagnosis, or overtreatment
  o Harms of treatment

In contrast, the outcomes that will be evaluated for the peripartum female sample include:

Mental Health Outcomes

  o Symptoms of depression (continuous or dichotomous) or diagnosis of MDD (using a validated diagnostic interview)
  o Health-related quality of life (validated tools)
  o Suicidality (suicide ideation, attempt, or completion)
  o False positive screens (positive screens in absence of depressive disorder), overdiagnosis, or overtreatment
  o Harms of treatment

Parenting Outcomes

  o Relationship with partner and other supports
Reported/observed capacity to parent (attachment, responsiveness to infant, positive regard of infant/fetus

Mother-child interactions including mutual touching, smiling, vocalizations, and impact on other children

**Infant Outcomes**

Infant health and development (i.e., developmental delay; failure to thrive) cognitive, emotional, motor and neural functioning and development

Infant responsiveness

Although it makes sense that the two sets of outcomes cannot be isomorphic, as of course the child and parenting related outcomes in the context of the peripartum female sample make perfect sense. However, it was not clear why some other outcomes from the general population are not being evaluated in the female peripartum sample (i.e., day-to-day functionality; lost time at work/school; impact on lifestyle behaviour). To ensure maximal comparability of results between the two sets of literature reviews, these three outcomes should be added to the list that will be evaluated for the female peripartum sample.

Response: This is the list of outcomes the working group has decided on for the two separate reviews. A working group meeting was organized separately for each population to discuss the rankings and come to consensus about overall assessments of outcomes and harms as critical, important, or not important. The initial GRADE rankings range from 1 to 9. Outcomes and harms ranked from 7 to 9 are critical for decision-making, those ranked from 4 to 6 are important but not critical, and those ranked from 1 to 3 are not important. Only outcomes and harms considered important (rating 4–6) or critical (rating 7–9) are included in the evidence profile.

Finally, a few minor issues:

1. As the reviews and are intended to be done without restriction of language, it would be helpful to know how the researchers intend to review the literatures. Article translation may be needed, and so some indication of how this process would be handled would be welcome.
Response: There is no restriction on language when performing the search. However, as stated in the methods section, only English and French language articles and documents will be included. A list of potentially relevant articles in other languages will be provided in the appendices, under “Other language” as the reason for exclusion. To ensure this is more clearly stated, we have added a row in the PICOs tables to discuss language of inclusion.

2. The authors state that one of the eligibility criteria for articles will be that patients "are randomized prior to administering the screening test". This criterion is unclear, as randomization prior to screening may affect the results of screening. Whether or not this is a concern depends on the nature of the randomization. If the authors are referring to randomization to a screening versus no screening condition, or to different types of screening, then this criterion is not a consideration since by definition the no screening group will not provide any data relevant to the systematic review. However, if the authors are referring to randomization to treatment or no treatment, or to treatment conditions, that criterion might affect the results of screening, and so is potentially problematic. The authors should clarify exactly what this eligibility criterion is.

Response: We are referring to screening vs no screening, however it is possible for the no screening group to provide relevant data to the systematic review, as they can be identified as depressed through other means, for example, unaided clinician diagnosis or patient report. If a patient reports they are feeling depressed, then they would not be considered as screened, but who continue onto diagnostic assessment. Additionally, patients in comparator trial arms may be administered depression symptom questionnaires for the purpose of baseline or outcome assessments as long as scores are not provided to the patients or healthcare providers.

3. Given that this will be a very broad set of articles, I commend the authors for their intended subgroup analyses, that will examine a variety of potential moderators of screening outcomes (see pages 23 and 24 in the document). The proposed variables all make sense, and should be pursued. However, one of the variables that the authors do not propose to examine is culture. There have been suggestions that depression screening practices vary among different parts of the world, and in particular between high income and low and middle-income countries. A meaningful sub-analysis therefore may be to examine the effects of countries or regions as on screening outcomes. Indeed, although the author group is comprised predominantly of people from central Canada, subgroup analyses by country or region may make the results of the systematic review even more relevant to the various national organizations that have made depression screening recommendations, as they will be able to meaningfully interpret the impact of screening in their own particular geographical regions.
Response: Thank you for this suggestion. We’ve added country/region as an example in the geographic location subgroup. “… and geographical location (e.g., rural vs. urban settings, country/region).”

4. One of the authors is a co-editor-in-chief of the journal Systematic Reviews. It is to the authors’ credit that they list this competing interest.

Response: This is provided under competing interests “DM is co-editor-in-chief, Systematic Reviews.”