PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Gender differences in the mental health of unaccompanied refugee minors in Europe – a systematic review
AUTHORS	Mohwinkel, Lea-Marie; Nowak, Anna Christina; Kasper, Anne; Razum, Oliver

VERSION 1 – REVIEW

REVIEWER	Anna Sarkadi Uppsala University, Sweden
REVIEW RETURNED	05-Mar-2018

GENERAL COMMENTS	General comments Thank you for the opportunity to review this manuscript which is generally well written and of high interest. I have some comments on the included studies and the quality assessment; I also suggest a shift in the focus of the study, all of which I will elaborate on below. Specific comments I think the paper would greatly benefit from stricter inclusion criteria and a focus on mental health problems only, exclusively based on primary data from studies. The inclusion of a qualitative study on expert opinion without primary data in a systematic review seems to me very confusing and hard to justify. I would argue the same for the guardian reports in the Beat et al study from 2006, the reason being that guardians' perceptions of higher need of healthcare in females may have a number of reasons and do not necessarily reflect actual differences in health problems and should not serve as a proxy for those2. In terms of physical health, there seems to be only one pilot study, that provides the basis of quite far-reaching conclusions which I don't believe are justified. The abstract has a very strong focus on these results, and relatively little is said about the mental health problems for which there is quite good quality data available. I strongly suggest restructuring the abstract to reflect this. An additional reason not to include the Marquardt study is that the sample size is small and the URM population seems different (37% from South Asia?) from the general population of URM in Europe, where African or Afghani origin is far more common (as is well mirrored in the other studies selected). Therefore, my clear suggestion is to exclude physical health from the review. In terms of quality assessment, although a score is assigned to each study, it is not clear for the reader which aspect each study fared better or worse on. This is very important information for the reader when trying to figure out how to weigh these different studies. A clear description of quality assessment and criteria fulfillment (or not)

	for each aspect should be added to Supplementary file 2. Finally, I would like to see a discussion on gender differences in the use of self-report instruments that might affect study results and conclusions. For example, the Edinburgh Postnatal Depression Scale has different cut-offs for women and men due to differential acceptability of emotional expression between the genders. I don't think any studies have done this kind of gender comparison on self-report instruments used in URM, but there is a theoretical possibility that it is socially more acceptable for females or less acceptable for males to express e.g. feeling depressed.

REVIEWER	Tilman Reinelt University of Bremen, Germany, Center for Clinical Psychology and Rehabilitation
REVIEW RETURNED	11-Mar-2018

GENERAL COMMENTS	**REVIEW OF THE ARTICLE** "Gender differences in the physical and mental health of unaccompanied refugee minors in Europe – a systematic review" In their manuscript the authors aim at summarizing evidence regarding health differences between female and male unaccompanied refugee minors. The paper is well written in general; yet, there are several major and minor concerns. Major concerns 1. I do not understand the motivation to explore gender differences in the health of URM. The authors offer no model or theory of gender playing in important role in health. There are only associated risks and some arguments do not seem valid, as for instance gender influencing "the way in which experiences of displacement affect the physical and psychosocial well-being". I am not aware of any psychological models assuming differences in experience being dependent on gender. 2. The authors should clarify their language. While I agree

	that gender can be considered a risk factor, the authors offer no evidence for a causal risk factor. Therefore, wordings as “other effects of gender on aspects of the physical and mental health of URM” or “gender is only one of many determinants of health” should be formulated less causal. My main problem with the article is that it aims at unaccompanied refugee minors. However, already in the introduction differences between URM and accompanied refugee minors (ARM) are sometimes not clear. For instance, in the introduction the authors name “specific psychological burdens” of URM. However, most of the listed burdens afterwards apply not only to URM but refugees/refugee minors in general. The issue gets even more problematic regarding the literature review. Wouldn’t one need to explore the gender differences in URM as compared to gender differences in ARM? How are the reported results specific to *unaccompanied* refugee minors? Also regarding the discussion, the authors conclude that “female gender is a risk factor for the development of depression in URM”. However, right before they argue that the results are in line with the general higher vulnerability for depression in females. Thus, how specific are the results to URM? A similar point applies to the conclusion: “Results so far underline that services need to be gender-sensitive – in particular so when they are dealing with URM”. This conclusion can only be made when comparing gender differences in URM with gender differences in

	ARM. 3. I am not convinced that “health” is a sufficient key word. If the authors want to aim at specific physical or mental health outcomes, why not directly search for them (e.g., PTSD, trauma, anxiety, depression)? 4. The health outcomes are very diverse (e.g., iron deficiency anaemia, infectious diseases, under- or overweight, dental status). What is the theoretical foundation of these outcomes? The outcomes should be chosen based on theory or the authors should make clear that their review relies on a lot of exploratory data. In that case, they should discuss replicability issues regarding explorative results. 5. Given that gender differences are not the primary outcome of the included studies and thus are probably often only reported if significant, relying only on studies, which reported analyses on gender differences, should overestimate the effects. The authors should discuss how this might affect their results. Minor concerns 6. What is the problem with behavioral outcomes? How are they defined in the context and distinguished from externalizing disorders, which would clearly be mental health issues, but are not part of the review?. 7. “Mental health outcomes were assessed by using self-rating scales, which are not designed for clinical diagnostics and may have led to an underestimation of

	prevalences due to social desirability bias". Witt et al. (2015) reported that in general prevalence rates are reduced for clinical diagnoses as compared to self-reports. 8. I do not understand the second conclusion on resilience, as resilience was not the topic of the review. Also, there are different understandings of resilience and one could argue that URM who have no health issues can be considered as resilient. It seems that the authors want to focus on resilience processes (p. 10), however, then they should also focus on the processes behind health issues (e.g., detrimental emotion regulation strategies). I also do not understand, why services should be gender-sensitive given the results. This would only apply if different strategies were needed to treat female and male URM.

VERSION 1 – AUTHOR RESPONSE

Manuscript ID bmjopen-2018-022389 entitled "Gender differences in physical and mental health of unaccompanied refugee minors in European countries – a systematic review"

Thanks to the editor and to the 2 reviewers for the valuable comments. We were able to implement numerous changes to improve the paper.

Lea-Marie Mohwinkel

(for the authors)

Reviewer: 1 **Reviewer Name: Anna Sarkadi** General comments	We thank the reviewer for her positive comment on our paper. Thanks also for her helpful suggestions, all of which we could implement.

Thank you for the opportunity to review this manuscript which is generally well written and of high interest. I have some comments on the included studies and the quality assessment; I also suggest a shift in the focus of the study, all of which I will elaborate on below.	
Specific comments I think the paper would greatly benefit from stricter inclusion criteria and a focus on mental health problems only, exclusively based on primary data from studies. The inclusion of a qualitative study on expert opinion without primary data in a systematic review seems to me very confusing and hard to justify. I would argue the same for the guardian reports in the Beat et al study from 2006, the reason being that guardians' perceptions of higher need of healthcare in females may have a number of reasons and do not necessarily reflect actual differences in health problems and should not serve as a proxy for those2.	We have revised our inclusion criteria and excluded qualitative studies relying on secondary data.
In terms of physical health, there seems to be only one pilot study, that provides the basis of quite far-reaching conclusions which I don't believe are justified. The abstract has a very strong focus on these results, and relatively little is said about the mental health problems for which there is quite good quality data available. I strongly suggest restructuring the abstract to reflect this. An additional reason not to include the Marquardt study is that the	We agree that the data basis regarding physical health is very limited and have therefore followed your advice to focus exclusively on mental health problems. We restructured our abstract accordingly.

sample size is small and the URM population seems different (37% from South Asia?) from the general population of URM in Europe, where African or Afghani origin is far more common (as is well mirrored in the other studies selected). Therefore, my clear suggestion is to exclude physical health from the review.	
In terms of quality assessment, although a score is assigned to each study, it is not clear for the reader which aspect each study fared better or worse on. This is very important information for the reader when trying to figure out how to weigh these different studies. A clear description of quality assessment and criteria fulfillment (or not) for each aspect should be added to Supplementary file 2.	We have attached Supplementary file 3 showing the quality assessment of each study with detailed results.
Finally, I would like to see a discussion on gender differences in the use of self-report instruments that might affect study results and conclusions. For example, the Edinburgh Postnatal Depression Scale has different cut-offs for women and men due to differential acceptability of emotional expression between the genders. I don't think any studies have done this kind of gender comparison on self-report instruments used in URM, but there is a theoretical possibility that it is socially more acceptable for females or less acceptable for males to express e.g. feeling depressed.	Thank you for your useful suggestion. We completely agree on this point and have added a sentence as follows: "…it is questionable whether all instruments used were sensitive enough for gender and cultural differences. The social acceptability of expressing emotional problems may be influenced by socially constructed gender roles or the cultural background." See also our response to the first comment of Reviewer 2.

Reviewer: 2 **Reviewer Name: Tilman Reinelt** The paper is well written and addresses an understudied subject. However, I have several concerns, mainly regarding the specificity of the results for URM. Please, find attached a detailed document with my comments. In their manuscript the authors aim at summarizing evidence regarding health differences between female and male unaccompanied refugee minors. The paper is well written in general; yet, there are several major and minor concerns.	Thanks for the positive comment on our paper, but in particular thanks for the critical issues you raise. We agree with many of them and therefore made a number of changes to improve the paper.
Major concerns 1. I do not understand the motivation to explore gender differences in the health of URM. The authors offer no model or theory of gender playing in important role in health. There are only associated risks and some arguments do not seem valid, as for instance gender influencing “the way in which experiences of displacement affect the physical and psychosocial well-being”. I am not aware of any psychological models assuming differences in experience being dependent on gender.	Thank you for this comment. We have revised our argumentation in the introduction in order to clarify our motivation and explain the role of gender for health referring to a theoretical model by Faltermaier and Hübner (2015). We also agree with Reviewer 1 on potential gender differences to express mental health issues.
2. The authors should clarify their language. While I agree that gender can be considered a risk factor, the authors offer no evidence for a causal risk factor. Therefore, wordings as “other effects of gender on aspects of the physical and	We agree with your comment and have removed misleading wordings from the manuscript.

mental health of URM" or "gender is only one of many determinants of health" should be formulated less causal.	
3. My main problem with the article is that it aims at unaccompanied refugee minors. However, already in the introduction differences between URM and accompanied refugee minors (ARM) are sometimes not clear. For instance, in the introduction the authors name "specific psychological burdens" of URM. However, most of the listed burdens afterwards apply not only to URM but refugees/refugee minors in general. The issue gets even more problematic regarding the literature review. Wouldn't one need to explore the gender differences in URM as compared to gender differences in ARM? How are the reported results specific to *unaccompanied* refugee minors? Also regarding the discussion, the authors conclude that "female gender is a risk factor for the development of depression in URM". However, right before they argue that the results are in line with the general higher vulnerability for depression in females. Thus, how specific are the results to URM? A similar point applies to the conclusion: "Results so far underline that services need to be gender-sensitive – in particular so when they are dealing with URM". This conclusion can only be	Thank you for your critical and at the same time very helpful comment. We have revised the introduction to clarify the differences between URM and ARM. From our point of view, the lack of a familial system in a critical phase of physical and mental development (and consequently, coping with the "usual" burdens experienced by refugees without this important resource) constitutes the specific situation of URM. URM are also more often affected by mental health issues than ARM (cf. introduction). Therefore, we believe gender comparisons *within* the specific group of URM to be more informative than gender comparisons between different refugee subgroups.
made when comparing gender differences in URM with gender differences in	

ARM.	
4. I am not convinced that “health” is a sufficient key word. If the authors want to aim at specific physical or mental health outcomes, why not directly search for them (e.g., PTSD, trauma, anxiety, depression)?	We intentionally kept our outcome keyword broad in order to maximize the sensitivity of the search. Since we knew beforehand that the health research regarding URM relies on a lot of exploratory data and there is no theoretical foundation yet, we aimed at giving the reader an overview of the outcomes that have been studied so far. Therefore, no specific outcomes were addressed. We conducted a thorough manual search of reference lists of the included publications to identify publications we may have missed in the systematic search.
5. The health outcomes are very diverse (e.g., iron deficiency anaemia, infectious diseases, under- or overweight, dental status). What is the theoretical foundation of these outcomes? The outcomes should be chosen based on theory or the authors should make clear that their review relies on a lot of exploratory data. In that case, they should discuss replicability issues regarding explorative results.	Thank you for bringing up this important aspect. We entirely agree that it is a major limitation of the existing research on URM’s (physical) health that studies rely on exploratory data. Unfortunately, there is no theoretical model on refugee minor’s physical health developed so far, from which outcomes could be chosen. Following the suggestion of Reviewer 1, we excluded physical health outcomes from the paper and focus on mental health outcomes.
6. Given that gender differences are not the primary outcome of the included studies and thus are probably often only reported if significant, relying only on studies, which reported analyses on gender differences, should overestimate the effects. The authors should discuss how this might affect their results.	We have amended the discussion accordingly.
Minor concerns	We have revised our inclusion and exclusion criteria

1. What is the problem with behavioral outcomes? How are they defined in the context and distinguished from externalizing disorders, which would clearly be mental health issues, but are not part of the review?	(cf. Reviewer 1, comment 1). Although behavioral outcomes are often not clearly defined in the included studies, we have decided to report them under the condition that the used screening instruments clearly correspond with ICD-10 or DSM IV diagnoses. We still exclude potentially eligible studies or results when they report unspecific behavioral problems (measured by e.g. the Child Behavior Checklist before 2001).
2. "Mental health outcomes were assessed by using self-rating scales, which are not designed for clinical diagnostics and may have led to an underestimation of prevalences due to social desirability bias". Witt et al. (2015) reported that in general prevalence rates are reduced for clinical diagnoses as compared to self-reports.	We have added a sentence as follows: "However, Witt et al. (2015) reported that in general prevalence rates are reduced for clinical diagnoses as compared to self-reports".
3. I do not understand the second conclusion on resilience, as resilience was not the topic of the review. Also, there are different understandings of resilience and one could argue that URM who have no health issues can be considered as resilient. It seems that the authors want to focus on resilience processes (p. 10), however, then they should also focus on the processes behind health issues (e.g., detrimental emotion regulation strategies). I also	Thank you for this useful comment. By mentioning resilience in the conclusion, we wanted to point to the deficit orientation of health research on URM so far. Indeed, there are different understandings of this term, why we decided to remove the paragraph from our manuscript.
do not understand, why services should be gender-sensitive given the results. This would only apply if different strategies were needed to treat female and	

male URM.	

VERSION 2 – REVIEW

REVIEWER	Anna Sarkadi Uppsala University, Sweden
REVIEW RETURNED	14-May-2018

GENERAL COMMENTS	Appropriate revision, no further comments.

REVIEWER	Tilman Reinelt University of Bremen, Germany, Center for Clinical Psychology and Rehabilitation, Center for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany
REVIEW RETURNED	03-Jun-2018

GENERAL COMMENTS	**Review of the article** "Gender differences in the mental health of unaccompanied refugee minors in Europe – a systematic review" I think the authors did a great job in revising their manuscript. The paper is much more clear and I am only left with some smaller issues. If these are addressed, I generally recommend the manuscript to be published in BMJ Open. Minor issues: 1. "Gender (used here as an umbrella term for biological sex and the socially constructed gender role)" (p. 6): How well does this definition reflect the reviewed articles? 2. Please elaborate why you rely on the definitions of the DSM-IV (p. 8). The DSM-V has been released in 2013 and within the reviewed time span the DSM-III has been in clinical use as well. Does this have any implications on the interpretation of the results? 3. "The methodological quality varied, but was overall acceptable" (p. 12). Please elaborate and give references

	when possible on the underlying guidelines. How many criteria need to be fulfilled in order for a study to be acceptable? In addition, please elaborate in Supplementary File 3 on the guidelines for the rating. For example: How did you rate an adequate sample size? Did you perform power analyses to determine this? If so, how did you do it? How did you conclude a method is valid? Based on literature on construct validity? What cut-offs did you use for your decision? 4. "To reduce language barriers during the interviews, the respective questionnaires were provided …" (p. 12): It seems unusual that all questionnaires were applied in an interview setting. If that was the case than I would state this explicitly. 5. You gave mean levels on the Hopkins Scale"(1.90 vs. 1.75)" (p. 14): Without information on the possible range, this information is rather meaningless. You might want to present an effect size instead. 6. "Therefore, it can be assumed that female gender is a risk factor for the development of depression in URM. At the same time, it is possible that other factors, such as gender-specific reasons to flee or a higher vulnerability of girls towards sexual abuse,[2] account for part of the gender differences." (p. 16): I appreciate the theoretical considerations that were outlined in the theoretical part as well, however, I am wondering, whether these results just reflect the typical gender differences in depression and reveal nothing specific about URM. I suggest elaborating this in more detail. For instance, wouldn't one assume a larger difference in the prevalence rates between the

	genders in URM as compared to the general population if URM specific stressors were important? Did studies reporting more gender specific risk factors reveal larger gender differences in depression? (You reported results on this for PTSD, why didn't you do the same for depression?)

VERSION 2 – AUTHOR RESPONSE

Manuscript ID bmjopen-2018-022389.R1 entitled "Gender differences in the mental health of unaccompanied refugee minors in Europe – a systematic review"

Thanks to the editor and to the 2 reviewers. We were able to implement a few minor changes to further refine the paper.

Lea-Marie Mohwinkel

(for the authors)

Reviewer: 1 **Reviewer Name: Anna Sarkadi** Appropriate revision, no further comments.	We thank the reviewer for her positive final comment on our paper.

Reviewer: 2 **Reviewer Name: Tilman Reinelt** I think, the authors did a great job revising the manuscript. I still have some minor aspects (see file attached), but after addressing them, I would generally recommend publishing the article.	Thank you for your positive comment and recommendation.
I think the authors did a great job in revising their manuscript. The paper is much more clear and I am only left with some smaller issues. If these are addressed, I generally recommend the manuscript to be published in BMJ Open.	Thanks, we have tried to address the issues you raised as explained below.

Minor issues: 1. “Gender (used here as an umbrella term for biological sex and the socially constructed gender role)“ (p. 6): How well does this definition reflect the reviewed articles?	The included studies did not focus explicitly on gender differences in the mental health of URM. Therefore, they gave no definitions of gender, but categorized the study populations into ‘male’ and ‘female’, merely reflecting the (controversial, if not incorrect) biological dichotomy. Our definition of gender is used in the awareness that sex and the socially constructed gender role both have the potential to influence health.
2. Please elaborate why you rely on the definitions of the DSM-IV (p. 8). The DSM-V has been released in 2013 and within the reviewed time span the DSM-III has been in clinical use as well. Does this have any implications on the interpretation of the results?	Thank you for raising this question, which refers to our inclusion and exclusion criteria. The mental health outcomes reported in the included studies relied on screening questionnaires or checklists (all listed in Supplementary File 2). These instruments are usually based on the ICD or DSM criteria. The diagnostic criteria for PTSD were revised in the DSM-V compared to the DSM-IV, but since the release of the current version was 2013, it seemed unlikely that new versions of the screening questionnaires were already available in different languages and in scientific use (cf. Stammel and Böttche 2017). Therefore, we rely on the definitions of the DSM-VI.
3. “The methodological quality varied, but was overall acceptable” (p. 12). Please elaborate and give references when possible on the underlying guidelines. How many criteria need to be fulfilled in order for a study to be acceptable? In addition, please elaborate in Supplementary File 3 on the guidelines for the rating. For example: How did	Thank you for your helpful comment. We used checklists developed by the Joanna Briggs Institute which did not include a scoring system. In order to give the reader an overall result of the critical appraisal of each study, we rated a study as ‘acceptable’ if it fulfilled at least half of the criteria (cf. Supplementary Files 2 and 3). We have added

you rate an adequate sample size? Did you perform power analyses to determine this? If so, how did you do it? How did you conclude a method is valid? Based on literature on construct validity? What cut-offs did you use for your decision?	further details on the guidelines for the rating in Supplementary File 3.
4. “To reduce language barriers during the interviews, the respective questionnaires were provided …” (p. 12): It seems unusual that all questionnaires were applied in an interview setting. If that was the case than I would state this explicitly.	Thanks, we have corrected this mistake and changed “interviews” to “data collection process”.
5. You gave mean levels on the Hopkins Scale”(1.90 vs. 1.75)” (p. 14): Without information on the possible range, this information is rather meaningless. You might want to present an effect size instead.	We apologize for this mistake and have added information on the possible range.

6.“Therefore, it can be assumed that female gender is a risk factor for the development of depression in URM. At the same time, it is possible that other factors, such as gender-specific reasons to flee or a higher vulnerability of girls towards sexual abuse,[2] account for part of the gender differences.” (p. 16): I appreciate the theoretical considerations that were outlined in the theoretical part as well, however, I am wondering, whether these results just reflect the typical gender differences in depression and reveal nothing specific about URM. I suggest elaborating this in more detail. For instance, wouldn’t one assume a larger difference in the prevalence rates between the genders in URM as compared to the general population if URM specific stressors were important? Did studies reporting more gender specific risk factors reveal larger gender differences in depression? (You reported results on this for PTSD, why didn’t you do the same for	Thank you for bringing up these important questions. We followed your advice to elaborate on this in more detail in the discussion. We believe it is not possible (based on the current evidence) to give a definite answer to the question if the reported gender differences in depression (partly) result from URM-specific influencing factors or just reflect typical patterns also found in the general population. This should be investigated in further studies.

depression?)	

Literature:

Stammel N, Böttche M. Psychodiagnostik (Psychodiagnostics). In: Liedl A, Böttche M, Abdallah-Steinkopff B, Knaevelsrud C eds. Psychotherapie mit Flüchtlingen – neue Herausforderungen, spezifische Bedürfnisse. Das Praxisbuch für Psychotherapeuten und Ärzte. (Psychotherapy with refugees – new challenges, specific needs. The practice book for psychotherapists and physicians.)

VERSION 3 – REVIEW

REVIEWER	Tilman Reinelt University of Bremen, Germany
REVIEW RETURNED	05-Jul-2018

GENERAL COMMENTS	The authors adequately addressed all minor aspects. I have no further comments.