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)	Development of a Brief Measure of Intimate Partner Violence Experiences: The Composite Abuse Scale (Revised) – Short Form (CAS_R-SF)
AUTHORS	Ford-Gilboe, Marilyn; Wathen, Nadine; Varcoe, Colleen; MacMillan, Harriet; Scott-Storey, Kelly; Mantler, Tara; Hegarty, Kelsey; Perrin, Nancy

VERSION 1 - REVIEW

REVIEWER	Kjersti Alsaker Kjersti Alsaker Associate Professor Department of Social Education and Social Work Faculty of Health and Social Sciences Høgskolen i Bergen/ Bergen University College Møllendalsveien 6-8 Postboks 7030 5020 Bergen Norway
REVIEW RETURNED	05-Jul-2016

GENERAL COMMENTS	BMJ Open Development of a Brief Measure of Intimate Partner Violence Experiences: The Composite Abuse Scale (Revised) – Short Form (CASR-SF) This is a very important and well-designed study. However I miss some critical gaps that are not addressed in the discussion of the measure. 1. Title The original CAS and this revised version focuses specifically on women's experiences. This might be clarified by taking "women" into the title and in keyword (Intimate partner violence towards women's experiences) 2. Discussion A. Pregnancy As the focus was on evolving a brief version of the Composite Abuse Scale (CAS) I miss some discussion about how CAS may confine the ability to address all critical gaps. Most women get pregnant and violence during pregnancy is not mentioned at all. Violence directed at the pregnant woman's abdomen and in general during pregnancy is not presented in any questions and this should be discussed. Violence during pregnancy is associated with increased rates of spontaneous abortion, preterm birth and low birth weight as well as miscarriages or pregnancy complications due to abuse. Forced sex

	and abuse during pregnancy were found to be identifiable risk factors for intimate partner femicide (Campbell JC, Webster D, Koziol-McLain J, Block C, Campbell D, Curry MA, et al. Risk factors for femicide in abusive relationships: results from a multisite case control study. Am J Public Health. 2003 Jul; 93(7):1089-97)(Martin SL, Macy RJ, Sullivan K, Magee ML. Pregnancy-associated violent deaths: the role of intimate partner violence. Trauma Violence Abuse. 2007 Apr;8(2):135-48). B. Education Financial abuse is addressed in the new brief version regarding historical, cultural and situational context. However, obstacles for education or being denied education is not addressed.

REVIEWER	Mariana Dufort Karolinska Institutet, Sweden
REVIEW RETURNED	09-Jul-2016

GENERAL COMMENTS	The article is well written and findings reported here are very interesting and add to the existing knowledge in the field. There are however a few points that might need to be clarified before publication. The article is well written and findings reported here are very interesting and add to the existing knowledge in the field. There are however a few points that might need to be clarified before publication. Please find some suggestions hereunder that could help improve the quality of the paper. Introduction In the second paragraph authors state that studies tend to focus on physical violence and mention the General Social Survey, which is based on the CTS as an example. The CTS includes five dimensions of conflict solving in intimate relationships where both acts of physical, psychological and sexual violence are considered. Please consider revising that example as an argumentation for the need of better measures that include other types of violence than physical violence. Please add a reference to the statement: "/…/contrary to the statement that there is a single experience of IPV/…/" Methods In phase 1, experts are asked to rate CAS-items according to their importance, clarity and appropriateness. A convenience sapling was used to identity 31 participating experts of which 25 are academic

	researchers. Does the sampling method affect the variance of possible input on the studied measure? Another question due to participant's characteristics is if ratings of items differ between the different professions (eg researchers vs others)? Please ad a sentence on whether you have considered this or not. This should also be mentioned in the discussion section as possible circumstances that may influence the expert's suggestions regarding the CAS. In Phase 2 In the second paragraph under "Data sources", second sentence: it should be 10.7% of the pooled sample and not 9.7 Please motivate the choice of pooling samples from five different studies since the clinical sample of women is large (n=5608) and constitutes the vast majority of the pooled total sample of 6278. A more thorough elaboration of the motives of using all five samples would be helpful. The pooled sample was randomly divided into a development sample and a confirmatory sample. It would be interesting if authors report the distribution of original study samples since four of them are very small. Further, authors state that the samples where stratified by project. Why pool all samples into one to then stratify if by project? Please motivate. Also, if analyses where performed stratified by project this does not show in the result section or in the rest of the paper. In order to assess concurrent validity of the CAS_R-SF, authors compare with measures of depression, PTSD and coercive control. Results show similar concurrent validity as the CAS, which is good. However, considering that the main aim of the present study is to develop a brief measure that adequately captures the complexity of IPV, how come concurrent validity is not assessed in relation to other measures that consider physical sexual and psychological IPV rather than depression and PTSD? Please motivate Please ad a sentence concerning ethical considerations regarding the present study.

	Discussion Please revise discussion section according to the comments above. Also, please elaborate if there are any limitations related to the use of secondary data analyses.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Comment: The original CAS and this revised version focuses specifically on women's experiences. This might be clarified by taking "women" into the title and in keyword (Intimate partner violence towards women's experiences)

Thank you for pointing out this issue. We have added the keyword "Intimate partner violence against women" (the term used by the World Health Organization). However, we have not changed the title as we feel that it is already quite long.

Comment: As the focus was on evolving a brief version of the Composite Abuse Scale (CAS), I miss some discussion about how CAS may confine the ability to address all critical gaps. Most women get pregnant and violence during pregnancy is not mentioned at all. Violence directed at the pregnant woman's abdomen and in general during pregnancy is not presented in any questions and this should be discussed. Violence during pregnancy is associated with increased rates of spontaneous abortion, preterm birth and low birth weight as well as miscarriages or pregnancy complications due to abuse. Forced sex and abuse during pregnancy were found to be identifiable risk factors for intimate partner femicide (Campbell JC, Webster D, Koziol-McLain J, Block C, Campbell D, Curry MA, et al. Risk factors for femicide in abusive relationships: results from a multisite case control study. Am J Public Health. 2003 Jul; 93(7):1089-97)(Martin SL, Macy RJ, Sullivan K, Magee ML. Pregnancy-associated violent deaths: the role of intimate partner violence. Trauma Violence Abuse. 2007 Apr;8(2):135-48).

We are also aware of the evidence showing that IPV has significant negative health consequences for both the pregnant woman and her fetus/child. This is an important contribution to the literature. The CASR-SF is a general measure of IPV experiences appropriate for use with any women, including those who are pregnant. There is no reason to believe that the CASR-SF would perform differently for women who are pregnant that for those who are not pregnant, and, therefore, we have not specifically discussed violence during pregnancy.

Comment: Financial abuse is addressed in the new brief version regarding historical, cultural and situational context. However, obstacles for education or being denied education is not addressed.

Given our goal of limiting this brief scale to no more than 15 items, and considering the need to capture a range of acts that are core to the complex concept of IPV, we included only a single item about financial abuse. As noted in the manuscript, the content of the item we propose is based both on expert feedback and on current literature and the goal of ensuring that would be broadly applicable to all women. Based on feedback during a WHO consultation, we revised this item since the initial submission to make is even more general and applicable (from Kept me from having access to a job, money or credit cards, to Kept me from having access to a job, money or financial resources). While being denied education is an important aspect of economic sabotage, it would be applicable only to a subset of women who wish to continue their education. Furthermore, providing too many examples

within a single item reduces item clarity and makes item completion more difficult.

Reviewer 2
Comment: In the second paragraph authors state that studies tend to focus on physical violence and mention the General Social Survey, which is based on the CTS as an example. The CTS includes five dimensions of conflict solving in intimate relationships where both acts of physical, psychological and sexual violence are considered. Please consider revising that example as an argumentation for the need of better measures that include other types of violence than physical violence.

The point we intended to raise here is the tendency to privilege physical violence over other types of violence, particularly the lack of attention given to adequately measuring psychological violence. Like many population surveys, the GSS is based, in part, on the CTS. While the CTS includes some items that tap into aspects of psychological aggression, these items are framed as actions partners use to manage conflict, many of which do not reach the threshold for acts of IPV that are linked to poor health outcomes. This is a critical issue that has been well discussed in the literature. Our intention here is not to provide a detailed critique of the CTS specifically, but to point out more general limitations of measuring IPV in ways that do not fit with current definitions which favour complexity, including the importance of psychological abuse, coercive control and gender. We have done some editing of this section to make these points clearer, without slipping into a detailed critique of the CTS as this is a secondary issue and not the focus of our paper.

Comment: Please add a reference to the statement: "/…/contrary to the statement that there is a single experience of IPV/…/"
We have added references to support this statement.

Comment: In phase 1, experts are asked to rate CAS-items according to their importance, clarity and appropriateness. A convenience sampling was used to identity 31 participating experts of which 25 are academic researchers. Does the sampling method affect the variance of possible input on the studied measure? Another question due to participant's characteristics is if ratings of items differ between the different professions (e.g. researchers vs others)? Please ad a sentence on whether you have considered this or not. This should also be mentioned in the discussion section as possible circumstances that may influence the expert's suggestions regarding the CAS.

With any convenience sample, there is a risk of bias. The experts we approached as participants were known to our research team and represented a range of disciplinary and professional backgrounds. They were invited to take part because they are regarded as experts in the field and we believed that they would provide thoughtful feedback based on this expertise. While it is possible that the feedback would be different with a sample that includes a higher percentage of policy experts, the high level of agreement in the feedback provided by this group suggests that this is not likely the case. We have not looked at possible differences in the ratings and feedback of researchers versus practice/policy experts because the groups are too small to allow meaningful comparisons. We have added a note in the discussion that the composition of the expert sample may have affected their ratings and feedback, but that the type of impact is now known.

Comment: Phase 2, in the second paragraph under "Data sources", second sentence: it should be 10.7% of the pooled sample and not 9.7.
This correction has been made.

Comment: Please motivate the choice of pooling samples from five different studies since the clinical sample of women is large (n=5608) and constitutes the vast majority of the pooled total sample of 6278. A more thorough elaboration of the motives of using all five samples would be helpful.

Our primary rationale for pooling these samples was to maximize diversity of sample such that the brief measure would be appropriate for women from diverse backgrounds. We see this as a strength of our study. The largest sample included women recruited into an IPV screening trial conducted in health care settings in one Canadian province. While the other 4 samples represent only ~10% of the overall sample, women in these volunteer, community-based samples all self-identified as having experienced IPV (and, therefore, may vary from the clinical sample in their stage of help seeking). The 4 smaller studies also include women from different geographic contexts (both provinces and rural/urban settings), and one study included only Indigenous women (an important but often under-represented population in studies of IPV). We have added a brief rationale for including the 4 studies with non-clinical samples in this analysis.

Comment: The pooled sample was randomly divided into a development sample and a confirmatory sample. It would be interesting if authors report the distribution of original study samples since four of them are very small. Further, authors state that the samples where stratified by project. Why pool all samples into one to then stratify if by project? Please motivate.

Assignment of cases to the developmental or confirmatory sample was stratified by project in order to reduce the risk of introducing systematic bias into these two analysis samples based on some differences in the characteristics of women who took part in the 5 original studies (see response above for more information). We have noted this in the text, but have not included a detailed description of the characteristics of each of the 5 study samples, since this paper is already at the maximum allowable length and the other reviewer did not find this point to be problematic. We don't feel that there is benefit in adding this point, but leave this to the discretion of the editor.

Comment: Also, if analyses where performed stratified by project this does not show in the result section or in the rest of the paper.
Analyses were not stratified by project but reported for the pooled sample.

Comment: In order to assess concurrent validity of the CASR-SF, authors compare with measures of depression, PTSD and coercive control. Results show similar concurrent validity as the CAS, which is good. However, considering that the main aim of the present study is to develop a brief measure that adequately captures the complexity of IPV, how come concurrent validity is not assessed in relation to other measures that consider physical sexual and psychological IPV rather than depression and PTSD? Please motivate.

We agree that it would have been ideal to use additional measures of IPV for validation purposes. As a secondary analysis of existing data, we used available data with priority given to measures that were common across all data sets. There were no common measures of IPV across these studies (other than the CAS). We have reported associations with coercive control (based on the WEB), an aspect of IPV, but the sample size is quite small as this measure was used in only a few studies. As we have noted in the discussion, additional validation of the CASR-SF is required.

Comment: Please ad a sentence concerning ethical considerations regarding the present study.
The original studies from which the data were drawn for this analysis were all reviewed and approved by appropriate Research Ethics Boards (REBs). We have noted this under "data sources" in the methods section for Phase 2.

Comment: Please revise discussion section according to the comments above. Also, please elaborate if there are any limitations related to the use of secondary data analyses.

As noted above, the use of secondary analysis placed limits on the data available for analysis. In

future validation studies of the CASR-SF, consideration should be given to incorporating a wider range of standardized measures for use in order to examine concurrent validity of the scale.

VERSION 2 – REVIEW

REVIEWER	KJersti Alsaker Bergen University College Norway
REVIEW RETURNED	28-Sep-2016

GENERAL COMMENTS	Development of a Brief Measure of Intimate Partner Violence Experiences: The Composite Abuse Scale (Revised) – Short Form (CASR-SF) This is a very important and well-designed article. However, there are still a few critical gaps that should be addressed. 1. Title The original CAS and this revised version focuses specifically on womens' experiences. A new keyword, "Intimate partner violence against women" is now added, which is good. However, "women" could also be mentioned in the abstract. 2. Discussion A. Pregnancy It is not about changing the new CAS-F only to discuss a limitation. Arguing that not all women get pregnant is weak, when it is important to discuss critical gaps and the serious effects of violence during pregnancy is well known. As you argue, whether a woman is pregnant or not should not make a difference to the way CARS-F performs, but in research it may be important to address this form of violence directly, asking specifically for violence directed at the pregnant woman's abdomen. Education It is good that financial abuse is addressed in the new brief version of CARS-F. However, obstacles for education or being denied education is not addressed. I agree that addressing financial abuse is a strength in the new CARS-F and that in many societies being denied education may be closed linked to financial abuse. However, being denied education may be linked to other issues as well, including being denied the possibility to develop one's own resources. That being said, I accept that there is a limit of 15 items and appreciate the fact that financial abuse is addressed.

VERSION 2 – AUTHOR RESPONSE

The following changes had be made to this manuscript based on the reviews provided:

1. The abstract has been revised to include the term 'women'' based on the reviewer's feedback and a conclusion added (as requested by the editor).

2. Dr. Alsaker also asked that we acknowledge in the discussion that failure to ask about injury to a pregnant women's abdomen due to abuse is a limitation of the CASR-SF. We do not agree with Dr. Alsaker that this is a limitation of the scale. In fact, her suggestion is not congruent with our goal of developing a brief general scale that is appropriate for diverse contexts (including pregnancy) and is based on experiences of abusive acts, and not the effects or impacts of those acts (such as injuries of

health problems). This is a different purpose. If researchers wish to capture the impacts of IPV (such as injuries or health problems, including spontaneous abortion or complications of pregnancy), this would need to be assessed separately from CASR-SF items. We have added a statement to the discussion to indicate that the CASR-SF measures abusive acts, and not the effects of these acts on the woman (in terms of injuries and/or health for example) in case this point is unclear to other readers.