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)	Incidence of unintended pregnancy among female sex workers in low- and middle-income countries: a systematic review and meta-analysis
AUTHORS	Ampt, Frances Willenberg, Lisa Agius, Paul Chersich, Matthew Luchters, Stanley Lim, Megan

VERSION 1 - REVIEW

REVIEWER	Alberto Madeiro
REVIEW RETURNED	11-Feb-2018

GENERAL COMMENTS	This is a systematic review and meta-analysis to determine the incidence of unintended pregnancy among female sex workers in low- and middle-income countries. Twenty-five eligible studies were identified between 2000 and 2016 in six online databases. The main results showed that methodological quality was low and unintended pregnancy incidence revealed high heterogeneity. This is an interesting and well written paper whose methodology is explained at all stages of the development of the manuscript in a consistent manner. The literature review is comprehensive, but it would be more interesting to develop a broader perspective. Furthermore, the authors could discuss in more depth the role of induced abortion within the context of sexual and reproductive health of this group of women.

REVIEWER	Eileen Yam
REVIEW RETURNED	21-Feb-2018

GENERAL COMMENTS	This is an under-researched topic, and the authors conduct a meticulous review of literature on a critical global public health challenge: unintended pregnancy among FSWs. My overarching recommendation is to reconsider the analysis of two separate outcomes (i.e., pregnancy with intention not determined, and unintended pregnancy) and, instead, simply use the primary outcome of pregnancy (i.e., examine all of the studies together). The analysis of just the 10 studies with unintended pregnancy as an outcome could be a secondary objective, for instance. I suggest this because although the article is framed largely from a

	family planning perspective with an eye towards reducing unintended pregnancy, we would be remiss if we did not acknowledge the substantial HIV-related concerns surrounding pregnancy among FSWs (regardless of whether intended or not). HIV prevalence is extremely high among FSWs. Yet there is little to no discussion of how, once pregnant, it is imperative that FSWs have access to PMTCT services, for example. Since the authors recognize how problematic it is to measure/define pregnancy "intent," it seems odd to then look at unintended pregnancy as a primary outcome. Regardless of whether intended or not, an analysis of pregnancy among FSWs has important programmatic and policy implications for HIV and FP programs alike. METHODS Analysis I felt this article was a rather dense read, and one reason for that is the large number of sub-analyses, some of which don't seem to warrant as much attention as they receive in this paper. For example, the nature of the various interventions is so varied, and it seems vastly oversimplified to categorize studies as simply "intervention vs no intervention." Even if there is statistical heterogeneity, this does not necessarily mean there is a conceptual basis for doing a subgroup analysis. It would help if the authors could provide clearer rationales for why the six different subgroup analyses were conducted. Alternatively, consider leaving some of them out. Table 1 It would help if consistent language were used throughout. For example, in some cases it says "FSWs without HIV" but in others it says "HIV-uninfected." Likewise with "in sex work" as opposed to "FSWs." What do the single asterisks refer to? It is confusing because there are multiple footnotes that seem to correspond to a single asterisk. In the row for Behets 2008, the column for "consistent condom use" says "0% in last 2/52." What does this mean? Likewise with Deschamps 2016. What does "partners in last 6/12" refer to? On p. 18 line 268-269, the text describing number of dependents and education levels indicates that these figures can be found in Table 1. But this table does not contain this information.

REVIEWER	Andrew Hinde
REVIEW RETURNED	23-Mar-2018

GENERAL COMMENTS	This is a well-written and extensively documented systematic review and meta-analysis of the incidence of unintended pregnancy among female sex workers in low and middle income countries. It is an important and little-researched topic, the latter partly because the study population is hard to reach, and implementing rigorous sampling methods is often not possible.

	I am supposed to be a statistical reviewer, but as a demographer I do know something about the measurement of fertility intentions, so I have a couple of substantive comments to contribute as well. Generally speaking, the statistical methods used in the paper (DerSimonian and Laird random effects models) seem appropriate for the range and size of studies being considered. At any rate, it is not clear to me that a more complicated approach is justified. I do have some detailed comments about their implementation, though. Statistical and methodological points p. 7, ll. 1-3 I agree with your decision to exclude cross-sectional studies, but the justification you give here is weak. A stronger justification would run as follows. There is a large literature about the measurement of whether pregnancies were intended or wanted. Few firm conclusions are available, but among the firmest is that prospective and retrospective measurement give different results, as women change their minds. This is particularly true of women with unstable expectation patterns, and I suspect many sex workers are in this category. Since cross-sectional studies must be retrospective, to include them would make an already complex analysis more so, as you would need to sub-set your studies into longitudinal and cross-sectional, as well as into cohort and randomised controlled trials. p. 9, ll. 21-22 '97 papers, with responses received for 54'. 97 – 54 = 43, but my interpretation of Figure 1 is that there were 48 studies from which responses were not received. I may be missing something here. Table 1, p. 11, bottom row. You say that the study by Gaffoor (2013) had 3% female sex workers. Should it have been included with this small proportion of the population of interest? Table 1, p. 12, fourth row. Can you justify including the study by Peterson (2007), when the population was not female sex workers, but a group of women who were working in areas where female sex work was common? My concern about the inclusion of this study is amplified by Figure 3, p. 41, where it is clearly an outlier among the randomised controlled trials. What would the I2 statistic be if the Petersen (2007) study was not included? p. 20, ll. 10-11 'The three studies of less than one year duration were more homogenous (I-squared=59.1%)'. They were not necessarily more homogenous, but because they had small exposure, the confidence intervals around the incidence rates were large and so it could not be determined statistically that the incidence rates were different. p. 18, ll. 2-3 'a median of five years of education, and the majority of women were supporting a least one financial dependent (table 1)'. Where is this in Table 1? I do not see any information about education or dependents. p. 21, ll. 9-10 'in the third study … over 85% of women who became pregnant… reported an abortion'. This is a remarkable statistic. So remarkable did I find it that I went back and looked at

	the paper by Van Damme et al. (2002) which was cited as the source. I could not find any mention of abortion (or termination) in this paper. Further, on p. 973 of Van Damme et al. (2002) there is a flow chart which indicates that 10 women withdrew because of pregnancy out of 892 in the study. Now, 85% of 10 is 8.5 which seems implausible, as I do not see how a woman can have half an abortion, or how half a woman can have an abortion. Moreover, there were 892 women in the Van Damme et al. study, and to reconcile the reported number of 10 pregnancies this with your incidence rate of 8.6 per 100 person years means that the total exposure time in the Van Damme study was 116 person years, giving an average exposure time for each woman in the study of 0.13 years. This seems very short. It may be that Van Damme or one of his colleagues provided more information when you contacted them, but I think this whole issue needs clarification. p. 24, ll. 2-3 'there was a marked absence of well-described sampling and recruitment techniques'. Is this surprising? Constructing a sample frame for female sex workers is not easy. Other points p. 4, l. 21 'misconceptions'. Given the subject of the paper, it might be better to use a different word, such as 'misunderstandings'. p. 5, l. 14 Define 'LMICs'. The superscript numbers in Tables 1 and 2, pp. 11-15 and 17 do not relate to any key or list of references. For example, in Table 1 on p. 11, row 1, 'Behets 2005' has the superscript '1' after it. What does the '1' mean? It does not lead me to the relevant reference in the bibliography. The same applies to all the numbers in Tables 1 and 2.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Alberto Madeiro, State University of Piauí, Brazil
Comment: This is a systematic review and meta-analysis to determine the incidence of unintended pregnancy among female sex workers in low- and middle-income countries. Twenty-five eligible studies were identified between 2000 and 2016 in six online databases. The main results showed that methodological quality was low and unintended pregnancy incidence revealed high heterogeneity. This is an interesting and well written paper whose methodology is explained at all stages of the development of the manuscript in a consistent manner.
Response: We thank the reviewer for this constructive praise.

Comment: The literature review is comprehensive, but it would be more interesting to develop a broader perspective.
Response: This systematic review and meta-analysis was conceptualised and written up in a pre-developed protocol with specifically defined aims and objectives, a clearly defined population and with pre-defined outcomes of interest. Providing a broader perspective could indeed have been more interesting for a broader audience, but would have risked losing its focus and potentially introduced more heterogeneity. The methodology used in a systematic review incurs the weakness noted by the reviewer: the research question is, per force, relatively narrow and broadening the question post hoc, is very difficult to do if one adheres to the systematic review methodology. The paper is already

lengthy and covers a number of issues, we feel broadening it would detract from the main messages of the paper. Many other papers cover broader aspects of FSW reproductive health.

Comment: Furthermore, the authors could discuss in more depth the role of induced abortion within the context of sexual and reproductive health of this group of women.
Response: We agree that induced abortion, particularly in the context of restrictive legal environments and the global gag rule, is an important health and rights issue for female sex workers. Incidence of abortion was a secondary outcome of this review, however only three studies captured these data. The review was not designed to collect other measures of induced abortion (such as reported lifetime abortions or how they were procured), precluding a more thorough analysis of these issues. Furthermore, the paper already explores a number of issues related to pregnancy incidence, and broadening it to discuss abortion in more detail would detract from its main messages.

Reviewer: 2: Eileen Yam, Deputy Director, Project Strengthening Operational AIDS Research (SOAR), Population Council, USA.

Comment: This is an under-researched topic, and the authors conduct a meticulous review of literature on a critical global public health challenge: unintended pregnancy among FSWs. My overarching recommendation is to reconsider the analysis of two separate outcomes (i.e., pregnancy with intention not determined, and unintended pregnancy) and, instead, simply use the primary outcome of pregnancy (i.e., examine all of the studies together). The analysis of just the 10 studies with unintended pregnancy as an outcome could be a secondary objective, for instance. I suggest this because although the article is framed largely from a family planning perspective with an eye towards reducing unintended pregnancy, we would be remiss if we did not acknowledge the substantial HIV-related concerns surrounding pregnancy among FSWs (regardless of whether intended or not).
HIV prevalence is extremely high among FSWs. Yet there is little to no discussion of how, once pregnant, it is imperative that FSWs have access to PMTCT services, for example. Since the authors recognize how problematic it is to measure/define pregnancy "intent," it seems odd to then look at unintended pregnancy as a primary outcome. Regardless of whether intended or not, an analysis of pregnancy among FSWs has important programmatic and policy implications for HIV and FP programs alike.
Response: We appreciate the thoughtful comment of the reviewer. The biggest source of heterogeneity in the paper stems from study methodology (whether studies measured pregnancy intent or not). This characteristic separates the studies into 2 groups, which differ considerably by pregnancy incidence, by study quality and by frequency of measurement of the pregnancy, for example. Moreover, in the original study design and the registered study protocol, we stipulated that we would analyse these groups separately. We acknowledge the limitation of measures of pregnancy intent, and also agree that grouping these studies may have simplified the analysis and interpretation. However, in the end, we decided to retain the present approach, principally as we believe that mixing these studies would bias the findings given the large variation in quality of these 2 groups. The group with undefined pregnancy intention should be considered as more of an exploratory or hypothesis-generating analysis, as intended by labelling it a ‘secondary outcome’.
We agree with the comment regarding the paper addressing the programmatic and policy implications for HIV and FP programs, including around PMTCT. Much of the paper, as with all systematic reviews, dwells on the methods used, sources of heterogeneity and limitations, rather than implications of the findings. We have text in the introduction on the importance of FP for FSWs, and in the conclusions section. We have added to this text, highlighting the points raised by the reviewer. Most of the introduction focuses on these issues.

Comment: I felt this article was a rather dense read, and one reason for that is the large number of sub-analyses, some of which don't seem to warrant as much attention as they receive in this paper. For example, the nature of the various interventions is so varied, and it seems vastly oversimplified to

categorize studies as simply "intervention vs no intervention." Even if there is statistical heterogeneity, this does not necessarily mean there is a conceptual basis for doing a subgroup analysis. It would help if the authors could provide clearer rationales for why the six different subgroup analyses were conducted. Alternatively, consider leaving some of them out.
Response: We agree that the way the 6 groups were presented originally was suboptimal. We have now grouped these outcomes into 2 groups: covariates (geographical region and intervention/no intervention) and methodological causes of heterogeneity (study design, study duration, and method of pregnancy measurement (measured regularly vs. only when indicated). Given this revised approach, we hope it allows the reader to understand the rationale for the sub-groups.

Comment: Table 1; It would help if consistent language were used throughout. For example, in some cases it says "FSWs without HIV" but in others it says "HIV-uninfected." Likewise with "in sex work" as opposed to "FSWs."
Response: We have carefully reviewed the table for consistency in wording and have amended this where appropriate.

Comment: What do the single asterisks refer to? It is confusing because there are multiple footnotes that seem to correspond to a single asterisk.
Response: This has now been addressed and footnotes are now more clearly linked.

Comment: In the row for Behets 2008, the column for "consistent condom use" says "0% in last 2/52." What does this mean?
Response: This refers to no participants reporting consistent condom use in the past two weeks. This has now been clarified.

Comment: Likewise with Deschamps 2016. What does "partners in last 6/12" refer to?
Response: This refers to the number of partners in the past 6 months (447 partners in last 6 months) and has now been clarified

Comment: On p. 18 line 268-269, the text describing number of dependents and education levels indicates that these figures can be found in Table 1. But this table does not contain this information.
Response: Apologies for this oversight as this information has been omitted to reduce the size of the table. The text now clarifies that this data is not shown.

Reviewer 3: Andrew Hinde, University of Southampton

Comment: This is a well-written and extensively documented systematic review and meta-analysis of the incidence of unintended pregnancy among female sex workers in low and middle income countries. It is an important and little-researched topic, the latter partly because the study population is hard to reach, and implementing rigorous sampling methods is often not possible.
Response: We appreciate the positive feedback from the reviewer and agree this is an important and little-researched topic.

Comment: I am supposed to be a statistical reviewer, but as a demographer I do know something about the measurement of fertility intentions, so I have a couple of substantive comments to contribute as well. Generally speaking, the statistical methods used in the paper (DerSimonian and Laird random effects models) seem appropriate for the range and size of studies being considered. At any rate, it is not clear to me that a more complicated approach is justified. I do have some detailed comments about their implementation, though. Statistical and methodological points; p. 7, ll. 1-3 I agree with your decision to exclude cross-sectional studies, but the justification you give here is weak. A stronger justification would run as follows. There is a large literature about the measurement of

whether pregnancies were intended or wanted. Few firm conclusions are available, but among the firmest is that prospective and retrospective measurement give different results, as women change their minds. This is particularly true of women with unstable expectation patterns, and I suspect many sex workers are in this category. Since cross-sectional studies must be retrospective, to include them would make an already complex analysis more so, as you would need to sub-set your studies into longitudinal and cross-sectional, as well as into cohort and randomised controlled trials.
Response: We agree with these important reviewer inputs and have added a sentence to the manuscript further justifying the exclusion of cross-sectional studies. It now states: "Cross-sectional studies were included in the initial screen for this purpose, but were subsequently excluded as there were sufficient longitudinal studies measuring incidence. The addition of period-prevalence in the last 12 months as an outcome would have required additional sub-analyses; in addition, measurement of retrospective pregnancy intention in cross-sectional studies differs from prospective measurement as women may change their minds during the course of their pregnancy."

Comment: p. 9, ll. 21-22 '97 papers, with responses received for 54'. 97 – 54 = 43, but my interpretation of Figure 1 is that there were 48 studies from which responses were not received. I may be missing something here.
Response: We thank the reviewer for cross-checking the numbers and agree that the figures didn't add up in the way it was described, and was due to the fact that some authors weren't directly contacted but contact had been made about the same study in a different paper. This line has now been removed, as the authors feel that it may cause confusion to the readers and has limited benefits.

Comment: Table 1, p. 11, bottom row. You say that the study by Gaffoor (2013) had 3% female sex workers. Should it have been included with this small proportion of the population of interest?
Response: The authors wish to refer the reviewer to the 'Inclusion and exclusion criteria' on page 6 of the manuscript where we described that "FSWs had to account for at least two thirds of the sample, unless data could be disaggregated by sex work status." The Gaffoor (2013) paper meets the inclusion criteria as data could be disaggregated by sex work status. We, however, were only able to include data for the 3% who actually reported sex work. The 41 in Table 1 represents the number of sex workers, not the total number of study participants. The same was done for Watson-Jones, Kaewkungwal and Robb, as all had less than 2/3 sex workers.

Comment: Table 1, p. 12, fourth row. Can you justify including the study by Peterson (2007), when the population was not female sex workers, but a group of women who were working in areas where female sex work was common? My concern about the inclusion of this study is amplified by Figure 3, p. 41, where it is clearly an outlier among the randomised controlled trials. What would the I2 statistic be if the Petersen (2007) study was not included?
Response: With respect to the study population, the authors of the Peterson et al. manuscript provided justification for considering most of the women as sex workers (or practicing transactional sex) when contacted via email. Specifically, they noted that one of the key eligibility criteria for the trial was that a woman had to report four or more partners in the month prior to screening and more than 50% of women at each site reported more than six partners in the month prior to screening. Although women were not asked directly if they were sex workers, there was a common understanding among the study staff that most of the enrolled women were sex workers or otherwise engaged in transactional sex for food or goods.
Although we believe it is difficult to determine visually if the study is an outlier, we have removed the Peterson study from the analysis of unintended pregnancy in accordance with the reviewer's suggestion and found that the I-squared value remained high (93.7%), but the I-squared for RCTs was indeed lower (50.4%). Based on this, we feel including the Peterson study remains justified.

Comment: p. 20, ll. 10-11 'The three studies of less than one year duration were more homogenous (I-squared=59.1%)'. They were not necessarily more homogenous, but because they had small

exposure, the confidence intervals around the incidence rates were large and so it could not be determined statistically that the incidence rates were different.
Response: We thank the reviewer for these considerate inputs, but would like to argue that a wide confidence interval does not necessarily mean the test for homogeneity is problematic. However, indeed the number of events in this sub-analysis is quite small. To address this concern, we've included a sentence in the limitations section of the manuscript which encourages readers to exercise caution when interpreting the heterogeneity estimates from the forest plots, specifically where the number of cases are small, and that these are descriptive statistics, and do not provide inferential data. Also, we have tried to soften some of the wording for interpretation of these sub-analyses to reflect this.

Comment: p. 18, ll. 2-3 'a median of five years of education, and the majority of women were supporting a least one financial dependent (table 1)'. Where is this in Table 1? I do not see any information about education or dependents.
Response: Apologies for this oversight; this information had been omitted to reduce the size of the table. The text now clarifies that this data is only provided in text and not presented in the table.

Comment: p. 21, ll. 9-10 'in the third study … over 85% of women who became pregnant… reported an abortion'. This is a remarkable statistic. So remarkable did I find it that I went back and looked at the paper by Van Damme et al. (2002) which was cited as the source. I could not find any mention of abortion (or termination) in this paper. Further, on p. 973 of Van Damme et al. (2002) there is a flow chart which indicates that 10 women withdrew because of pregnancy out of 892 in the study. Now, 85% of 10 is 8.5 which seems implausible, as I do not see how a woman can have half an abortion, or how half a woman can have an abortion. Moreover, there were 892 women in the Van Damme et al. study, and to reconcile the reported number of 10 pregnancies this with your incidence rate of 8.6 per 100 person years means that the total exposure time in the Van Damme study was 116 person years, giving an average exposure time for each woman in the study of 0.13 years. This seems very short. It may be that Van Damme or one of his colleagues provided more information when you contacted them, but I think this whole issue needs clarification.
Response: We appreciate the detailed review and apologise that it was unclear. The total pregnancies in this study was 72; 10 reported in the paper as you noted, as well as 62 reports of abortion as an adverse event, which was reported by the author in an email. There were 837.5 person-years of follow up in total (calculated by adding the total in the two trial arms in table 2 of the van Damme paper). As 765 were included in the analysis, this equates to 1.09 years of follow up per woman. Thus the incidence rate was 72/837.5 = 8.6.
The estimate of the proportion reporting an abortion is 62/72=86%. We agree that this is disproportionately high, and perhaps indicates that a number of pregnancies were missed or not reported (maybe the pregnancy testing was not as regular as indicated, or not performed on all women?). However, the author did not agree with this hypothesis when we contacted her.
The source of data in table 3 has now been clarified.

Comment: p. 24, ll. 2-3 'there was a marked absence of well-described sampling and recruitment techniques'. Is this surprising? Constructing a sample frame for female sex workers is not easy.
Response: Although not easy, it seems more recently greater efforts have been made to obtain a sample frame through enumeration and hotspot mapping, with subsequent time-location sampling or other more complex sampling methods. Also, several techniques have been developed to enhance the ability to obtain a representative sample of sex workers, and overcome some of the difficulties in developing a sampling frame. Much work has been done in this area and various studies employ respondent driven sampling. A useful paper describing these methods is: 'JMIR Public Health Surveill. 2017 Sample Size Calculations for Population Size Estimation Studies Using Multiplier Methods With Respondent-Driven Sampling Surveys. Fearon E, Chabata ST, Thompson JA, Cowan FM, Hargreaves JR.

Other points
Comment: p. 4, l. 21 ‘misconceptions’. Given the subject of the paper, it might be better to use a different word, such as ‘misunderstandings’.
Response: We have rephrased the word misconceptions as misunderstandings as requested.

Comment: p. 5, l. 14 Define ‘LMICs’.
Response: The abbreviation of LMIC was written in full in the first sentence of the introduction stating:” Unintended pregnancy affects a large number of women in low- and middle-income countries (LMICs), and can have significant impacts on maternal and child health.” No further definition has been provided as this terminology is well known, but we are willing to reconsider if the reviewers feel this is more appropriate.

Comment: The superscript numbers in Tables 1 and 2, pp. 11-15 and 17 do not relate to any key or list of references. For example, in Table 1 on p. 11, row 1, ‘Behets 2005’ has the superscript ‘1’ after it. What does the ‘1’ mean? It does not lead me to the relevant reference in the bibliography. The same applies to all the numbers in Tables 1 and 2.
Response: The reference numbering has now been corrected and included in both tables 1 and 2

VERSION 2 – REVIEW

REVIEWER	Alberto Madeiro
REVIEW RETURNED	30-May-2018

GENERAL COMMENTS	The re-submitted version of the paper improved quite substantially compared to the original submission. The authors have addressed all of my comments adequately.

REVIEWER	Eileen Yam
REVIEW RETURNED	16-May-2018

GENERAL COMMENTS	I feel the authors' responses and the subsequent revision are sufficient and I recommend publication.

REVIEWER	Andrew Hinde
REVIEW RETURNED	31-May-2018

GENERAL COMMENTS	Thank you for addressing those of my comments on the previous version which I considered critical to the acceptance of the paper. The others were more points for reflection, and to address them would need further analysis, which I do not advocate. I think the paper is now a very good contribution to the literature on this important topic. Research on the subject of pregnancy among sex workers is inherently difficult, and it is unsurprising that the methods used are varied and often far from 'textbook'. This paper brings together a wide range of studies to provide a comprehensive and balanced overview of the state of knowledge.