

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)	Sexual violence at each stage of human-trafficking cycle and associated factors: a retrospective cohort study on Ethiopian female returnees via three major trafficking corridors
AUTHORS	Gezie, Lemma Derseh; Worku, Alemayehu; Kebede, Yigzaw; Gebeyehu, Abebaw

VERSION 1 - REVIEW

REVIEWER	Nicola Pocock London School of Hygiene & Tropical Medicine, UK; United Nations University International Institute of Global Health, Malaysia
REVIEW RETURNED	13-Aug-2018

GENERAL COMMENTS	<p>Thank you to the authors for this interesting paper and for conducting this study, we need more studies which are conducted outside of service settings. There are however clarifications that need addressing before this paper can be published.</p> <p>I understand there are word count constraints – my comments relate mainly to expanding detail in the Methods section. The authors can significantly tighten the Introduction and wording there (if someone can review the English and writing), e.g. last paragraph in Introduction lines 45-51 could be cut as you go into that detail in methods.</p> <p>Please carefully review the manuscript for English before re-submitting, I am not commenting on this here but there are several instances where grammar syntax etc need addressing e.g. “in line with the leveling of other most literature”</p> <p>Abstract</p> <ul style="list-style-type: none">• Please specify exactly where participants were recruited (please see methods comments below) <p>Methods</p> <p>Please specify:</p> <p>Participant recruitment & study design -</p> <ul style="list-style-type: none">• Where exactly were participants recruited?<ul style="list-style-type: none">o Border towns are mentioned. But was this a household survey, did you conduct interviews at immigration offices at the border, etc? Please specify. Please comment on how privacy of conducting interviews was ensured at this location.• How was the study instrument designed? Did you build on/adapt the survey from a previous trafficking study? <p>Ethics -</p>
-------------------------	---

	<ul style="list-style-type: none"> • I saw the ethics statement is after the conclusion but this should be upfront in the Methods section. I understand the word limit constraints, but this is very important, as I had concerns reading the whole paper that ethical and safety considerations were not adhered to, IRB clearance was not there, until the end. • Ethics – were women repeatedly asked the same sexual violence 3 items for each trafficking stage (i.e. 4-5 times?) Or were the sexual violence items asked once, and “when did this happen? Pre departure, transit, during, post”. Please specify. This has some implications for retraumatization, when repeatedly asking the same question. • Who conducted the interviews, study PIs or RAs or both? How were interviewers trained to ask sensitive questions? • Were there referral pathways to health, reintegration, support services in place for the women interviewed? There should have been as per ethics recommendations in the WHO guide. How did you decide who to refer and when (were participants who disclosed sexual violence referred for example?) How was this information communicated to participants? <p>Variables and measurement -</p> <ul style="list-style-type: none"> • Please insert references on page 7, line 56 at end of “in line with the leveling of other most literature” • Statistician should comment on use of model and methods. I thought the bivariable screen (page 14, line 11) has been discredited as a method for deciding predictors that should be included in a multivariable model but would prefer a statistician to comment, also on whether the mixed effect model was appropriate as I cannot comment on this. <p>Results</p> <ul style="list-style-type: none"> • Please use one decimal place for all % findings, this will be easier to read. <p>Discussion</p> <ul style="list-style-type: none"> • Please include a paragraph on study limitations. I see this is upfront in the summary next to the abstract, but these should appear somewhere in the manuscript.
--	---

REVIEWER	Dr Ella Cockbain UCL (University College London)
REVIEW RETURNED	04-Nov-2018

GENERAL COMMENTS	<p>Review of “Sexual violence at each stage of human-trafficking cycle and associated factors: a retrospective follow-up study on Ethiopian female returnees via three major trafficking corridors”.</p> <p>This was an interesting and original study that has the potential to make a good contribution to the literature on human trafficking and public health. There was a good size sample and a wide range of variables was examined. I think the paper needs some fairly substantial revisions in order to be ready for publication. I would particularly encourage the authors to consider the following points to ensure that their paper is as clear, nuanced and robust as possible.</p> <p>Strengths and limitations: I think the line ‘It is also the first to examine various risk factors...’ is a little misleading. While some of the risk factors considered here are indeed novel, there are many prior studies that have looked at risk factors. I would suggest rephrasing to avoid ambiguity. E.g.</p>
-------------------------	---

	<p>Some of the potential risk factors analysed here have not previously been addressed in the literature.</p> <p>Introduction: The opening sentence was a bit clunky and not entirely clear and I would suggest rephrasing. I think the authors are simply trying to say that cases of human trafficking have been reported in 152 countries. I'm not convinced that the idea that trafficking has been recognised as a public health concern is quite that widespread so would want to see further supporting evidence if the argument was to remain the same.</p> <p>The authors draw on a wide range of relevant literature on human trafficking and public health. It would probably be helpful to acknowledge more explicitly that evidence on this topic is limited & fragmented. Many of the studies rely by necessity on convenience samples and are small in scale. The various studies mentioned were conducted in numerous different places worldwide and local context may play an important role. These caveats are worth acknowledging explicitly, lest the reader otherwise get the misleading impression that we have a clear and generalisable body of evidence on the health effects of trafficking.</p> <p>I would like to see more clarity around exactly why the authors think it is so important to look at sexual violence on a stage-by-stage basis. What are the key benefits in breaking it down in this way as opposed to considering prevalence across the whole trafficking period? I'd also like more clarity around exactly what is meant by pre-departure – i.e. lifetime pre-departure or say the previous x months before departure?</p> <p>'The trafficking of women or girls could be for the purpose of' – several forms of exploitation mentioned here but this is not an exhaustive list. It would be worth treating these as examples rather than the full set.</p> <p>Design and methods: It would help to have a more explicit justification of the study sites and a reflection elsewhere in the paper on the possible implications for the results and their interpretation.</p> <p>One of the key things missing from my perspective was information on the number of people who were approached to participate but declined to do so. In order better to understand how the final sample compared to the broader population of relevant returnees, it would be helpful to report at a minimum on participation rate (any other information that the authors might be able to provide on how participants differed from non-participants would of course also be useful, although I appreciate that they may not have such information.)</p> <p>The authors also talk about how the Palermo Protocol definition was used to identify relevant participants for inclusion. It might be useful to provide some further context on the overall volume of potential participants approached who were not found to meet this inclusion parameter (i.e. some insights into the possible prevalence of trafficking victims among broader migratory flows). The Palermo Protocol is also notoriously broad – it would be interesting to hear whether the authors encountered any difficulties operationalizing it consistently. Did the volume of trafficked to not trafficked potential</p>
--	--

	<p>participants vary by study site? Again this would be useful information to be given.</p> <p>It would be helpful to have a brief but explicit justification of the design to include only potential participants who had been trafficked in the previous 3 to 24 months. It would also help to report on how many potential participants were excluded on this parameter.</p> <p>It would be useful to have more information on how the surveys were administered (oral or written?) and by whom (gender, nationality, professional affiliation & any other relevant background on those collecting the data). Also whether there were any incentives for participation (and if so, what). All of these have possible implications for the results/sources of bias.</p> <p>Meaning of the sentence 'other variables were categorized in line with the leveling of other most literature' unclear. Please clarify.</p> <p>The 'Oslo support scale' is mentioned but without a supporting reference, please provide.</p> <p>I noted that 'restricted freedom of movement' was limited to cases where victims were physically confined & was interested in the rationale behind this decision. Elsewhere in the trafficking literature it has been documented that victims' movements can be confined through various psychological control mechanisms (threats, warnings, orders not to go out etc) – not just keeping them under lock and key (see, e.g., https://academic.oup.com/policing/article/12/2/129/3884650). Such measures can be very powerful.. It might help to acknowledge here that freedom of movement can be confined in various ways, not limited to the physical measures under investigation here. To be more accurate, it might therefore be useful also to rename this variable to something like 'physically confined' or 'physically restricted freedom of movement'.</p> <p>The final sentence of this section (beginning 'A returnee'. ...) was very long and quite hard to follow. Would help to simplify and clarify. Effectively what I understand is someone is only being considered smuggled if her initial journey out of Ethiopia involved smuggling. Is this correct? It might then also be useful to provide a more precise name for this variable (e.g. 'initially smuggled' rather than the broader 'smuggled').</p> <p>It would be useful to provide (if at all possible) some information on who was responsible for the sexual violence experienced at the various stages.</p> <p>Results This section should be renamed 'Results' rather than 'Result'.</p> <p>In Table 1 it is clear that 'age' refers to age at departure – the same clarification should be provided in the body text.</p> <p>For several other variables it was unclear whether they related to status at departure or at return (including marital status, parents alive, residence, occupation). Clarification would be very useful.</p>
--	---

	<p>‘Either illiterate or never attended formal education’ – surely these two are not mutually exclusive. ‘And/or’ might be more appropriate than ‘or’.</p> <p>Sentence ‘Of the four stages..’ appears to be missing a noun (had an average duration or length perhaps?).</p> <p>I would have found it very helpful to have information on the purpose for which victims were trafficked in Table 1 (ie the breakdown across different industries/occupations). For me, one of the most interesting findings of this study is that the rates of sexual violence were very high despite few (any?) of the participants having been trafficked into the commercial sex industry. I was surprised that the authors did not explore this result and its implications more explicitly/in more detail.</p> <p>The authors report on the incidence rate as x events per person weeks. I found this a little confusing – to my understanding the study had dealt with experience of sexual violence as a binary (yes/no) variable and the authors had not collected information on the number of sexually violent events experienced. In this respect, it felt misleading to report on incidence rate in this way as it potentially obscures differences between stages in the volume (incidence) of sexually violent incidents.</p> <p>I though the section ‘Risk factors of sexual violence’ could do with some more work. Despite using a regression analysis, the full results did not seem to have been included. Odds ratios were presented but I’d have liked to see clearer and more comprehensive presentation of results, including specific p values for each of the variables and the overall significance and goodness of fit of the model.</p> <p>Discussion Given that many of the previous studies mentioned did include substantial numbers of sex trafficking victims (some focused only on sex trafficking), it seemed odd not to make more of a discussion point out of how high the current study’s prevalence rate is considering very few of the participants had been trafficked for commercial sexual exploitation.</p> <p>I wasn’t convinced by the argument that ‘Probably, this might suggest that by the time trafficked females were at home, they experienced a similar risk of sexual violence as that of the general population in the country’. First, the ‘by the time .. were at home’ implies this point relates to post-trafficking whereas it deals in fact with pre-departure experienced. Second, if 10% is the lifetime prevalence rate for sexual violence, I would argue that 10% pre-departure rates could be fairly high for this group given that many are not particularly old (unless of course most of those subject to sexual violence across their lifetime do in fact experience it at a young age).</p> <p>I had similar concerns around the discussion of incidence rate per person weeks as to those I have already raised in relation to the results. Feels rather confusing.</p> <p>The discussion felt a bit underdeveloped and it might be useful to reflect on the implications for the results for policy & practice in supporting returning victims of trafficking, for example. While future</p>
--	---

	<p>research directions are dealt with briefly in the abstract, they are not considered in the discussion, which was a shame.</p> <p>Language: The paper was generally well written but a little clunky in places and there were some spelling, grammatical and typographical errors that would need correction before publication (eg 'sages' instead of 'stages', 'evidences sowed' instead of 'evidence showed' 'who left abroad' rather than 'who are left abroad' or 'who remain abroad' , 'remains to be a challenging practice' instead of simply 'remains a challenging practice' etc).</p> <p>The phrase 'their willingness to return' might usefully be replaced with the term 'voluntary return'.</p>
--	---

REVIEWER	Dr Rama Jayaraj College of Health and Human Sciences, Charles Darwin University, Australia
REVIEW RETURNED	14-Nov-2018

GENERAL COMMENTS	<p>Dear editor,</p> <p>This manuscript titled 'Sexual violence at each stage of human-trafficking cycle and associated factors: a retrospective follow-up study on Ethiopian female returnees via three major trafficking corridors" addresses the prevalence of sexual violence at each stage of trafficking cycle and risk factors associated with it among trafficked females. Information about the prevalence of sexual violence is limited due to a range of risk factors including lack of reporting or under reporting due to the methodologies and definitions employed in the numerous surveys or data sources</p> <p>Statistical Review: Good. The authors have used a generalized linear mixed model (GLMM) is an extension to the generalized linear model (GLM) in which the linear predictor contains random effects in addition to the usual fixed effects for each stage of trafficking cluster in this case over the ordinary logistic regression. In general, they estimate odds ratio with only year.</p> <p>The authors are trend setters and high innovative gone above the outstanding , did many tables</p> <p>Table 2. Prevalence of sexual violence and 95% confidence interval .quite high values and prevalence seems very high</p> <p>Table 3; most fascinating table demonstrating characteristics, such as</p> <p>Predeparture stage - age, marital status, Education, residence, travelling stage, Trafficking corridor, smuggled at departure, Destination stage -type of job, restricted freedom of movement, detention stage, length of destination,</p> <p>Few questions:</p> <p>Age cut off points varies between table 1 and 3: is there any specific reasons for this change?</p> <p>Any definition for informal education?</p> <p>Some of the characteristics (eg- length of detention) – you have to give some explanation some where?</p> <p>Kindly mention list of abbreviations</p> <p>Table 4 Comparison of Crude odds ratios (COR) and Adjusted odds ratio (AOR) and their 95% confidence interval (CI) for factors association with events of sexual violence- well done.</p>
-------------------------	--

REVIEWER	Steven McPhail Queensland University of Technology, Australia
REVIEW RETURNED	21-Jan-2019

GENERAL COMMENTS	<p>As requested, I have focused this review on description of methods related to statistical analyses (and reporting of findings from analyses), although have also proffered a discretionary comment on the conclusion of abstract. Most comments are generally minor, but some may be considered major. Overall, my impression is that the research is generally sound and makes a useful scientific contribution to what is a very important and under-researched field. I encourage the authors to continue their work in this field.</p> <p>Abstract</p> <p>Methods:</p> <p>1. (Minor) - Could potentially reword the end of the last sentence of abstract methods to say: “... examined using generalised linear mixed models”</p> <p>Results:</p> <p>2. (Minor) – There seems to be a false level of precision by reporting percentages to a hundredth of a percent? Perhaps better to just report percentages to a maximum of one decimal place (this applies throughout the manuscript).</p> <p>3. (Minor) – Rather than using a dash – between two separate values (e.g., between confidence intervals) it may be better to use the word ‘to’ in order to avoid confusion with negative symbols (also a dash in front of a number). For example, 31.5% to 38.7% may be clearer than 31.5 -38.7.</p> <p>4. (Minor) – It would probably be better to just state the age ranges of the age group classifications than only call them ‘girls’ or the ‘oldest group’ (e.g., ‘Girls aged <18 years were approximately twice as likely (AOR=1.97; 95%CI 1.11 to 3.52) to experience sexual violence than women aged >25’ ... or whatever would be appropriate descriptors if I have misunderstood.</p> <p>Conclusions:</p> <p>5. (Discretionary) – This is just my opinion, the authors may choose to ignore this: I wonder whether the higher priority for research to mention in the limited space of the abstract conclusion should relate to reproductive health problems... or something more like the need for research regarding important government policies / other preventative actions to prevent sexual violence from occurring among these girls /women from Ethiopia? Both are important topics for future research, and I appreciate it is easier to conduct research investigating complications than uncover / highlight (with research) genuine preventive solutions for what are very complex issues and circumstances. Nonetheless, I just wanted to raise this question as a prompt for the authors to consider whether their research priorities are primarily driven by ‘genuine priority for research’ vs ‘availability of data’. This is not a criticism of the authors who are doing a great job shining a light on a very important problem. The authors will be better positioned than me to know about research priorities in this specific area.</p> <p>Main text</p>
-------------------------	---

	<p>Sample size:</p> <p>6. (Major) I think the authors should provide a clearer explanation (with references) of their sample size calculation / estimate. I am not concerned with the study sample size (n=671), which I believe is certainly useful for addressing the research aims. I am just asking for clarification regarding the calculation that they have mentioned.</p> <p>Variables and measurement</p> <p>7. (Major) 3rd paragraph of this section: "other variables were categorized in line with the levelling of other most literature". I don't think this sentence makes sense, or at the very least, I do not know what it means and therefore cannot judge whether this is appropriate. Please revise / clarify this sentence.</p> <p>8. (Major) 4th paragraph of this section: For the Oslo support scale, it is not clear to me about which period of time the participants were rating their perceptions of the 'level of care they were receiving from others' (e.g., care before, during or after the trafficking period).</p> <p>9. (Major) 5th paragraph of this section: Further information is required to clarify how the composite wealth index was determined at a household level using PCA? Also, has this approach been used before / validated (etc)?</p> <p>Statistical analysis</p> <p>10. (Minor) The first sentence of the statistical analysis section is poorly worded and likely redundant. The authors may wish to remove it.</p> <p>11. (Major) GLMM is referred to as a singular model (although it seems like more than one model was prepared), without sufficient description of what analyses were actually conducted. (e.g., I presume the authors used a logistic family GLMM when examining factors associated with a binary outcome of whether or not participants experienced sexual violence?).</p> <p>12. (Major) The authors state that the 'gain in fit over the ordinary logistic regression was insignificant', but importantly the authors did not describe in the analysis section how this was determined (e.g., LR test?). I recommend adding the description of any statistical analyses conducted into the statistical analysis section.</p> <p>13. (Major) For each linear model analysis, please state clearly what were the fixed effect(s) and random effect(s). For random effects models, please clarify whether random slope or random intercept. Did the 'univariable' analysis include one fixed effect and one random effect (and the outcome measure) or were the univariable models just fixed effects (no random effect). Also, perhaps it would be better to use a phrase like: 'factors with $p < 0.20$ were carried forward for inclusion in the multivariable analysis' rather than defining 0.20 as a 'level of significance'?</p> <p>Results</p> <p>14. (Minor) I advise against using the term 'average', which would perhaps be better replaced by the term "mean" throughout (assuming this matches the intended meaning).</p>
--	--

	<p>15. (Minor) At least some of the SD's are implausibly small. I strongly suspect there is an error and advise the authors to check this (be cautious no to confuse SEM with SD) Risk factors for sexual violence</p> <p>16. (Minor) I see the term 'bi-variable' analysis, perhaps re-word this for consistent use of terminology across the analysis and results sections?</p> <p>17. (Minor) Third sentence of this section: "As the (GLMM) did not show a gain in fit..." – has this already been described elsewhere / is it required here also? Perhaps just write it once if this is the case.</p> <p>18. (Major) The choice of age category cut-offs has not yet been justified? Was it consistent with prior research in the field or was there some other reason for selecting it?</p> <p>Table 4</p> <p>19. (Discretionary) I am not sure whether the 'referent' category is best expressed as 1.0 or simply described as 'referent'.</p> <p>General language comment – the manuscript would likely benefit from English language proof reading / editing.</p>
--	--

REVIEWER	Dr Marcos Vera-Hernandez University College London, UK
REVIEW RETURNED	05-Feb-2019

GENERAL COMMENTS	<p>I performed a statistical review so I did not look at the references outside the statistics or the supplementary reporting.</p> <p>Statistical review on the paper "Sexual violence at each stage of human-trafficking cycle and associated factors: a retrospective follow-up study on Ethiopian female returnees via three major trafficking corridors" submitted to BMJ Open.</p> <p>I found the statistics methods used in the paper to be correct and valid. I only have some minor comments.</p> <p>-Provide cite for formulae used to compute sample size.</p> <p>-On the statement "a composite wealth index was determined at household level using principal component analysis," clarify what household it was. The household the woman used to live in Ethiopia, or the household she is returning to?</p> <p>-Please review sentence "Descriptive and summary statistics were done using texts and tables"</p>
-------------------------	--

VERSION 1 – AUTHOR RESPONSE

Reviewer 1		
1	Introduction and wording there (if someone can review the English and writing), e.g. last paragraph in Introduction lines 45-51 could be cut as you go into that detail in methods	Thank you. We did it accordingly
2	Please carefully review the manuscript for English before re-submitting, I am not commenting on this here but there are several instances where grammar syntax etc need addressing e.g. “in line with the leveling of other most literature”	Efforts were made to improve its English.
3	Please specify exactly where participants were recruited (please see methods comments below)	Thank you. Now it has been provided at the end of the first paragraph of the method section.
4	Please specify: Participant recruitment & study design -	Now, it is Included
5	Where exactly were participants recruited? Border towns are mentioned. But was this a household survey, did you conduct interviews at immigration offices at the border, etc? Please specify. Please comment on how privacy of conducting interviews was ensured at this location.	As mentioned above, now it has been added at the end of the first paragraph of the methods section.
6	How was the study instrument designed? Did you build on/adapt the survey from a previous trafficking study?	For this and other related concerns, we have added a sub-section called Data Collection Tools and Procedures in the methods section.
6	I saw the ethics statement is after the conclusion but this should be upfront in the Methods section. I understand the word limit constraints, but this is very important, as I had concerns reading the whole paper that ethical and safety considerations were not adhered to, IRB clearance was not there, until the end.	We have placed the Ethics section at the end of the Methods section
7	Ethics – were women repeatedly asked the same sexual violence 3 items for each trafficking stage (i.e. 4-5 times?) Or were the sexual violence items asked once, and “when did this happen? Pre departure, transit, during, post”. Please specify. This has some implications for re-traumatization, when repeatedly asking the same question.	Returnees were contacted and asked only once at the time of return at the border. They were encouraged to respond by recalling (from their memory) their experiences at each stage of trafficking. We briefed them about sexual violence and ask whether they experienced similar problems during predeparture, travel, destination, and detention stages. They were not repeatedly asked.
8	Who conducted the interviews, study PIs or RAs or both? How were interviewers trained to ask sensitive questions?	Now we have added about it in the Data Collection Tools and Procedures section.
9	Were there referral pathways to health, reintegration, support services in place for the women interviewed? There should have been as per ethics recommendations in the WHO guide. How did you decide who to refer and when (were participants who	We informed them how to get into the reintegration system, especially how to be organized in groups, getting work places, and budget from the government. At

	disclosed sexual violence referred for example?) How was this information communicated to participants?	the end of the interview, data collectors also recommend what to do, especially if they experienced, unprotected sex, injuries, and other health problems. In some severe cases, we took them into the nearby health facility.
Variables and measurement		
10	Please insert references on page 7, line 56 at end of "in line with the leveling of other most literature"	So, revisions were made.
Results		
11	Please use one decimal place for all % findings, this will be easier to read.	Corrected.
Discussion		
12	Please include a paragraph on study limitations. I see this is upfront in the summary next to the abstract, but these should appear somewhere in the manuscript.	Now, we have given it a title at the end of the Discussion section so that it can be seen easily.
Reviewer 2		
Strengths and limitations:		
1	I think the line 'It is also the first to examine various risk factors...' is a little misleading. While some of the risk factors considered here are indeed novel, there are many prior studies that have looked at risk factors. I would suggest rephrasing to avoid ambiguity. E.g. Some of the potential risk factors analysed here have not previously been addressed in the literature.	Corrected
2	Introduction: The opening sentence was a bit clunky and not entirely clear and I would suggest rephrasing. I think the authors are simply trying to say that cases of human trafficking have been reported in 152 countries. I'm not convinced that the idea that trafficking has been recognised as a public health concern is quite that widespread so would want to see further supporting evidence if the argument was to remain the same.	We have reworded it.
3	Design and methods: It would help to have a more explicit justification of the study sites and a reflection elsewhere in the paper on the possible implications for the results and their interpretation.	Thank you. We have added in the middle of the first paragraph in the Methods section.
4	One of the key things missing from my perspective was information on the number of people who were approached to participate but declined to do so. In order better to understand how the final sample compared to the broader population of relevant returnees, it would be helpful to report at a minimum on participation rate (any other information that the authors might be able to provide on how participants differed from non-participants would of course also be useful, although I appreciate that they may not have such information.)	We have provided the reference (for the larger study) and the response rate as 90.95% in the larger study. Unfortunately, we do not have pre-organized data about the reasons for non-response. However, mainly it was not related to trafficking consequences. Instead, it is something external such as

		shortage of time to interview before they start moving home from border sites.
5	The authors also talk about how the Palermo Protocol definition was used to identify relevant participants for inclusion. It might be useful to provide some further context on the overall volume of potential participants approached who were not found to meet this inclusion parameter (i.e. some insights into the possible prevalence of trafficking victims among broader migratory flows). The Palermo Protocol is also notoriously broad – it would be interesting to hear whether the authors encountered any difficulties operationalizing it consistently. Did the volume of trafficked to not trafficked potential participants vary by study site? Again, this would be useful information to be given.	Again, we have provided our published paper as a reference that contain the detail and also reported that about half of the returnees were trafficked.
6	It would be helpful to have a brief but explicit justification of the design to include only potential participants who had been trafficked in the previous 3 to 24 months. It would also help to report on how many potential participants were excluded on this parameter.	Thank you. It is done accordingly.
7	It would be useful to have more information on how the surveys were administered (oral or written?) and by whom (gender, nationality, professional affiliation & any other relevant background on those collecting the data). Also whether there were any incentives for participation (and if so, what). All of these have possible implications for the results/sources of bias.	We have provided about all these under the sub-topic called Data Collection Tools and Procedures.
8	Meaning of the sentence 'other variables were categorized in line with the leveling of other most literature' unclear. Please clarify.	Now, we made it clearer
9	The 'Oslo support scale' is mentioned but without a supporting reference, please provide.	Now, we have inserted the reference
10	I noted that 'restricted freedom of movement' was limited to cases where victims were physically confined & was interested in the rationale behind this decision. Elsewhere in the trafficking literature it has been documented that victims' movements can be confined through various psychological control mechanisms (threats, warnings, orders not to go out etc) – not just keeping them under lock and key (see, e.g., https://academic.oup.com/policing/article/12/2/129/3884650). Such measures can be very powerful.. It might help to acknowledge here that freedom of movement can be confined in various ways, not limited to the physical measures under investigation here. To be more accurate, it might therefore be useful also to rename this variable to something like 'physically confined' or 'physically restricted freedom of movement'.	Dear reviewer, our intention was also just like your interest. If you see at the end of the definition we used, it also includes other types of restrictions (i.e. not to do what she wanted) but not on physical restriction only. The following is a part of the definition we used "... or never be free to go anywhere she wanted or do what she wanted" Other literature which was published in Lancet also used this measurement.
11	The final sentence of this section (beginning 'A returnee'. ...) was very long and quite hard to follow. Would help to simplify and clarify. Effectively what I understand is someone is only being	

	considered smuggled if her initial journey out of Ethiopia involved smuggling. Is this correct? It might then also be useful to provide a more precise name for this variable (e.g. 'initially smuggled' rather than the broader 'smuggled').	Thank you. After getting your comment, we have revised in as initially smuggled.
12	It would be useful to provide (if at all possible) some information on who was responsible for the sexual violence experienced at the various stages.	It was a very good idea. Unfortunately, we don't have the data.
13	Results This section should be renamed 'Results' rather than 'Result'.	Corrected
14	In Table 1 it is clear that 'age' refers to age at departure – the same clarification should be provided in the body text.	We have made it clear for all variables in the title of Table 1: 'Table 1: Predeparture characteristics of sampled female returnees ...'. We have also added a legend.
15	For several other variables it was unclear whether they related to status at departure or at return (including marital status, parents alive, residence, occupation). Clarification would be very useful.	We have clarified it in the title of Table 1. We also added a legend.
16	'Either illiterate or never attended formal education' – surely these two are not mutually exclusive. 'And/or' might be more appropriate than 'or'.	Corrected. However, we want you to be informed that it says, 'informal education' but not 'never attended formal education'. It is about religious education and we added it.
17	Sentence 'Of the four stages.' appears to be missing a noun (had an average duration or length perhaps?).	We rewrote it.
18	I would have found it very helpful to have information on the purpose for which victims were trafficked in Table 1 (ie the breakdown across different industries/occupations). For me, one of the most interesting findings of this study is that the rates of sexual violence were very high despite few (any?) of the participants having been trafficked into the commercial sex industry. I was surprised that the authors did not explore this result and its implications more explicitly/in more detail.	We put their occupations in Table 3 with the corresponding rates of sexual violence. Unfortunately, very few returnees reported commercial sex industry as their occupation, maybe because it is sensitive issue.
19	The authors report on the incidence rate as x events per person weeks. I found this a little confusing – to my understanding the study had dealt with experience of sexual violence as a binary (yes/no) variable and the authors had not collected information on the number of sexually violent events experienced. In this respect, it felt misleading to report on incidence rate in this way as it potentially obscures differences between stages in the volume (incidence) of sexually violent incidents.	Clarification is added at the end of the paragraph below Table 2.

20	I though the section 'Risk factors of sexual violence' could do with some more work. Despite using a regression analysis, the full results did not seem to have been included. Odds ratios were presented but I'd have liked to see clearer and more comprehensive presentation of results, including specific p values for each of the variables and the overall significance and goodness of fit of the model.	We have added the overall model adequacy test we used (likelihood ratio test) in the Methods section, under Statistical analysis. Regarding specific p-values, we have added them in the last column of Table 4. (Our intention was the CIs were sufficient).
21	Discussion Given that many of the previous studies mentioned did include substantial numbers of sex trafficking victims (some focused only on sex trafficking), it seemed odd not to make more of a discussion point out of how high the current study's prevalence rate is considering very few of the participants had been trafficked for commercial sexual exploitation.	Some discussions about the occupation of commercial sex trafficking is made under the Limitation section, at the end of the first paragraph.
22	I wasn't convinced by the argument that 'Probably, this might suggest that by the time trafficked females were at home, they experienced a similar risk of sexual violence as that of the general population in the country'. First, the 'by the time ... were at home' implies this point relates to post-trafficking whereas it deals in fact with pre-departure experienced. Second, if 10% is the lifetime prevalence rate for sexual violence, I would argue that 10% pre-departure rates could be fairly high for this group given that many are not particularly old (unless of course most of those subject to sexual violence across their lifetime do in fact experience it at a young age).	Now, further clarifications are provided; i.e. it has been discussed in the context of both prevalence and rate (i.e. taking the length of exposure time into account).
23	I had similar concerns around the discussion of incidence rate per person weeks as to those I have already raised in relation to the results. Feels rather confusing.	It has been cleared in the Result section, under the subtopic 'Sexual violence among trafficked females' at the end of the second paragraph.
24	The discussion felt a bit underdeveloped and it might be useful to reflect on the implications for the results for policy & practice in supporting returning victims of trafficking, for example. While future research directions are dealt with briefly in the abstract, they are not considered in the discussion, which was a shame.	Two paragraphs of implications were added, one for prevalence of sexual violence and another for its risk factors.
25	The paper was generally well written but a little clunky in places and there were some spelling, grammatical and typographical errors that would need correction before publication (eg 'sages' instead of 'stages', 'evidences sowed' instead of 'evidence showed' 'who left abroad' rather than 'who are left abroad' or 'who remain abroad', 'remains to be a challenging practice' instead of simply 'remains a challenging practice' etc). The phrase 'their willingness to return' might usefully be replaced with the term 'voluntary return'.	Corrected accordingly.
Reviewer 3		
1	Age cut off points varies between table 1 and 3: is there any specific reasons for this change?	Thank you; corrected.

2	Any definition for informal education?	Added. It means 'religious education'.
3	Some of the characteristics (eg- length of detention) – you have to give some explanation somewhere?	It means the length of time in weeks that returnees were detained by border or government securities.
4	Kindly mention list of abbreviations	Thank you. However, we are not sure the right place for the Abbreviations. So we put the list next to Conclusion (Discussion).
5	Table 4 Comparison of Crude odds ratios (COR) and Adjusted odds ratio (AOR) and their 95% confidence interval (CI) for factors association with events of sexual violence- well done	Thank you.
Reviewer 4		
1	Abstract Methods: (Minor) - Could potentially reword the end of the last sentence of abstract methods to say: "... examined using generalised linear mixed models"	Thank you and we have Paraphrased it.
2	Results: (Minor) – There seems to be a false level of precision by reporting percentages to a hundredth of a percent? Perhaps better to just report percentages to a maximum of one decimal place (this applies throughout the manuscript).	Revised.
3	(Minor) – Rather than using a dash – between two separate values (e.g., between confidence intervals) it may be better to use the word 'to' in order to avoid confusion with negative symbols (also a dash in front of a number). For example, 31.5% to 38.7% may be clearer than 31.5 -38.7.	Dear reviewer, that is possible. However, if it doesn't create any incontinence, can we leave it as it is? Because we don't have any negative values in the whole document as we reported OR, percent, or counts only. Thank you.
4	(Minor) – It would probably be better to just state the age ranges of the age group classifications than only call them 'girls' or the 'oldest group' (e.g., 'Girls aged <18 years were approximately twice as likely (AOR=1.97; 95%CI 1.11 to 3.52) to experience sexual violence than women aged >25' ... or whatever would be appropriate descriptors if I have misunderstood.	Corrected
5	Conclusions: (Discretionary) – This is just my opinion; the authors may choose to ignore this: I wonder whether the higher priority for research to mention in the limited space of the abstract conclusion should relate to reproductive health problems... or something more like the need for research regarding important government policies / other preventative actions to prevent sexual violence from	Following this important comment, we have recommended only 'prevention methods for sexual violence during trafficking' in the abstract. However, in the major text (under Discussion) we

	<p>occurring among these girls /women from Ethiopia? Both are important topics for future research, and I appreciate it is easier to conduct research investigating complications than uncover / highlight (with research) genuine preventive solutions for what are very complex issues and circumstances. Nonetheless, I just wanted to raise this question as a prompt for the authors to consider whether their research priorities are primarily driven by 'genuine priority for research' vs 'availability of data'. This is not a criticism of the authors who are doing a great job shining a light on a very important problem. The authors will be better positioned than me to know about research priorities in this specific area</p>	<p>recommended also 'investigation of consequences of violence'</p>
6.	<p>Main text Sample size: (Major) I think the authors should provide a clearer explanation (with references) of their sample size calculation / estimate. I am not concerned with the study sample size (n=671), which I believe is certainly useful for addressing the research aims. I am just asking for clarification regarding the calculation that they have mentioned.</p>	<p>We just used the usual sample size formula for proportion. We took a proportion of 50% as it gives the maximum sample size given a margin of error and a confidence level (alpha). The reason we used a design effect was because we used two stage sampling, i.e. cluster of trafficking corridors and individuals. Moreover, we have also inserted a reference as per your comment.</p>
7.	<p>Variables and measurement (Major) 3rd paragraph of this section: "other variables were categorized in line with the levelling of other most literature". I don't think this sentence makes sense, or at the very least, I do not know what it means and therefore cannot judge whether this is appropriate. Please revise / clarify this sentence.</p>	<p>Now, clarification has been added. Thank you.</p>
8.	<p>(Major) 4th paragraph of this section: For the Oslo support scale, it is not clear to me about which period of time the participants were rating their perceptions of the 'level of care they were receiving from others' (e.g., care before, during or after the trafficking period).</p>	<p>It was about 'before departure', and following your important comment, clarification has been added.</p>
9	<p>. (Major) 5th paragraph of this section: Further information is required to clarify how the composite wealth index was determined at a household level using PCA? Also, has this approach been used before / validated (etc.)?</p>	<p>We generated the composite wealth index by using the following steps: First, we identified the items that are common to both urban and rural dwellers. Fortunately, all the 11 items were common to both urban and rural as none of them were < 5% or > 95% in each of the two settings.</p>

		<p>Second, we generated the first component separately for urban and rural using the 'pca' command in Stata. In doing so we generated two separate factors (components) of wealth index scores, for urban and rural.</p> <p>Third, we merged the two components (factors) to get the composite wealth index for both urban and rural.</p> <p>Fourth, to get the wealth quantiles, we used the 'xtile' command in Stata so that we would have categorized (leveled) wealth index.</p> <p>However, because we have a word count constraint, we couldn't write all these in the manuscript.</p>
10.	<p>Statistical analysis (Minor) The first sentence of the statistical analysis section is poorly worded and likely redundant. The authors may wish to remove it.</p>	<p>Now, we have paraphrased it.</p>
11.	<p>(Major) GLMM is referred to as a singular model (although it seems like more than one model was prepared), without sufficient description of what analyses were actually conducted. (e.g., I presume the authors used a logistic family GLMM when examining factors associated with a binary outcome of whether or not participants experienced sexual violence?).</p>	<p>Thank you. Now clarification is added as: a logit family GLMM ...</p>
12.	<p>(Major) The authors state that the 'gain in fit over the ordinary logistic regression was insignificant', but importantly the authors did not describe in the analysis section how this was determined (e.g., LR test?). I recommend adding the description of any statistical analyses conducted into the statistical analysis section.</p>	<p>Yes, we have added it now. It was LR test.</p>
13.	<p>(Major) For each linear model analysis, please state clearly what were the fixed effect(s) and random effect(s). For random effects models, please clarify whether random slope or random intercept. Did the 'univariable' analysis include one fixed effect and one random effect (and the outcome measure) or were the univariable models just fixed effects (no random effect). Also, perhaps it would be better to use a phrase like: 'factors with $p < 0.20$ were carried forward for inclusion in the multivariable analysis' rather than defining 0.20 as a 'level of significance'?</p>	<p>We have made it clear now in the manuscript.</p> <p>The first model was Fixed effect and Random intercept (i.e. for the random effect)</p> <p>The second model was Fixed effect and both Random intercept and random slope. Both of the models were compared to each other as well as to the conventional logistic regression using LR test.</p>

14.	Results (Minor) I advise against using the term 'average', which would perhaps be better replaced by the term "mean" throughout (assuming this matches the intended meaning).	Corrected.
15.	(Minor) At least some of the SD's are implausibly small. I strongly suspect there is an error and advise the authors to check this (be cautious no to confuse SEM with SD)	Thank you very much. Corrections were made in two SDs. Really this is an amusing observation!
Risk factors for sexual violence		
16.	(Minor) I see the term 'bi-variable' analysis, perhaps re-word this for consistent use of terminology across the analysis and results sections?	Corrected.
17.	(Minor) Third sentence of this section: "As the (GLMM) did not show a gain in fit..." – has this already been described elsewhere / is it required here also? Perhaps just write it once if this is the case.	Corrected.
18.	(Major) The choice of age category cut-offs has not yet been justified? Was it consistent with prior research in the field or was there some other reason for selecting it?	Now, we provided reason for it in the Methods section, under subtitle "Variables and measurement" in the fourth paragraph, and first sentence
19.	(Discretionary) I am not sure whether the 'referent' category is best expressed as 1.0 or simply described as 'referent'.	Well, following this comment we changed '1' to 'Referent'. Perhaps, it may be more explanatory.
20	General language comment – the manuscript would likely benefit from English language proof reading / editing	Thank you. Efforts were made to improve it.
Reviewer: 5		
1	-Provide cite for formulae used to compute sample size.	Now, it has been provided
2	-On the statement "a composite wealth index was determined at household level using principal component analysis," clarify what household it was. The household the woman used to live in Ethiopia, or the household she is returning to?	Now, again clarification is given. It was about the household at the pre-departure stage.
3	-Please review sentence "Descriptive and summary statistics were done using texts and tables"	Corrected!

VERSION 2 – REVIEW

REVIEWER	Dr Ella Cockbain Department of Security & Crime Science, University College London (UCL), UK.
REVIEW RETURNED	03-Apr-2019

GENERAL COMMENTS	<p>This version of the manuscript is much improved and I was pleased to see the authors respond so thoroughly to the various reviewers' comments and recommendations. It's an interesting paper and I think it is nearly ready for publication. I think it still needs a final few changes but they should be quick to implement. List below:</p> <ul style="list-style-type: none"> • P.4 Change 'it also results in acquiring HIV' to 'it can also result' – the former is inaccurate and stigmatising. Similarly I didn't think the way of referring to victims as creating 'an extra burden' was very helpful – felt quite stigmatising too. More neutral language would be preferable. • I'm not convinced that 'detention' and 'integration/reintegration' are 'stages of the trafficking process' – surely 'the trafficking and post-trafficking process' would be more accurate? But also surely detention is not inevitable, which characterising it as a core stage of the process implies it is? This should be addressed throughout. I also wondered about the implications in terms of the overall prevalence rates presented for 'the whole trafficking period' (p.13).- surely this should be 'the whole pre-, during and post trafficking period' • The conclusions section of the abstract is more or less a direct repetition of the results section of abstract – I think it would be a better use of the space to talk about the implications here. • P 6 line 12 – I'm not convinced the authors can claim to get 'a representative sample'. Without clear sampling frame etc this is still a convenience sample (albeit a broader one than one that relies on support services/small groups alone). Perhaps 'a more representative sample' would be more accurate? • In the response to reviewers I note that the authors clarified that the 'restricted freedom of movement' variable covered not just physical restrictions but also psychological ones, threats etc to stop people going out. This is really important and needs to be made explicit in the paper. • P. 15 'univariable analysis' – I was a little confused here. Surely this was a bivariable analysis looking at their relationship to sexual violence? Or not? • P.18 the idea that conditions in Global North 'encourage them to report every event of sexual violence duly' just doesn't happen. Sexual offences are notoriously under-reported. I would adjust this line – perhaps to something like 'report sexual violence more often, although under-reporting remains a problem here too' • P. 19 How realistic is it for trafficking victims to be provided healthcare services 'at any time starting from the date of departure'? I suspect it's commendable but unlikely – in which case I'd add the caveat 'although it is recognised that this may be very difficult to do in practice' (or similar). • P.20. It would be good – as I think another reviewer noticed – to address (even just briefly) in the discussion the need for more research not just on responses to victims but how to prevent trafficking from Ethiopia in the first place. Given the high harms
-------------------------	--

	<p>caused, it would seem sensible to invest in research into better prevention too.</p> <ul style="list-style-type: none"> • English much improved but still some typographic/spelling/grammatical errors that need fixing (e.g. 'from Ethiopian' in abstract line 25, 'deny of wage' p.6.) and also a couple of places where the language was a barrier to comprehension: P.10 Line 'we assumed that the missing mechanism of data was non-informative' reads quite strangely and not intuitively understandable though I think I know what they mean (ie.. that unlikely to be systematic differences between those who agreed to and declined to participate). Could this be simplified please?; More clarity is needed similarly around the line p.13 'The sexual violence data were binary type (yes/no) but not about the number of episodes of sexual violence'. Could the latter part be simplified to 'but participants were asked about the number of incidents of sexual violence at each stage' (if that is indeed what is meant here). • P.4 line 27: recommend changing 'for sex trafficking' to 'for sexual exploitation' ('trafficked... for sex trafficking sounds odd!) • P.5 Line 38 'wider population' – presumably this should be 'wider sample' • p.18 both 'traffickers' should be changed to 'traffickers/smugglers' as that seems to be the actual group being referred to here.
--	---

REVIEWER	Steven McPhail Queensland University of Technology and Metro South Health
REVIEW RETURNED	29-Mar-2019

GENERAL COMMENTS	<p>I appreciate the authors' efforts to be responsive to most of the previous suggestions. My main remaining concern relates to my original comment: "Further information is required to clarify how the composite wealth index was determined at a household level using PCA? Also, has this approach been used before / validated (etc.)?" To which the authors gave an explanation in the response document, but in the end stated: "However, because we have a word count constraint, we couldn't write all these in the manuscript."</p> <p>While a long-winded explanation is not required in the manuscript, enough information is required to enable someone else to replicate the process as well as be able to make their own judgement about it's scientific merit if not a widely used approach for the items used in this study. I don't believe this is currently the case. Two potential solutions may be:</p> <ol style="list-style-type: none"> 1) Succinctly provide sufficient information for an independent reader to be able to replicate the process in the manuscript text and / or 2) add a supplementary table / appendix / whatever approach the publisher recommends, as an additional 'downloadable' file that provides sufficient information for someone to be able to replicate the process. <p>I will leave any other minor things to copy editors.</p> <p>Overall, I think this paper will make a good contribution to the field. The authors are to be congratulated.</p>
-------------------------	--

REVIEWER	Marcos Vera-Hernandez University College London United Kingdom
REVIEW RETURNED	25-Mar-2019

GENERAL COMMENTS	My previous comments have been adequately addressed.
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

Point by Point response

Reviewer 4:

Comment: I appreciate the authors' efforts to be responsive to most of the previous suggestions. My main remaining concern relates to my original comment:

"Further information is required to clarify how the composite wealth index was

determined at a household level using PCA? Also, has this approach been used before / validated (etc.)?" To which the authors gave an explanation in the response document, but in the end stated: "However, because we have a word count constraint, we couldn't write all these in the manuscript."

While a long-winded explanation is not required in the manuscript, enough information is required to enable someone else to replicate the process as well as be able to make their own judgement about its scientific merit if not a widely used approach for the items used in this study. I don't believe this is currently the case. Two potential solutions may be:

- 1) Succinctly provide sufficient information for an independent reader to be able to replicate the process in the manuscript text and / or
- 2) add a supplementary table / appendix / whatever approach the publisher recommends, as an additional 'downloadable' file that provides sufficient information for someone to be able to replicate the process. I will leave any other minor things to copy editors.

Overall, I think this paper will make a good contribution to the field. The authors are to be congratulated.

Response: Dear reviewer, thank you very much for your concern about the sufficiency of the description we provided regarding how the wealth index was calculated. Following your comment, we have preferred your second suggestion i.e. adding supplementary or an additional downloadable file that describes how we did it. We believed that it will make the manuscript bulky if we incorporate it. It

is also difficult to present it succinctly as it will not be clear for readers. Its file name is 'wealth_index_determination'.

Reviewer: 2

Comment: P.4 Change 'it also results in acquiring HIV' to 'it can also result' – the former is inaccurate and stigmatising. Similarly I didn't think the way of referring to victims as creating 'an extra burden' was very helpful – felt quite stigmatising too. More neutral language would be preferable.

Response: Accepted, and we made paraphrasing for the two statements.

Comment: I'm not convinced that 'detention' and 'integration/reintegration' are 'stages of the trafficking process' – surely 'the trafficking and post-trafficking process' would be more accurate? But also surely detention is not inevitable, which characterising it as a core stage of the process implies it is? This should be addressed throughout. I also wondered about the implications in terms of the overall prevalence rates presented for 'the whole trafficking period' (p.13). - Surely this should be 'the whole pre-, during and post trafficking period'

Response: Dear review thank you for your concern. As indicated by Zimmerman and colleagues (a well-known author in human trafficking), there are five stages of human trafficking including Pre-departure, Travel, Destination (Exploitation), Detention, and Integration/Re-integration. Here it doesn't necessarily mean that they are in trafficking situation in all these stages. It is just to examine trafficking victims in the whole range of their life. Authors also proposed that health and health related outcomes among trafficked persons can be well understood if studied considering these stages (Zimmerman et al, 2011). The reference is indicated at the end of this Response. Therefore, with this background we have tried to understand sexual violence at each stage of trafficking. Interestingly, we have got a considerably high magnitude of sexual violence in the Detention stage. However, this has a very important implication in that despite trafficked persons are victims (not criminals), they had been detained and then experienced sexual violence. If we see it neutrally, migrants should have been screened as trafficked and non-trafficked, and those trafficked individuals should have got assistance instead of being detained for up to 6 months and violated. Unfortunately, transit and destination countries do not screen migrants based on their trafficking status. Instead, the criminals (e.g. smuggled but not trafficked) and trafficked persons are treated equally. If we see it from the point of view of UN 2000 anti-human trafficking protocol, this should not have happened. Of course, though we have got considerable prevalence of sexual violence at the detention stage, in the analysis of risk factors, those who were detained were not significantly different from non-detained ones.

Moreover, these five stages have different types of exposure, are different milestones of the life of trafficking victims, and may be important to see their experience in those five stages which would be important for intervention. This is because, for different reasons, trafficked persons would encounter different mistreatments even during post-trafficking situations including detention and reintegration stages.

Reference

Zimmerman C, Hossain M, Watts C. Human trafficking and health: A conceptual model to inform policy, intervention and research. *Social science & medicine*. 2011;73(2):327-35.

Comment: The conclusions section of the abstract is more or less a direct repetition of the results section of abstract – I think it would be a better use of the space to talk about the implications here.

Response: We have added one implication in the conclusion section though there are already other more implications in the discussion sections (after discussing variables there are implications or suggestions for most factors).

Comment: P 6 line 12 – I'm not convinced the authors can claim to get 'a representative sample'. Without clear sampling frame etc. this is still a convenience sample (albeit a broader one than one that relies on support services/small groups alone). Perhaps 'a more representative sample' would be more accurate?

Response: Dear reviewer thank you again for raising the issue. As described in the manuscript, there are two stages of sampling. The primary sampling unit was border towns which involves the selection of three trafficking corridors out of four major trafficking corridors in the country. Once the three corridors (border towns) were determined, the second sampling units i.e. returnees (or individuals) via those border towns were recruited consecutively. Here consecutively mean every person who fulfilled the inclusion criteria (being trafficking) would be included into the study (of course, there were some individuals who were not included in the study for various reasons – but were smaller in number). In this case, it can be considered as a random sample because in the case of constative sampling what is important is the STARTING point of time to recruit participants. If the starting point of time were selected purposively (or had the starting point of time been selected that was convenient for us especially by considering the returning migrants ahead of time), then it could have been non-probability sampling as you mentioned. But in our case we were doing our preparatory works such as proposal writing, pilot study etc. and then when the preparatory work was completed we started the data collection in May 2016. Here the time lapsed for the preparatory work is nothing to do with the people who would return home. In other words, we don't know ahead of time who would come via those border towns, what their sexual experience is and so on. Moreover, because returnees were recruited from the border when they enter home, they were from every corner of the country but not limited to specific localities. As a result, we believe that our data reflect the country picture or a wider population of returnees who were trafficked. Convenience sampling is when we recruit people who are convenient for us haphazardly such what journalists do by interviewing anyone in the street.

However, following your important concern, we have added more clarifications about the sampling procedures we followed.

Comment: In the response to reviewers I note that the authors clarified that the 'restricted freedom of movement' variable covered not just physical restrictions but also psychological ones, threats etc. to stop people going out. This is really important and needs to be made explicit in the paper.

Response: Accepted, and extra clarification is added.

•Comment: P.15 'univariable analysis' – I was a little confused here. Surely this was a bivariable analysis looking at their relationship to sexual violence? Or not?

Response: Dear reviewer we acknowledge your concern as there are inconsistencies in naming the relationship between variables. In biostatistics/statistics, when someone says uni-variable and multi-

variable, it means there is one and many independent variables, respectively. In other words, when the suffix “variable” is used, the prefix indicates the number of independent variables only, not about the dependent variable or both. On the contrary, when someone says univariate and multivariate, it is to mean there are one and many dependent variables, respectively. In other words, when the suffix “variate” is used, (which is different from “variable”), the prefix indicates the number of dependent variables only. Univariate models are statistical methods such as simple linear regression, multiple linear regression, simple logistic regression, multiple logistic regression etc. On the other hand, multivariate models have many dependent variables such as in structural equation modeling when we try to fit complex or networked or concomitant relationships among variables. In summary, in biostatistics/statistics, when one says xxx-variable it is all about the number of independent variables whereas xxx-variate is about the number of dependent variables where xxx is either uni or multi.

Comment: P.18 the idea that conditions in Global North ‘encourage them to report every event of sexual violence duly’ just doesn’t happen. Sexual offences are notoriously under-reported. I would adjust this line – perhaps to something like ‘report sexual violence more often, although under-reporting remains a problem here too’

Response: Thanks, accepted and corrected accordingly.

Comment: P.19 How realistic is it for trafficking victims to be provided healthcare services ‘at any time starting from the date of departure’? I suspect it’s commendable but unlikely – in which case I’d add the caveat ‘although it is recognised that this may be very difficult to do in practice’ (or similar).

Response: following this comment, now further clarification is added. People who are under the control of traffickers could liberate (get free) from the condition of trafficking during the travel stage or exploitation stage. Therefore, at the time of liberation they should be examined and get treatment accordingly. Moreover, when they are detained by transit or destination countries, or when they were integrated (by destination country) or reintegrated (by home country), they should also be treated. Sometimes, traffickers themselves take victims to healthcare services when their health deteriorate and are less productive in the exploitation process. In that case, though it is difficult to identify them (because of close follow-up by the offender), the room to support them is not still zero.

Comment: P.20. It would be good – as I think another reviewer noticed – to address (even just briefly) in the discussion the need for more research not just on responses to victims but how to prevent trafficking from Ethiopia in the first place. Given the high harms caused, it would seem sensible to invest in research into better prevention too.

Response: Dear reviewer this is one of the many important suggestions we got from you. We can fight human trafficking in two phases: the first is to combat human trafficking itself. On the other hand, for those who are already trafficked, we can assist them by minimising the consequences of trafficking. Therefore, as you recommended, we have started by identifying the reasons for human trafficking using a qualitative study (reference number 27 which is under review) and a quantification of it (reference 22 which is already published). We have forwarded a number of recommendations in those two papers that may be helpful to curb human trafficking itself. Therefore, this manuscript, is as a part of the second phase i.e. minimizing the consequences of trafficking.

Comment: English much improved but still some typographic/spelling/grammatical errors that need fixing (e.g. 'from Ethiopian' in abstract line 25, 'deny of wage' p.6.) and also a couple of places where the language was a barrier to comprehension: P.10 Line 'we assumed that the missing mechanism of data was non-informative' reads quite strangely and not intuitively understandable though I think I know what they mean (ie.. that unlikely to be systematic differences between those who agreed to and declined to participate). Could this be simplified please? More clarity is needed similarly around the line p.13 'The sexual violence data were binary type (yes/no) but not about the number of episodes of sexual violence'. Could the latter part be simplified to 'but participants were asked about the number of incidents of sexual violence at each stage' (if that is indeed what is meant here).

Response: Now, we have revised it in a manner it meets both your request as well as one review who gave us another comment in the first round.

Comment: P.4 line 27: recommend changing 'for sex trafficking' to 'for sexual exploitation' ('trafficked... for sex trafficking sounds odd!')

Response: We have replaced it.

Comment: P.5 Line 38 'wider population' – presumably this should be 'wider sample'

Response: Now, we have paraphrased it differently. The aim was to say that we wanted to examine a larger (wider) population (country wide) by taking samples at the main gates of returnees at the border. With this design, it is possible to get returnees from every corner of the country which is not limited to specific localities. In short we recruited samples in such a way that it would help us to assess a wider population (country level).

Comment: p.18 both 'traffickers' should be changed to 'traffickers/smugglers' as that seems to be the actual group being referred to here.

Response: Added accordingly.

VERSION 3 - REVIEW

REVIEWER	Ella Cockbain University College London, UK
REVIEW RETURNED	15-May-2019

GENERAL COMMENTS	The final revisions were all sensible and adequate. Happy to recommend that the manuscript provides to publication. I look forward to seeing it in print and being able to share with others!
-------------------------	---