

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)	Prevalence of HIV at the Kokoyo informal gold mining site: What lies behind the glitter of gold with regard to HIV epidemics in Mali? A community-based approach (The ANRS-12339 Sanu Gundo cross-sectional survey)
AUTHORS	Sagaon-Teyssier, Luis; Balique, Hubert; Diallo, Fodié; Kalampalikis, Nikos; Mora, Marion; Bourrelly, Michel; Suzan-Monti, Marie; Spire, Bruno; Dembélé Keita, Bintou

VERSION 1 - REVIEW

REVIEWER	Dr Martin Weihs HSRC South Africa
REVIEW RETURNED	09-Mar-2017

GENERAL COMMENTS	<p>A very interesting article providing important knowledge about HIV and AIDS in IASGMs in West Africa. Some minor issues to be adressed:</p> <p>question 2: it is advisable to provide separate sections for setting and design. Include the qualitative design and findings in the abstract</p> <p>question 4: for the qualitative data include sampling method, criteria for sampling size, data analysis method. Provide more details of the qualitative sample in the results section.</p> <p>question 7: please indicate if general assumptions were met for the multivariate probit regression? Introduce p value in table 2 as you refer to this value in the text. Be consistent placing "Ref:..." in table 2</p> <p>question 15: article needs to be proof read by English native speaker.</p>
-------------------------	---

REVIEWER	Antoine Jaquet INSERM 1219, France
REVIEW RETURNED	12-Apr-2017

GENERAL COMMENTS	<p>In this present manuscript Sagaon-Teyssier et al conducted a cross-sectional study to estimate the prevalence of HIV infection and associated factors in people living in an informal artisanal small-scale gold mining (IASGM) site in Mali, West Africa. The aim is relevant as people living in IASGM sites constitute a vulnerable population, understudied, and potentially an important driver of the HIV epidemic, especially in Mali where the HIV prevalence in the general adult population is relatively low compared to other parts of West Africa.</p> <p>However, there are several limitations that need to be adressed: Major comment 1: the main objective of the present study was to</p>
-------------------------	--

provide an estimation of the HIV prevalence. Therefore, providing an unbiased estimation must be a major concern. Unfortunately, there is limited information on how this convenient sample was constituted. Authors explained that their sample was selected among people living in IASGM site attending prevention activities performed by the local NGO but no information is available on how participants were recruited. Page 4, line 22 authors claimed that "224 participants were randomly included in the study" but no information is available on how they proceeded to perform their randomization. This point is particularly crucial with regards to authors results. Indeed, they showed quite heterogeneous prevalence estimates according to categories of people living in the IASGM site. Therefore, the overall prevalence estimate might be significantly biased if over represented by a certain type of categories (women sex workers, Gold-diggers...).

Major comment 2: In the Method section: Authors gave no information on how patients were screened for HIV: which type of rapid test, type of confirmation test? How patients were concealed once results provided? Any referral if screened positive?

Major comment 3: The statistical analyses performed in the present manuscript are poorly described. Especially as authors choose to use a probit logistic model instead of a more commonly used logit logistic model to identify factors associated with HIV infection. Authors need to clearly state in their method how their model link the dependent variable with other covariates and how this link must be interpreted. They should explain how they take into account multivariable analysis; explain the strategy used to select a final multivariate model. In addition, authors should specify that 95% confidence intervals were computed for each estimated HIV prevalence. They should also cite which basic statistics were used to compare participant's characteristics (page 5, line 5: age comparison).

Major comment 4: In their ethical statement, authors gave no clear information on how patients were informed about the present research and how they gave their formal consent to participate.

Major comment 5: One of the main findings of the present study is the wide variation of HIV prevalence according to activity, but there is limited hypothesis trying to explain these differences in the discussion. In addition, authors should add 95% CI to each estimated prevalence to inform readers on the precision of these estimates.

Major comment 6: In the discussion section page 9, line 34, authors stated that HIV test was well accepted by participants but gave no results with regards to acceptability (rate of refusal...).

Minor comments

In the abstract: page 2, line 3; authors present an 98% CI instead of an 95% CI I suppose. Please correct and harmonize throughout the manuscript.

In the method section, page 4, lines 25 & 26: authors should clearly state which types of sociodemographic/socioeconomic characteristics were collected and how they were defined. Risky behaviors; how where they measured?

In the result section; page 6, line 39; authors start discussing their results comparing their findings with national estimates. This part should be moved to the "discussion" section. More generally, authors should ensure that they did not mix the different section of the manuscript as well as in the abstract.

In the discussion section; authors should check and remove redundant sentences: page 8, line 37, 38 and page 9 lines 15, 16

	<p>Table 1 The title is quite lapidary and should be more informative. I also suggest that authors present their results according to HIV status. That table would then provide a first indication of univariate association between HIV infection and available factors. In addition, authors should explain abbreviations (SD, IASGM) as a footnote table. Same comment for Table 2.</p> <p>Table 2: following my comment on the statistical analysis, need more explanations on how to interpret the table. What dF/dX and Coefficient stands for? The definition of income is different from table 1, authors should harmonize their definition of income and define in the method section. Self-perceived health status: how it was assessed and defined?</p>
--	--

VERSION 1 – AUTHOR RESPONSE

A very interesting article providing important knowledge about HIV and AIDS in IASGMs in West Africa. Some minor issues to be addressed:

question 2: it is advisable to provide separate sections for setting and design. Include the qualitative design and findings in the abstract

Authors comment:

The manuscript was modified by identifying the “settings and design” and “participants” section separately from other sub-sections.

The design and findings of the qualitative survey were included in the abstract

question 4: for the qualitative data include sampling method, criteria for sampling size, data analysis method.

Authors comment:

The methods section was modified as follows:

“For the qualitative survey, purposive sampling was implemented. This seemed to be the most suitable sampling method as the main objective of the qualitative survey was to collect information about prevention and access to care for HIV/STI among specific groups. The sample size for each group was fixed between 5 and 8 survey participants in order to ensure diversity among them. Five activity-specific focus groups were organized: Malian gold-diggers, Non-Malian gold-diggers, female sex-workers, female non sex-workers, and people guaranteeing the organizational functioning at the gold mine site (damantiguis and tombolomas).”

question 7: please indicate if general assumptions were met for the multivariate probit regression?

Introduce p value in table 2 as you refer to this value in the text. Be consistent placing "Ref:..." in table 2

Authors comment:

The general assumptions for Probit regression were met. We modified the “statistical analyses” section by including the following paragraph:

“Given the nature of the normal distribution assumption of the error term in the Probit estimation, coefficients are not affected by the presence of extreme values in independent variables.

Furthermore, Probit regression allows the computation of marginal effects that are more flexible and more informative than odd-ratios^{25,26} which tend to be larger in the presence of rare events and are in any case preferred for small samples²⁷. Marginal effects, dF/dX, are interpreted as instantaneous rates of change: for a dichotomous explanatory variable, a marginal effect shows how predicted probabilities change when the variable changes from 0 to 1.”

The necessary corrections were made in the table.

question 15: article needs to be proof read by English native speaker.

Authors comment:

The article was proof read by an English native speaker

Reviewer: 2

Reviewer Name

Antoine Jaquet

Institution and Country

INSERM 1219, France

Please state any competing interests or state 'None declared':

None declared

Please leave your comments for the authors below

In this present manuscript Sagaon-Teyssier et al conducted a cross-sectional study to estimate the prevalence of HIV infection and associated factors in people living in an informal artisanal small-scale gold mining (IASGM) site in Mali, West Africa. The aim is relevant as people living in IASGM sites constitute a vulnerable population, understudied, and potentially an important driver of the HIV epidemic, especially in Mali where the HIV prevalence in the general adult population is relatively low compared to other parts of West Africa.

However, there are several limitations that need to be addressed:

Major comment 1: the main objective of the present study was to provide an estimation of the HIV prevalence. Therefore, providing an unbiased estimation must be a major concern. Unfortunately, there is limited information on how this convenient sample was constituted. Authors explained that their sample was selected among people living in IASGM site attending prevention activities performed by the local NGO but no information is available on how participants were recruited. Page 4, line 22 authors claimed that "224 participants were randomly included in the study" but no information is available on how they proceeded to perform their randomization. This point is particularly crucial with regards to authors results. Indeed, they showed quite heterogeneous prevalence estimates according to categories of people living in the IASGM site. Therefore, the overall prevalence estimate might be significantly biased if over represented by a certain type of categories (women sex workers, Gold-diggers...).

Authors comment:

The decision of conducting our study on a convenience sample was taken as this kind of study had never had been conducted among populations of informal workers in IASGM sites in Mali. It was the first step of a research program that will extend our methodology to other IASGM sites in Mali with a larger and "more" representative sample to be conducted. Given that we focused on 5 categories of workers (Malian gold-diggers, non-malian gold-diggers, Sex workers, women not sex workers, and traditional authorities tombolomas/damantiguis) we fixed the sample of each category to be at least of 30 participants (we were expecting a sample of at least 150 participants). Both the NGO and the research teams were present in the IASGM site of Kokoyo during 10 days. The strategy to recruit participants was the following:

- 1) Collective conversations about global health topics were organized every day in the evening at different places in the village. At the end of these conversations participants were informed about the presence of the NGO activities (prevention discussions, condoms and gel distribution) including medical check-up. Furthermore, the information about the presence of the NGO activities was disseminated by the representatives of the IASGM site by "word of mouth".

2) NGO activities were conducted every day and the time slot was different for each day in order to adapt the presence of the NGO team to the different working-time of people living there: gold-diggers are available after 5pm, whereas sex-workers are available principally between 11am and 5pm, women not sex-workers are available early in the morning or late in the evening, etc.

So, the random character of our convenience sample lies on two facts: first, participants were invited to participate in the study at the time of their presence in the NGO activities without conditioning the activities to participation in the survey; second, the variety of time slots and places to conduct the NGO activities contributed to the reduction of the overrepresentation of a given category.

In order to clarify this issue, we added the following paragraph in the method section:

“For the quantitative survey, 224 participants constituted the convenience sample. In order to reduce any potential over-representation and under-representation sampling bias, ARCAD-SIDA activities were conducted every day at different time slots and locations at the site in order to adapt participant recruitment to the work schedule of the different categories of people living there.”

Major comment 2: In the Method section: Authors gave no information on how patients were screened for HIV: which type of rapid test, type of confirmation test? How patients were concealed once results provided? Any referral if screened positive?

Authors comment:

We apologize for this missing information, word limit makes sometimes difficult the introduction of details. However, in order to account for the reviewer comment we added the following sentence in the methods part of the manuscript indicating how patients were screened for HIV and the referral when positive:

“For HIV screening, ARCAD-SIDA follows the Malian Ministry of Health recommendations using Determine® rapid tests, and ImmunoComb® II for confirmation. ARCAD-SIDA provided counselling about the importance of being treated to participants testing HIV positive, and referred them to regional health care centres (Centre de Santé de Référence, CSREF) for blood assessment and inclusion in treatment and follow-up programmes.”

Concerning “concealing” of positive participants:

ARCAD activities were organized under the framework of a global health offer. It was not necessary to “conceal” positive participants as HIV screening was proposed at the time of the medical check-up in a setting ensuring confidentiality. Only the medical staff and the participant knew whether HIV screening occurred or not. The survey did not focus on HIV positive participants, and the results of the HIV screening were provided after the administration of the questionnaire for those accepting to participate in the survey. The information of the HIV screening was concatenated to the dataset of the survey through the anonym personal identifier attributed to each participant at the time of their presence to the ARCAD activities. There were no specific procedures allowing the identification of participants positive to HIV. The ARCAD protocol includes a time of counseling at the time of providing the results of the rapid test whatever the result is positive or not. HIV positive participants received detailed information about the procedure to be linked-to-care and included in caseloads for follow-up.

Major comment 3: The statistical analyses performed in the present manuscript are poorly described. Especially as authors choose to use a probit logistic model instead of a more commonly used logit logistic model to identify factors associated with HIV infection. Authors need to clearly state in their method how their model link the dependent variable with other covariates and how this link must be interpreted. They should explain how they take into account multivariable analysis; explain the strategy used to select a final multivariate model. In addition, authors should specify that 95% confidence intervals were computed for each estimated HIV prevalence. They should also cite which

basic statistics were used to compare participant's characteristics (page 5, line 5: age comparison).

Authors comment:

Concerning the choice of probit instead of logit.

We agree with the reviewer about the fact that the logit model is the most used model and that interpretation of coefficients is straightforward compared to probit model given the possibility of calculating odd-ratios. However, logistic distribution (functional distribution for logit model) has the property of being a distribution with "fat tails" which implies that logit regression associates higher probability to "extreme" values than normal distribution does (i.e. functional form of probit model). It has been demonstrated that logit and probit regressions produce similar results and the choice between them is mostly dependent on the discipline (Amemiya, 1981). In addition, probit estimation allows computing marginal effects that are more flexible and less sensitive than odd-ratios in presence of rare events, where odd-ratios tend to be large.

Concerning the link between covariates and independent variables:

We explicitly indicated the functional link between covariates and the dependent variable and we explained how the link is interpreted. The following paragraph was included in "statistical analysis" sub-section of the Methods:

"Given the nature of the normal distribution assumption of the error term in the Probit estimation, coefficients are not affected by the presence of extreme values in independent variables. Furthermore, Probit regression allows the computation of marginal effects that are more flexible and more informative than odd-ratios^{25,26} which tend to be larger in the presence of rare events and are in any case preferred for small samples²⁷. Marginal effects, dF/dX , are interpreted as instantaneous rates of change: for a dichotomous explanatory variable, a marginal effect shows how predicted probabilities change when the variable changes from 0 to 1."

Major comment 4: In their ethical statement, authors gave no clear information on how patients were informed about the present research and how they gave their formal consent to participate.

Authors comment:

For the information about the research, participants coming to the ARCAD NGO activities were informed about the presence of a research team and they proposed to participants to be included in the qualitative and/or the quantitative surveys. The content of the research was detailed in an "information note" provided to those participants interested in the research. The Malian research team provided in-depth information to participants requiring it. Participants that accepted to participate in the qualitative and/or quantitative surveys provided formal consent by signing the "consent letter". The following paragraph was included in the ethical statement of the manuscript:

"Participants in the conversations were then invited to participate in the qualitative and quantitative surveys (they could choose either or both) which formed the basis for the ANRS-12339 Sanu Gundo survey. Community-based agents sent those interested to the Malian team of researchers, who in turn provided detailed information about the content of the survey, its main objectives, and the advantages/risks of participating in this kind of survey. Survey participants provided written formal consent to participate in the survey(s) by signing a letter of consent."

Major comment 5: One of the main findings of the present study is the wide variation of HIV prevalence according to activity, but there is limited hypothesis trying to explain these differences in the discussion. In addition, authors should add 95% CI to each estimated prevalence to inform readers on the precision of these estimates.

Authors comment:

95% CI were added to estimated prevalences and the discussion section was modified in order to account for the reviewer comment about the need of potential explanation of prevalence differences across groups. For not-significant prevalences (i.e. 95% confidence interval including zero) the 90% confidence interval was also added in order to show that some prevalences are not-significant

because the lack of statistical power.

We modified the discussion in order to present some hypothesis explaining the differences between the different groups in terms of prevalences as follows:

“The different HIV prevalence rates found across groups may reflect the lack of prevention programs adapted to key populations other than sex-workers in Mali. Most of the efforts in the fight against HIV/AIDS in Mali are concentrated on female sex-workers and men having sex with other men (MSM), who are identified as the most vulnerable groups. This could explain - at least in part - the low prevalence rate among female sex-workers in Kokoyo IASGM and the higher prevalence among other groups, including female non-sex workers and gold-diggers. Indeed participants from these two categories expressed during the focus groups that prevention campaigns are mostly directed at female sex-workers, who seem to be more informed about the risk of HIV contamination and about prevention tools.”

Major comment 6: In the discussion section page 9, line 34, authors stated that HIV test was well accepted by participants but gave no results with regards to acceptability (rate of refusal...).

Authors comment:

The rate of refusal was reported in the discussion:

“Among 236 persons to which HIV test has been proposed, only 5% refused.”

Minor comments

In the abstract: page 2, line 3; authors present an 98% CI instead of an 95% CI I suppose. Please correct and harmonize throughout the manuscript.

Authors comment:

This typewriting error has been corrected.

In the method section, page 4, lines 25 & 26: authors should clearly state which types of sociodemographic/socioeconomic characteristics were collected and how they were defined. Risky behaviors; how where they measured?

Authors comment:

The required information has been added to the manuscript

In the result section; page 6, line 39; authors start discussing their results comparing their findings with national estimates. This part should be moved to the “discussion” section. More generally, authors should ensure that they did not mix the different section of the manuscript as well as in the abstract.

Authors comment:

We removed the comparison of prevalences with national estimates from the “results” section, and these were reported in the “discussion” section as suggested.

In the discussion section; authors should check and remove redundant sentences: page 8, line 37, 38 and page 9 lines 15, 16

Authors comment:

The necessary modifications have been done.

Table 1 The title is quite lapidary and should be more informative.

Authors comment:

The table was modified

In addition, authors should explain abbreviations (SD, IASGM) as a footnote table. Same comment for

Table 2.

Authors comment:

The necessary modifications have been done.

Table 2: following my comment on the statistical analysis, need more explanations on how to interpret the table. What dF/dX and Coefficient stands for? The definition of income is different from table 1, authors should harmonize their definition of income and define in the method section. Self-perceived health status: how it was assessed and defined?

Authors comment:

The necessary modifications have been done. Tables were harmonized and the definition was provided in the method section. Similarly for self-perceived status.

VERSION 2 – REVIEW

REVIEWER	Dr. Martin Weihs Human Sciences Research Council Port Elizabeth South Africa
REVIEW RETURNED	24-May-2017

GENERAL COMMENTS	Question 2 and 4: One or two sentences about the study design in the abstract as well as in the main document/methodology are still missing. Please refer to the guidelines for authors http://www.bmj.com/about-bmj/resources-authors/article-types/research Question 4: Information about the data analysis method used to analyse the qualitative data is still missing. Add one or two sentences.
-------------------------	--

REVIEWER	Antoine Jaquet INSERM 1219, ISPED, Bordeaux University
REVIEW RETURNED	12-May-2017

The reviewer completed the checklist but made no further comments.

VERSION 2 – AUTHOR RESPONSE

Following the comments of the Reviewer 1, we did the following corrections to the manuscript:

Question 2 : One or two sentences about the study design in the abstract as well as in the main document/methodology are still missing.

Authors: we modified the abstract in order to add information about the study design and we restructured the abstract following the guidelines for authors as suggested.

Question 4: Information about the data analysis method used to analyse the qualitative data is still missing. Add one or two sentences.

Authors: we did the necessary correction in order to add information about the qualitative data

analysis method. For this purpose we add the following paragraph to the "Methods" section in the manuscript:

"Qualitative data analysis

Focus groups were audio-recorded and transcribed. The principle of thematic encoding developed by Flick²⁹ for the analysis of interviews was applied as for other studies with focus groups³⁰. Therefore, thematic and pragmatic dimensions were combined into a dialogical unit to highlight patterns specific to the different groups. Lexical analysis was carried out using both Alceste and MAXQDA software. The different dimensions were cross-classified to generate new insights about the organisation of the data (i.e. how common themes are linked together, specific interactions intra- and inter-groups, etc.). "