

## PCI Infections Article #90 HIV Self-testing Recommender's Comments (Round 2)

Title: HIV self-testing positivity rate and linkage to confirmatory testing and care: a telephone survey in Côte d'Ivoire, Mali and Senegal  
Arsène Kouassi Kra et al. doi: <https://doi.org/10.1101/2023.06.10.23291206>

I thank the authors for speedy and thorough revision of this manuscript, which I found to be of much improved quality but still in need of substantial revisions. Below are some remaining concerns that need to be addressed by the authors before I can make my recommendation. Apologies for the level of detail, but I think it will help to make sure the revised version is much closer to a recommendation.

1. Please change the keywords to the following: AIDS; HIV; Self-Testing; Key Populations; MSM; sex-workers; phone-based survey; West Africa; confirmatory testing; follow-up care; public health program evaluation
2. Line 34: This still says "reactive" - did phase two include those who said "reactive" but who had just one line? And doesn't it include those with two lines but who said "not reactive"? It's still not clear? I thought this should be "positive" - either two lines, reactive, or both. No?
3. Line 36. I think the use of the word "consistent results" may be confusing. In my opinion, it should be "consistent response". This last option is ideal for what appears to be a set of questions used to address self-consistency in reporting. It should be stated somewhere in the methods!! Otherwise, "consistent results" has a connotation that implies consistency between multiple results, whereas what the authors mean here is coherence, or agreement ("concordance") or "consistency" between the result (1 or 2 lines) and its interpretation (reactive or non-reactive), which is response to questions and not results of a test. So, for this line in the abstract, one might say "X% reported an HIVST interpretation (reactive vs non-reactive) that was consistent with the test results (2 lines vs 1 line, respectively)". Obviously, if it is changed here, it should be changed throughout the manuscript. Even in Table 1 the authors use the term "consistent answer". So I think it really should be "consistent response". And also change in the tables and figures as needed.
4. Line 38: "depending on calculations" is not specific enough. Either

choose one figure and report on that, or say, "depending on whether or not incoherent results were included... ". Or "between 2.4% and 9.1%, varying by age groups, countries, response coherence, etc. " Whatever is true.

5. Lines 46-47: What does this mean? Linkage was suboptimal because only 1% of original study participants responded? I think the authors mean that confirmatory testing following a positive HIVST result was low, but what is optimal? This is subjective. Please clarify. Address also in line 415 (concluding paragraph of discussion).
6. Line 68: It's → It is. We try not to use contractions in formal writing. Or better yet, marry to the previous sentence. "Individuals..... result, and is therefor widely accepted.... "
7. Line 73 should read: "The HIV Self-Testing in Africa (STAR) initiative carried out in ... (<https://www.psi.org/fr/project/star/>). " (Spelling out the acronym and also moving the link up from where it currently is a few lines down).
8. Line 78: "where HIV epidemics differ, ..." But also: HOW do they differ? More concentrated how? Need also a reference to this.
9. Line 81: type-o delete "were"
10. Line 90: period (.) at end of sentence.
11. Line 93: This sentence should have been worded as authors stated in their response to reviewers: "...second phase....conducted among those with an HIVST reactive result or had reported two lines in the first survey". I would go one step further and say "conducted among those with an HIVST positive result or interpretation (reporting either reactive, two lines, or both during the first survey)." It's stated differently in different parts of the manuscript, and is confusing.
12. Line 99: Probably best to follow a comma-delimited number convention. 397 367 → 397,367
13. Line 102: "...no previous experience **with** HIVST".
14. Lines 116-119: This reference should be inserted as a normal reference. " Figure 1. ATLAS delivery channels (adapted from [29]). FSW= .... "
15. Line 131: Please edit as follows → "should be interpreted as follows:

"reactive" if two lines (C and T) are visible (even barely), "non-reactive" if only the C (control) line is visible, and "invalid" if no line is visible or if only the T (test) line is visible."

16. While the manufacturer's instructions instruct the tester to interpret the results as either "positive" or "negative" (or invalid), the survey responses appeared to be "reactive" or "non-reactive" (or invalid) results. Is this just a translation thing? I understand the academic jargon is "reactive" or "non-reactive" but this is hardly as clear to the general public as "positive" or "negative". Maybe it would be best to share a copy of the survey, at least in English, to understand what could be improved to limit inconsistent/incoherent results? I expected to see it in the data management plan, but either missed it or it wasn't there.
17. Line 140: needs space between "HIVST" and "and"
18. Line 155: A link to the survey given should be included here, and explicit details about the data collected. (Judging from the rest of the materials, this should be: "gender (identified as man or woman), age group (24 years or less, 25-34 years, 35 years or more), marital status (single, living with partner/married, divorced/separated/widowed...), education level (none/primary, secondary, higher...), and whether or not this was the first time they had [?? used an HIVST, been tested for HIV, ... whatever the authors mean by "first time tester"]. As a rule, any data on which analyses or results rely should be explicitly stated here.
19. Line 157: move ref 37 into the sentence (before the period(.))
20. Lines 159-162: Somewhere here the authors need to state that they advertised the financial incentive on the flyer.
21. Line 175: The authors reference a poster here that supposedly reports on these same data. A poster, in my opinion, is not appropriate here as a reference. In addition, it draws further scrutiny because the numbers do not match. I would very much like to know the total number of kits distributed FOR THIS STUDY, as the poster provides (44,598). I also want to know why the poster says there was a total of 2405 participants while the manuscript claims 2615 participants??
22. Line 183: "completed **the** phase 2 questionnaire." Also, what specific data did the phase 2 questionnaire collect? The questionnaire needs to be provided, as with phase 1. Were responses multiple choice or open-ended, for instance?

23. Lines 195-199: The wording of this paragraph/sentence could be better: "...we distinguish between those for which the reported number of visible lines was consistent with the reported self-interpretation (2 visible lines reported as reactive, one line reported as non-reactive, or no/one line and interpreted as invalid), those for which the test results were inconsistent with the interpretation, or those who returned only a partial response (refusal to answer or answered "I don't know" to one or both questions).
24. Lines 200-204: I think here, the three hypotheses would be much more clear if they are presented slightly differently, with the "central" hypothesis being the base calculation for positivity (the authors use it for the category tests later on), then saying that because of inconsistent interpretations and the eligibility of those with positive results or interpretation for phase two, they also calculated the lowest possible positivity (assuming all DK-R responses as non-reactive), and the highest possible positivity (assuming all DK-R responses as reactive) of the population surveyed. This allows the reader to understand why one would make such outrageous assumptions (neither of which is likely to be true).
25. Paragraph 206-209 (participation bias) should go before the positivity rates (after Line 199).
26. Lines 206-209 (participation bias): "... **sociodemographic and key population source** characteristics were... **using... ??**" In Table S1 results, the authors perform multiple chi-sq tests on non-independent data without correcting for multiple tests. The more correct test here would be a log-likelihood ratio test with all response variables as arguments in an additive multinomial multiple regression model (e.g., VGAM package `vglm(response group ~ country + sex_and_distribution_channel + age group + marital_status + education + first_time_tester, family = "multinomial")`), giving the test statistic for each variable or a results table for model comparisons. Also, authors need to explain why they group gender and distribution channels? I don't see the reason for this, since men and women can both be reached by all channels - it implies some hidden analysis that has not been made explicit. If there's an a-priori reason for separating the PWUD vs different outreach facilities, please detail it. In my opinion, it would be better to separate Men vs Women ("gender") and then facility vs community outreach-based? Then if the authors want to test which outreach-based facilities differed, they could run a post-hoc test.

27. Line 205 - This is still not clear. Why? Just for description? For testing? If testing, with what statistical test? And the authors do not mention age class here, either. I think the best thing to do is to run the same test as for the participation bias (although a simple binomial multiple regression (glm) will suffice in this case), using only the central hypothesis. Then, depending on those results, it would make more sense how the results of the 3 hypotheses are 'stratified' (displayed in the figure = is it more important to stratify based on age? Country? Sex and channel?). Authors are at liberty to include results as they feel are important or were intended a priori. However, it currently feels very random as to which combinations of results are being displayed.
28. Line 220: (*prop.test* function in the 'stats' package)
29. Line 238 - this should still be a part of the prior paragraph.
30. Line 245 Table 1: Partial answer formula "P7" needs subscripting.
31. Lines 259, 332: As the third reviewer mentioned, the categories here do not match with the rest of the manuscript. The authors do not seem to have understood this comment. So, explicitly, is the lower age class 15-24, or 24 years and under ? Is it "24 and under" or "under 24"? And how does discussion of "25 and older Were respondents under 15 years old excluded?
32. Figure 3, Tables S1, S2, etc: It is not always clear what distribution channels are being referenced. There are Community-based MSM and FSW channels, but also facility based MSM and FSW channels, and then PWUD community-based channels, and then another whole list of facility-based channels. Please be specific/explicit.
33. Line 318: why "secondary" distribution? Figure 1 shows it was primary and secondary...?
34. Lines 319 -320: Be sure to report here only the central hypothesis, unless explicitly stating how the lower and upper numbers were attained (because they are not valid results, but lower and upper limits).
35. Lines 327-335: Here, again, just present the central hypothesis (which is a valid figure). The lower and upper limits can then be discussed as relevant (e.g., this is high, even if we had considered all missing responses to have been negative).

36. Lines 341-342: How do your results suggest this? Were you able to correct your estimates for treatment? If not, how is this part of the discussion relevant. It's missing something.
37. Lines 342-345: Reference [43]'s modified explanation does not answer the concern here. I've read through and it appears this is an inappropriate reference for the figures cited. There is no mention of the figures stated, nor the methods on 184 population surveys. This reference is interesting because it talks about what could be expected if key populations were targeted, so maybe the authors meant to reference it elsewhere? The authors should double-check this and adjust accordingly. The methods we wanted to see were not the modeling methods ("shiny90" doesn't actually tell us anything) but the "conventional HIV testing" methods referenced. I think it's important to have these figures of "traditional" testing, because those should be fairly high - when people have a reason to get tested, there is typically a bias towards positivity ("passive surveillance"), because it tells you that your active surveillance method was identifying a hard-to-reach population and at positivity levels at least as high as with passive surveillance. This is really the crux of how you will make the case that this program was effective.
38. Line 347: in **line** with
39. Lines 353-355: How can 90% interpret correctly, but only 2% interpret incorrectly? And why is it those who reported two visible lines that suggests a problem with interpretation? Please double-check and re-word to clarify.
40. Lines 368-372: I would add here that the instructions given with the test (Figure 2) give conflicting instructions. They say to interpret as "Positive" or "Negative", instead of "Reactive" and "Non-reactive". For instance, hearing two different terms used, I would have thought that 2 lines (C + / T +) = reactive and positive, one line (C + / T -) = reactive and negative (as in, the control worked, it reacted), and one or no line (C - / T x) = non-reactive/invalid. Throughout this manuscript, I was under the distinct impression that the use of the term "reactive" or "non-reactive" was purely academic for reporting here, and not a colloquial usage for the surveys. This highlights why it is so important for the surveys - in the languages delivered to participants - need to be shared here, so that further improvements can be made.
41. Lines 374-383: State first the results, then compare to other estimates as you have done, then conclude with the first sentence at the end of

the paragraph to say that while this study's numbers were low, they still showed a lower rate of linkage to confirmatory testing. I would maybe also provide the linkage by country, because a country's system for availability of traditional testing facilities may have a role to play, or in the case of this study, due to the language used the the interviewer (which was the clear leader of reasons why confirmatory testing was not sought).

42. Line 403: the cost - not necessary, but do the authors have any information on the cost of this study? In reference to the overall cost of HIVST programs?
43. Lines 407-408: Again, remove this reference to a poster which gives conflicting results, and could be considered as a prior publication of these results, which would mean we would have to reject this article. Please explain in your response (not in the article) why this poster's analysis showed different numbers and also be sure that all authors from that poster are included in this peer-reviewed submission.
44. Line 416 : remove the term "reactive". Maybe say "a positive self-test result" ? Again, be consistent with the rest of the data/methods.
45. Lines 426 (Table S1): Two important questions: 1) Why were there 3 people with 2 lines and reactive, non-reactive, or DK-R for reactivity in the "Not eligible for Phase 2) column?? Does "Not eligible" mean they also did not consent to be recontacted? If so, please change the language of this column. 2) Because eligibility is dependent on the results of the test, authors should not be testing "Result and number line" differences between the columns. Instead, I think "results consistency" could be tested here, because the column does not depend on this (either 2 lines or a "reactive" response qualifies for eligibility, if I understood correctly).
46. Lines 439 (Table S4) : the percentages on the column-totals should not be there. They don't make sense with the rest of the table. \*\*  
Check to make sure there is discussion about how a respondent could have given consistent responses but said their reason for no followup test was because their test was non-reactive...
47. Just a suggestion: if available, Efficacy could be estimated by understanding the % of new infections from the countries reported for 2021 that came from the STAR/ATLAS program...?

48: Table S7: Apologies - I originally said this comment was for table S5. In fact, it is for Table S7: please arrange from shortest to longest time periods.

49. Please check on how the term "sex" was used. Did the survey use the terms "man" and "woman" for "sex" (which should be "male" and "female", as man and woman typically refer to genders). And also comment on whether there was any ambiguity about e.g., transgender identity.