





13 September 2024

Dr. Target Editor

Editor

Target Journal

Re:  Manuscript EDE24-0055

Dear Professor Editor:

Thank you for the opportunity to revise our manuscript and thanks to the reviewers for their helpful comments. We believe the revised paper is improved due to the comments. Please see our responses below, where page numbers refer to the revision with tracked changes.

Sincerely,

A black background with a black square

Description automatically generated with medium confidence

Mark Klose

Reviewer Comments

**Reviewer 1**

1.0. *“…A simulation would also help to illustrate this and may make this paper more interesting.”*

We have added a simulation study in the supplement to illustrate the finite properties of the population attributable fraction estimator using M-estimation. In this simulation, we examine two settings, when the study is a random sample of the target population and when the study is not a random sample of the target population. Please see sentence on page 8 and the new section in the Supplemental Materials.

1.1. *“Please state the year of data used in the text for CDC data (This is not mentioned except in Table 1, that you used 2021 data). Also please explain why this year was chosen as the target population, when the WIHS data is from 1995. Would the fact that treatments for HIV have changed over time change dramatically during this long time period matter for your choice of the target population?”*

We provide clarification regarding the year on page 4 under “Data Sources”. Based on comments from the other reviewer (2.0 and 2.1 below), we now use the CDC data from 2008.

1.2. *“Your link for reference 10 needs to be updated-vol 33 is for years 2016-2020, it does not link to the 2021 data that you use (I believe you need to link to vol -34). Can you explain the discrepancy in your Table 1 vs the CDC table, which says "6,554" female in 2021 and "1075 with injection drug use"- was the CDC table updated since your analysis?”*

We now use CDC data from 2008 and the data in our Table 1 match that in (now) references 10 and 11.

1.3. *“You say 'Table 1 highlights differences in baseline characteristics between the study sample, target population, and samples from the target population." It does not provide characteristics for samples from the target population.”*

We have reworded this sentence to reflect the updated Table 1 on pages 4 and 5.

1.4. *“You say "From the 6587 incident HIV cases...we take random samples without replacement of size 200, 500, 1000, and 500..." Can you explain how you are generating these samples? The CDC data is an aggregated dataset. Do you generate individual patient data from the aggregated CDC table, then draw samples from this? If so, please explain. This step is not clear to me.”*

We have clarified on pages 4-5 to reflect that samples were selected with weights proportional to the number of individuals in each stratum. This is equivalent to generating individual patient data from the aggregated CDC dataset and drawing a random subset of size .

1.5. *“In Table 1, are these all the possible covariates with similar measurements between the WIHS and CDC? Or are there others? Why did you choose to collapse race into 2 categories, when the CDC lists many more categories?”*

We now incorporate this information on pages 4-5 under “Data Sources”. Race was collapsed into two categories due to how it was coded in the WIHS data set. We have added a sentence noting this in the Methods on page 7 and the Discussion on page 10.

1.6. *“In equation (1), your notation (italics of the Pr) is not consistent with the notation you use elsewhere. Normally I would not comment on this, but it is distracting throughout the entire paper. Please review your notation (e.g., once again on page 6 the font is not even consistent within the numerator and denominator in the same equation). Consider for example you use both, " = and A=a.””*

We apologize and have rewritten the equations to ensure consistency in notation and fonts; changes are primarily on pages 5-6.

1.7. *“In the first paragraph, the variable "W" is undefined (I saw in the appendix that it means covariates). In the application, it would help readers if you say what A and Y are. For example, remind readers that Y= aids death or diagnoses and that A= no history of drug use.”*

We added the definition of and now define each variable as suggested on page 5 in the section titled “Population Attributable Fraction”.

1.8. *“Can you have at least a paragraph where you discuss some of the identification conditions in Table 2, on page 6? And justify some of these conditions using your example. This would make it easier to follow. Or at least highlight some of the conditions so that this statement "Given the identification conditions in Table 2" flows better.”*

We have added a discussion regarding the identification conditions given in Table 2 on page 6. In addition, we reworded the supplement on page 18 to improve clarity.

1.9. *“You say you model using logistic regression "with nadir CD4 cell count" - Can you list this covariate in Table 1, so we can see the distribution of it? And you can say CD4 count is unavailable in the CDC (I assume this is why you omit it from Table 1?)”*

We have added the nadir CD4 cell count to Table 1, and added a sentence saying that CD4 count is unavailable in the CDC data on page 7.

1.10. *“I do not follow your interpretation of the results. So, 0.31 is the numerator in the WIHS. "If the data was not transported across populations" you get 0.24 (What population is this in?) Maybe say for the numerator you get 0.24 (Pr[Y|S=1] and the denominator you get 0.31. In the first paragraph you give results without transportability and the second paragraph you give results with transportability, for the full CDC (target population) sample? I would make the connection of these results clearer to Table 3 as well. Mention Table 3 sooner.”*

On page 8 and 9, we have restructured the results section to incorporate these suggestions, which also allows us to better connect the results with the notation. We restructured the results section to begin with the correctly transported results, which is shown in Table 3. Then, we introduce the naïve approach of making inference to the observed study population, which differs from the desired target population.

1.11. *“In the results I would also comment on the similarity of confidence intervals between the 3 different ways you calculated it.”*

We updated the results such that we now state the standard errors clearly on page 8, and on page 9 we have added to the discussion to highlight the difference in the estimators of the standard error.

1.12. *“On page 19, do you mean to say consider if (W, Y, A=a,S=1) was available? Then you can identify Pr[Y|S=1]. I think you also need to observe A.”*

We have clarified this point in the appendix on page S3; when are both available, then the quantity is trivially identified.

1.13. *“On page 20, I think you wrote the same step twice for the proof. You can delete the second line. The described steps seem out of order otherwise.”*

We removed the repeated line on supplemental pages S3 and S4.

1.14. *“"Based on the examined data...we would expect to see a caseload reduction of 14% if injection drug use was completely prevented..." I was expecting to see this type of interpretation in the results (I think it fits better there).”*

We rewrote this sentence and added the interpretation on pages 8 and 9.

1.15. *“I thought the discussion of limitations was good. But I think a stronger claim needs to be made, that it is extremely unlikely you have exchangeability give the 3 covariates. Can you also discuss needing more covariates for exchangeability over A versus exchangeability over S? Was it possible you could have adjusted for different sets of covariates?”*

On page 9, we have added emphasis on how our assumptions are unlikely to hold with so few (and coarse) covariates and discuss variable selection. Also on page 9, we now discuss the other assumptions made in Table 2.

1.16. *“For censoring, can you give an idea of how much censoring there is in your application?”*

We added that 6% of patients were right censored by end of follow up on page 4.

1.17. *“In Table 2, would be easier to read if single-spaced. Should the conditional sampling exchangeability assumption be about the counterfactuals, not the observed data? I think this condition is not written correctly.”*

We corrected the conditional sampling exchangeability assumption in Table 2. We also reformatted all Tables for consistent spacing and readability (see pages 11-13).

**Reviewer 2**

2.0. *“…However, the example chosen for the application case study is not appropriate. In particular, the assumption of sampling exchangeability is almost certainly violated due to the difference in time periods of the sample and target data. It is well-established that the risk of death or AIDS progression among people living with HIV was much higher in 1995 than it now is due to advancements in medical treatments.”*

In line with this suggestion and comment 1.1, we changed the target population to be from 2008, which we detail on page 4. Furthermore, we emphasize the limits of this case study in the discussion section.

2.1. *“In order to achieve sampling exchangeability, the inverse odds weights would need to account for specific treatments used but doing so would lead to an inevitable sampling positivity assumption violation since current treatments were not available to the sample population. I strongly recommend redoing the example using a target population of all women living with HIV in 1995 rather than the current target population of 2021.”*

We have reduced the potential of sampling positivity violations resulting from different HIV treatment regimens across time periods by changing the target population to be from 2008, as now noted on page 4.

2.2. *“The variable W is not defined in the main text (only the appendix).”*

We have added the definition of on page 5 in the section titled “Population Attributable Fraction”, see also Reviewer 1 comment 1.7.

2.3. *“The interpretation of the results is misleading. The text says "... we would expect to see a caseload reduction of 14% ..." but the outcome is AIDS or death, not 'caseload'. In fact, if the results are true then the number of people living with HIV should be expected to increase -- i.e., the 'caseload' and need for healthcare resources would potentially increase. Also, note that if the target population time period is updated to 1995 then the interpretation should reflect that this is a statement about the past. I suggest rewording this to something more explicit. For example" " .... we would have expected to see 14% fewer people developing AIDS or dying over 10 years if injection drug use were completely prevented...."”*

Thank you, we have rewritten and restructured the results section to match our composite outcome definition. A clearer interpretation is now provided on page 9.

2.4. *“Overall, the assumptions required are very minimally discussed here, despite being very strong and unlikely to hold--are nadir CD4, age, and binary race really sufficient for exchangeability here? When is "nadir" CD4 count measured -- is that before or after IDU use? Is it appropriate to use the same covariates for sampling exchangeability and treatment exchangeability? Similarly, the violation of causal consistency seems to be rather more important than it is presented to be.”*

In line with comment 1.15, we added emphasis on the validity of our assumptions from Table 2 in the discussion (pages 9 and 10), ensuring that the violation of causal consistency is better highlighted. We further clarified notation of covariates for sampling exchangeability and treatment exchangeability on pages 5 and 6. We added a definition of when nadir CD4 count is measured on page 7. Although covariate sets for sampling exchangeability and treatment exchangeability need not be identical, we clarify this in the discussion on page 10.