

Response to reviewers for “Spatio-temporal estimates of HIV risk
group proportions for adolescent girls and young women across 13
priority countries in sub-Saharan Africa”

Corresponding author: Adam Howes (ath19@ic.ac.uk)

Contents

Reviewer 1	2
Reviewer 2	3
References	4

Reviewer 1

We thank the reviewer for their helpful comments regarding the statistical modelling.

1. In the manuscript, explain why you use the INLA not WINBUGS coding? The multinomial regression could be modeled in WINBUGS directly.

The reviewer is right to note that multinomial logistic regression models can be implemented in probabilistic programming languages like **WinBUGS** directly. However, for this application, Markov chain Monte Carlo approaches would be prohibitively expensive. For this reason, we chose to use integrated nested Laplace approximations via **R-INLA**, which have been shown to have comparable accuracy for latent Gaussian models in the realistic, non-asymptotic regime. We have added the following text to the manuscript to clarify this point: *Insert added text here.*

2. Is there no other potential covariate that could be used for better modeling?

Many of the covariates which would be most predictive of risk group proportions are themselves difficult to accurately measure, and make only modest at best improvements to model performance. For example, despite the case for their being a clear link between the “proportion clients of FSW” covariate and the “proportion FSW” outcome we found only marginal benefits to inclusion.

3. Please explain sub-national effect more clearly. Why and how you used it?

Using district-level spatial random effects allowed us to account for subnational variation in risk group proportions. We considered both spatially unstructured (IID) and spatially structured (Besag) random effects, which can be implemented in **R-INLA** by setting `model = "iid"` or `model = "besag"` respectively.

4. Why the interaction term for spatiotemporal effect didn't consider in the modeling framework?

We found that the spatiotemporal interaction terms substantially increased the computational burden of the model, so decided to exclude them.

Reviewer 2

We thank the reviewer for their kind comments.

My little concern is about using different data from UNAIDS Key Population Atlas apart from the DHS, which is the may source data for the analyses. I believe the two variants of data are based on different designs, and combining them may not result in dependable results. It would have been more attainable if the UNAIDS data had been used in their sensitivity analysis to confirm the results from the DHS data.

The UNAIDS KP Atlas data and household survey data are different in their design. We believe the household data to be more biased.

Figure 1 is not clear. I recommend that the authors use a table as an alternative visualization.

Table B.3 in the appendix provides an alternative tabulation of the surveys we used, with sample size further broken down by age group. We are unsure how Figure 1 might be made clearer. *Look for other papers which have similar plots. Is there anything they do that we could add to make it clearer.*

References