

Association of Medical Male Circumcision and Sexually Transmitted Infections in a Population-Based Study: A Targeted Maximum Likelihood Estimation Approach

Lateef Amusa (✉ amusasuxes@gmail.com)

University of KwaZulu-Natal

Temesgen Zewotir

University of KwaZulu-Natal

Delia North

University of KwaZulu-Natal

Ayesha Kharsany

Centre for the AIDS Programme of Research in South Africa

Lara Lewis

Centre for the AIDS Programme of Research in South Africa

Research Article

Keywords: STI, Male circumcision, TMLE, IPTW, South Africa

Posted Date: January 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-137122/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Association of medical male circumcision and sexually transmitted infections in a population-based study: a targeted maximum likelihood estimation approach

Lateef Amusa^{1,2*}, Temesgen Zewotir¹, Delia North¹, Ayesha B M Kharsany³, Lara Lewis³

* Corresponding Author

Email: amusasuxes@gmail.com

¹ Department of Statistics, School of Mathematics, Statistics and Computer Science, University of Kwazulu-Natal, Durban, South Africa

² Department of Statistics, University of Ilorin, Ilorin, Nigeria

³ Nelson R. Mandela School of Medicine, Centre for the AIDS Programme of Research in South Africa, Durban, South Africa.

Abstract

Background: Epidemiological theory and many empirical studies support the hypothesis that there is a protective effect of male circumcision against some sexually transmitted infections (STIs). However, there is a paucity of randomized control trials (RCTs) to test this hypothesis in the South African population. Due to the infeasibility of conducting RCTs, estimating marginal or average treatment effects with observational data, are of increasing interest. Using targeted maximum likelihood estimation (TMLE), a doubly robust estimation technique, we aim to provide evidence of association between medical male circumcision (MMC) and two STI outcomes.

Methods: We investigated the associations between MMC and the two STI outcomes, HIV and HSV-2, using data from the HIV Incidence Provincial Surveillance System (HIPSS) study in KwaZulu-Natal, South Africa. We estimated marginal odds ratios using TMLE and compared estimates with those from propensity score full matching and inverse probability of treatment weighting (IPTW).

Results: TMLE estimates suggest that MMC was associated with 46.9% lower odds of HIV (OR: 0.531; 95% CI: 0.455, 0.621) and 20.5% for HSV-2 (OR: 0.795; 95% CI: 0.694, 0.911). The propensity score analyses also provided evidence of association of MMC with lower odds of HIV and HSV-2. For full matching: HIV (OR: 0.546; 95% CI: 0.402, 0.741), and HSV-2 (OR: 0.705; 95% CI: 0.545, 0.910). For IPTW: HIV (OR: 0.541; 95% CI: 0.405, 0.722), and HSV-2 (OR: 0.694; 95% CI: 0.541, 0.889).

Conclusion: Using a TMLE approach, we present further evidence of a protective effect of MMC against HIV and HSV-2 in this hyper-endemic South African setting. TMLE has the potential to enhance the evidence base for recommendations that embrace the effect of public health interventions on health or disease outcomes.

Keywords: STI; Male circumcision; TMLE; IPTW; South Africa.

Background

Numerous public health initiatives to better control the prevalence of HIV/AIDS and other sexually transmitted infections (STIs) have been implemented. One such public health intervention has been medical male circumcision (MMC), which focuses on the anatomical structure of the penis. It is well established that the inner foreskin of the penis is highly susceptible to infection and that the surgical removal of the foreskin, or the retractable fold of tissue covering the head of the penis, reduces susceptibility to infections. Therefore, MMC is recognized as being one modifiable vector of STIs, including HIV in men. Among men, MMC has a protective effect against HIV infection and some sexually STIs via heterosexual transmission [1-3]. Evidence from three randomized controlled trials (RCTs) showed that MMC decreased heterosexual acquisition of HIV by 53% to 60%, herpes simplex virus type-2 (HSV-2) by 28% to 34% and genital ulcer disease among men [4-6].

The numerous studies highlighted above, amongst others, underline the importance of the relationship between MMC and the acquisition of STIs. However, these studies investigating the associations between MMC and STIs have estimated conditional effects, usually by using traditional regression models. To date, none has estimated average or marginal treatment effects, usually estimated by RCTs and propensity score analyses. By mimicking a RCT, where a marginal treatment effect is obtained by contrasting the outcomes between the exposed and non-exposed groups, there is increasing interest in estimating marginal treatment effects using observational data, not adjusted or conditional treatment effects [7].

Besides estimating marginal treatment effects, model misspecification is another problem for the assessment of association between the exposure (or treatment) and outcome. A misspecification of the model terms could substantially bias the estimated effects, as well as the statistical inference [8]. Machine learning methods, using automated data-adaptive strategies that capture important patterns and interactions among variables, can typically overcome these limitations [9, 10]. Though machine learning has traditionally focused on risk prediction or classification, its utility has been extended to effect estimation and inference [11, 12]. Targeted maximum likelihood estimation (TMLE) is a doubly-robust semiparametric method that estimates exposure effects or associations without relying on model specifications [13]. It combines semiparametric estimation, using machine learning algorithms, with an additional estimation process to optimize a parameter of interest (e.g. risk difference, risk ratio, and odds ratio) [12].

The goal of this analysis is to investigate the associations between MMC and two STI outcomes. Specifically, we used a population-based study to estimate the association between MMC and two STIs; namely, HIV and herpes simplex virus type-2 (HSV-2), among males in the KwaZulu-Natal region of South Africa. We obtained marginal odds ratios using TMLE and further compared our results with estimates from propensity score analyses, including full matching and inverse probability of treatment weighting (IPTW) methods.

Methods

Study design and participants

We used data from the HIV Incidence Provincial Surveillance System (HIPSS), a detailed and robust surveillance project that monitored HIV prevalence and incidence trends in KwaZulu-Natal, South Africa. The HIPSS study aimed to assess the impact of programmatic intervention efforts,

including HIV-related prevention and treatment programmes on HIV prevalence, uptake of antiretroviral therapy (ART), CD4 cell counts and viral suppression, in a real-world non-experimental setting. Survey weights that adjust for varying selection probabilities and differential non-response rates were included in the study design. The HIPSS study design, source population and recruitment procedures, have been described previously [14, 15].

Briefly, HIPSS was a household population-based study conducted in the Vulindlela and the Greater Edendale areas, in the uMgungundlovu District of KwaZulu-Natal, South Africa. The study had two cross-sectional surveys of randomly selected individuals, aged 15-49 years, conducted one year apart. For each survey, a multi-stage cluster sampling method was used to choose enumeration areas, households and individuals. All participants completed questionnaires had peripheral blood samples collected and were allocated a unique identification number, with a unique number allocated to link the household, respondents' questionnaire and laboratory data.

This study utilized the HIPSS household survey comprising a total of 9812 men and women enrolled between June 2014 and June 2015. It had an overall individual participation rate of 69.1% among inhabitants of occupied households and 86.7% of enrolled households. Details on the variables for which data were collected have been previously published [14, 15].

Variables and Inclusion criteria of participants

We included men who self-reported their MMC status and were sexually active. The main exposure of interest was the MMC status; i.e. whether a participant had MMC or not. Those who reported being uncircumcised or traditionally circumcised (represents partial removal of the foreskin), or did not know their circumcision status, were classified as not having MMC. The two outcomes of interest in our analysis were the HIV test result (+ve = 1, -ve = 0) and HSV-2 test result (+ve = 1, -ve = 0). Covariates collected included age (in years), marital status (married,

widowed/divorced/separated, single), education (no education, primary/ not completed high school, completed high school, degree/diploma), number of lifetime sexual partners (one, multiple), condom usage (always/sometimes, never), and had sex in the last 12 months (yes or no). These variables are epidemiologically plausible or possible confounders for the relationship between MMC status and the HIV and HSV-2 outcomes.

From the original 3547 male participants, we removed 692 participants who reported never having had sex. We further excluded participants who had missing values for MMC status ($n = 5$). Our analytic sample consisted of 2850 male participants.

Statistical Analysis

We contrasted the marginally adjusted odds of the HIV and HSV-2 outcomes that would be observed for the MMC exposure. In other words, we compared the odds, for each of the two outcomes, when the men were medically circumcised with not being circumcised. Further, all contrasts were adjusted for the predefined set of important confounders, which include age, marital status, educational level, number of lifetime sexual partners, condom usage, sexual activity in the last 12 months. We estimated associations of MMC with HIV and HSV-2 using TMLE, full matching on the propensity score [15] and inverse probability of treatment weighting (IPTW) [16]. The implementation of TMLE is straightforward. Let T , Y denote the exposure (or treatment) indicator and observed outcome (MMC status and STI outcome, respectively, in this context), and let W be a vector including the identified confounders for the effect of T on Y . We first estimated the initial conditional odds of the STI outcome Y , given the MMC status and covariates, $Q_0(T, W) = E_0(Y | T, W)$. The estimate $Q_n^0(T_i, W_i)$ and the predictions $Q_n^0(1, W_i)$ and $Q_n^0(0, W_i)$ were estimated with Super Learner. Super Learner is an ensemble learner of a pre-specified library of algorithms with parameters. It uses cross-validation to adaptively create an optimally weighted

combination of estimates from candidate algorithms [17]. Optimality was defined based on each ensemble learner fit using 10-fold cross-validation, thereby reducing the chance of overfitting.

These estimates $Q_n^0(T_i, W_i)$, $Q_n^0(1, W_i)$, and $Q_n^0(0, W_i)$ form additional columns in our data matrix. We then plugged-in our estimates $Q_n^0(1, W_i)$, and $Q_n^0(0, W_i)$ into our substitution estimator of the parameter of interest, log odds ratio, to obtain an untargeted estimate:

$$\Psi_{MLE,n} = \text{logit} \{Q_n^0(1, W_i)\} - \text{logit} \{Q_n^0(0, W_i)\},$$

Where $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$.

We next estimated the conditional distribution of MMC given covariates W) $g_0 = P(T | W)$ with Super Learner, using the same set of algorithms. The predictions $g_n(1 | W_i)$ and $g_n(0 | W_i)$ were added to our data matrix. Initial estimates of $Q_0(T, W)$ were then updated along a path of some fluctuation parameters, incorporating additional information from the propensity score function to reduce residual confounding in $Q_0(T, W)$. This updating involves two steps: Firstly, g_n , was used in a clever covariate $H_n^*(T, W)$ to define a parametric working model to fluctuate $Q_0(T, W)$.

$$H_n^*(T, W) = \left(\frac{I(T=1)}{g_n(1 | W)} - \frac{I(T=0)}{g_n(0 | W)} \right)$$

For each individual with $T_i = 1$ and $T_i = 0$, the clever covariates are calculated as $H_n^*(1, W_i) = \frac{1}{g_n(1 | W_i)}$ and $H_n^*(0, W_i) = \frac{-1}{g_n(0 | W_i)}$, respectively. In addition to adding the columns $H_n^*(1, W_i)$ and $H_n^*(0, W_i)$, these values are then combined to form a column $H_n^*(T_i, W_i)$ in the data matrix.

In the second and final step, we estimated the fluctuation parameter ε_n by fitting an intercept-free logistic regression of Y on $H_n^*(T, W)$ with the logit of $Q_n^0(T, W)$ being an offset (fixed quantity), where is the resulting coefficient of the clever covariate $H_n^*(T, W)$. We next updated the estimate Q_n^0 into a new estimate Q_n^1 of $Q_1(T, W)$:

$$\text{Logit } Q_n^1(T, W) = \text{Logit } Q_n^0(T, W) + \varepsilon_n H_n^*(T, W).$$

We calculated

$$\text{Logit } Q_n^1(1, W) = \text{Logit } Q_n^0(1, W) + \varepsilon_n H_n^*(1, W),$$

for all individuals, and then

$$\text{Logit } Q_n^1(0, W) = \text{Logit } Q_n^0(0, W) + \varepsilon_n H_n^*(0, W),$$

for all individuals and included additional columns of $Q_n^1(1, W_i)$ and $Q_n^1(0, W_i)$ to our data matrix.

The updated estimates $Q_n^1(1, W)$ and $Q_n^1(0, W)$ were then used to compute the targeted estimator:

$$\psi_{TMLE,n} = \text{logit} \{Q_n^1(1, W_i)\} - \text{logit} \{Q_n^1(0, W_i)\},$$

Our Super Learner library algorithms included generalized linear model (GLM), least absolute shrinkage and selection operator (LASSO) regularized GLM, generalized additive models, random forests, neural networks, k-nearest-neighbours, and the simple mean.

For the two propensity score methods, full matching and IPTW, we defined the propensity score as the conditional probability that a participant was circumcised, given the covariates. As suggested by [16], we also included the survey weight as an additional covariate in the propensity score model. Secondly, we used the estimated propensity scores to create two sets of weights, each derived from full matching and IPTW. These induced weights, for each of the two propensity score methods, are then incorporated in a logistic regression model, which involves regressing the STI outcome on the MMC status.

Analyses accounted for the survey design by incorporating the survey weights in their final estimation. Only a few variables had missing values and are shown in Table 1. In the multivariable analyses, ‘missing’ was made a separate category for the variable capturing the number of partners; for education level, missing values (n=1) were excluded. TMLE was implemented using the *tml*

package [17] in R version 4.0.0. Full matching and IPTW were implemented R packages *MatchIt* [18] and *WeightIt* [19], respectively.

Results

In the analytical sample of 2840 men, 29.1% reported receiving MMC. These men were more likely than their uncircumcised counterparts to be younger and single. They also had a majority who had completed high school, wore condom with a recent partner, had sex in the last one year, and had more than five sexual partners (Table 1).

To examine possible violations of the positivity assumption for estimators that rely on the propensity score, including the TMLE, we examined the distribution of the estimated propensity score. The histogram of the estimated propensity score by the exposure groups is shown in Figure 1. As shown in Figure 1, the lower tail of the distribution does not have tiny, close to 0 values in both the exposed (range: 0.058 – 0.681; median: 0.379) and unexposed groups (range: 0.039 – 0.712; median: 0.239), while the upper tail of the distribution does not have values close to 1. There are, therefore, no indications of the near-positivity violation.

Figure 1: Histogram of estimated propensity score

The prevalence of HIV and HSV-2 were lower among men who had received MMC than those who did not (Table 2). HSV-2 prevalence was higher (53.2%) than HIV (32.4%). Estimates of the unadjusted odds ratios showed a significant effect of MMC for each of the two STI outcomes.

After adjusting for the identified confounders, we found evidence of protective associations between MMC and HIV when the propensity score techniques were utilized (Figure 2). For full matching: HIV (OR: 0.546; 95% CI: 0.402, 0.741), and HSV-2 (OR: 0.705; 95% CI: 0.545, 0.910).

For IPTW: HIV (OR: 0.541; 95% CI: 0.405, 0.722), and HSV-2 (OR: 0.694; 95% CI: 0.541, 0.889). Though the TMLE estimates were in the same direction as the estimates from the propensity score techniques, the TMLE estimates were more precise. For example, the TMLE estimates suggest that, for those who had MMC compared to the uncircumcised, the odds of acquiring HIV and HSV-2, was 46.9% and 20.5% lower, respectively (OR: 0.531; 95% CI: 0.455, 0.621) and HSV-2 (OR: 0.795; 95% CI: 0.694, 0.911).

<insert Table 1 here>

For the Super Learner ensemble of algorithms, there was no single algorithm that produced the best data fit, as measured by the lowest cross-validated mean squared error (results not shown). A further assessment of the weights (or coefficients) of each of the learning algorithm showed that only a subset of them contributed to the Super Learner predictions. For instance, in the Super Learner estimation of the relationship between HIV and MMC, generalized additive models contributed the most with a weight of 0.60, followed by random forest (0.21), and k-nearest-neighbours (0.19). The other algorithms had no contributions (weight = 0). A similar pattern was observed in the relationship between HSV-2 and MMC (results not shown).

<insert Table 2 here>

Discussion and conclusion

We examined the utility of a relatively new methodology, targeted maximum likelihood estimation technique (TMLE) to estimate the association between MMC and sexually transmitted infections among males in a South African population-based study. This study adds to the body of growing knowledge providing evidence of the benefits of MMC in STI prevention. Specifically, we found that for men, MMC has a protective effect against HIV and HSV-2. Though the utilization of

TMLE did not indicate a null effect nor alter the direction of the association, we found evidence of more precise effects.

Figure 2: Associations between MMC and STI outcomes among men in the HIPSS study (n = 2850).

Legend: (A) Adjusted odds ratios (95% CIs) of HIV among men who did MMC Versus those who did not. (B) Adjusted odds ratios (95% CIs) of HSV-2 among men who did MMC Versus those who did not.

Note: 11 men were missing for HSV-2.

Abbreviations: TMLE, Targeted maximum likelihood estimation; IPTW, Inverse probability of treatment weighting; PSM, Propensity score full matching.

The utilized HIPSS study also collected data on other STIs like *Chlamydia trachomatis*, *Neisseria gonorrhoea*, *Mycoplasma genitalium*, *Trichomonas vaginalis*, syphilis, hepatitis B, and human papillomavirus (HPV) infection. However, we selected HIV and HSV-2 because of their relatively high prevalence compared to other STIs, and the application of TMLE for rare outcomes is still in its infancy [20]. Moreover, previous reports [1], have shown that the associations of MMC with STIs other than HIV and HSV-2 were unreliable as the study was underpowered to detect rare STI outcomes.

Public health interventions for HIV and HSV-2 are critically important to study, especially in African settings with high burden syndemics. South Africa as a country - with over seven million HIV positive individuals - has the highest number of people living with HIV in the world, and the KwaZulu-Natal province is the worst hit, with a prevalence of 27% as recorded at the end of 2017 [15]. There was an estimated 417 million cases of HSV-2 globally in 2012 [21]. The currently reported prevalence of HSV-2 in sub-Saharan Africa is as high as 80% among men and women aged 35 and older [22]. Biological and epidemiological evidence further suggests a cofactor effect

of HIV and HSV-2. In other words, HSV-2 infection increases the odds of HIV acquisition [23, 24].

Parametric models require the correct specification of the functional form of the relationship between the exposure and the confounders, or the outcome-confounders relationship. This requirement is challenging and not usually satisfied in practice. The most attractive and unique property of TMLE is its double-robustness, which reduces bias due to model misspecification [12]. This doubly-robust property ensures that TMLE estimates are unbiased if either one of the exposure or outcome model is consistently estimated. TMLE, like other doubly-robust techniques, offers an opportunity to rely on nonparametric methods (like machine learning) in its estimation process, thereby increasing efficiency [13]. Previous theoretical and simulation studies have shown that TMLE has greater efficiency and less bias when compared with mis-specified parametric and nonparametric singly robust methods [11, 28]. This was also evident from the result of our TMLE estimates and confidence intervals in this study.

The proportion of refusals or non-participation of the utilized HIPSS study, both at the household and the individual level was lower than most community-based surveys [25]. Although the utilized data source is robust, it is cross-sectional. We are thus limited by the ability to conclude the temporal relationship between the self-reported factors with the STI outcomes. In other words, it cannot be determined whether observed associations existed before the STI outcomes or vice-versa. Data other than from the STI outcomes came from self-reports; hence, our work is likely to suffer from self-recall bias due to differential recall or social desirability. Misclassification of circumcision status is also a possibility. Not controlling for important risk factors such as a history of narcotics usage and additional comorbidities, which were not in the HIPSS database, is another limitation of this study.

For HIV, we did not exclude key subpopulations whose odds of acquisition would not result from heterosexual transmission. Our inclusion of these subpopulations would most likely bias associations towards the null since there will be less impact of their circumcision status on their HIV acquisition risk. Most of these limitations will be partly addressed by a planned analysis of a longitudinal cohort study capturing STI incidence, thereby validating findings from this study as well as others that have utilized the HIPSS study.

Our TMLE results provide further evidence of the protective effect of MMC against some STIs in men. This study has important practical implications for studies using nonparametric estimation techniques. Notably, TMLE estimates should be interpreted in light of a careful assessment of the propensity score distribution among the exposed and unexposed, and be compared with results from alternative parametric and nonparametric techniques. Due to its double robustness, TMLE, in comparison to its competitors, often results in efficiency gains and bias reduction of estimated exposure effects. In general, the TMLE method has the potential to advance the field of epidemiology and public health, enhancing the evidence base for recommendations that embrace the effect of public health interventions on health or disease outcomes.

List of abbreviations

HIV: human immunodeficiency virus; STI: sexually transmitted infection; HSV-2: herpes simplex virus type-2; MMC: medical male circumcision; TMLE: targeted maximum likelihood estimation; IPTW: inverse probability of treatment weighting

Declarations

Ethics approval and consent to participate

The University of Kwazulu-Natal (UKZN) Biomedical Research Ethics Committee (BF 269/13), the Associate Director of Science of the Centre for Global Health (CGH) and the Provincial Department of Health (KwaZulu-Natal; HRKM 08/14), approved the HIPSS study protocol and informed consent. All methods were carried out in accordance with relevant guidelines and regulations

Informed consent

Informed consent was obtained from all participants

Availability of data and materials

The dataset used in this study is available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

Not applicable. No funding was received for this research

Authors' contributions

LA designed the study and analyzed the data. LL and AK provided the data. TZ, LL, AK, and DN critically reviewed the manuscript and gave constructive comments which improved the manuscript. All authors have read and approved the manuscript.

Acknowledgement

Not Applicable

References

1. Davis S, Toledo C, Lewis L, Maughan-Brown B, Ayalew K, Kharsany AB. Does voluntary medical male circumcision protect against sexually transmitted infections among men and women

- in real-world scale-up settings? Findings of a household survey in KwaZulu-Natal, South Africa. *BMJ Global Health*. 2019;4(3):e001389.
2. Prodger JL, Kaul R. The biology of how circumcision reduces HIV susceptibility: broader implications for the prevention field. *AIDS Research and Therapy*. 2017;14(1):1-5.
 3. Tobian AA, Kacker S, Quinn TC. Male circumcision: a globally relevant but under-utilized method for the prevention of HIV and other sexually transmitted infections. *Annual review of medicine*. 2014;65:293-306.
 4. Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. *PLoS medicine*. 2005;2(11):e298.
 5. Bailey RC, Moses S, Parker CB, Agot K, Maclean I, Krieger JN, et al. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *The lancet*. 2007;369(9562):643-56.
 6. Gray RH, Kigozi G, Serwadda D, Makumbi F, Watya S, Nalugoda F, et al. Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *The Lancet*. 2007;369(9562):657-66.
 7. Austin PC. The performance of different propensity score methods for estimating marginal odd ratios. *Stat Med*. 2007;26:3078-94.
 8. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health*. 2006;60(7):578-86.
 9. Grömping U. Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*. 2009;63(4):308-19.
 10. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*: Springer Science & Business Media; 2009.
 11. Bahamyirou A, Blais L, Forget A, Schnitzer ME. Understanding and diagnosing the potential for bias when using machine learning methods with doubly robust causal estimators. *Statistical methods in medical research*. 2019;28(6):1637-50.
 12. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *American journal of epidemiology*. 2017;185(1):65-73.
 13. Van Der Laan MJ, Rubin D. Targeted maximum likelihood learning. *The international journal of biostatistics*. 2006;2(1).
 14. Kharsany AB, Cawood C, Khanyile D, Grobler A, Mckinnon LR, Samsunder N, et al. Strengthening HIV surveillance in the antiretroviral therapy era: rationale and design of a longitudinal study to monitor HIV prevalence and incidence in the uMgungundlovu District, KwaZulu-Natal, South Africa. *BMC public health*. 2015;15(1):1149.
 15. Kharsany AB, Cawood C, Khanyile D, Lewis L, Grobler A, Puren A, et al. Community-based HIV prevalence in KwaZulu-Natal, South Africa: results of a cross-sectional household survey. *The Lancet HIV*. 2018;5(8):e427-e37.
 16. DuGoff EH, Schuler M, Stuart EA. Generalizing observational study results: applying propensity score methods to complex surveys. *Health services research*. 2014;49(1):284-303.
 17. Gruber S, Van der Laan MJ. *tmle: An R package for targeted maximum likelihood estimation*. 2011.
 18. Ho DE, Imai K, King G, Stuart EA. *MatchIt: nonparametric preprocessing for parametric causal inference*. *Journal of Statistical Software*. 2011;42(8):1-28.
 19. Griefer N. *WeightIt: Weighting for Covariate Balance in Observational Studies*. R package version 0.9.0.; 2020.

20. Pearl M, Balzer L, Ahern J. Targeted Estimation of Marginal Absolute and Relative Associations in Case–Control Data. *Epidemiology*. 2016;27(4):512-7.
21. Looker KJ, Magaret AS, Turner KM, Vickerman P, Gottlieb SL, Newman LM. Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012. *PloS one*. 2015;10(1):e114989.
22. Rajagopal S, Magaret A, Mugo N, Wald A, editors. Incidence of herpes simplex virus type 2 infections in Africa: a systematic review. *Open forum infectious diseases*; 2014: Oxford University Press.
23. Bradley J, Floyd S, Piwowar-Manning E, Laeyendecker O, Young A, Bell-Mandla N, et al. Sexually transmitted bedfellows: exquisite association between HIV and herpes simplex virus type 2 in 21 communities in southern Africa in the HIV prevention trials network 071 (PopART) study. *The Journal of infectious diseases*. 2018;218(3):443-52.
24. Looker KJ, Welton NJ, Sabin KM, Dalal S, Vickerman P, Turner KM, et al. Global and regional estimates of the contribution of herpes simplex virus type 2 infection to HIV incidence: a population attributable fraction analysis using published epidemiological data. *The Lancet Infectious Diseases*. 2020;20(2):240-9.
25. Larmarange J, Mossong J, Bärnighausen T, Newell ML. Participation dynamics in population-based longitudinal HIV surveillance in rural South Africa. *PloS one*. 2015;10(4):e0123345.

Figures

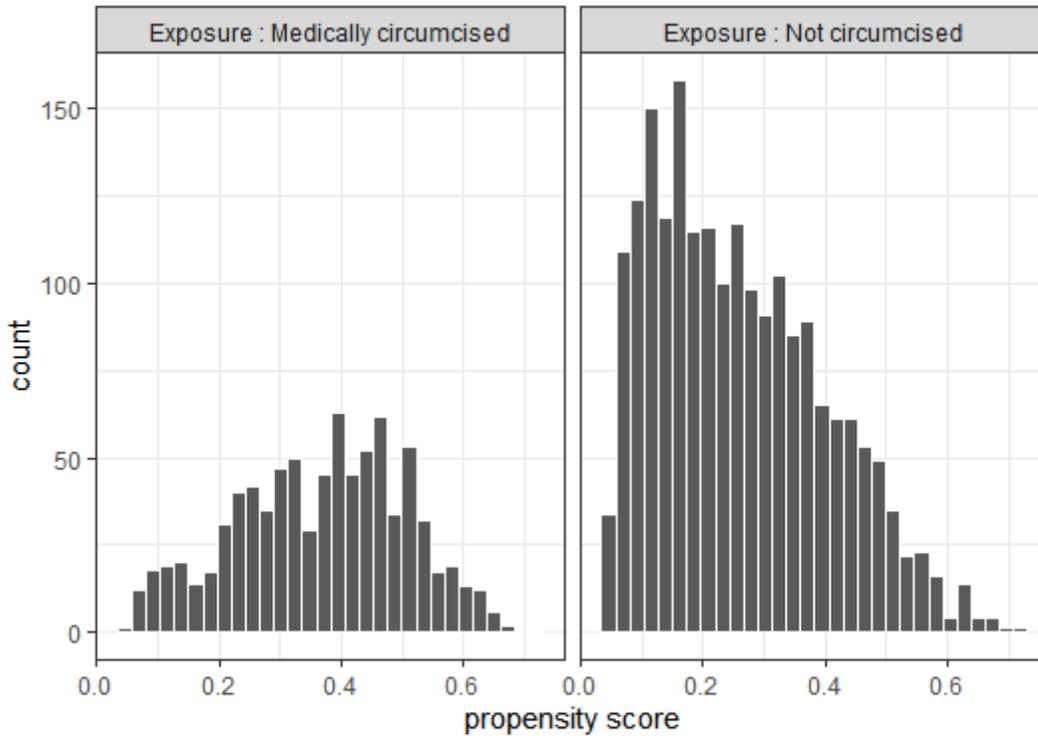


Figure 1

Histogram of estimated propensity score

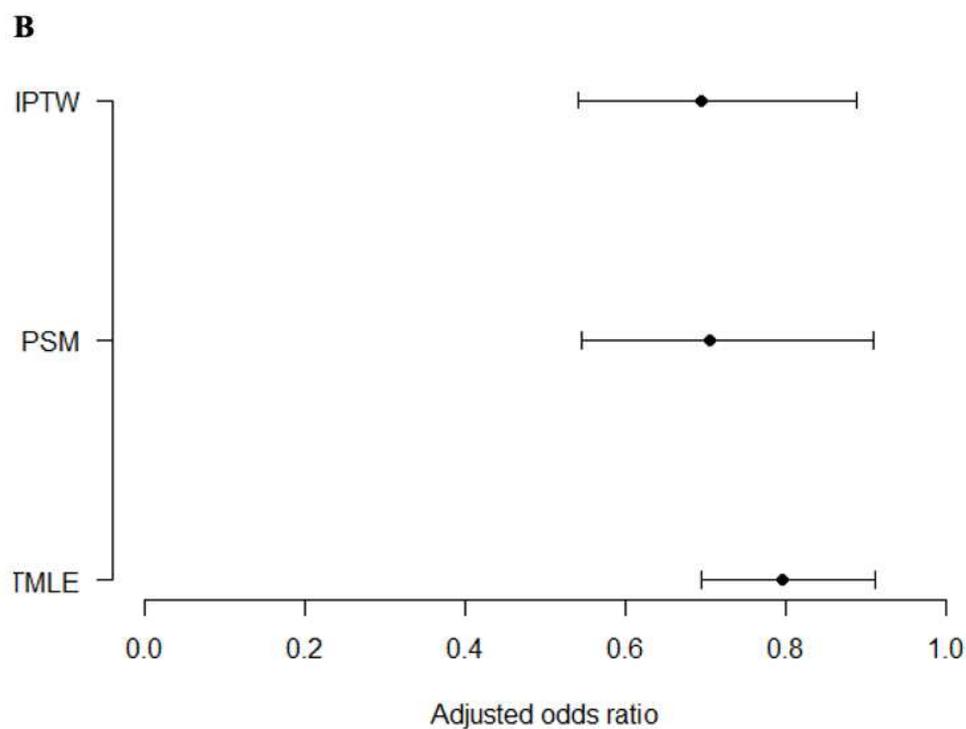
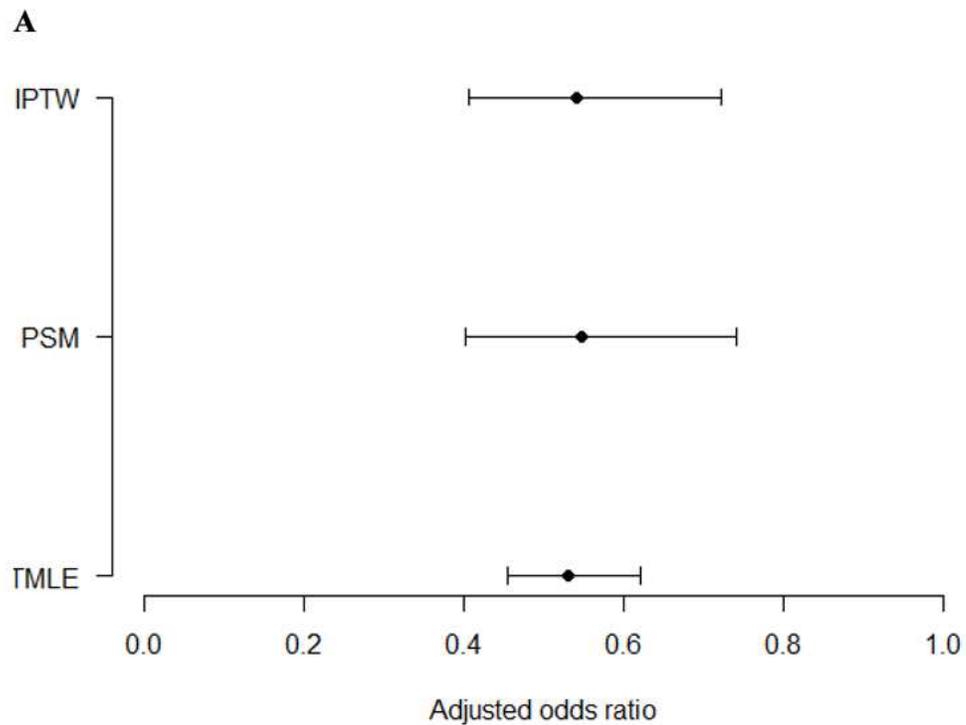


Figure 2

Associations between MMC and STI outcomes among men in the HIPSS study (n = 2850). (A) Adjusted odds ratios (95% CIs) of HIV among men who did MMC Versus those who did not. (B) Adjusted odds ratios (95% CIs) of HSV-2 among men who did MMC Versus those who did not. Note: 11 men were missing for HSV-2. Abbreviations: TMLE, Targeted maximum likelihood estimation; IPTW, Inverse probability of treatment weighting; PSM, Propensity score full matching.