Eric W. Djimeu
Jeffrey E. Korte
Flor A. Calvo

Male circumcision and HIV acquisition

Reinvestigating the evidence from young men in Kisumu, Kenya

September 2015

Replication Paper 8

Health

3ie International Initiative for Impact Evaluation

About 3ie

3ie is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding and producing high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie Replication Paper Series

The 3ie Replication Paper Series is designed to be a publication and dissemination outlet for internal replication studies of development impact evaluations. Internal replication studies are those that reanalyse the data from an original paper in order to validate the results. The series seeks to publish replication studies with findings that reinforce an original paper, as well as those that challenge the results of an original paper. To be eligible for submission, a replication study needs to be of a paper in 3ie's online Impact Evaluation Repository and needs to include a pure replication. 3ie invites formal replies from the original authors. These are published on the 3ie website together with the replication study.

The **3ie Replication Programme** also includes grant-making windows to fund replication studies of papers identified on the candidate studies list. Requests for proposals are issued one to two times a year. The candidate studies list includes published studies that are considered influential, innovative or counterintuitive. The list is periodically updated based on 3ie staff input and outside suggestions. The aim of the 3ie Replication Programme is to improve the quality of evidence from development impact evaluations for use in policymaking and programme design.

About this report

This paper was funded through 3ie's Replication Window with generous funding from the Bill & Melinda Gates Foundation. All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or the 3ie Board of Commissioners, or the Bill & Melinda Gates Foundation. Any errors and omissions are the sole responsibility of the authors. Comments and queries should be directed to the corresponding author, Eric W Djimeu, at edjimeu@3ieimpact.org.

Suggested citation: Djimeu, EW, Korte JE and Calvo, FA, 2015. *Male circumcision and HIV acquisition: reinvestigating the evidence from young men in Kisumu, Kenya, 3ie Replication Paper 8*. Washington, DC: International Initiative for Impact Evaluation (3ie).

3ie Replication Paper Series executive editor: Annette N Brown
Managing editors: Benjamin DK Wood and Anna Heard
Assistant managing editor: Jennifer Ludwig
Production manager: Brigid Monaghan
Copy editor: Jaime L Jarvis
Proofreader: Yvette Charboneau
Cover design: John F McGill
Printer: VIA Interactive

Male circumcision and HIV acquisition: reinvestigating the evidence from young men in Kisumu, Kenya

Eric W Djimeu
International Initiative for Impact Evaluation (3ie)
CEREG, University of Yaoundé II

Jeffrey E Korte
Department of Public Health Sciences
Medical University of South Carolina

Flor A Calvo
Independent consultant

3ie Replication Paper 8
September 2015

Acknowledgments

The authors would like to acknowledge the assistance of Robert C Bailey, principal investigator of the original study, as well as Benjamin Wood, Anna Heard, Annette N Brown and two anonymous external reviewers, who all provided valuable comments on this work.

Abstract

This paper presents findings from our replication study of a flagship 2007 study on the impact of medical male circumcision on HIV acquisition, one of three studies that were influential in providing evidence for scale-up of male circumcision in eastern and southern Africa. Using the same methodology for our pure replication as the original study, we found results similar to those presented in the original paper, namely a protective effect of male circumcision of about 60 per cent. In our measurement and estimation analyses, we used an econometric approach including ordinary least squares, fixed-effects and instrumental variables estimation to assess the robustness of results. As in the pure replication, we found that male circumcision significantly reduced HIV acquisition. In general, the effect sizes were close to the effect size presented in the original study. Our pure replication and measurement and estimation analyses found no evidence of heterogeneous treatment effects by age, similar to the original study. Unlike the original study, however, although we closely replicated the results, we interpret the results as showing some evidence of risk compensation, and our measurement and estimation analyses found strong evidence of risk compensation. Male circumcision reduced the probability of staying abstinent by 17 per cent and increased the probability of not using protection during last sexual intercourse by 12 per cent. Finally, in our sensitivity analysis, we demonstrated that it is not mathematically plausible that the missing data could have biased the results sufficiently to wipe out the observed association. Our findings reinforce the importance of male circumcision for HIV reduction but highlight the need for associated messages for reduction of risk compensation in any strategies aiming to increase the uptake of male circumcision in eastern and southern African countries.

Contents

List of figures and tables

Figures

Tables

Abbreviations and acronyms

CI	Confidence interval
ELISA	Enzyme-linked immunosorbent assay
MEA	Measurement and estimation analysis
OLS	Ordinary least squares
RR	Relative risk
STI	Sexually transmitted infection
UNAIDS	Joint United Nations Programme on HIV/AIDS
WHO	World Health Organization

1. Introduction

The impact of HIV prevention methods is mixed. Although behavioural interventions – including social and behaviour change communication, HIV testing and counselling, and HIV care and treatment – have led to declines in HIV prevalence and incidence in several African countries, HIV prevalence remains extremely high in many sub-regions (Cohen *et al.* 2011; Granich *et al.* 2010; Buchbinder and Liu 2011; UNAIDS 2013). There is an urgent and continuing need to identify cost-effective and practical interventions to prevent HIV transmission.

In the last decade, three studies found that medical male circumcision was effective in reducing HIV acquisition (Auvert *et al.* 2005; Bailey *et al.* 2007; Gray *et al.* 2007). These studies were so influential that they led the World Health Organization (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS) to recommend male circumcision as an efficacious intervention, justifying their recommendation by citing the existence of compelling evidence – a 60 per cent reduction in the risk of heterosexually acquired HIV infection in men (WHO and UNAIDS 2007). WHO/UNAIDS also emphasised that male circumcision should be considered an efficacious intervention for HIV prevention in countries and regions with heterosexual epidemics, high HIV prevalence and low rates of male circumcision. Consequently, 13 countries were selected as priorities for promotion and scale-up of male circumcision: Botswana, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, South Africa, Swaziland, Uganda, the United Republic of Tanzania, Zambia, and Zimbabwe.

The goal was to circumcise 20.34 million men by 2015 in order to have a high epidemiologic impact and avert 3.36 million new HIV infections through 2025 (Njeuhmeli *et al.* 2011). By the end of 2013, close to 6 million circumcisions had been completed in the priority countries (Sgaier *et al.* 2014). Moreover, Njeuhmeli *et al.* (2011) found that the scale-up of male circumcision would cost a total of US$2 billion between 2011 and 2025 and would result a net savings (due to averted treatment and care costs) amounting to US$16.51 billion. The intervention of medical male circumcision therefore shows clear potential for a major impact in reducing HIV transmission and the associated human and economic costs. Finally, male circumcision is now a key component of cutting-edge research on HIV prevention (Hayes *et al.* 2014; Iwuji *et al.* 2013; Havlir 2013; Moore *et al.* 2013). The importance of medical male circumcision for HIV prevention and the magnitude of effort required to scale up any evidence-based intervention underscore the importance of carefully reviewing, understanding and confirming the study results.

Two main reasons motivated us to conduct a replication study of one of the three studies showing the impact of male circumcision on HIV acquisition. First, it appears clear that all three continue to be influential. These studies have significantly shaped HIV prevention in Africa, given that male circumcision is the only intervention that not only can be entirely completed at a single time point and but also significantly reduces HIV acquisition for men. Second, given that the original analysis of the three studies was based on epidemiological approaches, we can make a valuable contribution by

using an econometric approach to examine the same data, including ordinary least squares (OLS) estimation, fixed-effects estimation and instrumental variable estimation.

In fact, the decision of most men in the intervention arm and a few men in the control arm to undergo male circumcision (surgery) might be due not only to the random assignment but also might depend on many other individual characteristics. The individual characteristics that may influence men to undergo male circumcision after being randomly assigned to one of the two study arms may or may not be observable and may or may not change over time. The use of intention-to-treat and as-treated analysis in the epidemiological approach may suffer from endogeneity due to individual characteristics that could affect the decision to undergo circumcision. In order to re-estimate the impact of male circumcision (surgery) on HIV acquisition, we use an analytic approach that takes into account individual characteristics that could affect both the decision to undergo circumcision and the probability of acquiring HIV. Fixed-effects estimation and instrumental variable estimation enable us to control for these unobserved individual characteristics. Specifically, fixed-effects estimation enables us to obtain an alternative and potentially less-biased intention-to-treat estimate, and instrumental variable estimation enables us to obtain an alternative and potentially less-biased treatment on the treated estimate.

The three studies on male circumcision assessed risk compensation by comparing the proportion of different risky sexual behaviours at different points of time over the course of the study between the intervention and the control with Fisher exact tests or $\chi 2$ tests. Although this approach is useful to describe the pattern of the evolution of risky sexual behaviours in the intervention group and the control group to assess the causal relationship between male circumcision and risky sexual behaviours, we prefer an alternate approach. When imbalance exists at baseline in important covariates between the two study arms, as is the case in this study, the analytic approach used must control for these imbalanced covariates in order to draw any conclusions regarding risk compensation due to male circumcision.

To overcome this limitation, we used OLS to estimate the relationship between risky sexual behaviour variables and male circumcision and control for covariates that were imbalanced at baseline. OLS allows to test if the difference in the proportion of different risky sexual behaviours between the intervention arm and the control arm is statistically significant when controlling for covariates which were imbalanced at baseline, and allows us to estimate the effect size (magnitude) of the relationship. Fixed-effects estimation and instrumental variable estimation enable us to fully control for unobserved characteristics that change (or not) over time and which might affect both the decision to undergo male circumcision (the selection process) and the probability of engaging in risky sexual behaviours.[1]

[1] A detailed description of the original study and our replication plan can be found at http://www.3ieimpact.org/media/filer_public/2014/04/09/djimeu_published_replication_plan.pf.

The remaining sections of this paper are organised as follows. Section 2 presents the literature review and our paper's contribution. Section 3 presents our pure replication. Specifically, we present data provided by the authors, methods used in the original paper and findings of our pure replication. We compare our findings with findings presented in the original paper by highlighting similarities and differences. Section 4 presents our measurement and estimation analysis (MEA).[2] In the first part of our MEA, given that the Cox model is the main estimator used in the original paper to model the relationship between the intervention (medical male circumcision) and HIV acquisition, we explicitly test whether the proportional hazards assumption is verified, since this assumption underpins the Cox model. Moreover, taking advantage of the panel structure of the data, we justify the use and present findings from fixed-effects estimation of the main results presented in the original paper. We also justify and present findings from instrumental variables estimation. In particular, we compare results from instrumental variables estimation with findings from as-treated analysis presented in the original paper. In the second part of our MEA, we conducted several sensitivity analyses. Specifically, we discuss the critique related to supportive bias and present findings of the heterogeneous treatment effects by risky sexual behaviour, thereby addressing the issue of selection and sampling bias. Finally, we present results of the sensitivity analysis, addressing the issue of missing data, in which we determined the number of additional individuals in treatment and control groups that would be required to offset the significant observed impact of male circumcision on HIV acquisition. Section 5 presents the discussion of our findings and implications for medical male circumcision as HIV prevention. Section 6 concludes the paper.

2. Literature review

Despite the fact that since the 1990s, two ecological studies and more than 40 observational studies have shown a strong link between circumcision and reduced HIV prevalence (Moses *et al.* 1990; Bailey *et al.* 1999; Siegfried *et al.* 2005) the emergence of male circumcision as a key HIV prevention method in Africa comes from three randomised controlled trials that demonstrated the efficacy of medical male circumcision in reducing HIV acquisition and transmission by approximately 60 per cent (Auvert *et al.* 2005; Bailey *et al.* 2007; Gray *et al.* 2007).

The first trial, conducted in South Africa from July 2002 to February 2004, randomly assigned 3,274 uncircumcised men aged 18–24 years to a control or an intervention group (medical male circumcision), with follow-up visits at months 3, 12 and 21. Using intention-to-treat analysis, Auvert *et al.* (2005) found a relative risk of 0.40 (95 per cent confidence interval [CI]: 0.24–0.68 per cent; $p<0.001$) that corresponds to a protective effect of male circumcision against HIV acquisition of 60 per cent (95 per cent CI: 32–76 per cent).

[2] The pure replication is to re-conduct the original analyses using data provided by the authors. The MEA goes beyond the pure replication to further test the robustness of the original findings beyond the checks described in the original article (Brown *et al.* 2014).

The second study, conducted in Uganda from 2004 to 2006, randomly assigned 4,996 uncircumcised, HIV-negative men aged 15–49 years to receive immediate circumcision (n=2,474) or circumcision delayed for 24 months (n=2,522). In a modified intention-to-treat analysis, Gray *et al.* (2007) found that HIV incidence over 24 months was 0.66 cases per 100 person-years in the intervention group and 1.33 cases per 100 person-years in the control group (estimated efficacy of intervention 51 per cent; 95 per cent CI: 16–72; p=0.006).

The third study, which is the focus of this replication study, was done in Kenya. The choice of Bailey *et al.* (2007) is simple: of the three studies, we were able to obtain access to data for this study. The authors evaluated the impact of male circumcision on HIV-1 acquisition, through random assignment of 2,784 men aged 18–24 years in an intervention group (circumcision; n=1,391) and a control group (delayed circumcision; n=1,393). Participants were enrolled from 4 February 2002 through 6 September 2005. Using intention-to-treat with about 87 per cent of the follow-up experience accrued, the authors found that male circumcision reduced HIV acquisition by 53 per cent. Furthermore, in the as-treated analysis the authors found that the relative risk for male circumcision was 0.40, corresponding to a 60 per cent protective effect of circumcision. There was no heterogeneous treatment-effect-by-age group (18–20 and 21–24 years). Finally, the authors analysed the impact of male circumcision on behavioural variables by treatment at baseline, month 6, month 12, month 18 and month 24.

Bailey *et al.* (2007) found that men in both study arms tended to decrease their risky sexual behaviours, with one behaviour (two or more sexual partners in the previous six months) showing a significantly larger decrease in the control arm compared with the intervention arm over the study period. At the 24-month visit, two other behaviours (unprotected sexual intercourse and consistent condom use) were significantly better in the control arm than the intervention arm. Nevertheless, the authors conclude that, in general, there was no evidence of risk compensation because risky sexual behaviours were not seen to increase in the intervention group. Overall, it appears clear that the three studies were influential; given the magnitude of effort required for scaling up medical male circumcision and the millions of dollars already spent (and yet to be spent) for this procedure, it is important to carefully review, understand and confirm the study results.

This replication study makes three main contributions. First, we further illuminate the relationship between male circumcision and HIV acquisition among men by using fixed-effects estimation and instrumental variables estimation. By re-estimating the impact of male circumcision on HIV acquisition and risky sexual behaviours, our study reinforces the importance of male circumcision for HIV acquisition. Second, using an econometric approach – different from the methodological approach commonly used to assess risk compensation (Gray *et al.* 2007; Agot *et al.* 2007; Mattson *et al.* 2008; Kong *et al.* 2012; Westercamp *et al.* 2014; Wilson *et al.* 2014) – we shed new light on the relationship between male circumcision and risky sexual behaviour. Third, we conducted several sensitivity analyses by discussing and addressing some criticisms addressed to these three trials and related to supportive bias, selection and sampling

bias, and missing data. In short, our replication study offers new insight into the relationship between male circumcision and HIV acquisition among men.

3. Pure replication

3.1 Data used by the authors

The database used for this replication study was the 30 September 2010 Public Release Analysis Database, prepared by a data coordinating centre at RTI International. This dataset comprises all the information collected during the trial and used in the original study and extended follow-up data up to 30 September 2010. It is important to note certain differences between the data used in the original paper and the data used in our replication paper.

3.1.1 Sample size, follow-up time and data files

The public release database reflects information on 2,781 of the 2,784 randomly assigned individuals who were part of the original study. To protect anonymity, all records were deleted for three individuals outside the target age range of 18–24 years at the time of enrolment. The public release database contains only 12 of the 21 collection forms used in the study. The remaining forms were excluded to protect subjects' anonymity; information in the forms was not necessary for our replication.

The authors also provided three ancillary files containing information on final known circumcision status (CircStatus), HIV seropositivity (HIVPos) and urban/rural residence designation (Residence). The data we used for our replication study contains final, corrected data on seroconversion timing and circumcision dates that might not have been available or matched exactly with the information used for the original report. Finally, as reported in the original publication, which we also found in our replication analysis, three individuals were classified as having been HIV-positive at baseline. In our replication analysis, similar to the original publication, we conducted secondary analyses excluding these individuals. Although the original publication included follow-up time of 24 months after randomisation, the public release database includes extended follow-up time of up to 54 months for the 2,781 individuals (depending on their enrolment date). For this replication analysis, we limited our analysis to 24 months in order to follow the original publication more closely.[3]

3.1.2 Altered data

To protect subjects' anonymity, certain information that could be used to identify individuals was altered in the public release database, such as the three individuals

[3] The public release database included circumcision status of men regardless of group assignment (treatment or control) during the trial. Thus, we found in the database that nearly 40 per cent of men in the control group were circumcised at the 24-month visit. However, through communication with the original authors we determined that all those circumcised at 24 months (and probably all of those circumcised at 18 months) were circumcised after December 2006 (the date the trial ended). Therefore, as advised by the original authors, in this replication we classify men circumcised at 18- and 24-month visits as uncircumcised during the trial.

excluded from the database because they fell outside the age range. In addition, records associated with second enrolments of two participants were deleted. Furthermore, a random change (+/- up to 5 days) was generated for each individual, and this change made to all dates for that individual, so that for each individual, the time between specific study visits and the time to seroconversion should be unchanged. Some sparse participant responses were suppressed to protect anonymity related to ethnic origin, categories for number of sexual partners in the previous 30 days and the previous 6 months, and some adverse events after surgery. This generated slight differences in our sample and in the pure replication results.

3.1.3 Manipulation and construction of variables

The data provided by the public release database was obtained in SAS format and converted to Stata 12.1.[4] The sample coding for merging the different data files and construction of time-to-event variables was also provided for SAS. After careful review of the SAS coding and logic used to construct the variables and communication with the authors on issues related to the classification of HIV-positive individuals, we used all the data from all forms up to 31 October 2006 in our analysis. This dataset resembles, as closely as possible, the data used for the original paper, while taking into account the difference in sample size and variable alterations.

For the main variable used for our replication study, 'HIV classification and time of seroconversion', we followed the coding explained in the paper: individuals were classified as HIV-positive if they had double-positive rapid test results from Collection Form 10 (from the public release database) or if their ELISA tests results were positive after discordant rapid tests. This provided us with 60 individuals. After comparing with the HIVPos ancillary file, we confirmed the classification of these 60 individuals as HIV-positive. To establish the timing of seroconversion, we used the first visit at which an individual tested positive. Seven other individuals were classified as HIV-positive at the cut-off date, all of whom were confirmed through polymerase chain reaction tests. Since the HIVPos file contained updated information, we included these individuals, with a resulting 67 individuals classified as HIV-positive for the two-year follow-up period after their enrolment in the study.

3.2 Statistical methods used in the original paper

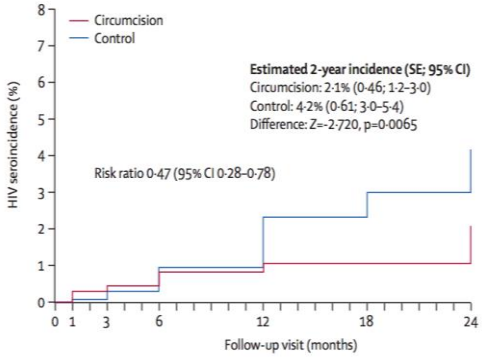
Bailey *et al.* (2007) employed an epidemiological data analysis for their original study, including tabular analyses and Cox proportional hazards modelling, in addition to Kaplan-Meier survival analysis. Kaplan-Meier survival curves and Cox proportional hazards regression models are both used to examine differences between groups in time-to-event outcomes – in this case, the time until seroconversion to HIV. (Individuals not yet seroconverted are considered still at risk, and individuals who are lost to follow-up are excluded from the analysis at the appropriate point.) For the epidemiologic component of our replication analysis, we conducted a pure replication

[4] Stata 12.1 refers to version 12.1 of StataCorp LP's statistical software package.

(i.e., following results presented in the publication) and further analyses expanding the analytic scope and investigating other questions. We used Stata 12.1 for all analyses.

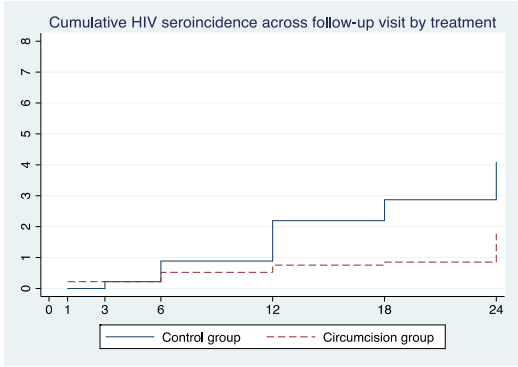
To examine unadjusted survival curves between the two study arms, we used the Kaplan-Meier product-limit estimator to assess each successive follow-up period (Figure 1); this is a nonparametric method to calculate the cumulative survival over time, taking into account differing risk sets at each time point with individuals lost to follow-up, still at risk, or having already experienced the outcome (Kaplan and Meier 1958). We used contingency tables to examine unadjusted relationships between categorical variables in preliminary analyses and Cox proportional hazards regression to estimate time-to-event for HIV seroconversion between study arms. The primary analysis of intervention impact was based on an intention-to-treat analysis; i.e., exposure status was defined by study arm random assignment.

Figure 1: Time-to-event analysis: original vs. replication results



Panel A: Original results

Reprinted from Bailey, RC, Moses, S, Parker, CB, Agot, K, Maclean, I, Krieger, JN, Williams, CFM, Campbell, RT and Ndinya-Achola, JO, 2007. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *The Lancet*, 369(9562), pp. 643–56.



Panel B: Replication results

Note: Time to HIV-positive status is taken as the first visit when a positive HIV test result is noted. Time is credited as the follow-up visit month. Participants without HIV-positive status are censored at the last regular follow-up visit where HIV testing was done, credited specifically as months 1, 3, 6, 12, 18 and 24.

Finally, to examine the relationship over time between circumcision status and risky sexual behaviours, we employed a time-dependent measure of circumcision status to evaluate the possibility of behavioural risk compensation among circumcised men. In these analyses, we fit generalised estimating equations models to allow the analysis of multiple observations per person over time and to appropriately separate inter- and intra-individual variation in behaviour. Using these models we were able to examine possible behavioural differences over time between men receiving or not receiving circumcision, by fitting 'treatment assignment', 'visit' and 'interaction between treatment and visit' as independent variables in the model. Each behaviour of interest was fit as the dependent variable in separate models. Similar to analyses performed by the original authors, we used a logit link to allow estimation of odds ratios for the strength of association and p-value between treatment and each risky sexual behaviour across the study period. In addition, as in the original publication, we

conducted a specific test for differential effect of treatment by group in the change from baseline to month 24, using a one-degree-of-freedom test in the generalised estimating equations model.

3.3 Pure replication results

We present our results in tables showing our pure replication results and the original authors' results to enable comparison of the two sets of results. This is done for each of the four tables and one figure from the original paper. When needed, we shade results from our replication to indicate discrepancies we detected between the original results and results from our reanalysis.[5]

We present baseline demographic characteristics from the original publication with corresponding results based on our own analysis (Table A1 in the Appendix). In characterising the population and checking for imbalance between study arms in covariates of interest, we note that the two study arms were similar on most demographic and clinical characteristics, with few substantial group differences (occupation, infection with herpes, chlamydia trachomatis and sex with other men). Our baseline results show similar numbers to those in the original paper, with minor differences due to the three individuals missing from the dataset. Shading indicates findings with a greater discrepancy from the authors' results; however, all differences were small, with slightly fewer participants noted as positive for herpes or negative for gonorrhoea and slightly more noted as not having given gifts or money to a woman for sex and not having used a condom in the previous six months, in comparison with the original publication.

[5] The original paper has two figures; however, we are unable to replicate the first figure, representing the trial profile, because we lack information such as people registered at clinic, people tested for HIV and people screened according to the protocol.

Table 1: Time to HIV-positive status: original vs. replication results

Panel A: original paper results							
Follow-up visit (months)	1	3	6	12	18	24	Total
Circumcision (n=1,391)							
Number at risk+	1,367	1,351	1,323	1,287	1,029	764	
Number HIV-positive	4	2	5	3	0	8	22
Control (n=1,393)							
Number at risk	1,380	1,368	1,350	1,302	1,035	740	
Number HIV-positive	1	3	9	18	7	9	47
Panel B: replication results							
Circumcision (n=1,390)							
Number at risk+	1,363	1,347	1,322	1,289	1,029	763	
Number HIV-positive	3	0	4	3	1	7	18
Control (n=1,391)							
Number at risk	1,377	1,366	1,348	1,300	1,036	740	
Number HIV-positive	0	3	9	17	7	9	45

Note: Number at risk+ is defined as individuals who have not seroconverted, have dropped out of the programme or do not have censored data. (Sources: authors' calculations using public release analysis database)

In the intention-to-treat analysis, we used Kaplan-Meier survival curves to examine unadjusted HIV seroconversion trajectories in the two study arms (Figure 1). The corresponding number of participants at risk and seroconverting to HIV-positive status during each follow-up interval are summarised in Table 1. 'Number at risk' during each interval was defined as individuals who had not seroconverted before that interval and who were not censored by the end of the interval because they had dropped out of the study. As shown in Table 1, Panel B (our replication results), 41 individuals across the 2 treatment groups who could not be evaluated for HIV seroconversion due to early censoring and missing data were not included in any interval as 'at risk for HIV seroconversion', thereby reducing the number of participants in the analysis from 1,390 to 1,363 in the circumcision group for the first interval, and from 1,391 to 1,377 in the control group.

The replication results for the 24-month follow-up period demonstrate 18 participants seroconverting to HIV-positive status in the circumcision group versus 45 in the control group (Table 1, Panel B). HIV seroconversion incidence rates for each study arm in specific follow-up periods and across the entire 24-month follow-up period are provided in Table 2, Panel B. For the entire 24-month study follow-up period, our replication analysis indicated incidence rates were 1.8 per cent in the circumcision group versus 4 per cent in the control group (Table 2, Panel B), with an overall relative risk of 0.40 (0.21, 0.71). These results do not differ in any notable way from the results presented by Bailey *et al.* (2007).

Table 2: Incidence rates for intervals of follow-up: original vs. replication results

	Panel A: original paper results			Panel B: replication results		
	Circumcision group	Control group	Total	Circumcision group	Control group	Total
0–6 months*	0.8% (0.3–1.3)	1.0% (0.4–1.5)	0.95% (0.5–1.2)	0.5% (0.2–1.1)	0.9% (0.5–1.6)	0.7% (0.4–1.1)
6–12 months	0.2% (0.1–0.7)	1.4% (0.8–2.2)	0.8% (0.5–1.3)	0.2% (0.1–0.7)	1.3% (0.8–2.1)	0.8% (0.5–1.2)
12–18 months	0.0% (0.0–0.5)	0.7% (0.3–1.5)	0.3% (0.1–0.7)	0.1% (0.01–0.6)	0.7% (0.3–1.5)	0.4% (0.2–0.8)
18–24 months	1.0% (0.5–2.1)	1.2% (0.6–2.4)	1.1% (0.7–1.8)	0.9% (0.4–2.0)	1.2% (0.6–2.4)	1.1% (0.6–1.8)
0–24 months*	2.1% (1.2–3.0)	4.2% (3.0–5.4)	3.1% (2.4–3.9)	1.8% (1.2–2.6)	4.0% (3.1–5.2)	2.9% (2.3–3.6)

Note: * Based on Kaplan-Meier methods. (Source: authors' calculations using public release analysis database)

Using Cox proportional hazards regression, we fit several sets of models. First, to assess the main study findings in intention-to-treat analyses, we compared time to HIV seroconversion in the two study groups. In all Cox regression analyses, we defined follow-up for each participant based on their latest study data in the 24-month follow-up period, whether or not they had missed intermediate study visits. Among participants known to have seroconverted to HIV-positive, time to seroconversion was defined based on the time between the baseline visit and the visit at which the participant was observed to have seroconverted. Participants who were observed to remain HIV-negative were censored at their last study visit or at the end of the 24-month follow-up period. The unadjusted hazard ratio was 0.40 (95 per cent CI: 0.23, 0.69), i.e., a 60 per cent reduction in HIV seroconversion among the circumcision study arm, corresponding closely to the results published in the original paper. As expected in a randomised controlled trial, we did not find confounding by any covariate; therefore, adjusting for covariates did not substantially change the hazard ratio.

The second set of Cox proportional hazards models examined the impact of circumcision on time to HIV seroconversion in an as-treated analysis, in which surgery status (circumcised or not circumcised) was used as the main predictor of time to HIV seroconversion. In these analyses, we found that the unadjusted hazard ratio was 0.386 (95 per cent CI: 0.22, 0.67), and the adjusted hazard ratio was similar after accounting for covariates of interest (Table 3, Panel B). As in the intention-to-treat analyses, herpes status was the only covariate showing significant prediction of time to HIV seroconversion.

Overall, the results of our Cox proportional hazards regression models closely tracked with the results presented in the original publication. Our conclusion based on these analyses was the same conclusion reached in the original publication – i.e., that circumcision resulted in an approximately 60 per cent decrease in the risk of HIV

seroconversion over the 24-month follow-up period. We found no evidence of effect heterogeneity by age, with a p-value of 0.48 for the interaction term between group assignment and dichotomous age (18–20 versus 21–24 years).

Table 3: Main findings using Cox proportional hazards regression

Panel A: from the original paper			
	As-treated analysis		
	(1)	(2)	(3)
Hazard ratio	0.45	0.32	0.4
95% CI	0.27–0.68	0.18–0.58	0.23–0.68
Controls	No	No	Yes
Excluding subjects found as positive at baseline	No	Yes	No
Panel B: from our replication paper			
Hazard ratio	0.386	0.377	0.391
95% CI	0.221–0.673	0.216–0.658	0.226–0.676
Controls	No	No	Yes
Excluding subjects found as positive at baseline	No	Yes	No

Note: Controls include occupation, infection with herpes, chlamydia trachomatis and sex with other men.
(Source: authors' construction using public release analysis database)

We conducted analyses of adverse events as part of our replication analysis, finding results similar to but not exactly matched to those in the original publication. For some adverse events, we found slight differences in the number recorded: higher in some cases (e.g., infection) and lower in some cases (e.g., disruption). Nevertheless, the overall conclusion is similar to the original findings. Some of the small differences observed between our analysis and the original could be attributable to data corrections that were made after the original publication was completed or to the minor changes that were made to prepare the data for public release. In addition, as noted in Table 4, Panel B, we did not have access to several variables, including anaesthetic-related event, wound at base of penis, pubic abscess, and folliculitis.

Table 4: Adverse events recorded by severity and relatedness to the surgery: original vs. replication results

	Panel A: original paper results			Panel B: replication results	
	Number of occurrences	Severity	Related to surgery?	Number of occurrences	Severity
Bleeding	5	2 mild, 3 moderate	Definitely	5	2 mild, 3 moderate
Infection	5	2 mild, 3 moderate	Definitely	7	3 mild, 4 moderate
Disruption	4	Mild	Definitely	3	Mild
Delayed healing	3	Mild	Definitely	4	Mild
Swelling	2	1 mild, 1 moderate	Definitely	2	Mild
Anaesthetic-related event	1	Moderate	Definitely	0	
Wound at base of penis	1	Moderate	Probably	0	
Pubic abscess	1	Moderate	Possibly	0	
Folliculitis	1	Mild	Possibly	0	
Erectile dysfunction	1	Moderate	Possibly	1	Moderate

Note: * No code was provided for the anaesthetic-related event, wound at base of penis, pubic abscess or folliculitis variables. (Source: authors' calculations using public release analysis database)

We examined the distribution of self-reported risky sexual behaviours at study time points between the two groups (Table 5). For the variable reflecting two or more partners, our results were essentially the same as those in the original publication. Other variables differed in several ways. In examining the numerator of the risk for each sexual behaviour, we found more participants reporting the 'unprotected intercourse' behaviour than in the original publication. In examining the denominator of the risk, we found that for two variables ('casual partner' and 'inconsistent condom use'), we included more participants in the denominator at each study visit. For the remaining two variables ('unprotected intercourse' and 'abstinence'), we found that we included the same number of participants at baseline, but at later study visits we included more participants in the denominator than the original publication. We confirmed our results, but were not able to ascertain whether an error was present in the original paper or the public release dataset, or whether the public release dataset had more complete data than that published in the original paper.

These differences resulted in a higher observed per cent reporting unprotected intercourse and a lower observed per cent reporting casual partners, non-abstinence and inconsistent condom use. Overall, however, our observed differences between intervention and control were similar in pattern to the results presented in the original paper, with risky sexual behaviours decreasing in both treatment groups.

Table 5: Sexual history with women reported at baseline and follow-up visits: original vs. replication results

	Panel A: original paper		Panel B: replication results	
	Circumcision	Control	Circumcision	Control
Unprotected sexual intercourse with any partner in previous 6 months				
	p=0.1666+		p=0.1104+	
Baseline	867/1,385 (63%)	872/1,387 (63%)	931/1,389 (67%)	947/1,387 (68%)
Month 6	623/1,231 (51%)	623/1,262 (49%)	674/1,242 (54%)	668/1,274 (52%)
Month 12	631/1,227 (51%)	585/1,228 (48%)	684/1,235 (55%)	628/1,235 (51%)
Month 18	505/985 (51%)	495/988 (50%)	534/994 (54%)	522/1,008 (52%)
Month 24	381/741 (51%)	331/727 (46%)	406/754 (54%)	350/742 (47%)
	p-value=0.0349^		p-value=0.003^	
Last time had sexual relations with a casual partner				
	p=0.8044+		p=0.7345+	
Baseline	211/1,053 (20%)	227/1,053 (22%)	211/1,389 (15%)	227/1,387 (16%)
Month 6	180/929 (19%)	192/955 (20%)	180/1,242 (14%)	192/1,274 (15%)
Month 12	199/1,014 (20%)	204/1,007 (20%)	199/1,235 (16%)	203/1,235 (16%)
Month 18	198/985 (20%)	196/988 (20%)	198/994 (20%)	196/1,008 (19%)
Month 24	140/741 (19%)	125/729 (17%)	140/754 (19%)	123/742 (17%)
	p-value=0.2174^		p-value= 0.183^	
Sexual abstinence in previous 6 months				
	p=0.4287+		p=0.5196+	
Baseline	192/1,388 (14%)	194/1,389 (14%)	192/1,389 (14%)	192/1,387 (14%)
Month 6	191/1,232 (16%)	216/1,263 (17%)	191/1,242 (15%)	214/1,274 (17%)
Month 12	188/1,227 (15%)	203/1,229 (17%)	188/1,235 (15%)	203/1,235 (16%)
Month 18	155/985 (16%)	166/988 (17%)	153/994 (15%)	166/1,008 (16%)
Month 24	104/741 (14%)	132/728 (18%)	103/754 (14%)	132/742 (18%)
	p-value=0.0825^		p-value= 0.042^	
Consistent condom use in previous 6 months				
	p=0.1143+		p=0.4083+	
Baseline	265/1,193 (22%)	254/1,193 (21%)	266/1,389 (19%)	254/1,387 (18%)

	Panel A: original paper		Panel B: replication results	
	Circumcision	Control	Circumcision	Control
Month 6	370/1,040 (36%)	378/1,046 (36%)	370/1,240 (30%)	378/1,234 (31%)
Month 12	358/1,039 (34%)	398/1,025 (39%)	357/1,235 (29%)	399/1,234 (32%)
Month 18	296/830 (36%)	304/822 (37%)	296/993 (30%)	304/741 (41%)
Month 24	231/637 (36%)	246/595 (41%)	230/753 (31%)	245/742 (33%)
	p-value=0.0326^		p-value= 0.184^	
Two or more partners in previous 6 months				
	p=0.0383+		p=0.0409+	
Baseline	585/1,388 (42%)	579/1,389 (42%)	584/1,387 (42%)	579/1,387 (42%)
Month 6	409/1,232 (33%)	443/1,263 (35%)	408/1,230 (33%)	443/1,261 (35%)
Month 12	360/1,227 (29%)	408/1,229 (33%)	359/1,225 (29%)	408/1,227 (33%)
Month 18	294/985 (30%)	300/988 (30%)	294/982 (30%)	300/990 (30%)
Month 24	225/741 (30%)	199/728 (27%)	225/738 (30%)	199/726 (27%)
	p-value=0.2044^		p-value=0.158^	

Note: n/N (%). ^ Test for difference between the treatment groups in change from baseline to month 24. + Global test for any difference between the treatment groups in change from baseline to follow-up visits.

We do not concur with the original authors that the results do not provide evidence of disinhibition in the circumcision group. The significant differences in three outcomes including unprotected sexual intercourse with any partner in previous six months, sexual abstinence in the previous six months, and two or more partners in the previous six months between groups consistently favour safer sexual practices in the control group, suggesting that men receiving circumcision had a perception of decreased risk.

Moreover, the overall trend even for non-significant difference between groups suggests, in general, that men in the control group had safer sexual practices than men in the treatment group. We recognise that although there are some statistically significant differences in risky sexual behaviours between groups, the point estimates in most cases are relatively small. Therefore, although we conclude that there is evidence of risk compensation, this pure replication indicates that this evidence is weak; the findings might be interpreted as the original authors did.

Overall, our pure replication shows that we are able to successfully replicate the original results by using the 30 September 2010 data from the public release database used to produce the original results, without codes used to obtain results by the authors and applying analytical approaches described in the original paper. To the best of our knowledge, except for one variable (‘last time had sexual relations with a casual partner’) reported in Table 5, the findings reported in the published paper are accurate and free of coding errors. In general, our interpretation of our pure replication

findings is similar to the authors' interpretation of their original findings. The only difference in the interpretation of results is related to whether or not risk compensation is observed. Although we are able to replicate findings related to the existence of risk compensation, our interpretation is different. We think the statistically significant greater presence in risky behaviour, combined with the overall (albeit non-significant) trend, indicates that there is some manifestation of risk compensation associated with male circumcision.

4. Measurement and estimation analysis

4.1 Proportional hazards assumption test

One important assumption with Cox models is the proportional hazards assumption. If the hazard curves have fairly similar shapes over time for different exposure groups and if the hazard ratio remains fairly constant over time, then the proportional hazards assumption of the model is met. If this assumption is met, then the model residuals should be fairly constant over time; therefore, we can test for violations of the proportional hazards assumption by testing for non-zero slope of the modelled Schoenfeld residuals over time (Grambsch and Therneau 2000; Schoenfeld 1982). A second method for evaluating this assumption is to test time-dependent covariates to evaluate whether the strength of association (for example, the treatment effect) changes over time (Grambsch and Therneau 2000). For both our intention-to-treat and as-treated analyses, we used these two techniques to evaluate the proportional hazards assumption.

4.2 Robustness checks: econometric approach

As part of our MEA, we used an econometric approach that was not used in the original paper. This approach allowed us to assess the robustness of the original findings and to re-estimate the impact of male circumcision on HIV acquisition in the presence of a potential endogeneity problem. Specifically, we used fixed-effects estimation to take advantage of the panel structure and to account for time-invariant, unobserved individual characteristics. Unobserved characteristics such as personality traits, attitudes and expectations may also influence whether or not men engage in risky sexual behaviours leading to HIV acquisition.

When these unobserved characteristics are constant over time, they are called time-invariant individual heterogeneity. Not properly controlling for these unobserved characteristics could bias intent-to-treat and treatment-on-the-treated estimates. For example, men who are more prone to use condoms or have less risky sexual behaviours to avoid HIV acquisition may also be more likely to undergo male circumcision for further protection against HIV acquisition. Using an experiment that randomly offered varying-priced subsidies and comprehensive information to affect the uptake of male circumcision in Malawi, Chinkhumba *et al.* (2014) found that those who used a condom at last sex were significantly more likely to get circumcised. Thus, we may need to control for unobserved confounders which are constant over time. As Chinkhumba *et al.* (2014) pointed out, selection based on *ex ante* risk would

significantly affect the efficacy of male circumcision rollout. Fixed-effects estimation allows us to control for these unobserved time-invariant characteristics.

Moreover, even though we control for unobserved time-invariant individual characteristics through fixed-effects estimation, unobserved characteristics that change over time and affect both the decision to undergo male circumcision and risky sexual behaviours leading to HIV acquisition may still exist. These unobserved characteristics that vary over time could bias the estimate of the impact of male circumcision. The preference to engage in riskier or less-risky sexual behaviours and the motivation to seek male circumcision may be dynamic. For example, the length of men's sexual relationships, their partners' past sexual experiences and their partners' or friends' current attitudes towards condom use might change men's preference to engage in riskier or less-risky sexual behaviours, as well as their decision to undergo male circumcision.

Instrumental variables estimation allows us to control for these unobserved characteristics that change over time and obtain the potentially less-biased treatment-on-the-treated estimate of the impact of male circumcision. Instrumental variables estimation also controls for unobserved characteristics that do *not* change over time. Assignment to the treatment group is a valid instrument for surgery because the probability to undergo surgery is strongly correlated with the random assignment and affects HIV infection exclusively through surgery. Thus, since instrumental variables estimation controls for unobserved time variant and invariant characteristics, an appropriate instrument, such as assignment to the treatment group, can provide an unbiased treatment-on-the-treated estimate of the impact of male circumcision on HIV acquisition in the presence of these unobservables.[6]

We estimated the effect of treatment on the treated using an instrumental variables approach. Specifically, we instrumented the variable surgery (1 for those who got the surgery and 0 otherwise) by the random assignment of participants to the trial. Lastly, in the fixed-effects estimation and the treatment-on-the-treated instrumental variable estimation, we adjusted for variables that were imbalanced at baseline.

4.3 Risk compensation analysis

For the behavioural outcomes, we used a different perspective to evaluate risk compensation. Whereas the original authors used Fisher exact tests or $\chi 2$ tests to compare differences in the proportion of risky sexual behaviours between treatment and control groups at different follow-up visits, we used OLS estimation to estimate the magnitude of differences in risky sexual behaviours between treatment and control groups and formally test whether these differences were statistically significant, while controlling for covariates that were imbalanced at baseline. Despite the randomisation used in this study to balance observed and unobserved characteristics across the two study arms, several covariates were imbalanced at baseline: occupation, infection with

[6] An instrument is appropriate when it is highly correlated with the potential endogenous explanatory variable but independent of error term of the structural equation.

herpes, chlamydia trachomatis and sex with other men. These covariates are related to risky sexual behaviours and should be controlled when assessing the relationship between male circumcision and risky sexual behaviours. Furthermore, these variables are not endogenous to assignment because they were collected at baseline before the intervention.

Failure to control for important variables related to outcomes (risky sexual behaviours) that are imbalanced at baseline will produce a biased estimate of the impact of male circumcision on outcomes, even though the allocation in the two study arms is random. For example, at baseline the proportion of men with herpes simplex 2 virus and *Chlamydia trachomatis* is statistically greater in the circumcision group than the control group. Thus, men in the circumcision group given this feature might be more likely or less likely to engage in risky sexual behaviours even in the absence of circumcision. Changes in sexually risky behaviours might also be different in the control group, even in the absence of male circumcision. Failing to control for these initial differences between the two groups will produce a biased estimate of the impact of male circumcision on risky sexual behaviours.

Moreover, we used fixed-effects estimation and instrumental variables estimation to control for unobserved characteristics that might affect both the decision to undergo male circumcision (the selection process) and the probability of engaging in risky sexual behaviours. These analytic approaches allow us to estimate the actual impact of male circumcision on risky sexual behaviours, and not the impact of offering male circumcision, as is the case in the original paper. For example, individual risk aversion attitude towards HIV transmission and sexually transmitted infections (STIs) – an attitude that affects whether someone will take precautions to avoid contracting HIV and other STIs – will affect both the number of sexual partners and the decision to undergo male circumcision.

Oster (2012) found that in Sub-Saharan Africa, individuals living in a given area changed their risky sexual behaviours in reaction to HIV mortality in the area. This means that individual risk aversion attitude towards risky sexual behaviours that lead to HIV and other STI acquisition might be constant (or not) over time and may depend on contextual factors or shocks. Fixed-effects estimation and instrumental variables estimation allow us to control for these unobserved confounders, which are not controlled when using OLS estimation.

Our MEA of the impact of male circumcision on HIV acquisition aims to assess the robustness of the original results and estimate the impact of male circumcision (surgery) on HIV acquisition and risky sexual behaviours, accounting for unobserved individual characteristics. Thus, in addition to the fact that data used in this replication study differ slightly from data used in the original paper, we do not expect results using the econometric approach (fixed effects and instrumental variables) to be identical to the original findings, because the epidemiological approach used in Bailey *et al.* (2007) and the econometric approach used for our MEA rely on different assumptions. However, we do expect results from our MEA to be similar to the original findings in sign and significance.

For behavioural outcomes, however, in our MEA we may arrive at different conclusions, because the perspective we used to evaluate risk compensation was more appropriate for this purpose than the method used in the original paper. We think the methods in the original paper are more appropriate for understanding the pattern of behavioural outcomes over the course of the study than for assessing the causal relationship between male circumcision and behavioural outcomes.

4.4 Measurement and estimation analysis results

4.4.1 Proportional hazards assumption test

A main assumption of the Cox model is proportional hazards among the exposure groups; in other words, that the hazard ratio remains stable over study follow-up. To test this assumption, we performed the likelihood ratio test and the Shoenfeld residuals test. As Table 6 illustrates, neither test was significant. For the likelihood ratio test, none of the time-dependent variables included was significant; therefore, we did not reject the proportional hazards assumption. Similarly, the Schoenfeld residuals test results were not significant. Therefore, we can conclude that the proportional hazards assumption was not violated.

Table 6: Proportional hazards assumption test

		Period Surgery	Treatment
Likelihood ratio test	Prob > chi2	0.7928	0.9919
Schoenfeld residuals test	Prob > chi2	0.7934	0.9919

Note: Source is authors' estimates using public release analysis database

4.4.2 OLS, fixed-effects and instrumental variables results

Our MEA consists primarily of an econometric approach to estimate the impact of male circumcision on HIV acquisition. Table 7 presents the impact of male circumcision on HIV acquisition estimated through OLS, with endline data, fixed-effects estimation with longitudinal data and instrumental variables estimation with endline data.[7] In these estimations, we did not include 234 participants (attritors) for whom there was incomplete information on HIV status.

In Table 7, Panel A, Column 1, shows that male circumcision significantly reduced the probability of acquiring HIV. Specifically, the absolute risk reduction of HIV acquisition due to male circumcision was 2.2 per cent. The relative risk reduction was 57.89 per

[7] To model the relationship between HIV acquisition and explanatory variables, for OLS we chose to use the linear probability model because it is easy to estimate and the coefficients are easily interpretable (Wooldridge 2002). In linear probability, the slope coefficient of a variable measures the predicted change in the probability of failure (HIV acquisition) when the variable increases by one unit.

cent (0.022/0.038).[8] Column 2 presents the results when we included observables that were imbalanced at the baseline. The findings remained unchanged.

Columns 3 and 4 present results of the impact of male circumcision on HIV acquisition using fixed-effects estimation – taking into account time-invariant individual heterogeneity. Column 3 shows that the absolute risk reduction of HIV acquisition due to male circumcision was 1.3 per cent. This represents a 40.90 per cent smaller absolute risk reduction than in Column 1. The corresponding relative risk reduction of the impact of male circumcision was 56.52 per cent (0.013/0.023). This finding suggests that unobserved time-invariant individual characteristics also affect HIV acquisition. More specifically, although the relative risk reduction remained relatively unchanged when using fixed-effects estimation, when accounting for unobserved individual characteristics that were constant over time, male circumcision had a lesser effect on reduction of HIV acquisition. Column 4 presents the result when we included observables that were imbalanced at the baseline. The findings remained unchanged.

The findings using fixed-effects estimation suggest that the absolute risk reduction of HIV acquisition due to male circumcision obtained through the Kaplan-Meier and OLS models may be biased upwards, but the relative risk reduction is largely unaffected.

Columns 5 and 6 present the impact of male circumcision on HIV acquisition using instrumental variables estimation with endline data. These columns show that male circumcision significantly reduces HIV acquisition.[9] The absolute risk reduction of HIV due to male circumcision is 2.3 per cent. The corresponding relative risk reduction was 61.82 per cent (0.023/0.0372). The estimate using the instrumental variable approach shows a slightly higher protective effect than that obtained using the epidemiologic approach (intention-to-treat analyses), OLS and fixed effects. For the econometric approach, instrumental variables estimation yields an unbiased estimate of the impact of male circumcision if the decision to undergo male circumcision is endogenous. This case may be unlikely in this study, since the decision to undergo male circumcision is determined mainly by random assignment rather than self-selection.

However, the F-test of the joint significance of the fixed effects (Columns 3 and 4) shows that the fixed unobserved individual characteristics are jointly significant. This suggests that unobserved characteristics that are constant over time also affect HIV status. Thus, unobserved individual characteristics that change over time and which are not captured by fixed-effects estimation might also affect HIV status and the probability of accepting circumcision, whether randomised to the circumcision arm or

[8] This relative risk reduction is the absolute reduction of HIV acquisition due to male circumcision (0.022) divided by the proportion of HIV infection in the control group (0.038). Table A3 presents the total number of HIV positive individuals at endline by treatment status.

[9] As predicted, we find a strong correlation between the random assignment and the choice to undergo surgery. An individual assigned to the treatment group has a 93.86 per cent probability of undergoing surgery. This result shows that random assignment in the treatment group and the control group is a valid instrument. The instrumental variables estimation consists of two stages. In the first stage, we regress the treatment assignment on the surgery status. In the second stage, the predicted value of the surgery status is regressed on HIV status using endline data.

not. When this is the case, instrumental variables estimation yields an unbiased estimate of the impact of male circumcision on HIV status.

The estimate using the instrumental variables estimation does not suggest this possibility, because of the very small difference in effect size between estimates from OLS and instrumental variables. To test whether this difference between the estimates from OLS and instrumental variables is statistically significant, we performed a Hausman test.[10] As shown in Table 7, the p-value of the Hausman test was 0.0903.[11] Therefore, the OLS estimate is an unbiased estimate of the relationship between male circumcision and HIV status. This result suggests that the decision to undergo surgery (male circumcision) is mainly determined by the random assignment and not correlated with unobserved individual characteristics that might affect HIV acquisition. In the strict econometric approach, this result suggests that OLS is our preferred estimate of the relationship between male circumcision and HIV status.

The findings from OLS estimation, fixed-effects estimation, and instrumental variables estimation have three main implications. First, we found that male circumcision reduced HIV acquisition, regardless of the estimation strategy. The relationship is robust. Second, the three estimation strategies provide fairly similar relative risks estimating the protective impact of male circumcision on HIV acquisition. However, the three estimation strategies provide different estimates of absolute risk reduction. In particular, when accounting for unobserved individual characteristics that are constant over time, the absolute risk reduction of the impact of male circumcision on HIV acquisition is reduced. The relative risk and absolute risk difference are informative and interpretable in different ways; therefore, we feel they are both important when assessing group differences in clinical studies.

[10] For the Hausman test, under the null hypothesis, OLS and instrumental variable estimates are consistent, but only the OLS estimate is efficient. Under the alternative hypothesis, the OLS estimate is inconsistent, whereas the instrumental variable estimate is consistent.
[11] When p-value is less than 5 per cent, we reject the null hypothesis that the difference in coefficient is not systematic and accept the alternative hypothesis (instrumental variable estimate is the less unbiased estimate of the relationship).

Table 7: Impact of treatment allocation on HIV acquisition at endline

	OLS results		Fixed-effects results		Instrumental variable (IV) results	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Treatment	-0.022***	-0.022***	-0.013***	-0.013***	-0.023***	-0.024***
	(0.006)	(0.005)	(0.003)	(0.003)	(0.007)	(0.007)
Controls	No	Yes	No	Yes	No	Yes
Observations	2,547	2,511	11,443	11,283	2,547	2,511
R-squared	0.005	0.009			0.002	0.006
Number of Individuals	2,547	2,511	2,547	2,511	2,547	2,511
			F(2547, 8665)=5.85 Prob > F=0.000	F(2511, 8542)=5.84 Prob> F=0.000	Hausman (OLS vs. IV) Chi square 9.51 p-value 0.0903	
Panel B	*Age group (18–20 and 21–24)*					
Treatment	–0.023***	–0.024***	–0.012***	–0.013***	–0.046***	–0.046***
	(0.008)	(0.008)	(0.005)	(0.005)	(0.016)	(0.017)
Age group	0.003	0.005	0.002	0.003	0.002	0.005
	(0.008)	(0.008)	(0.005)	(0.005)	(0.018)	(0.018)
Treatment X age group	0.003	0.003	0.000	0.000	0.004	0.004
	(0.012)	(0.012)	(0.007)	(0.007)	(0.024)	(0.024)
Controls	No	Yes	No	Yes	No	Yes
Number of Individuals	2,547	2,511	2,547	2,511	2,547	2,511

Note: Standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1
Controls include occupation, infection with herpes, *Chlamydia trachomatis* and sex with other men.
(Source: authors' estimates using public release analysis database)

In short, our results using econometric approaches suggest that the relative risk reduction is very close to the original authors' finding of 60 per cent risk reduction conferred by male circumcision. Our results from the econometric approach show that this protective effect is independent from unobserved individual characteristics that may or may not be constant over study follow-up.

Regarding the comparison of estimates obtained from the epidemiologic approach (as-treated analysis) and the econometric approach (instrumental variables estimation), the relative risk reduction of the impact of male circumcision on HIV acquisition is very close. The relative risk reduction from the as-treated analysis was 61.4 per cent, compared to 61.82 per cent from the instrumental variables estimation. The two approaches gave the same results because the protective effect of male circumcision in this study is independent from unobserved individual characteristics.

Our Hausman test showed that unobserved individual characteristics that change over time did not affect participation or the probability of HIV acquisition. In the context of this study, this is easily explained: in general, men who agreed to participate in the study were already willing to be circumcised, whether or not they were randomised to the circumcision arm. This is underscored by the nearly 100 per cent uptake of male circumcision in the circumcision arm. Men who were not ready or unprepared to accept circumcision if assigned to the circumcision arm were not included in the study. This means that the protective effect observed is the impact of male circumcision (surgery) on HIV acquisition, and this impact is not biased by unobserved characteristics.

The Kenya demographic and health survey 2008–2009 (2010) reported that the frequency of sexual activity was lower among the youngest age group (15–19). More than half of the men in this age group (15–19) had never had sex. One might reasonably think the frequency of sexual activity could affect the impact of male circumcision on HIV acquisition. In order to assess this plausible relationship, we therefore conducted heterogeneous treatment analysis by age, as shown in Table 7, Panel B. We distinguished two age groups: 18–20 and 21–24. The interaction (treatment × age group) coefficient allows us to assess whether the impact of male circumcision was different in the two age groups.

Results from OLS estimation (Columns 2 and 3), instrumental variables estimation (Columns 3 and 4) and fixed-effects estimation (Columns 5 and 6) show that the coefficient of the interaction term was not significant. Consequently, we concur with the original authors that there is no heterogeneous treatment effect by age. The finding that the effect of male circumcision is not different by age group and the fact that frequency of sexual activity differs by age groups suggest that the impact of male circumcision on HIV acquisition does not change with the frequency of sexual activity. The lack of heterogeneous treatment effect of male circumcision on HIV acquisition therefore increases the external validity of the finding of this study to other age groups.

4.4.3 Results for risk compensation analysis

In the presence of a widespread belief and understanding that male circumcision reduces the risk of HIV infection, one important concern for scale-up is that it may lead to an increase of risky sexual behaviours among men receiving circumcision. Men who assume they are partially protected from HIV acquisition by male circumcision may be more likely to engage in risky sexual behaviours. This behaviour is called ‘risk compensation due to male circumcision’. In this section, we use an econometric

approach, controlling for covariates that were imbalanced at the baseline, to assess the evidence for risk compensation in this study.

We present the effect of male circumcision on behavioural outcomes using OLS, instrumental variable and fixed-effects estimation in Table 8. Our findings suggest that male circumcision reduced abstinence and increased the probability of having unprotected sexual intercourse with any partner in the previous 6 months (Panel C and Panel E). Reduction of abstinence due to male circumcision is not a risky sexual behaviour *per se*. However, when the reduction of abstinence is associated with an increase in the probability of having unprotected sexual intercourse, it suggests that circumcised men are having more unprotected sex. This is a manifestation of risk compensation, because circumcised men may believe that they are fully protected against HIV acquisition and that they can have more sex without protection.

Specifically, male circumcision reduced the probability of abstinence by 3 per cent (Panel E) and increased the probability of having unprotected sexual intercourse by 5.7 per cent (Panel C). The results were similar, regardless of whether we controlled for covariates that were imbalanced at baseline. In addition, the latter result (the effect of male circumcision on the probability of having unprotected sexual intercourse) was robust to different analytic approaches, including OLS, instrumental variable and fixed-effects estimation.

For each risky sexual behaviour outcome, we also performed the Hausman test to test for the presence of unobserved individual characteristics that changed over time, and affected both the decision to undergo male circumcision and to engage in risky sexual behaviours.[12] For all risky sexual behaviours considered, we rejected the hypothesis that the unobserved individual characteristics changing over time affected both the decision to undergo male circumcision and risky sexual behaviours. These results suggest that unobserved individual characteristics that change over time did not change as a result of male circumcision and did not affect both the decision to undergo male circumcision and risky sexual behaviours. As in the case of HIV acquisition, this is explained by the fact that men who agreed to participate in the study were already willing to be circumcised whether or not they were randomised to the circumcision arm. Thus, male circumcision did not seem to function as a shock that might have affected their unobserved characteristics and affected their decisions to undergo male circumcision and engage in risky sexual behaviours. In other words, the decision to undergo male circumcision had been taken before recruitment into the study, so they were ready for the procedure when randomly assigned to the circumcision arm.

Overall, these findings suggest that uncircumcised men who became circumcised changed their behaviours and were engaging in more risky sexual behaviours. Consequently, we conclude that there is strong evidence of risk compensation among

[12] For the Hausman test, when p-value is less than 5 per cent, we reject the null hypothesis that the difference in coefficient is not systematic and accept the alternate hypothesis (instrumental variable estimate is the less unbiased estimate of the relationship).

men assigned to the circumcision arm. Male circumcision increased the probability of having sex, and this increase was associated with a lack of protection during the sexual intercourse. Specifically, male circumcision reduces the probability of staying abstinent by 3 percentage points. In relative terms, this represents a reduction of 16.66 per cent of the probability of staying abstinent due to circumcision. Male circumcision also increases the probability of 'using no protection during the last sexual intercourse' by 5.7 percentage points. In relative terms, this represents an increase of 12.12 per cent the probability of using no protection during last sexual intercourse.[13]

Table 8: The effect of male circumcision on risky sexual behaviour using OLS, fixed-effects and instrumental variable

	OLS results		Instrumental variable (IV) results		Fixed-effects results	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A	*Consistent condom use in previous 6 months*					
Treatment	–0.030	–0.028	–0.0322	–0.0301	–0.0088	–0.0068
	(0.019)	(0.019)	(0.020)	(0.020)	(0.011)	(0.011)
Controls	No	Yes	No	Yes	No	Yes
Observations	2,514	2,479	2,514	2,479	10,408	10,263
R-squared	0.001	0.006	0.001	0.007		
Number of individuals	2,514	2,479	2,514	2,479	2,547	2,511
	Hausman test				F-test results	
		Chi square	2.07		F(2546, 8359) = 1.98	
		p-value	0.1498		Prob > F =0.0000	
Panel B	*Two or more partners in previous 6 months*					
Treatment	0.0087	0.0091	0.0093	0.0097	–0.0054	–0.0063
	(0.019)	(0.019)	(0.020)	(0.020)	(0.013)	(0.013)
Controls	No	Yes	No	Yes	No	Yes
Observations	2,472	2,439	2,472	2,439	10,314	10,173
R-squared	0.000	0.003	0.000	0.003		
Number of individuals	2,472	2,439	2,472	2,439	2,547	2,511
	Hausman test				F-test results	
		Chi square	0.22		F(2546, 8359) = 2.80	
		p-value	0.6450		Prob > F =0.0000	
Panel C	*Sexual abstinence in previous 6 months*					
Treatment	–0.029**	–0.030**	–0.030**	–0.031**	–0.015	–0.016
	(0.015)	(0.015)	(0.015)	(0.015)	(0.010)	(0.010)
Controls	No	Yes	No	Yes	No	Yes

[13] The relative reduction of abstinence due to male circumcision is the absolute reduction due to male circumcision (–0.03) divided by the proportion of men in the control group who abstained from sex (0.18) at endline. The relative increase of unprotected sex due to male circumcision is the absolute increase due to male circumcision (0.057) divided by the proportion of men in the control who have unprotected sex (0.47) at endline.

	OLS results		Instrumental variable (IV) results		Fixed-effects results	
	(1)	(2)	(3)	(4)	(5)	(6)
Observations	2,516	2,481	2,516	2,481	10,415	10,270
R-squared	0.002	0.021	0.001	0.021		
Number of individuals	2,516	2,481	2,516	2,481	2,547	2,511
	Hausman test				F-test results	
		Chi square	4.2		F(2546, 8359) =2.85	
		p-value	0.0410		Prob > F = 0.0000	
Panel D	*Last time had sexual relations with a casual partner*					
Treatment	0.013	0.013	0.013	0.016	–0.0007	0.0010
	(0.016)	(0.016)	(0.017)	(0.017)	(0.009)	(0.009)
Controls						
Observations	2,516	2,481	2,516	2,481	10,415	10,270
R-squared	0.000	0.004	0.000	0.004		
Number of individuals	2,516	2,481	2,516	2,481	2,547	2,511
	Hausman test				F-test results	
		Chi square	0.89		F(2546, 8359) = 1.93	
		p-value	0.3447		Prob > F = 0.0000	
Panel E	*Unprotected sexual intercourse with any partner in previous 6 months*					
Treatment	0.057***	0.054***	0.060***	0.057***	0.024*	0.023*
	(0.020)	(0.020)	(0.021)	(0.021)	(0.013)	(0.013)
Controls						
Observations	2,516	2,481	2,516	2,481	10,415	10,270
R-squared	0.003	0.028	0.003	0.028		
Number of Individuals	2,516	2,481	2,516	2,481	2,547	2,511
	Hausman test				F-test results	
		Chi square	7.13		F(2546, 8359) = 2.54	
		p-value	0.0076		Prob > F = 0.0000	

Note: Standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1
Controls include occupation, infection with herpes, *Chlamydia trachomatis* and sex with other men. (Source: Authors' estimates using public release analysis database)

4.5 Sensitivity analyses: discussion of supportive bias, methods to address sampling bias and missing data

4.5.1 Discussion of supportive bias

This section addresses several criticisms that have been formulated against the findings of the three trials of male circumcision. The first, 'supportive bias', relates to the fact that participants in the treatment group were advised to abstain from sexual activity for at least 30 days after circumcision; in addition, they interacted with clinicians during three post-circumcision visits (day 3, day 8 and day 30 post-surgery). Critics have considered these clinical interactions to constitute an additional treatment component of treatment. One way to address this critique would be to calculate the number, rate or proportion of attendance at the three post-circumcision visits among

circumcised individuals and add this measure of attendance as a covariate in the analysis in order to adjust (control) for the visits. This could have led to erroneous findings, however, because the variable is highly correlated with being circumcised.

Another approach would be to determine whether the effectiveness of male circumcision differed with the number of post-circumcision visits a participant attended. We looked at the distribution of attendance among circumcised men in the treatment group, and found that 92.16 per cent attended all three post-circumcision visits, as required by the study. This does not provide enough variability to assess the impact of the post-surgery visits on the effectiveness of male circumcision.[14] Given the difficulty of disentangling the effect of the post-circumcision visits using an epidemiological approach or an econometric approach, we feel that such post-surgery clinical visits will inevitably be a part of medical male circumcision even at a larger scale, whether or not the visits have any effect on the outcome of HIV seroconversion. The potential impact of these visits is essentially inseparable from – and should be factored with – the impact of circumcision itself.

We are not able to fully address the criticism related to sampling bias. Sampling bias relates to the fact that people who chose to participate in the trials might have been different from the general population (Boyle and Hill 2011). Thus, those who were randomly assigned to the treatment group and the control group might be different from the general population. For instance, the trial in Kenya was located in an area (Nyanza) with a high HIV prevalence relative to the rest of Kenya and where male circumcision was uncommon. This feature of the location could have affected the findings because of the dynamics of the HIV epidemic in the area. In fact, we might reasonably think that the impact of male circumcision would differ depending on the local prevalence of HIV; this poses a problem of external validity that can be solved only by conducting trials in different areas with different HIV prevalence. We agree with this critique and recognise that it cannot be addressed by this replication study.

4.5.2 Methods to address sampling bias

Another criticism formulated against the findings of the three trials and related to sampling bias concerns the fact that most individuals who agreed to participate in the trials were poorly educated, impoverished and unemployed men who knew they would be compensated for participating. We agree that if sexual behaviours of poorly educated, impoverished and unemployed men are different from the general population, then one may expect to see different effects of male circumcision in this group than in the general population. The correlation between poverty and HIV in African countries is mixed. The World Bank states in its ‘Confronting AIDS’ report published in 1997, ‘widespread poverty and unequal distribution of income that typify underdevelopment appear to stimulate the spread of HIV’ (Ainsworth and Mead 1997).

[14] To better understand the role of the post-circumcision visits, we asked the original authors, who stated, ‘They were visits to remove the bandage (Day 3) and to check the wound for infection and healing progress (Day 8 and 30). The participant did not see a counselor at those visits and their purpose was not for counseling’.

However, the argument that poverty fuels HIV has been challenged by recent studies based on statistical correlations of epidemiologic and socioeconomic data. Thus, without thorough studies taking into account the endogeneity problem of this relationship, it is difficult to know what the sense of the relationship is. Nevertheless, assuming that this critique is true, and that poorly educated and impoverished men engage in high risk sexual behaviours, one way to empirically evaluate the assertion that the impact of male circumcision will be less important for men engaging in risky behaviours is to assess whether the impact of male circumcision on HIV acquisition is modified by risky sexual behaviours. We hypothesised that the impact of male circumcision might be reduced for men engaging in risky sexual behaviours.

To shed light on this plausible relationship, we evaluated the heterogeneous treatment effects by risky sexual behaviour. To assess the heterogeneous treatment effects by risky sexual behaviour, we constructed a 'risky sexual behaviour' variable by looking at our five risk behaviour variables: two or more partners, consistent condom use, sexual intercourse with non-regular partner, practicing abstinence in the previous six months and having unprotected sexual intercourse with any partner in the previous six months. To identify individuals with riskier behaviour versus less risky behaviour, we considered participants to have high-risk behaviour if they 1) had two or more partners in the previous six months and did not use a condom consistently; 2) had sexual relations with a non-regular partner during the previous six months and did not use a condom consistently or 3) used no protection during last sexual intercourse and had sexual relations with two or more partners in previous six months. Individuals who did not fall into these categories were considered as having safe sexual behaviours.

In order to assess the heterogeneous treatment effects by risky sexual behaviour at baseline, in our regression we included an interaction term, which is an interaction between the treatment variable and a dummy representing our risky behaviour indicator. If the interaction was significant, it would mean that the effect of male circumcision differed significantly in magnitude between men who did versus men who did not engage in risky sexual behaviours at baseline.

4.5.3 Missing data

A final concern is missing data. If there is a differential attrition rate between the circumcision arm and the control arm, and if men lost to follow-up are significantly different on the observables (covariates) at baseline, this could bias the estimated effect of male circumcision on HIV acquisition and risky sexual behaviours outcomes. To address this concern, we assessed whether there was a differential attrition rate between the two study arms. We also performed a test of mean differences for baseline covariates between those retained and those lost to follow-up. Furthermore, using an epidemiologic approach to perform a sensitivity analysis, we determined the number of additional individuals with certain characteristics in the treatment or control groups that would be required to offset the significant observed impact of male circumcision on HIV acquisition. Finally, using an econometric approach, we assessed how the impact of male circumcision on HIV acquisition would change in a scenario where all men lost to follow-up became HIV-positive or remained HIV-negative. These

extreme scenarios, though unlikely to reflect what actually occurred among the men lost to follow-up, allow us to understand how the magnitude of the impact would be affected and whether there would still be a significant impact of male circumcision on HIV acquisition.

4.5.4 Results of sensitivity analyses

In Table 9 we present intention-to-treat analyses with endline data, using OLS and Cox proportional hazards models to examine whether the impact of circumcision on HIV acquisition differed by sexual risk behaviour classification. In these analyses, the interaction coefficient (treatment X risky behaviour indicator) was not significant in either OLS or Cox models, showing that the impact of male circumcision did not differ between men who did and did not engage in risky sexual behaviours at baseline. The fact that male circumcision was not less effective for individuals engaging in risky behaviour rules out the second critique related to the sampling and selection, i.e. the argument that the effect of male circumcision might be biased because the sample consisted of mostly poorly educated, impoverished and unemployed men who were more likely to engage in sexual risky behaviour.[15]

Table 9: OLS and Cox estimation of heterogeneous effects of impact of male circumcision on HIV status by risky sexual behaviour at baseline

	OLS	Cox	
Variables	Beta coefficient	Hazard ratio	95% CI
Treatment	–0.021***	0.330***	(0.148–0.739)
	(0.008)		
Risky behaviour indicator	0.000	1.242	(0.692–2.229)
	(0.009)		
Treatment X risky behaviour indicator	0.002		
	(0.013)	1.48	(0.493–4.438)
Constant	0.034***		
	(0.006)		
Controls	No		No
Observations	2547		2740
R-squared	0.006	Time at risk	53382

Note: Standard errors and CI in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1.
(Source: Authors' estimates using public release analysis database)

In view of our findings showing no heterogeneous effect of male circumcision on HIV acquisition by risky sexual behaviour at baseline, we sought to extend this finding by investigating whether the impact of male circumcision on risky sexual behaviour outcomes varied depending on risky sexual behaviour at baseline. One would expect

[15] We conducted the same analysis by controlling for covariates that were imbalanced at the baseline; our results were unchanged.

that a lack of differential impact of male circumcision on HIV acquisition by risky sexual behaviour at the baseline would imply an absence of differential impact on risky behaviours by sexual behaviour at the baseline, because the main drivers of HIV acquisition in the study population are the men's risky sexual behaviours.

Using a range of analytic approaches including OLS, instrumental variable and fixed-effects estimation (Table A2 in Appendix), we found no evidence of heterogeneous effects of male circumcision on risky sexual behaviour outcomes, by baseline risky sexual behaviours. The interaction term (treatment X baseline risky behaviour indicator) coefficient, which allows us to assess whether the impact of male circumcision was different in the two groups, was not significant for any risky sexual behaviour outcomes.

In short, our results suggest that there is some risk compensation with male circumcision, so that men who are circumcised are somewhat more likely to engage in risky sexual behaviour. This effect does not vary if the man was engaging in high-risk sexual behaviour at baseline. Low-risk and high-risk men exhibited the same degree of risk compensation. The fact that the impact of male circumcision on HIV acquisition and risky sexual behaviours is not moderated by the level of risky sexual behaviours at baseline is a strong indicator that the protective effect observed in this study has a high external validity. In other words, this finding suggests that the protective effect of male circumcision is likely to be similar in contexts where men engage in riskier sexual behaviours or less-risky sexual behaviours. Consequently, the critique related to selection and sampling bias and more generally on external validity of the impact of male circumcision on HIV acquisition may not be relevant.

Table 10 present the results of our sensitivity analysis, in which we assess the robustness of our study results by determining how many individuals with contrary results would be required to offset our observed associations. We assessed the number of individuals required to attenuate our results to the point where they were no longer statistically significant, and assessed the number of individuals required to completely attenuate our results to the point where there was no difference between the study arms. As Table 10 illustrates, we began with the unadjusted study results from Table 1, producing a relative risk (RR) of 0.40 (0.21, 0.71), based on 18 of 1,390 seroconverting in the treatment group and 45 of 1,391 in the control group. When 10 hypothetical individuals are added as seroconverters to the treatment group, the RR becomes no longer significant at 0.621 (0.37, 1.01). Similarly, by adding 30 individuals as seroconverters to the treatment group, the RR rises to 1.06 (0.68, 1.62).

Examining the hypothetical impact of additional individuals in the control arm who did not seroconvert, we found that 500 such individuals would have been required to attenuate the association to non-significance (RR=0.58 [95 per cent CI: 0.32, 1.02]), and 1,800 additional individuals would have been required to observe a RR of 1.05 (95 per cent CI: 0.57, 1.86). Our conclusion based on these analyses is that it is implausible to speculate that bias resulting from missing data may explain the observed association. The imbalance in missing data that would be required to offset the strength of the observed circumcision effect is so lopsided that such an imbalance is extremely unlikely to have occurred.

Table 10: Sensitivity analysis: additional HIV-positive individuals in treatment group or additional HIV-negative individuals in control group required before losing significance of impact of male circumcision on HIV acquisition

	Treatment		Control			
	HIV+	Sample	HIV+	Sample	Relative Risk	CI
Results	18	1,390	45	1,391	0.40	0.21–0.71
Treatment group						
(+) 5 individuals	23	1,390	45	1,391	0.5118	0.30–0.86
(+) 10 individuals*	28	1,390	45	1,391	0.621	0.37–1.01
(+) 15 individuals*	33	1,390	45	1,391	0.73	0.45–1.17
(+) 20 individuals*	38	1,390	45	1,391	0.84	0.53–1.32
(+) 30 individuals*	48	1,390	45	1,391	1.06	0.68–1.62
Control group						
(+) 5 individuals	18	1,390	45	1,396	0.40	0.22–0.71
(+) 10 individuals	18	1,390	45	1,401	0.40	0.22–0.72
(+) 15 individuals	18	1,390	45	1,406	0.40	0.22–0.72
(+) 20 individuals	18	1,390	45	1,411	0.41	0.22–0.72
(+) 400 individuals	18	1,390	45	1,791	0.52	0.28–0.72
(+) 500 individuals	18	1,390	45	1,891	0.54	0.30–0.96
(+) 1000 individuals*	18	1,390	45	2,391	0.69	0.37–1.22
(+) 1800 individuals*	18	1,390	45	3,191	0.92	0.50–1.63

Note: * No longer significant. (Source: Authors' calculations using public release analysis database)

Except for age, we found no significant difference in baseline observable characteristics between men who were retained versus those lost to follow-up. Men lost to follow-up were younger (by less than 0.42 years) than men retained in the study. We also found no difference in the rate of attrition between the two study arms.[16] In Table 11, we present the findings of an extreme case where all men lost to follow-up are either all HIV-positive or all HIV-negative. Column 1 shows that if all men lost to follow-up were HIV-positive, male circumcision would not have any impact on HIV acquisition. However, Column 2 shows that if all men lost to follow-up remained HIV-negative, male circumcision would still show a reduction in HIV acquisition. The relative risk is similar to what we found in Table 7.

These results suggest that if all men lost to follow-up became HIV-positive, male circumcision would not have any impact on HIV acquisition. This scenario is unlikely, because the incidence of HIV for the men retained in the study, who had baseline observable characteristics similar to those lost to follow-up, was 2.66 (68/2,547). Given the fact that the men retained were similar to those lost to follow-up, it is plausible that the incidence rate for men lost to follow-up was close to 2.66 per cent, and not 100 per cent, as considered in this extreme scenario. In short, we think our results are robust to attrition.

[16] These results are available upon request.

Table 11: Effect of male circumcision on HIV acquisition, accounting for attrition

	All attritors are HIV-positive		All attritors are HIV-negative	
	(1)		(2)	
	OLS	Instrumental variable	OLS	Instrumental variable
	(1)	(2)	(1)	(2)
Treatment	-0.013	-0.0136	–0.0201***	–0.0213***
	(0.0142)	(0.0150)	(0.0058)	0.0062
Observations	2,781	2,781	2,781	2,781
R-squared	0.0003	0.0003	0.0042	0.0043

Note: Standard errors in parentheses. Significance *** p<0.01, ** p<0.05, * p<0.1. (Source: Authors' estimates using public release analysis database)

5. Discussion and implication

The results of this replication study have several important implications. Our pure replication suggests that the results reported in the original paper do not suffer from any errors that might come from different sources, such as construction of variables, data cleaning and codes used to obtain findings. Thus, we are able to confirm that the data and methods described by Bailey *et al.* (2007) are those used to produce the findings reported in the original paper.

The fact that we found no difference between findings obtained from instrumental variable (treatment-on-the-treated estimate) and the original findings (as-treated estimate) means that in this study, contrary to many impact evaluations of social interventions, unobserved individual characteristics that may or may not have changed over time appeared to play no role in men's decision to undergo male circumcision. The self-selection problem appears to have been addressed during the recruitment process, so that men who agreed to participate in the trial had already implicitly agreed to be circumcised if randomly assigned to the treatment group. This explains the very high uptake of male circumcision in that group. Thus, findings from instrumental variable estimation do not differ from the original findings because there appears to be no problem with self-selection in this study. Instrumental variable estimation would have produced different findings if self-selection had been present in the decision to undergo male circumcision.

This means that the protective effect of male circumcision against HIV acquisition obtained in this study has a strong external validity. This protective effect is similar to an effect that can be obtained through a vaccine or a medicine. The relative risk reduction of 60 per cent conferred by male circumcision means that once a man is circumcised, his risk of acquiring HIV is reduced by 60 per cent, compared with a man who is not circumcised – whatever the context. This protective effect is not confounded by unobserved individual characteristics. Moreover, this finding is reinforced by the fact that we found no heterogeneity of the impact of male circumcision by age group or baseline risky sexual behaviours. The protective effect of 60 per cent is neither a function of the frequency of sexual activity nor a function of risky sexual behaviours, as we found these factors do not diminish the impact of male circumcision.

We found some evidence of risk compensation using the original authors' methods, as well as stronger evidence from OLS, fixed-effects and instrumental variable estimation. We have evidence that men who are circumcised are less abstinent and engage in more risky sexual behaviours – i.e., unprotected intercourse. We have already found that risky sexual behaviours do not diminish the impact of male circumcision. Even if we assume that risky sexual behaviours partially mediate the impact of male circumcision on HIV acquisition, this mediation is already factored into the observed impact, given the existence of risk compensation. However, our heterogeneous effects analysis showed that risky sexual behaviours did not affect the impact of male circumcision. Since male circumcision provides only partial protection, this means that circumcised men engaging in risky sexual behaviours still have greater risk of acquiring HIV relative than men practicing safer sexual behaviours. Therefore, if circumcision leads to risk compensation, as observed in this study, the accumulation of risky sexual behaviours naturally increases the probability of acquiring HIV even though the impact of male circumcision is still 60 per cent. Finally, our results suggest that the relative decrease in risk is similar (60 per cent) in all behavioural risk groups.

By confirming results found in the original paper, this replication study reinforces the importance of male circumcision for HIV reduction and the need to find innovative strategies to increase the uptake of male circumcision in eastern and southern African countries. Our findings regarding the existence of risk compensation highlight the need to carefully evaluate the impact of male circumcision on risky sexual behaviour in different populations and outside of clinical trials in order to formulate appropriate messages in cases with a confirmed presence of risk compensation. Finally, we discovered although the econometric and epidemiologic approaches seem different at first glance, the results are in many cases very similar. This similarity is reinforced in this study when the uptake of the intervention is close to 100 per cent. We think that when possible these two methodological approaches should be used at least to assess the robustness of results, especially when there is self-selection (the actual uptake of the intervention is not high), even though participants are randomly assigned to the intervention and the control groups.

6. Conclusion

In this paper, we conducted a replication study of 'Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial' (Bailey *et al.* 2007), using the public release analysis database. This study was one of the three studies that led to the scale-up of male circumcision in eastern and southern Africa. We first conducted a pure replication using the same methods as the original authors, with the aim of reproducing and reconciling the findings published in the original study. Our pure replication confirmed the findings of the original report. Despite a few small differences in the details of analysis results (possibly attributable to changes made to the dataset to protect the anonymity of participants in preparation for public release), our overall conclusions are in concordance with those of Bailey *et al.* (2007). Our finding of a roughly 60 per cent reduction in HIV acquisition was robust to intention-to-treat or as-treated analysis; we also found that the intervention effect was similar for different age groups.

Second, in our MEA, we mainly used an econometric approach including OLS, fixed-effects and instrumental variables estimation to assess the impact of male circumcision on HIV acquisition. Our MEA revealed that male circumcision significantly reduced HIV acquisition. The relative risk effect size through OLS, fixed-effects and instrumental variables was very similar to what was found in the original study. Moreover, treatment-on-the-treated estimates through instrumental variables showed that male circumcision reduced HIV acquisition by 61.82 per cent, similar to the as-treated analysis of the original authors. Finally, using an econometric approach (OLS and instrumental variables estimation), we found strong evidence of the existence of risk compensation among circumcised men.

Third, using both an econometric approach and an epidemiologic approach, we conducted additional analysis to address some criticisms formulated towards the three trials of medical male circumcision. Our sensitivity analysis using an epidemiologic approach and an econometric approach showed that there was no evidence of heterogeneous impact of male circumcision on HIV acquisition and risky sexual behaviours by the level of risky sexual behaviours at the baseline. This finding partially rules out the critique related to sampling and selection bias. Results from our sensitivity analysis also show that it is implausible to speculate that bias resulting from missing data may explain the observed association. This finding rules out the critique related to missing data.

Our results provide strong support for the hypothesis that male circumcision can have a major impact on reducing the acquisition of HIV. However, given the state of knowledge about the relationship between male circumcision and risky sexual behaviours, scale-up of male circumcision should be associated with strategies to evaluate risk compensation. More studies are needed to assess the long-term impact of male circumcision on risky sexual behaviours outside of clinical trials.

Appendix

Table A 1: Baseline characteristics: original paper vs. replication results

Demographic characteristics	Circumcision group		Control group		Overall	
	Original	Replication	Original	Replication	Original	Replication
Age (years)	20 (19–22; 18–28; 1391)	20 (19–22; 18–24; 1390)	20 (19–22; 17–24; 1393)	20 (19–22; 18–24; 1391)	20 (19–22; 18–24; 2781)	20 (19–22; 18–24; 2781)
Ethnic group						
Luo	1,360 (97.84%)	1,360 (97.84%)	1,376 (98.92%)	1,376 (98.92%)	2,739 (98%)	2,736 (98.38%)
Other	30 (2.16%)	30 (2.16%)	15 (1.08%)	15 (1.08%)	45 (2%)	45 (1.62%)
Education level						
Less than secondary	468 (33.67%)	468 (33.67%)	478 (34.36%)	478 (34.36%)	946 (34.02%)	946 (34.02%)
Any secondary or above	922 (66.33%)	922 (66.33%)	913 (65.64%)	913 (65.64%)	1,835 (65.98%)	1,835 (65.98%)
Employment status						
Employed and receiving a salary	128 (9%)	128 (9.21%)	134 (10%)	134 (9.63%)	262 (9%)	262 (9.42%)
Self-employed	374 (27%)	374 (26.91%)	355 (25%)	355 (25.52%)	729 (26%)	729 (26.21%)
Unemployed	889 (64%)	888 (63.88%)	904 (65%)	902 (64.85%)	1,793 (64%)	1,790 (64.37%)
Occupation						
Professional/managerial	25 (2%)	25 (1.80%)	39 (3%)	39 (2.80%)	64 (2%)	64 (2.30%)
Skilled worker	141 (10%)	141 (10.14%)	113 (8%)	113 (8.12%)	254 (9%)	254 (9.13%)
Semi-skilled worker	95 (7%)	95 (6.83%)	86 (6%)	86 (6.18%)	181 (7%)	181 (6.51%)
Unskilled worker	698 (50%)	698 (50.22%)	758 (54%)	757 (54.42%)	1,456 (52%)	1,455 (52.32%)
Farm labourer/fisherman	107 (8%)	107 (7.70%)	90 (6%)	90 (6.47%)	197 (7%)	197 (7.08%)
Student	325 (23%)	324 (23.31%)	307 (22%)	306 (22%)	632 (23%)	630 (22.65%)
Marital status						
Not married (no live-in partner)	1,296 (93%)	1,296 (93.51%)	1,291 (93%)	1,289 (93.14%)	2,587 (93%)	2,585 (93.12%)
Not married (with live-in partner)	9 (0.6%)	9 (0.65%)	11 (0.8%)	11 (0.79%)	20 (0.7%)	20 (0.72%)
Married (not living with wife)	11 (0.8%)	11 (0.79%)	19 (1%)	19 (1.37%)	30 (1%)	30 (4.86%)

Demographic characteristics	Circumcision group		Control group		Overall	
	Original	Replication	Original	Replication	Original	Replication
Married (living with wife)	71 (5%)	70 (5.05%)	65 (5%)	65 (4.70%)	136 (5%)	135 (4.86%)
Physical and laboratory findings						
Weight (kg)	63 (59–68; 42–91; 1391)	63 (59–68; 42–91; 1390)	62 (58–67; 40–100; 1392)	62 (58–67; 40–100; 1390)	63 (59–67; 40–100; 2783)	63 (59–67; 40–100; 2780)
Haemoglobin (g/L)	154 (143–163; 90–199; 1386)	154 (143–163; 90–211; 1385)	153 (142–164; 83–201; 1391)	153 (142–164; 83–201; 1389)	153 (142–163; 83–201; 2777)	153 (142–163; 83–211; 2774)
Herpes simplex virus 2						
Positive	405 (29%)	387 (27.90%)	363 (26%)	348 (25.09%)	768 (28%)	735 (26.50%)
Negative	980 (71%)	1,000 (72.10%)	1,029 (74%)	1,039 (74.91%)	2,009 (72%)	2,039 (73.50%)
Syphilis						
Positive	19 (1%)	19 (1.37%)	9 (0.6%)	8 (0.58%)	28 (1%)	27 (0.97%)
Negative	1,369 (99%)	1,368 (98.63%)	1,379 (99.4%)	1,378 (99.42%)	2,748 (99%)	2,746 (99.03%)
Trichomonas vaginalis						
Positive	27 (2%)	27 (1.96%)	31 (2%)	31 (2.25%)	58 (2%)	58 (2.10%)
Negative	1,351 (98%)	1,351 (98.04%)	1,350 (98%)	1,348 (97.75%)	2,701 (98%)	2,699 (97.90%)
Neisseria gonorrhoeae						
Positive	32 (2%)	32 (2.33%)	25 (2%)	25 (1.81%)	57 (2%)	57 (2.07%)
Negative	1,351 (98%)	1,342 (97.67%)	1,355 (98%)	1,353 (98.19%)	2,697 (98%)	2,695 (97.93%)
Chlamydia trachomatis						
Positive	73 (5%)	73 (5.32%)	55 (4%)	55 (3.99%)	128 (5%)	128 (4.65%)
Negative	1,300 (95%)	1,300 (94.68%)	1,325 (96%)	1,323 (96.01%)	2,625 (95%)	2,623 (95.35%)
Haemophilus duereyi						
Positive	0	.	0	.	0	.
Negative	21 (100%)	21 (100%)	8 (100%)	8 (100%)	29 (100%)	29 (100%)
Sexual history with women						
Age at first sexual encounter (years)	16 (14–17; 5–23; 1346)	16 (14–17; 6–23; 1345)	16 (14–17; 6–24; 1354)	16 (14–17; 6–24; 1352)	16 (14–17; 5–24; 2700)	16 (14–17; 6–24; 2697)

Demographic characteristics	Circumcision group		Control group		Overall	
	Original	Replication	Original	Replication	Original	Replication
Sexual intercourse with any partner in previous 6 months						
Yes	1196 (86%)	1195 (86.16%)	1195 (86%)	1195 (86.16%)	2391 (86%)	2390 (86.16%)
No	192 (14%)	192 (13.84%)	194 (14%)	192 (13.84%)	386 (14%)	384 (13.84%)
Number of partners in previous 6 months						
0	192 (14%)	192 (13.84%)	192 (13.84%)	192 (13.84%)	384 (13.84%)	384 (13.84%)
1	611 (44%)	611 (44.05%)	616 (44.41%)	616 (44.41%)	1227 (44.23%)	1227 (44.23%)
2+	584 (42%)	584 (42.11%)	579 (41.74%)	579 (41.74%)	1163 (41.93%)	1,163 (41.93%)
Number of partners over lifetime	4 (3–7; 1–120; 1290)	4 (3–7; 1–50; 1289)	4 (3–7; 1–390; 1303)	4 (3–7; 1–50; 1301)	4 (3–7; 1–390; 2593)	4 (3–7; 1–50; 2590)
Gave gifts or money to a woman for sexual intercourse in previous 6 months						
Yes	195 (16%)	195 (16.24%)	210 (18%)	210 (17.46%)	404 (17%)	405 (16.85%)
No	1,002 (84%)	1,006 (83.76%)	985 (82%)	993 (82.54%)	1,987 (83%)	1,999 (83.15%)
Drank alcohol at last time of having sexual intercourse						
Yes	142 (10%)	141 (10.15%)	150 (11%)	150 (10.81%)	291 (11%)	291 (10.48%)
No	1,248 (90%)	1,248 (89.85%)	1,239 (89%)	1,237 (89.19%)	2,487 (89%)	2,485 (89.52%)
Used a condom with sexual intercourse in previous 6 months						
Always	265 (22%)	266 (22.22%)	254 (21%)	254 (21.15%)	519 (22%)	520 (21.68%)
Inconsistent	620 (52%)	620 (51.80%)	632 (53%)	632 (52.62%)	1,252 (52%)	1,252 (52.21%)
Never	308 (26%)	311 (25.98%)	307 (26%)	315 (26.23%)	615 (26%)	626 (26.11%)
Last occurrence of sexual intercourse was with regular partner						

Demographic characteristics	Circumcision group		Control group		Overall	
	Original	Replication	Original	Replication	Original	Replication
Yes	842 (80%)	841 (79.94%)	826 (78%)	826 (78.44%)	1,668 (79%)	1,667 (79.19%)
No	211 (20%)	211 (20.06%)	227 (22%)	227 (21.56%)	438 (21%)	438 (20.81%)
Trouble achieving/maintaining erection in previous 6 months (participants with partner in previous 6 months)						
Yes	80 (7%)	80 (6.71%)	89 (7%)	89 (7.46%)	169 (7%)	169 (7.09%)
No	1,112 (93%)	1,112 (93.29%)	1,104 (93%)	1,104 (92.54%)	2,215 (93%)	2,216 (92.91%)
Sexual history with men						
Ever had sexual relations with a boy or man						
Yes	5 (0.4%)	5 (0.36%)	1 (0.01%)	1 (0.07%)	6 (0.2%)	6 (0.22%)
No	1,385 (99.6%)	1,384 (99.64%)	1,388 (99.9%)	1,386 (99.93%)	2,773 (99.8%)	2,770 (99.78%)
Injection history						
Received an injection for any reason in previous 6 months						
Yes	391 (28%)	391 (28.17%)	360 (26%)	360 (25.96%)	751 (27%)	751 (27.06%)
No	998 (72%)	997 (71.83%)	1,029 (74%)	1,027 (74.04%)	2,027 (73%)	2,024 (72.94%)

Note: *Numbers presented are median (interquartile range, range, n) for continuous variables and n (per cent) for categorical variables.

Table A 2: OLS estimation of heterogeneous effects of impact of male circumcision on risky sexual behaviour by risky sexual behaviour at baseline

	OLS results		Instrumental variable (IV) results		Fixed-effects results	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Consistent condom use in previous 6 months						
Treatment	–0.008	–0.006	–0.008	–0.006	–0.003	0.001
	(0.0246)	(0.0248)	(0.0264)	(0.0266)	(0.0140)	(0.0140)
Risky behaviour indicator	–0.0248	–0.0250	–0.0238	–0.0240	–0.0857***	–0.0831***
	(0.0268)	(0.0271)	(0.0272)	(0.0274)	(0.0151)	(0.0151)
Treatment X risky behaviour indicator	–0.0566	–0.0556	–0.0594	–0.0584	–0.0216	–0.0253
	(0.0380)	(0.0383)	(0.0403)	(0.0405)	(0.0214)	(0.0215)
Controls	No	Yes	No	Yes	No	Yes
Observations	2,514	2,479	2,514	2,479	10,408	10,263
R-squared	0.005	0.010	0.005	0.011	0.0451	0.0614
Number of individuals					2,547	2,511
Total effect	–0.064	–0.0615	–0.0676	–0.0647	Hausman test (IV vs. OLS)	
Standard errors	(0.0290)	(0.0292)	(0.0305)	(0.0306)	Chi square	0.00
p-value	0.027	0.035	0.027	0.035	p-value	1.0000
Panel B: Two or more partners in previous 6 months						
Treatment	0.0228	0.0207	0.0245	0.0222	–0.00163	–0.00363
	(0.0237)	(0.0239)	(0.0254)	(0.0256)	(0.0140)	(0.0141)
Risky behaviour indicator	0.226***	0.221***	0.226***	0.221***	0.361***	0.357***
	(0.0258)	(0.0261)	(0.0262)	(0.0265)	(0.0151)	(0.0152)
Treatment X risky behaviour indicator	–0.0239	–0.0187	–0.0257	–0.0201	0.00487	0.00738
	(0.0366)	(0.0369)	(0.0388)	(0.0391)	(0.0215)	(0.0216)
Controls	No	Yes	No	Yes	No	Yes
Observations	2,472	2,439	2,472	2,439	10,314	10,173
R-squared	0.053	0.054	0.053	0.054	0.1372	0.1404

	OLS results		Instrumental variable (IV) results		Fixed-effects results	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of individuals					2,547	2,511
Total effect	–0.0011	0.00198	–0.0011	0.00209	Hausman test (IV vs. OLS)	
Standard errors	0.0279	(0.0281)	(0.0294)	(0.0296)	Chi square	0.00
p-value	0.969	0.944	0.969	0.944	p-value	1.0000
Panel C: Sexual abstinence in previous 6 months						
Treatment	–0.0437**	–0.0440**	–0.0469**	–0.0472**	–0.0207*	–0.0217*
	(0.0188)	(0.0189)	(0.0202)	(0.0202)	(0.0123)	(0.0123)
Risky behaviour Indicator	–0.120***	–0.109***	–0.120***	–0.109***	–0.124***	–0.117***
	(0.0205)	(0.0206)	(0.0208)	(0.0209)	(0.0133)	(0.0133)
Treatment X Risky behaviour Indicator	0.0310	0.0301	0.0335	0.0326	0.0131	0.0126
	(0.0291)	(0.0291)	(0.0308)	(0.0308)	(0.0189)	(0.0188)
Controls	No	Yes	No	Yes	No	Yes
Observations	2,516	2,481	2,516	2,481	10,415	10,270
R-squared	0.022	0.038	0.021	0.037	0.0258	0.0375
Number of individuals					2,547	2,511
Total effect	–0.0127	–0.0139	–0.0134	–0.0146	Hausman test (IV vs. OLS)	
Standard errors	(0.0222)	(0.0222)	(0.0233)	(0.0233)	Chi square	0.00
p-value	0.566	0.531	0.565	0.530	p-value	1.0000
Panel D: Last time had sexual relations with a casual partner						
Treatment	0.0253	0.0270	0.0272	0.0290	0.00467	0.00613
	(0.0203)	(0.0205)	(0.0218)	(0.0220)	(0.0112)	(0.0113)
Risky behaviour indicator	0.107***	0.107***	0.107***	0.108***	0.162***	0.164***
	(0.0222)	(0.0224)	(0.0225)	(0.0227)	(0.0121)	(0.0122)
Treatment X risky behaviour indicator	–0.0256	–0.0241	–0.0275	–0.0260	–0.00755	–0.00797
	(0.0314)	(0.0317)	(0.0333)	(0.0335)	(0.0172)	(0.0173)
Controls	No	Yes	No	Yes	No	Yes
Observations	2,516	2,481	2,516	2,481	10,415	10,270

	OLS results		Instrumental variable (IV) results		Fixed-effects results	
	(1)	(2)	(3)	(4)	(5)	(6)
R-squared	0.015	0.018	0.015	0.018	0.0413	0.0432
Number of individuals					2,547	2,511
Total effect	–0.0003	0.0027	-0.00032	0.0030	Hausman test (IV vs. OLS)	
Standard errors	(0.0240)	0.0241	0.0252	0.0253	Chi square	0.00
p-value	0.990	0.905	0.990	0.905	p-value	1.0000
Panel E: Unprotected sexual intercourse with any partner in previous 6 months						
Treatment	0.0466*	0.0423	0.0500*	0.0455	0.0206	0.0176
	(0.0258)	(0.0258)	(0.0277)	(0.0277)	(0.0160)	(0.0159)
Risky behaviour indicator	0.141***	0.128***	0.140***	0.127***	0.206***	0.196***
	(0.0281)	(0.0281)	(0.0286)	(0.0286)	(0.0173)	(0.0171)
Treatment X risky behaviour indicator	0.0316	0.0345	0.0322	0.0355	0.0168	0.0211
	(0.0399)	(0.0398)	(0.0423)	(0.0422)	(0.0246)	(0.0243)
Controls	No	Yes	No	Yes	No	Yes
Observations	2,516	2,481	2,516	2,481	10,415	10,270
R-squared	0.027	0.049	0.027	0.049	0.0451	0.0614
Number of individuals					2,547	2,511
Total effect	0.0781	0.0769	0.0822	0.0809	Hausman test (IV vs. OLS)	
Standard errors	(0.0304)	(0.0303)	0.0320	0.0319	Chi square	0.00
p-value	0.010	0.011	0.010	0.011	p-value	1.0000

Note: Standard errors in parentheses. Significance *** p<0.01, ** p<0.05, * p<0.1

Total effect is the coefficient obtained from linear combination (Treatment + Treatment X Sexual risk indicator).

Source: Authors' estimates using public release analysis database.

Table A 3: Total number of HIV individuals at endline, by treatment status

Status	Control group	Circumcision group	Total
HIV-negative	1,227	1,252	2,479
HIV-positive	49	19	68
Total	1,276	1,271	2,547

Note: Authors' calculations using public release analysis database

Table A 4: Total number of HIV-positive observations per individual

Status	Control group	Circumcision group	Total
HIV-negative observations	5,326	5,334	10,660
HIV-positive observations	128	47	175
Total observations	5,454	5,381	10,835

Note: Authors' calculations using public release analysis database

Table A 5: Total number of HIV-positive observations, by circumcision status

Status	Not circumcised	Circumcised	Total
HIV-negative	1,267	1,212	2,479
HIV-positive	49	19	68
Total observations	1,316	1,231	2,547

Note: Authors' calculations using public release analysis database

References

Ainsworth, M and Mead, OA, 1997. Confronting AIDS: public priorities in a global epidemic. A World Bank policy research report. Washington, D.C.: The World Bank. Available at: http://documents.worldbank.org/curated/en/1997/10/694016/confronting-aids-public-priorities-global-epidemic [Accessed 4 August 2015].

Auvert, B, Taljaard, D, Lagarde, E, Sobngwi-Tambekou, J, Sitta, R and Puren, A, 2005. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 trial. *PLOS Medicine,* 2(11)**,** E298.

Bailey, RC, Moses, S, Parker, CB, Agot, K, Maclean, I, Krieger, JN, Williams, CFM, Campbell, RT and Ndinya-Achola, JO, 2007. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *The Lancet*, 369(9562), pp. 643–56.

Bailey, RC, Neema, S and Othieno, R, 1999. Sexual behaviors and other HIV risk factors in circumcised and uncircumcised men in Uganda. *Journal of Acquired Immune Deficiency Syndromes*, 22(3), pp. 294–301.

Brown, AN, Cameron, DB and Wood, BDK, 2014. Quality evidence for policymaking: I'll believe it when I see the replication. *Journal of Development Effectiveness*, 6(3), pp. 215–35.

Buchbinder, SP and Liu, A, 2011. Pre-exposure prophylaxis and the promise of combination prevention approaches. *AIDS and Behavior*, 15(Suppl 1), pp. 72–79.

Chinkhumba, J, Godlonton, S and Thornton, R, 2014. The demand for medical male circumcision. *American Economic Journal: Applied Economics*, 6(2), pp. 152–77.

Granich, R, Crowley, S, Vitoria, M, Smyth, C, Kahn, JG, Bennett, R, Lo, Y-R, Souteyrand, Y and Williams, B, 2010. Highly active antiretroviral treatment as prevention of HIV transmission: review of scientific evidence and update. *Current Opinion in HIV and AIDS*, 5(4), pp. 298–304.

Gray, RH, Kigozi, G, Serwadda, D, Makumbi, F, Watya, S, Nalugoda, F, Kiwanuka, N, Moulton, PLH, Chaudhary, MA, Chen, MZ, Sewankambo, NK, Wabwire-Mangen, F, Bacon, MC, Williams, CF, Opendi, P, Reynolds, SJ, Laeyendecker, O, Quinn, TC and Wawer, MJ, 2007. Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *The Lancet*, 369(9562), pp. 657–66.

Havlir, D, 2013. Sustainable east Africa research in community health (SEARCH). Available at: www.searchendaids.com [Accessed on 4 September 2015].

Hayes, R, Ayles, H, Beyers, N, Sabapathy, K, Floyd, S, Shanaube, K, Bock, P, Griffith, S, Moore, A, Watson-Jones, D, Fraser, C, Vermund, S, Fidler, S and Team, THS, 2014. HPTN 071 (PopART): rationale and design of a cluster-randomised trial of the population impact of an HIV combination prevention intervention including universal testing and treatment – a study protocol for a cluster randomised trial. *Trials*, [online] Available at: <http://www.trialsjournal.com/content/15/1/57> [Accessed 4 September 2015.

Iwuji, CC, Orne-Gliemann, J, Tanser, F, Boyer, S, Lessells, RJ, Lert, F, Imrie, J, Bärnighausen, T, Rekacewicz, C, Bazin, B, Newell, M-L and Dabis, F, 2013. Evaluation of the impact of immediate versus who recommendations-guided antiretroviral therapy initiation on HIV incidence: the ANRS 12249 TASP (treatment as prevention) trial in Hlabisa sub-district, Kwazulu-Natal, South Africa: study protocol for a cluster randomised controlled trial. *Trials*, [online] Available at: <http://www.trialsjournal.com/content/14/1/230> [Accessed 4 September 2015].

Kaplan, E. L. and Meier, P, 1958. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282), pp. 457–81.

Moore, JS, Levine, WC, Essex, M, Degruttola, V, Lebelonyane, R and El Halabi, S, 2015. Botswana Combination Prevention Project [Online] Available at: https://clinicaltrials.gov/ct2/show/NCT01965470 [Accessed 4 September 2015].

Moses, S, Bradley, JE, Nagelkerke, NJD, Ronald, AR, Ndinya-Achola, JO and Plummer, FA, 1990. Geographical patterns of male circumcision practices in Africa: association with HIV seroprevalence. *International Journal of Epidemiology*, 19(3), pp. 693–97.

Njeuhmeli, E, Forsythe, S, Reed, J, Opuni, M, Bollinger, L, Heard, N, Castor, D, Stover, J, Farley, T, Menon, V and Hankins, C, 2011. Voluntary medical male circumcision: modeling the impact and cost of expanding male circumcision for HIV prevention in eastern and southern Africa. *PLOS Medicine*, [online] Available at: < http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001132> [Accessed 4 September 2015].

Oster, E, 2012. HIV and sexual behaviour change: why not Africa? *Journal of Health Economics*, 31(1), pp. 35–49.

Schoenfeld, D, 1982. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1), pp. 239–41.

Siegfried, N, Muller, M, Deeks, J, Volmink, J, Egger, M, Low, N, Walker, S and Williamson, P, 2005. HIV and male circumcision – a systematic review with assessment of the quality of studies. *The Lancet Infectious Diseases*, 5(3), pp. 165–73.

Therneau, TM and Grambsch, PM, 2000. *Modeling survival data: extending the Cox model*. New York: Springer-Verlag.

UNAIDS, 2013. Global Report. Geneva, Switzerland: The Joint United Nations Programme on HIV/AIDS.

WHO and UNAIDS, 2007. New data on male circumcision and HIV prevention: policy and programme implications. [Online] Available at: <http://data.unaids.org/pub/Report/2007/mc_recommendations_en.pdf> [Accessed 3 February 2015].

Wooldridge, JM, 2000. *Introductory econometrics: a modern approach*. Cincinnati, OH: South-Western College Publishing.

Publications in the 3ie Replication Paper Series

The following papers are available from http://www.3ieimpact.org/en/publications/3ie-replication-paper-series/

Quality evidence for policymaking: I'll believe it when I see the replication, 3ie Replication Paper 1. Brown, AN, Cameron, DB and Wood, BDK (2014)

TV, female empowerment and demographic change in rural India, 3ie Replication Paper 2. Iversen, V and Palmer-Jones, R (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 1: A pure replication, 3ie Replication Paper 3, part 1. Aiken, AM, Davey, C, Hargreaves, JR and Hayes, RJ (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 2: Alternative analyses, 3ie Replication Paper 3, part 2. Aiken, AM, Davey, C, Hayes, RJ and Hargreaves, JR (2014)

The long and short of returns to public investments in fifteen Ethiopian villages, 3ie Replication Paper 4. Bowser, WH (2015)

Recalling extra data: A replication study of finding missing markets, 3ie Replication Paper 5. Wood, BDK and Dong, M (2015)

The impact of India's JSY conditional cash transfer programme: A replication study, 3ie Replication Paper 6. Carvalho, N and Rokicki, S (2015)

Walking on solid ground: a replication study on Piso Firme's impact, 3ie Replication Paper 7. Basurto, MP, Burga, R, Toro, JLF and Huaroto, C (2015)

Replication Paper Series

International Initiative for Impact Evaluation
1625 Massachusetts Ave., NW
Suite 450
Washington, DC 20036
USA

replication@3ieimpact.org
Tel: +1 202 629 3939

www.3ieimpact.org