The Power to Protect: Household Bargaining and Female Condom Use*

Rachel Cassidy
Institute for Fiscal Studies

Marije Groot Bruinderink
Amsterdam Institute for Global Health and Development

Wendy Janssens
Amsterdam Institute for Global Health and Development

Karlijn Morsink†
University of Oxford
JOB MARKET PAPER

December 4, 2017
Click here for latest version

Abstract: Women may have a higher willingness than men to adopt certain household technologies. If use of a technology must be agreed upon by both partners, then women with low bargaining power may not convince their partners to adopt. Introducing a version of the technology that is more acceptable to men, even if somehow second-best from the perspective of a social planner, may therefore improve adoption and welfare in contexts where women have low bargaining power. Female condoms offer marginally lower protection and higher unit cost than male condoms, but lower discomfort and stigma especially to men. We conduct a field experiment in the slums of Maputo, Mozambique, where 30% of women have HIV. The intervention offers free access to and information about female condoms. We find strongest adoption among women with low household bargaining power, who are previously having unprotected sex. There is no substitution away from male condoms. Results are corroborated by high-frequency sexual diary data. We estimate that free provision of this second-best technology is therefore cost-effective under a range of simulated scenarios. The findings highlight how policy should take into account bargaining power and asymmetry in the costs and benefits of technology adoption across household members.
JEL classification: C78, J16, I12, O15, O33

---

*We thank Dan Anderberg, Nava Ashraf, Martin Browning, James Fenske, Glenn Harrison, Eline Korenromp, Emily Oster, Simon Quinn and Chris Woodruff for extremely helpful comments. We also thank audiences at the Oxford Development Economics Workshop 2017, NEUDC 2017, the University of Warwick, and the Institute for Fiscal Studies. We are grateful Lene Boehnke and Emilie Berkhout for excellent research assistance, and to Balthazar Chilundo from Eduardo Modlane University, Pathfinder, Oxfam Novib, and WeConsult for research support. This research project was funded by the Netherlands Organisation for Scientific Research (NWO) WOTRO Science for Global Development (grant no. W 07.40.203) and the Universal Access to Female Condoms (UAFC) Joint Programme (grant no. A-02974-02-01/506671). The IRB approval was given by the National Research Ethics Committee of Mozambique, the Comité Institucional de Bioética em Saúde da Faculdade de Medicina/Hospital Central de Maputo, Mozambique.

†Corresponding author. Email: karlijn.morsink@economics.ox.ac.uk. Website: www.karlijnmorsink.com

1 Introduction

Partners may differ systematically in preferences that determine the adoption of household technologies. For example, there is evidence to suggest that women have a stronger preference for risk reduction (Agnew et al., 2008) and household public goods (Duflo, 2003), and are less overconfident (Barber and Odean, 2001). Women may also bear more of the costs of non-adoption, through responsibility for domestic chores, or greater exposure to certain health or safety risks (Miller and Mobarak, 2013; Stopnitzky, 2017). The extent to which households adopt welfare-enhancing technologies has been shown to depend on women's bargaining power and control over household resources (Attanasio and Lechene, 2014; Duflo and Udry, 2004). Bargaining is likely to be especially important in cases where technologies cannot be adopted unilaterally, but instead require the agreement of both partners (Ashraf et al., 2014b). In such cases, introducing a technology that is second-best from the perspective of a social planner but more acceptable to men may still increase welfare.

This paper uses a field experiment to examine how intra-household bargaining affects adoption of condoms: a technology that is observable to both parties, and requires joint agreement for use.[1] We do so in a context where women disproportionately bear the costs of unprotected sex, but have low bargaining power, whilst men place a higher weight on the loss of pleasure associated with using male condoms, due to discomfort and stigma. We model and test the impacts of offering female condoms, which are second-best from the perspective of a social planner who cares about cost and health, due to marginally lower effectiveness than male condoms (Farr et al., 1994; Trussell, 2011; Beksinska et al., 2012)[2] and higher production costs (Mantell et al., 2015).[3] However, there is evidence that female condoms provide greater pleasure than male condoms, especially to men (Philpott et al., 2006; Wanyenze et al., 2011; Koster et al., 2015). We show that women with lower bargaining power, *who were previously having unprotected sex,* convince their partners to adopt female condoms when they are made freely available.

[1] Both male and female condoms are non-concealable. Female condoms can be inserted by women prior to intercourse, although they still remain visible to both partners. In contrast, contraceptive technologies that protect solely against pregnancy, such as the pill and injectables, can be used by women alone, and are often sufficiently concealable such that women may adopt them unilaterally. A large economic literature has documented the consequent impact of the pill on women's economic empowerment (Goldin and Katz, 2002; Bailey, 2006; Chiappori and Oreffice, 2008).

[2] Trussell (2011) finds in the US that first-generation female condoms have 95% efficacy at preventing pregnancy in the first year of use if perfectly used, but 79% effectiveness in ordinary use. Meanwhile male condoms are 98% effective if perfectly used, and 85% effective in typical use. On the other hand, Macaluso et al. (2007) find in a 20-week RCT that female condoms are associated with a much higher rate of failure or incorrect use than male condoms (34% versus 9%), but that this does not lead to a significant difference in efficacy between male and female condoms.

[3] The unit production cost for female condoms at current volumes is $0.57 compared to $0.03 for male condoms. This is due to a monopoly on the production of WHO-approved female condoms, and consequent low production volumes (Peters et al., 2010). Lower-cost female condoms have been developed in India and approved by the EU, although have been waiting for several years for WHO approval (*ibid.*). Estimates suggest that increases in the scale of female condom production could lead to a cost-neutral increase in coverage relative to male condoms (Dowdy et al., 2006).

Meanwhile, we see little substitution away from male condoms. Thus when women hold low bargaining power, the relevant cost comparison for the public health system is not between female and male condoms, but rather between female condoms and antiretroviral therapies.[4] Using an epidemiological model that takes into account the negative externalities of unprotected sex in terms of HIV/AIDS transmission, we estimate that free provision of this second-best technology is cost-effective, given the zero effect on male condoms and the cost of antiretroviral therapy. This demonstrates how policy design should take into account asymmetry in the costs and benefits of technology adoption across household members.

Condoms are an important technology from a public health perspective, since they are the only well-established technology that protects against sexually transmitted infections (STIs): that is, contraceptives which protect not only against unwanted pregnancy but also against HIV/AIDS and other STIs. There are several reasons to believe that women may have a higher willingness than men to adopt condoms. First and foremost, women are more vulnerable to HIV infection: in 2015, women accounted for 59% of all individuals aged 15 and over living with HIV, and the rate of new infections among young women aged 15-24 was double that among young men (UNAIDS, 2016a).[5] Second, women may also be more risk-averse than men (Eckel and Grossman, 2008). Third, women also disproportionately bear the costs of unwanted pregnancy, increasing their demand for contraceptives across the board. The observation that many women hold low bargaining power in the household, especially in the developing world, may therefore partly explain why condoms are subject to low adoption from the perspective of policymakers who care about health and costs.[6]

To examine these issues we evaluate a condom programme in the slums of Matola, Mozambique, which seeks to increase condom use by offering female condoms alongside male condoms. The female HIV prevalence rate in Matola has been estimated at 29.6% (Ministério da Saúde, 2015), and thus condom adoption is a key public health concern. In the programme, women attend a series of group sessions that provide information about contraceptives. Female condoms are also added to the set of products carried by local health workers, which already includes male condoms, which participants can access freely and discreetly at the end of each session.[7] The intervention thus allows us to study which women, if any,

[4]In principle the costs of severe ill health and caring for dependents should also be taken into account, especially given that women with low bargaining power are less likely to have access to antiretroviral therapy (Ruger, 2004).

[5]The reason for this gender disparity is that women tend to have older partners, lower access to sexual and reproductive health services, and a higher biological risk than men of becoming infected from heterosexual intercourse (UNAIDS, 2016b)

[6]For example, in 2015 alone, an estimated 3.3 billion risky sex acts took place without condoms in Sub-Saharan Africa, leading to 910,000 new HIV infections (UNAIDS, 2016a). The total estimated need is an estimate of the number of sex acts between sex workers and their clients (38%), non-regular partners (20%), partners requiring condoms as a method of family planning (15%), HIV-affected couples (14%), and men who have sex with men (8%). Of the 2.7 billion condoms which were used, female condoms accounted for just 1.6%.

[7]The female condoms distributed are of the FC2 variety, which are made of synthetic nitrile and have been found to have greater acceptability among users. See `https://fc2femalecondom.com/fc2-global-home/` for details.

adopt female condoms when informational, access and price constraints are alleviated.

To formalise our predictions, we introduce a collective model of the household where partners jointly choose STI protection methods. Both men and women value the levels of pleasure and of health protection associated with different contraceptive technologies. However, for the reasons outlined above, we argue that the relative marginal utility from health compared to pleasure is greater for women than for men. When the only STI protection options available are male condoms or unprotected sex,[8] the model predicts that women prefer to use male condoms, but that those with low bargaining power are unable to convince their partners to do so.

When female condoms are introduced, three effects are predicted by the model. First, on the intensive margin, some women with low bargaining power who were previously having unprotected sex are now able to convince their partners to adopt female condoms (but not male condoms), increasing condom coverage.[9] Second, some women with intermediate bargaining power who were previously using male condoms also substitute into using female condoms, decreasing average condom effectiveness. Third, on the extensive margin, some couples who were not previously having sex now have sex with female condoms. The relative magnitudes of each of these margins of response are important to determine empirically, in order to establish total effects on HIV transmission and on welfare.

We use a phased-in randomised control trial (RCT) to compare women who attended the condom training to women who were randomised to wait an additional six months before beginning the course. In addition to baseline and endline data we collect weekly sexual diary data for a subsample of the women. This allows us to investigate impacts at the sex-act level and mitigate concerns about recall bias and mis-reporting. To mitigate the latter concern we also record the number of condoms that participants take with them after each session. To measure bargaining power, we collect information about assets brought by the woman to the relationship, and also enumerate two different survey modules covering decision-making and power dynamics in the relationship (Donald et al., 2017).[10]

The results show a large impact of treatment on female condom use: an increase of 18.4 percentage points in the proportion of women who have ever used female condoms (equivalent to 209% of the endline mean in the control group) and an increase of 7.7 percentage points (385%) in the proportion of those currently using female condoms. Importantly, we see no significant evidence of substitution away

[8]This includes sex protected by pure contraceptives such as the pill, but not by an STI protection method.

[9]Inserting female condoms prior to intercourse may also allow women with low bargaining power to change the default from unprotected sex to female condom use as partners enter into bargaining over condom use. It could also potentially act as a commitment device, for example if arousal leads to changes in risk taking behaviour (Ariely and Loewenstein, 2006). In principle, both of these mechanisms could increase the likelihood that women with a given bargaining power are able to convince their partners to use female condoms.

[10]Whilst we also sought to include incentivized measures of remaining STI-free and of bargaining power, unfortunately the National Research Ethics Committee of Mozambique's procedures do not allow any financial remuneration of survey participants.

from male condoms. Moreover, the diary data show that treatment leads to an increase of 9.1 percentage points in the probability that an individual has sex each week (19% of the endline mean in the control group). As predicted by the model these results are driven by women with lower baseline bargaining power *and* those who are having unprotected sex at baseline. These results exist across our measures of bargaining power. Furthermore, the heterogeneous treatment effect by bargaining power is robust to controlling for access to male condoms and distance to health centres, use of other contraceptives, HIV status, and beliefs about HIV risk. Moreover, we see no impacts of our treatment on knowledge of HIV and on our measures of bargaining power. Results are particularly strong among women who are not in a stable relationship at baseline and women who are HIV-positive, suggesting increased coverage of the sex acts with the highest risk of HIV transmission.

To establish if the intervention is cost effective, we conduct simulations of the scale-up of our intervention, taking into account the increase in condom coverage, but also the decrease in average condom effectiveness compared to pure use of male condoms, and the observed increase in the number of sex acts. The results suggest that scaling up our full intervention is not cost-effective based on WHO cost-effectiveness standards[11], but that scaling up the free distribution of female condoms, with information provision coming through existing sex education programs, could be *highly cost-effective* or even *cost-saving* compared to the cost of antiretroviral therapy for the individual infections averted.[12]

To our knowledge, this is the first experimental study to consider intra-household bargaining over condom use, and how the introduction of female condoms might interact with the bargaining process.[13] A broader literature has examined the relationship between intra-household bargaining and technology adoption in the form of cookstoves (Miller and Mobarak, 2013), savings accounts (Schaner, 2015), saving through ROSCAs (Anderson and Baland, 2002) and microfinance (Van Tassel, 2004). With respect to contraception adoption, Gertler et al. (2005) model bargaining over male condom use between female sex workers and male clients in Mexico, and use observational data to establish that unprotected sex is associated with a higher price.

Our contribution is to model the repeated household bargaining process that takes place within couples, rather than a one-shot or finite-horizon interaction mediated by price. The existing literature on bargaining within couples focuses on fertility (Eswaran, 2002; Rasul, 2008; Ashraf et al., 2014b). In contrast, we consider bargaining over STI protection methods that also guard against HIV/AIDS and other STIs, which by their nature are not concealable protection methods. Regarding female condoms, a number of medical studies have shown positive effects on the likelihood that sex acts are protected as a consequence of introducing female condoms alongside male condoms (Vijayakumar et al., 2006; Fontanet

[11] http://www.who.int/choice/en/

[12] The Government of Mozambique's National Strategic Plan on HIV/AIDS includes funding antiretroviral therapy for all infected individuals; see Section 6 for details.

[13] For qualitative evidence in the sexual health literature, see Choi et al. (2004).

et al., 1998; Mantell et al., 2015; Coman et al., 2013), but have not considered the role of intra-household bargaining. In the economics literature, Ashraf et al. (2014a) examine the effect of incentives to sell female condoms on sales by agents, but do not study impacts on end users. Finally, we add to a broader economic literature which studies the impact of information and incentive interventions on reducing HIV transmission and broader risky sexual behaviour (Thornton, 2008; Dupas, 2011; De Walque et al., 2012; Baird et al., 2012, 2014; Bandiera et al., 2015; Galárraga et al., 2014; Thirumurthy et al., 2014; Bjorkman Nyqvist et al., 2015; Duflo et al., 2015; Godlonton et al., 2016).

Section 2 describes the context and intervention. Section 3 outlines the theoretical framework and predictions for the intervention's effects depending on female bargaining power. Section 4 describes the study design and data sources. Section 5 details the results and robustness checks. Section 6 describes the cost-benefit and cost-effectiveness analyses. Section 7 concludes.

2 Context and intervention

2.1 HIV and condom use in Maputo

Our study took place in Matola, the capital of Maputo Province, Mozambique. Matola lies approximately 10km west of Maputo City, and with 893,000 inhabitants is the second most populated area of Mozambique (Instituto Nacional de Estatistica Moçambique, 2014). HIV prevalence in Maputo Province is high and disproportionately affects women, at an estimated 29.6% for women and 15.8% for men (Ministério da Saúde, 2015). Concurrency has been identified as a contributing factor: having multiple concurrent sexual partners is regarded as acceptable and even desirable for men, although this practice is more frowned-upon for women (Macia et al., 2011). This is true both inside and outside of stable relationships. Out of the women in our sample 85% are in stable relationships, with an average duration of nine years and no break-ups during the length of our study. Nonetheless, 36% of the women in stable relationships report that they believe their partner is involved with other people. In such a climate, STI protection methods, which protect against transmission of HIV and other STIs, are not close substitutes for pure contraceptive methods such as the pill. As a result, STI protection methods may be used in addition to pure contraceptive methods: in our baseline sample, 39% of respondents are currently using pure contraceptive methods (mainly the pill or injectables), and of those 40% are also currently using male condoms.

Both male and female condoms are available in Matola, but male condoms are far more accessible, especially for men. Female condoms are generally only available at health facilities, which are often difficult to reach: subjects in our sample report taking on average 60 minutes to get to the nearest health facility. Even at these health facilities, female condoms are subject to frequent stock-outs (Pilz, 2014). Meanwhile, male condoms are readily available, both for free at health facilities and cheaply on the private market. Women nevertheless report feeling uncomfortable purchasing male condoms, and are also less able than men to access places where male condoms are sold more discreetly, such as bars and nightclubs.

Despite the widespread availability of male condoms, at least for men, there is evidence that men's preferences constrain adoption.[14] Of the women in our study who are currently sexually active but not using any form of protection at baseline, by far the most common reason is that their partner does not like using male condoms or simply refuses to use them (45% of responses).[15] This reluctance appears to be less of a problem for women with higher bargaining power: the probability of a woman reporting that her partner obstructs use of male condoms in this way is negatively and highly significantly correlated with four of our six measures of women's bargaining power; see Section 4.4 for details.[16]

2.2 Female condom intervention

The intervention we study is run by Pathfinder International, and is aimed at women in populations with high HIV risk. The programme consists of six group sessions lasting ninety minutes each, held every two weeks over a three month period. The facilitators are trained female health workers from the local area, and thus are socially proximal to the participants. The sessions cover information on female condoms and demonstration of their use on pelvic models; information about other contraceptive methods; information on HIV/AIDS and other STIs; and discussions around issues of consent, negotiation of contraceptive use, intimate partner violence, and women's rights. These discussions are a key ethical component of the programme design, in order to mitigate the risk of these women facing increased violence when introducing new contraceptives into the home. Group sizes range from a minimum of five to a maximum of twelve women per facilitator, which are thresholds set by the NGO for creating an environment conducive to discussion. Female condoms are also added to the set of products carried by local health

[14]In addition to the fact that women in the study area are more vulnerable to HIV/AIDS, there is also evidence that women have lower desired current fertility than their male partners: 12% of women in stable relationships say that they want another child now, whereas 23% claim that their partner does. Men may also have higher desired total fertility, as 68% of women claim their partner wants another child whereas only 55% of the women say they want another child.

[15]The other main reasons are having given up condoms when they wanted to become pregnant (27%), and discomfort (12%).

[16]All bargaining power variables are constructed from principal component analysis, and ordered such that a higher value implies higher bargaining power. The correlation between "partner refuses male condoms =1" and a one standard deviation increase in the principal component of bargaining power in the following domains are: power dynamics -0.11 (p-value 0.09); assets 1 -0.13 (p-value 0.03); assets 2 -0.10 (p-value 0.002); assets 3 -0.18 (p-value <0.01). We also observe a positive correlation between assets brought by the woman to the relationship and baseline use of male condoms in the last thirty days. Specifically, the correlation is a 7.3 percentage point increase in the likelihood of use per one standard deviation in the assets 1 (p-value 0.035).

workers, which already includes male condoms, and that participants can access freely and discreetly at the end of each session.

The intervention thus allows us to study which women, if any, adopt female condoms when both informational and supply constraints are alleviated. We do not attempt to disentangle these two channels. Moreover, the estimated treatment effect may also include the effect of simply coming together in a group with other women to discuss personal issues. However, since any standard sex education programme would likely involve all of these components, the effect of this whole package is arguably of most interest to policymakers.

3 Theoretical framework

In this section we introduce a simple model of intra-household bargaining over STI protection technologies. We abstract from pure contraceptive technologies such as the pill, in light of the evidence discussed above that these are not close substitutes for STI protection methods in contexts such as our study setting. We use the model to formalise two main predictions about the effects of making female condoms freely available: first, that their introduction will lead to an increase in the probability that couples have sex; and second, that female condoms will be adopted by women with intermediate bargaining power. For ease of representation, we present the model here without the possibility of intra-household transfers — for example, if one partner offers to do more household chores in order to compensate the other partner for a given choice of contraceptive technology. However, Appendix A.1 shows that all of the predictions are robust to generalising the model to allow for transfers, as long as those transfers are not perfectly frictionless.

Preferences: Consider a population of heterosexual couples each consisting of a male m and a female f. When considering the choice of STI protection technology, individual i has preferences over the levels of pleasure and health that the technology yields on average to the population, $u_i(P, H)$, which is quasi-concave and increasing in each argument. For example, P may include the average level of discomfort associated with the material used to produce the technology, and H may include the average level of HIV transmission risk provided by the technology. Idiosyncratic and gender-specific heterogeneity in preferences over P and H are allowed for through the utility functions. For example an individual may place a larger weight on health if she is particularly risk-averse, or believes that she has a particularly high risk of HIV infection, due to her own health status, her beliefs about her partner's sexual behaviour, and so on. However, we assume that on average, couples' preferences satisfy the following single-crossing property:

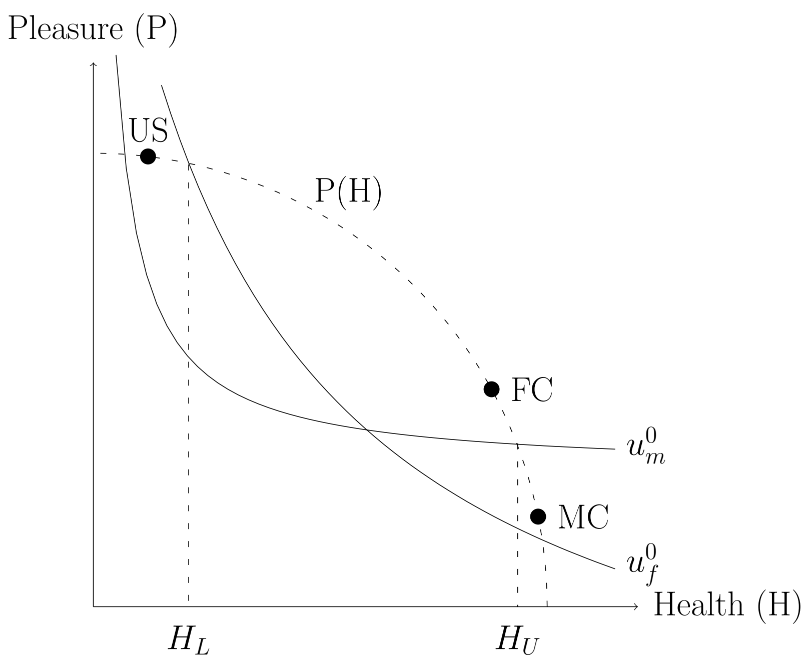
Assumption 1.

$$\frac{\partial u_m (P, H) / \partial P}{\partial u_m (P, H) / \partial H} > \frac{\partial u_f (P, H) / \partial P}{\partial u_f (P, H) / \partial H} \tag{1}$$

That is, we argue that the relative marginal utility from health compared to pleasure is greater for women than for men. This assumption is motivated by the fact that on average women face greater HIV risk and greater costs from pregnancy than men do, and that men have stronger reported displeasure from male condom use; as discussed in Sections 1 and 2.1.

Technologies: In general, let the STI protection technology frontier be represented by a continuously-differentiable function $P(H)$ for $H \in \left[\underline{H}, \overline{H}\right]$. By definition of being on the frontier, $P'(H) < 0$, and let $P''(H) \leq 0$ such that the frontier is weakly concave. This is illustrated in Figure 1.

Figure 1: STI protection technology frontier



In reality, only certain points on the frontier are easily accessible to couples, depending on the technologies that are readily available.[17] For simplicity, we assume that prior to our intervention, the set of readily available technologies is just the binary set of points on the frontier $\{US, MC\}$. Male condoms (MC)

[17] Of course, couples can mix their use of two or more technologies such as to obtain a wider range of points on the frontier. However, this is likely costly, due to increased transaction costs and costs of becoming familiar and comfortable with more than one technology. Thus we are effectively assuming that the range of points available on the frontier is larger the greater the number of technologies available, and indeed that couples will prefer to adopt a new technology that yields a certain point on the frontier rather than mixing two other technologies to obtain that point.

offer greater health than unprotected sex (US) because of their protection against HIV/AIDS and other STIs, but offer lower pleasure. Our treatment expands the set of readily available technologies to the ternary set of points on the frontier $\{US, FC, MC\}$.[18] As discussed in Section 1, female condoms (FC) provide lower effectiveness and thus lower health than male condoms, but are considered more pleasurable especially by men. They hence represent an intermediate option between male condoms and unprotected sex, as shown in Figure 1.

Co-operative decision-making: The decision to use an STI protection method must be taken jointly, since use of both male and female condoms is observable. It must also be taken each time a couple has sex, unlike decisions over longer-acting contraceptives. Therefore in stable couples, contraception can be thought of as a repeated game with a long time horizon. As such, it is natural to make the following modelling assumption:

Assumption 2. Decisions over STI protection technologies are taken co-operatively, resulting in choices that are Pareto efficient.[19]

This implies that if a couple decides to have sex, they choose H (and consequently P) to maximise the following household utility function

$$V = \alpha u_f \left(P(H), H\right) + (1-\alpha)\, u_m \left(P(H), H\right), \tag{2}$$

where $\alpha_i \in [0, 1]$ is the woman's Pareto weight in the couple's sharing rule (Browning and Chiappori, 1998). We interpret this weight as her bargaining power, and thus Equation 2 represents a collective model of the household (Chiappori, 1992). The weight α may depend on factors such as the woman's relative contribution to the couple's income and housework, and her options outside of the relationship.

Extensive margin: Let $s \in \{0, 1\}$ indicate the choice of whether to have sex or not. The no-sex option $s = 0$ can be enforced by either partner, and gives reservation utility u_i^0 to each partner. This can be thought of as the utility from their best immediate alternative, for example in terms of time use. Along with $s = 1$ is a choice of contraception from the available sets as described above.

The couple will choose $s = 1$ if and only if there exists some readily available technology on the frontier

[18] As a simplification, we also assume that the financial and opportunity costs of acquiring any of the technologies is zero. This is true in our experimental setting, since both male and female condoms are made available for free by the intervention. Nonetheless, the same qualitative predictions that we derive below hold for any intervention which expands the supply of female condoms, and thus reduces the direct and opportunity costs of acquiring them.

[19] The possibility of credibly committing to transfers is a mechanism sustaining this efficiency, see Appendix A.1.

(P, H) such that $u_i(P, H) \geq u_i^0 \;\; \forall i = m, f$. Define

$$I^0 = \left\{(P, H) \in R^2 | u_i(P, H) \geq u_i^0, i = m, f\right\} \tag{3}$$

as the set of all points (P, H) that satisfies both partner's participation constraints.[20] Prior to the introduction of female condoms, the couple will only choose $s = 1$ if the set $\{US, MC\} \cap I^0$ is non-empty. Meanwhile, following the introduction of female condoms, the couple will choose $s = 1$ if the set $\{US, FC, MC\} \cap I^0$ is non-empty. Since FC is an intermediate option between US and MC, and since I^0 is a quasi-convex set, the latter condition is more likely to be satisfied. Put differently, there is a weakly positive probability that there exist couples for whom US and MC lie outside of I^0, but for whom $FC \in I^0$. Indeed, Figure 1 illustrates a couple whose reservation utilities satisfy this property. This leads to our first prediction:

Prediction 1. Making female condoms freely available increases the probability that couples have sex.

Note that α does not enter a couple's decision as to whether to have sex or not: whether $\{US, MC\} \cap I^0$ or $\{US, FC, MC\} \cap I^0$ are non-empty depends only on individual reservation utilities, individual preferences over pleasure and health (which determine how the reservation utilities map into the pleasure-health space), and the set of readily available points on the technology frontier. Thus the effect of the intervention on the extensive margin decision is not predicted to vary by female bargaining power, conditional on these factors.

Intensive margin: In contrast, it is straightforward to show that as long as Assumption 1 holds, the optimal choice of health is increasing in α. Moreover, Appendix A.1 shows that this is still the case if partners can make transfers to one another, as long as these transfers are not perfectly frictionless.

To see why the optimal choice of health is increasing in α, assume that the intersection $\{US, FC, MC\} \cap I^0$ is non-empty, and thus that sex with some readily available technology provides greater utility to both members of the couple than no sex. Consider then the unconstrained household maximisation problem

$$\max_H \{\alpha u_f(P(H), H) + (1 - \alpha) u_m(P(H), H)\} \tag{4}$$

Since each $u_i(P(H), H)$ is quasi-concave, the objective function is also quasi-concave and has a unique solution. Denote this solution $\widetilde{H}(\alpha)$. It follows straightforwardly from the single crossing property in Assumption 1 that $\widetilde{H}'(\alpha) > 0$; see Online Appendix for proof. The intuition is simple: if the female places relatively greater weight on health than the male does, then the more bargaining power she

[20] Specifically, $I^0 = I_m^0 \cap I_f^0$, where $I_i^0 = \left\{(P, H) \in R^2 | u_i(P, H) \geq u_i^0\right\}$ is the upper contour set of the indifference curve corresponding to the reservation utility u_i^0.

holds, the more the household's choice of STI protection technology will be tilted towards health (and consequently away from pleasure).

However, it is possible that $\widetilde{H}(\alpha)$ does not lie on the intersection of I^0 and the technology frontier. By the single crossing assumption, the left-most endpoint H_L of this intersection is defined by $u_f(P(H_L), H_L) = u_f^0$, while the right-most endpoint H_U is defined by $u_f(P(H_U), H_U) = u_m^0$. This is illustrated in Figure 1. It could therefore be that $u_f\left(P\left(\widetilde{H}(\alpha)\right), \widetilde{H}(\alpha)\right) < u_f^0$ or that $u_m\left(P\left(\widetilde{H}(\alpha)\right), \widetilde{H}(\alpha)\right) < u_m^0$ (but not both). Consider the case in which her participation constraint binds, such that $u_f\left(P\left(\widetilde{H}(\alpha)\right), \widetilde{H}(\alpha)\right) < u_f^0$. The couple then instead chooses the closest incentive-compatible choice, which solves the incentive-constrained household utility maximisation problem

$$\max_H \left\{u_m(P(H), H) \,|\mu_f \left[u_f(P(H), H) - u_f^0\right]\right\}. \tag{5}$$

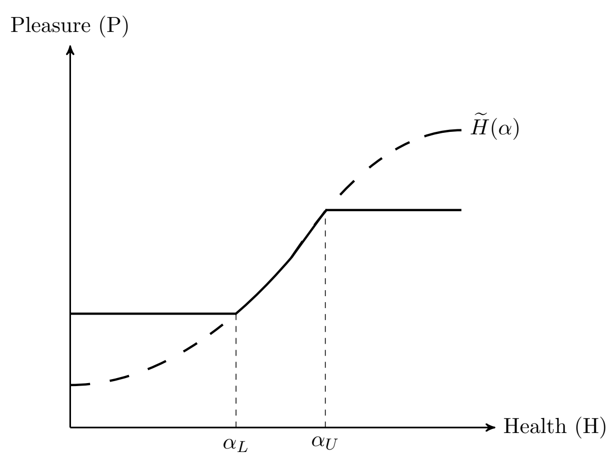
They hence choose H_L, which is independent of α. Vice versa, if his participation constraint binds they choose H_U. If neither partner's participation constraint binds, they choose $\widetilde{H}(\alpha)$ as before.

Given that $\widetilde{H}(\alpha)$ is increasing in α, this implies that there are threshold values for α defined by $\widetilde{H}(\alpha_j) = H_j$ for $j = L, U$ such that

$$H^*(\alpha) = \begin{cases} H_L & \text{if } \alpha < \alpha_L \\ \widetilde{H}(\alpha) & \text{if } \alpha \in [\alpha_L, \alpha_U] \\ H_U & \text{if } \alpha > \alpha_U. \end{cases} \tag{6}$$

It follows that $H^*(\alpha)$ is weakly increasing in α: $H^*(\alpha)$ is constant below α_L and above α_H, and is strictly increasing inbetween. This is illustrated in Figure 2.

Figure 2: Interior optimum health choices by female bargaining power



Now let us return to the case in which only certain points on the technology frontier are readily available. When only the binary set $\{US, MC\}$ is available, it follows directly from the weakly increasing nature of $H^*(\alpha)$ that there will be cut-off values of α such that[21]

$$H^*(\alpha) = \begin{cases} H_L & \text{if } \alpha < \alpha_L \\ H_{US} & \text{if } \alpha \in [\alpha_L, \alpha'] \\ H_{MC} & \text{if } \alpha \in [\alpha', \alpha_U] \\ H_U & \text{if } \alpha > \alpha_U. \end{cases} \tag{7}$$

The introduction of female condoms expands the available technologies to the ternary set $\{US, FC, MC\}$. Given that $H_{MC} > H_{FC} > H_{US}$, it follows directly that there will threshold values of α such that

$$H^*(\alpha) = \begin{cases} H_L & \text{if } \alpha < \alpha_L \\ H_{US} & \text{if } \alpha \in [\alpha_L, \alpha''] \\ H_{FC} & \text{if } \alpha \in [\alpha'', \alpha'''] \\ H_{MC} & \text{if } \alpha \in [\alpha''', \alpha_U] \\ H_U & \text{if } \alpha > \alpha_U. \end{cases} \tag{8}$$

This leads us to our second prediction:

Prediction 2. Women with intermediate bargaining power will adopt female condoms.

Note that this does not mean that we will necessarily observe an "inverse-U" relationship between bargaining power and female condom adoption in our experimental sample. This will depend on the extent to which women from the full distribution of bargaining power sign up for the condom programme. Women with the lowest bargaining power may not sign up if their partners disapprove of them participating, or indeed if they predict that they will not be able to convince their partners to use female condoms. If so, then we may not observe the upward-sloping portion of the inverse-U relationship between bargaining power and female condom adoption, but instead observe a simple negative relationship. On the other hand, women with the highest levels of bargaining power may not sign up, if they are already able to persuade their partners to use male condoms. If so, then we may observe a simple positive relationship between bargaining power and female condom adoption.

Margins of adoption: Intuitively, both couples who were previously having unprotected sex and couples who were previously using male condoms may adopt female condoms if this interior option allows them to get closer to the optimal choice on the technology frontier they would decide on if a continuum

[21] Of course, some of these sets may be empty.

of technologies were available.[22] From inspection of Equation 8, we have the following predictions:

Prediction 3a. Among the women who are engaging in unprotected sex at baseline, women with relatively higher bargaining power — although still relatively low bargaining power compared to the whole distribution — will take up female condoms.

Prediction 3b. Among women using male condoms at baseline, women with relatively lower bargaining power will switch from male to female condoms.

It is an important question whether effect *3a* or *3b* dominates empirically. If take-up of female condoms mainly comes from women who were engaging in unprotected sex at baseline, then female condoms represent a welfare-enhancing addition to the set of contraceptive technologies used to prevent HIV/AIDS and other STIs. On the other hand, if female condoms are mainly used as substitutes for male condoms, then offering female condoms will not lead to an increase in condom coverage. Moreover, whilst couples who switch to female condoms must be better off in terms of their private utility, the marginal loss of efficacy is likely to reduce welfare from the perspective of a social planner who cares about cost and health, given the negative externalities inherent in HIV/STI transmission.

4 Study design and data sources

4.1 Study design

Pathfinder International began its female condom programme in Matola in 2011, and in 2014 sought to expand this project to four additional neighbourhoods. We exploited this expansion via a phased-in experimental design. The healthcare workers who facilitate the programme first conducted door-to-door recruitment to identify women willing to participate. The eligibility criteria were that women needed to be between 18 and 49 years of age, sexually active, and not pregnant. To avoid spillovers, if a woman signed up with a friend or relative then they were both included in the programme but were automatically assigned to a separate set of groups which would not form part of the study but would still receive the intervention at a later stage. Otherwise, the average population of these neighbourhoods was 20,000 inhabitants, thus spillovers between unconnected individuals are unlikely to be a cause for concern. Overall 317 women were recruited into the study, and were subsequently administered a baseline

[22] If couples have uncertainty about the pleasure and health associated with female condoms, then many couples may try female condoms but then switch back to their original technology choice. In what follows we abstract from such uncertainty and consider the permanent adoption decision, once any uncertainty about female condoms has been resolved.

survey in August 2014.[23] Shortly after the baseline, one facilitator fell severely ill, and there was nobody sufficiently trained to replace her. Thus all 35 individuals recruited by her (treatment and control) were dropped from the programme and the study. This leaves a baseline sample of 298 respondents.

At the end of the baseline survey, it was explained to each participant that assignment to the first or second wave of the programme would be determined by computer randomisation. Once the entire sample had responded to the baseline survey, the research team randomly allocated half of the respondents recruited by each facilitator to the treatment group and half to the control group.[24] The reason for stratifying on facilitator was that each facilitator recruited in a separate area of the neighbourhoods, and we wanted to achieve balance within each area. This also ensured that there would be enough space for treatment and control participants to attend sessions close to their home.

The treatment group then received the intervention in September-December 2014. Following this, the endline survey was conducted in February-March 2015. A total of 232 respondents were traced and administered the endline survey. The retention rate was thus 78%, which although imperfect is reasonably high for a study in an urban slum area. Whilst attrition is higher in the control group (9.0 percentage points, p-value 0.049), Table 1 shows that the predictors of attrition are not differential across treatment and control. Following the endline survey, the control group then received the intervention from March-May 2015. The control group for one facilitator had already begun their sessions by the time they were administered the endline survey, hence these five observations are dropped from all estimations of treatment effects, leaving a final estimating sample of 227.

4.2 Survey data

Table 2 shows measures of key covariates and contraceptive use elicited for the full baseline sample.[25] All are balanced across treatment and control. Out of all respondents 85% report being in a stable relationship, of whom 63% are married.[26] The rest of the sample (15%) are sexually active but not in a stable relationship. The vast majority of respondents report having had just one sexual partner in the last twelve months, with 10% reporting zero partners and 3% reporting two partners. A third of respondents report being HIV-positive. This may be an underestimate given its self-reported nature,

[23] Respondents signed an informed consent form prior to the data collection, and it was explained that they could leave the study at any time.

[24] The randomisation was done in private, given the sensitive nature of participating in our intervention. Specifically, a member of the research team took the list of respondents for each facilitator, assigned each a random number in Excel, sorted respondents on this number, and assigned the first half to treatment and the second half to control.

[25] All baseline covariates are also balanced in the balanced panel, excluding attritters (see online Appendix).

[26] The latter includes traditional marriages and respondents who describe themselves as "living as married" but not legally married, which is common in this region due to the high bride price and costs of obtaining a marriage certificate.

Table 1: Predictors of attrition – treatment and control

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Treatment		**Control**		**Test $\beta_1 = \beta_2$**			**N**	
	Mfx	**p-val**	**Mfx**	**p-val**	χ^2	**p-val**	**T**	**C**	**All**
Demographics									
Age	-0.01	0.12	-0.01	0.18	0.01	0.91	152	146	298
Years of education	-0.01	0.41	-0.01	0.52	0.03	0.87	149	146	295
Literate	-0.09	0.27	-0.09	0.32	0.02	0.88	151	144	295
HH head	-0.05	0.49	-0.01	0.90	0.19	0.66	152	146	298
Income									
Has job	-0.03	0.67	0.04	0.57	0.47	0.50	151	144	295
Personal income last 30 days	-0.00	0.08	-0.00	0.26	1.13	0.29	152	146	298
Relationships									
In a stable relationship (incl. married)	-0.08	0.32	-0.01	0.95	0.45	0.50	152	146	298
Married (officially or unofficially)	-0.02	0.78	0.07	0.40	0.59	0.44	151	146	297
Years relation	-0.01	0.21	-0.01	0.11	0.04	0.84	121	114	235
Sexual behaviour									
HIV positive (self-report)	0.12	0.07	0.04	0.63	1.00	0.32	131	129	260
STI last 3 months (self-report)	0.06	0.47	-0.19	0.20	2.06	0.15	135	124	259
Names FC as contraceptive	-0.04	0.53	-0.06	0.43	0.00	0.97	150	146	296
Baseline use									
Ever used FC	0.05	0.60	0.12	0.32	0.07	0.80	152	146	298
Ever used MC	0.08	0.28	-0.02	0.78	1.04	0.31	152	146	298
Ever used other	-0.07	0.27	0.03	0.69	1.15	0.28	152	146	298
Used MC last 30 days	-0.04	0.53	-0.07	0.42	0.01	0.92	152	146	298
Current use MC	0.07	0.23	0.02	0.82	0.55	0.46	152	146	298
Current use other	-0.03	0.68	0.08	0.26	1.09	0.30	152	146	298

Notes: N=298 prior to attrition. Lower sample sizes reflect observations that are missing or not applicable. "Treatment" contains all individuals assigned to the treatment group, whether or not they attended the sessions. Columns 1-4 show marginal effects (Mfx) and p-values (p-val) for logit regressions of the probability of attritting on each covariate, in the treatment and control group respectively. Columns 5 and 6 show the χ^2 statistic and p-value for the test that the marginal effects are equal across the treatment and control groups. Columns 7-9 show sample sizes. "Ever used other" and "Current use other" refer to use of any other modern contraceptive method apart from condoms, e.g. the pill, injectables, or an IUD.

although the figure is close to the average figure of 29.4% for Maputo Province (Ministério da Saúde, 2015). Slightly more than 10% of respondents report having had an STI in the last month; but again this may be under-reported, especially since some further respondents report having had symptoms associated with STIs (not shown).[27] Fewer than half, 41%, are able to name the female condom when asked to list contraceptive methods.[28]

Our primary outcome variables are the use of contraceptive methods, disaggregated by female condoms, male condoms and other modern contraceptive methods (mainly the pill and injectables). For each method, we ask respondents whether they have ever used that method, and whether they are currently using it.[29] For male and female condoms, we also ask whether they have used that method in the last thirty days. These different measures allow us to distinguish whether respondents have tried a method at some point (ever use), have used recently (last 30 days) or consider it to be a part of their current portfolio (current use). Table 2 describes the baseline values of each of these measures. Baseline use of female condoms is low: 9% of the respondents have ever used a female condom, 3% have used one in the last 30 days, and 2% are currently using female condoms. Male condom use is substantially higher: around three quarters of women have ever used a male condom, 32% have used one in the last 30 days, and 39% percent say they are currently using male condoms. Altogether 39% are currently using non-condom methods at baseline, comprising 20% using the pill and 14% using injectables, and a small number using intrauterine devices (IUDs), the diaphragm, and sterilisation. These women may have signed up to the female condom programme as a way to switch out of these methods, or because they are seeking additional protection against HIV/AIDS and other STIs.

Finally, Table 3 compares our sample to a representative urban sample of women from Maputo Province, as described in the 2011 Demographic Health Survey data (DHS, 2011). Our sample is close to the overall adult female population of Maputo Province in terms of demographic characteristics such as age, years of education, marital status, pregnancy, and desired fertility. One exception is that the women in our sample are much less likely to have a job, which makes sense if women with a lower opportunity cost of time are more willing to participate in a time-intensive programme. On the other hand, the women in our study appear to have greater bargaining power than the representative sample: they began to have sex at a later age, are more likely to have used a condom the last time they had sex, and report greater decision-making power. This is in line with the suggestion in Section 3 that women with very

[27]We do not test for HIV, since we would have low power to measure an impact over the length of our study: official guidelines require a minimum of four weeks after exposure before the first test, and a second test three months later in the case of a negative result for high-risk cases (`https://www.gov.uk/government/publications/time-period-for-hiv-testing-position-statement`). We also opted not to test for STIs such as chlamydia, given the already sensitive nature of participation in the study and the budgetary implications of providing treatment to those who test positive.

[28]At endline this increases to 71% for the treatment group. However, it also increases to 65% for the control group, since they hear about female condoms by virtue of being enrolled in the programme. Hence we do not focus on this as an outcome variable.

[29]This is in line with the Demographic and Health Survey questions on contraceptive use.

Table 2: Baseline balance of covariates and use – full sample

	Mean	Control Mean	Treatment Mean	t-test	Total N	Control N	Treatment N
Demographics							
Age	30.32	30.12	30.52	-0.42	298	146	152
Years of education	6.21	6.26	6.17	0.27	295	146	149
Literate	0.84	0.84	0.85	-0.17	295	144	151
HH head	0.22	0.21	0.24	-0.51	298	146	152
Income							
Has job	0.38	0.42	0.33	1.64	295	144	151
Personal income last 30 days	745.85	854.52	641.46	1.41	298	146	152
Relationships							
In a stable relationship (incl. married)	0.85	0.85	0.84	0.17	298	146	152
Married (officially or unofficially)	0.63	0.64	0.62	0.37	297	146	151
Years relation	8.66	8.62	8.70	-0.08	235	114	121
# Partners last 12 months	0.92	0.92	0.93	-0.23	298	146	152
Sexual behaviour							
Pregnant	0.05	0.05	0.06	-0.42	297	145	152
HIV positive (self-report)	0.33	0.35	0.31	0.75	260	129	131
STI last 3 months (self-report)	0.13	0.13	0.13	-0.10	259	124	135
Names FC as contraceptive	0.41	0.44	0.39	0.90	296	146	150
Baseline use							
Ever used FC	0.09	0.09	0.09	0.11	298	146	152
Ever used MC	0.74	0.76	0.73	0.59	298	146	152
Ever used other	0.72	0.72	0.72	0.04	298	146	152
Used FC last 30 days	0.03	0.01	0.04	-1.39	298	146	152
Used MC last 30 days	0.32	0.28	0.35	-1.26	298	146	152
Current use FC	0.02	0.02	0.03	-0.33	298	146	152
Current use MC	0.39	0.37	0.41	-0.79	298	146	152
Current use other	0.39	0.41	0.37	0.75	298	146	152

Notes: N=298 in the baseline sample. Lower sample sizes reflect observations that are missing or not applicable. “Treatment Mean” contains all individuals assigned to the treatment group, whether or not they attended the sessions. Column 5 presents the test statistic for the null hypothesis that the mean in the treatment group is equal to the mean in the control group. “Ever used other” and “Current use other” refer to use of any other modern contraceptive method apart from condoms, e.g. the pill, injectables, or an IUD.

low bargaining power might not select into our study. That said, the women in our sample report higher rates of emotional, physical and sexual violence than the representative average.

Table 3: Comparison of Study Sample to DHS Representative Sample

	Study Mean	**DHS Mean**	**t-test**	**Study N**	**DHS N**
Age	30.18	29.47	1.28	276	1007
Years of education	6.35	6.72	-1.82	273	1007
Literate	0.85	0.76	3.82	273	1007
Income					
Has job	0.37	0.58	-6.33	273	1007
Relationships					
Married (officially or unofficially)	0.63	0.61	0.58	275	871
Pregnant	0.05	0.07	-0.76	275	1007
Wants another child in future	0.57	0.57	0.14	260	961
Sexual Behaviour					
Age of sexual debut	16.62	16.16	3.26	273	955
Used condom during last time sex	0.54	0.31	6.32	243	871
Decision-making visiting family	0.62	0.39	6.43	272	580
Decision-making spending earnings	0.59	0.21	10.85	275	569
Decision-making her health	0.53	0.39	4.04	275	580
Experienced emotional violence	0.47	0.37	2.50	212	372
Experienced physical or sexual violence	0.22	0.13	3.34	276	1007

Notes: Column 2 displays the mean from our study, N=298. Column 3 shows the 2011 DHS mean for women in urban areas of Maputo Province, N=1007. Lower sample sizes reflect observations that are missing or not applicable. Variables selected for comparison are those that appear in both our study and the DHS, with similar or identical wording.

4.3 Diary data

A subsample of the survey respondents also took part in weekly coital diaries. The diaries recorded detailed information on all of the respondents' sexual encounters in the seven days prior to each interview, including: type of sexual activity; relationship to the partner; whether any contraceptives were used, and if so, which ones; whether the two partners discussed the use of contraceptives prior to sex; and if so, who initiated this discussion. Diary interviews took place in a private place chosen by the respondent.[30] The same enumerator interviewed a given respondent each week, to maximise trust and confidentiality.

[30] Enumerators were carefully trained, including on first responder and referral procedures if respondents reported experiences of sexual or physical violence.

The diary sampling took place as follows. At baseline, respondents were asked about their willingness to participate in the diaries. This gave a diary sample of 57 respondents: 28 in the treatment group and 29 in the control group. Tables A.7 and A.8 in Appendix A.3 show that there is balance in baseline covariates and use across the treatment and control groups in the diary sample. Diary interviews took place over a period of 17 weeks, beginning four weeks prior to the first group receiving its first session and ending one week after the last group received its last session.[31] Each week on average 75% of the diary sample participated. Individual respondents took part in the diaries an average of 15 times, with a minimum of three weeks and a maximum of 17 weeks. There are no significant differences in participation between the treatment and control group. Tables A.9 and A.10 in Appendix A.3 also show that the final sample of diary participants are representative of the balanced panel of all survey participants in terms of baseline covariates and bargaining power, with a few exceptions: the diary participants have been in a relationship for longer than the average study participant, and no diary respondents are pregnant. All of the results from the diary subsample presented below are robust to re-weighting to make the diary subsample representative of the full sample.

The diaries are a more complex and granular instrument than the baseline and endline surveys. The fact that the diaries are administered very shortly after the sex acts that they ask about also almost eliminates potential recall bias (Das et al., 2012). Although the time period covered by the surveys was slightly different to that covered by the diary interviews, we are able to cross-check with the surveys to reduce concerns about misreporting in either instrument.[32] As far as there are inconsistencies, there is some evidence of under-reporting in the surveys: 5 out of 57 diary participants report never having used a female condom during the endline survey but do report using them in the diaries; whilst for male condoms the figure is 4 out of 57 respondents.[33] This strengthens our confidence that there is no systematic over-reporting of condom use in the survey data compared to the diary data. Nevertheless, we run the main impact analyses using both the survey data and the diary data, to show that the results hold across these two different survey instruments; see Section 5.2 for details.

[31] For the diary data, the baseline period for each respondent is taken to run from the start of the diary data collection until one week after the facilitator that the respondent was assigned to began her first meeting for her treatment-group participants. The length of the baseline and endline period therefore varies depending on when a respondent's group held its first meeting. The average in the baseline period is 5.6 weeks. For robustness, we run all estimations setting the end of the baseline period to one week before this, i.e. the exact week of the first treatment-group meeting, and to two weeks after the first treatment-group meeting. Results are similar and available on request. Meanwhile the endline period is taken to run from the end of the baseline period (i.e. one week after the first treatment group meeting) until the end of the diary data collection, comprising 8.9 weeks on average.

[32] Medical literature has shown that subjects may over-report incidence of sex and use of condoms in surveys compared to diaries (Stalgaitis and Glick, 2014), although this varies across survey populations (Hoppe et al., 2008).

[33] We cannot make the opposite comparison given that the endline survey took place two months after the end of the diaries, so if a respondent reports using condoms in the survey but not the diaries it may be that they adopted them during those two months.

As well as constructing variables for contraceptive use at the respondent level, the diary data allow us to analyse the impact of the intervention at the level of the sex act. Altogether respondents report a total of 349 sex acts during the endline period. This represents an average of 6.1 sex acts per respondent, with a minimum of zero and a maximum of 30. For each sex act and each contraceptive method we set "sex act full endline period" equal to one if that method was used and the sex act took place at any point during the endline period. Similarly, we set "sex act use last 30 days" equal to one if that method was used and the sex act occurred in the last four weeks prior to a respondent's last diary observation. Finally, we set "sex act last 14 days" equal to one if the method was used and the sex act occurred 14 days prior to a respondent's last diary observation.[34]

4.4 Bargaining power

The women who are in a stable relationship at baseline are on average 5.7 years younger than their partner and earned 3,360 MZN (approximately 110 USD) less than him in the last 30 days. This may point towards these women having lower bargaining power than their male partners, but may also proxy a host of other features of the marriage and labour markets. To obtain a more direct proxy of the woman's outside options at the time of the start of the relationship, we include a survey module about assets brought by the woman to the relationship. To proxy other features of her bargaining power, we also include two further sets of survey questions, covering decision-making and power dynamics in the relationship. Table 4 provides summary statistics for each of the individual questions.

Since the assets module contains multiple questions that are correlated, we perform a principal component analysis to construct indices. Principal component analysis (PCA) is an approach that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. [35] The first three principal components of the assets module are presented in Table A.11 in Appendix A.3 and jointly explain 68.5% of the variance in the assets questions. Similarly, we run a principal component analysis including the questions from the decision-making and power dynamics modules. Table A.12 in Appendix A.3 shows that the different survey modules each load a different principal component. Together these two components explain about 40% of the variance in the survey questions. Thus altogether we take five principal components as our preferred measures of bargaining power: three from the assets module, and one each from the decision-making and power dynamics modules. For ease of comparability, we scale the components so that the woman with least bargaining power on that measure has a score of zero, and normalise

[34]When the sex act is the unit of observation we construct the sex-act level variables during the endline period only.

[35]This approach is preferable to testing for heterogeneity by each question separately and then correcting p-values for multiple hypothesis testing, since the principal component approach takes into account which measures are proxying the same concept as opposed to different concepts.

Table 4: Bargaining power – summary statistics

	Mean	sd	Min	Max	Total
Did you bring the following assets to your relationship...					
...jewellery?	0.08	0.27	0.00	1.00	264
...animals?	0.00	0.06	0.00	1.00	264
...land?	0.01	0.11	0.00	1.00	264
...electronics?	0.02	0.14	0.00	1.00	264
...money?	0.07	0.25	0.00	1.00	264
...mobile phone?	0.07	0.25	0.00	1.00	264
...kitchen utensils?	0.10	0.30	0.00	1.00	263
Who decides about...					
...buying clothes for you?	0.80	0.40	0.00	1.00	297
...buying phone credit?	0.76	0.43	0.00	1.00	297
...education for the children?	0.49	0.50	0.00	1.00	288
...health expenses for you?	0.55	0.50	0.00	1.00	297
...health expenses for the children?	0.41	0.49	0.00	1.00	291
...if you are allowed to work?	0.59	0.49	0.00	1.00	296
...how earnings are used?	0.60	0.49	0.00	1.00	297
...visits to friends?	0.64	0.48	0.00	1.00	296
...visits to family?	0.64	0.48	0.00	1.00	294
Who usually has more say when you talk about serious things	0.47	0.50	0.00	1.00	250
In general, who do you think has more power in your relationship	0.39	0.49	0.00	1.00	249
Power dynamics					
Most of the time, we do what my partner wants to do	2.33	1.08	1.00	4.00	250
My partner won't let me wear certain things	2.61	1.11	1.00	4.00	250
When my partner and I are together, I'm pretty quiet	3.07	0.96	1.00	4.00	250
My partner has more say about important decisions that affect us	2.39	1.09	1.00	4.00	250
My partner tells me who I can spend time with	2.79	1.09	1.00	4.00	249
I feel trapped or stuck in our relationship	3.20	0.86	1.00	4.00	250
My partner does what he wants, even if I do not want him to	2.86	1.00	1.00	4.00	249
I am more committed to our relationship than my partner is	2.74	1.08	1.00	4.00	250
My partner is involved with other people apart from me	2.77	1.02	1.00	4.00	249
My partner always wants to know where I am	2.16	1.10	1.00	4.00	250
When my partner and I disagree, he gets his way most of the time	2.73	1.06	1.00	4.00	248

Notes: All values taken from the baseline survey (N=298). All variables are coded such that a higher value proxies greater bargaining power for the respondent. The assets module was enumerated to all women who lived with their partner at baseline, including a few who did not claim to be in a stable relationship (N=264). The decision-making module was enumerated to all respondents (N=298), except the questions "who has more influence" and "who has more power" which were asked only of women in a stable relationship at baseline (N=250). "Power dynamics" questions were also only asked to women who were in a stable relationship at baseline (N=250). Lower observation numbers reflect values missing or unwillingness to answer.

them such that a one point increase in each measure represents an increase of one standard deviation. Table A.13 in Appendix A.3 shows that these principal components are balanced across treatment and control.

These measures of bargaining power are correlated with certain baseline demographic characteristics in the way that one would expect. Specifically, both the second and the third principal components of assets brought by the respondent to the relationship are strongly correlated with her education. Meanwhile, the respondent's decision-making power is strongly positively correlated with her personal income in the last thirty days, age, whether she is the household head and whether she has a job; whilst her decision-making power is strongly negatively correlated with whether the couple is married. One anomaly is that her decision-making power is negatively correlated with her education. To avoid the measures of bargaining power spuriously proxying the effects of any of these demographic characteristics, we therefore add these characteristics as controls when estimating the effects of bargaining power on condom adoption; see Section 5.3 for details.

5 Results

5.1 Main impact results

Our preferred estimations are analysis of covariance (ANCOVA) linear probability models of the following form:[36]

$$Pr\left[Y_{if1} = 1 | Y_{if0}, treat_{if}, \eta_f\right] = \alpha + \delta Y_{if0} + \beta treat_{if} + \eta_f, \tag{9}$$

where Y_{if1} is the outcome variable of interest at endline, and Y_{if0} is its value at baseline. $treat_{if}$ is a dummy for being assigned to the treatment group, i.e. to receiving the programme in the first rather than the second phase. β represents the intent-to-treat effect, since not all individuals assigned to treatment attended the programme: the participation rate was around 65% for each individual session, with 20 women (17.7% of the control group) not attending any of the six sessions. η_f is a facilitator fixed effect, which is included for inference since randomisation was blocked on the seventeen facilitators (Bruhn and McKenzie, 2009). Standard errors are robust to individual-level heteroskedasticity, as this was the level of randomisation (Abadie et al., 2017).

Table 5 displays the main impacts of the intervention. The programme has a substantial and highly

[36] All of the results from the full sample are robust to using a logit specification; tables available on request. We employ a linear probability model to allow comparison with the treatment effects estimated for specific subsamples, such as the diary respondents. In such subsamples, appealing to the large-sample properties of maximum likelihood estimators becomes questionable, especially conditional on seventeen facilitator fixed effects.

significant effect on the use of female condoms. Specifically, we observe an 18.4 percentage point increase in the proportion of women who have ever used female condoms (equivalent to 209% of the endline mean in the control group); a 4.7 percentage point (470%) increase in the proportion who have used a female condom in the last thirty days; and a 7.7 percentage point (385%) increase in the proportion who are currently using female condoms. We see no evidence of changes in behaviour in the control group, as the control group endline means of female condom use are identical to those at baseline (see Table 2).[37] This finding reinforces the idea that our treatment expanded access to female condoms which are otherwise difficult to obtain in the study area. Indeed, reported use of female condoms in the treatment group at endline is highly correlated with the number of free female condoms that a respondent took from the sessions, which was discreetly recorded by facilitators.[38] This finding also weighs against concerns that reported use of female condoms among the treatment group might represent response bias. The fact that the treatment effect on ever use is higher than the treatment effect on use in the last thirty days and current use suggests that many women in the treatment group try female condoms at the start of the intervention, then a smaller although sizeable fraction continue to use them. This is a natural adoption pattern if couples experiment with female condoms and thereby learn more about their costs and benefits, then some return to their original contraceptive method while others adopt female condoms more permanently.

We do not observe any significant evidence that respondents substitute away from or increase their use of male condoms. However, our power to detect impacts on male condom use is lower given the high baseline rates of male condom use.[39] Table 6 shows that when we split the sample by women who are using or not using male condoms at baseline, in line with the model we see take-up from both groups. The strong take-up of female condoms among women not using male condoms at baseline (columns 1, 3 and 5) is particularly important from a policy perspective. It implies that the treatment decreases the number of women having sex unprotected from HIV/AIDS and other STIs, rather than generating pure substitution away from other STI contraceptive methods. In the full sample, the estimated treatment effect on current use of either male or female condoms is an increase of 7.9 percentage points; but we have lower power to detect changes in this aggregate outcome, and the point estimate is marginally insignificant with a p-value of 0.176 (not shown). We do not observe any significant impacts on male condom use when we split the sample by those using and not using male condoms at baseline (see online Appendix).

[37] There is thus no evidence that the control group experienced anticipation effects, disappointment effects, or spillovers.

[38] The correlation between the total number of condoms taken and endline measures of female condom use are: ever use 0.38 (p-value < 0.01); use last 30 days 0.21 (p-value 0.02); current use 0.29 (p-value <0.01).

[39] We have 80% power to detect the following minimum detectable effect sizes at the 5% level in a two-tailed test: ever use – female condoms 9.5 percentage points, male condoms 15.6 percentage points; use last 30 days – female condoms 5.2 percentage points, male condoms 17.0 percentage points; current use – female condoms 6.3 percentage points, male condoms 18.1 percentage points.

Table 5: Treatment effects – primary outcome variables

	(1) **Ever FC**	(2) **Ever MC**	(3) **Ever other**	(4) **Last 30 days FC**	(5) **Last 30 days MC**	(6) **Current FC**	(7) **Current MC**	(8) **Current other**
Treatment	0.184***	-0.012	0.020	0.047**	-0.052	0.077**	0.060	0.030
	(0.042)	(0.041)	(0.042)	(0.023)	(0.057)	(0.030)	(0.058)	(0.053)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓	✓	✓
Observations	227	227	227	227	227	227	227	227
Control mean endline	0.088	0.824	0.735	0.010	0.363	0.020	0.353	0.412

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms, male condoms and other modern contraceptive methods. Columns 1-3 refer to whether the respondent has ever used the method, columns 4 and 5 to whether she has used it in the last 30 days (this was only asked for condoms, not for other contraceptive methods), and columns 6 to 8 to whether she is currently using it. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table 5 shows that we also see no increase in or substitution away from other contraceptive methods such as the pill and injectables. This suggests that women who adopt female condoms were either previously using no contraceptives, or use female condoms in addition to other methods in order to protect against HIV/AIDS and other STIs. Indeed, of the women who are currently using female condoms at endline, 42% are also using other contraceptive methods (mainly the pill or injectables). Nevertheless, Table A.14 in Appendix A.4 shows that when we split the sample by women who are using or not using any method of contraception at baseline, the observed impacts on the use of female condoms are particularly strong among women who are not using any method at baseline (although they are not significantly higher than the impacts in the rest of the sample). We again do not observe any significant impacts on male condom use, either among those not using or using any contraceptive method at baseline; tables available on request.

We would expect women who are not in a stable relationship to place a larger weight on the health offered by STI protection technologies, and so to have a higher demand for condoms, since their risk of contracting HIV or other STIs may be higher and they may be more concerned about the risk of pregnancy. Table A.15 in Appendix A.4 shows that in line with this prediction, the treatment effect on ever use of female condoms is stronger for women who are not in a stable relationship at baseline.[40] Nonetheless, when we restrict the sample to just those women in a stable relationship the estimated treatment effect is a 16.4 percentage point increase in ever use of female condoms (p-value < 0.01), a 5.6 percentage point increase in use in the last 30 days (p-value 0.042), and a 7.9 percentage point increase in current use (p-value 0.019). Thus the treatment does also lead to take-up of female condoms among women in stable relationships. This may be rational if the partners are serodiscordant (i.e., one partner is HIV-positive while the other is HIV-negative) or if one or both partners have relations with others, or at least suspect that their partners have relations with others.[41]

5.2 Diary results

5.2.1 Impacts at the sex-act level

We can use the diary data to measure treatment effects at the sex act level. We examine impacts on the proportion of total sex acts protected by male and female condoms, since these are key variables of interest for policymakers and for cost-benefit analyses. Table A.16 in Appendix A.4 shows the results.

[40] This is perfectly correlated with being in a stable relationship at endline, as we do not observe any break-ups of stable relationships over the study period.

[41] In serodiscordant couples, use of condoms is still rational even if the partners do not expect to follow through on using them all of the time, since the chance of becoming infected from any one unprotected act is approximately 0.001 for female to male transmission and 0.002 for male to female transmission (Dowdy et al., 2006).

Table 6: Treatment effects on female condom use, by baseline male condom use

	(1) **Ever FC** **No MC** **at baseline**	(2) **Ever FC** **Current MC** **at baseline**	(3) **Last 30 days FC** **No MC** **at baseline**	(4) **Last 30 days FC** **Current MC** **at baseline**	(5) **Current FC** **No MC** **at baseline**	(6) **Current FC** **Current MC** **at baseline**
Treatment	0.169***	0.232***	0.073**	0.030	0.085***	0.049
	(0.047)	(0.074)	(0.030)	(0.034)	(0.031)	(0.057)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	141	86	141	86	141	86
Control mean endline	0.092	0.081	0.000	0.027	0.000	0.054
$\chi^2(1) : (a) = (b)$		0.52		0.91		0.30
Pr $> \chi^2$		0.471		0.340		0.583

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms. Columns 1-2 refer to whether the respondent has ever used that method, columns 3-4 to whether she has used it in the last 30 days, and columns 5-6 to whether she is currently using it. Odd-numbered columns present results for the subsample of individuals not currently using male condoms at baseline; even-numbered columns present results for the subsample of individuals who are currently using male condoms method at baseline. The bottom two rows present chi-squared statistics and their p-values for the test that the treatment effect is the same across the two subsamples. These are obtained from seemingly unrelated estimations. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

There is a significant increase in the proportion of sex acts protected by a female condom in the full endline period, the last 30 days, and the last 14 days (5.1, 8.5 and 6.7 percentage points respectively, all significant at the 5% level). Consistent with the respondent-level results from the survey data, there is no significant effect on the proportion of sex acts protected by male condoms, and the point estimates are small.

5.2.2 Individual-level diary results

To further understand the mechanisms for the observed treatment effects, we use the diaries to examine the likelihood of sex acts per respondent per week, the proportion of protected sex acts per respondent, and the proportion of sex acts per respondent in which the partners discussed condom use. Taking advantage of the weekly nature of the diaries, we estimate the following fixed effects panel specification:

$$Pr\left[Y_{ift}=1|treat_{if},\gamma_i,\eta_f\right]=\beta treat_{if}\times endline+\eta_f\times endline,\qquad t=1,2,......T \tag{10}$$

where Y_{ift} is the outcome variable of interest for individual i assigned to facilitator f in week t. The unit of observation is thus the respondent-week. *endline* is an indicator equal to one if the week falls in the endline period, i.e. one week or more after programme sessions led by i's facilitator have begun for the treatment group. The rest of the terms are defined as in Equation 21, and standard errors are again clustered at the individual level, since this was the level of randomisation (Abadie et al., 2017).

Table 7 shows that, in line with Prediction 3 of the model, the treatment leads to a significant increase in the likelihood of sex acts per week per respondent. In the full endline period, respondents in the treatment group were on average 9.1 percentage points more likely to report a sex act per week, compared to a control group mean of 0.469. In the last 30 and 14 days, the treatment effect on the likelihood of sex acts per week was 0.113 and 0.158 respectively, (compared to 0.471 and 0.491 in the control group). The fact that we observe an increase in the likelihood of sex acts per week in the treatment group indicates that there are couples in which one or both partners' participation constraints are binding when the only options are male condoms or unprotected sex, but where both find sex with female condoms preferable to not having sex. The introduction of female condoms therefore increases utility for such couples.

Table A.17 in Appendix A.4 shows the results for the proportion of sex acts protected per individual per week. In the full endline period, the treatment led to an increase of 9.7 percentage points in the proportion of sex acts per week protected by a female condom (over a control group mean of 1.0%). In the last 30 days the figure is 8.4 percentage points (compared to 1.0% in the control group), while the estimate for the last 14 days is not significant. Consistent with the respondent-level results from the survey data and with the overall sex-act-level results, there is no significant effect on the proportion of

Table 7: Impacts on likelihood of sex acts per respondent week – diary subsample

	(1) **Sex act per week full endline period**	(2) **Sex act per week last 30 days**	(3) **Sex act per week last 14 days**
Treat × endline	0.091**	0.113**	0.158*
	(0.045)	(0.057)	(0.086)
Treat × Facilitator f.e.'s	✓	✓	✓
Observations	863	536	367
Control mean	0.469	0.471	0.491

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are binary indicators for whether a respondent had a sex act in a particular week. "Treat × endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat × endline" is the intent-to-treat effect. All regressions are respondent-level LPM fixed effects models with the respondent-week as the unit of observation. All specifications include individual fixed effects (N=56), and facilitator × endline fixed effects (N=17) for inference since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation.

sex acts protected by male condoms.

Table A.18 in Appendix A.4 shows that we observe a large and highly significant reduction in the proportion of sex acts in which a discussion about condoms takes place: 19.4% of sex acts during the last 30 days and 35.6% of sex acts in the last 14 days respectively (compared to a control mean of 27.5% and 41.1% respectively). These results suggests that the availability of female condoms reduces the number of disagreements about the use of STI protection methods between partners. This is in line with our model, if the expansion from a binary to a ternary choice allows the couple to choose an STI protection technology that is closer to their preferred choice on the technological frontier.

Table 8 shows that the treatment effect is driven by those respondents who are not using MC at baseline: For these respondents the treatment effect for the full endline period is 15 percentage points higher, relative to the average treatment effect of 9.1 percentage points. For last 30 days this effect is 20 percentage points. The effect for the last 14 days is not significant.

5.3 Heterogeneity by bargaining power

We now examine which women adopt female condoms in terms of their bargaining power. Recall that Prediction 2 of the model is an inverse-U relationship between bargaining power and adoption of female condoms in the population. However, the descriptive evidence in Table 3 demonstrates that women with the lowest bargaining power in the population do not select into the condom programme. We would therefore expect the relationship between bargaining power and female condom adoption in our sample to be negative, as we should predominantly capture the downward-sloping portion of the inverse-U relationship predicted for the whole population. Table 9 displays the relationship between endline current use of female condoms and each of the baseline principal component measures of bargaining power, both alone and interacted with treatment.[42] We control for baseline female condom use and for the demographic factors which are significantly correlated with baseline bargaining power: age, education, income in the last thirty days, whether the respondent is in a stable relationship, and whether she is the household head.[43] This is to prevent the bargaining power measures from spuriously capturing the effect of variables which are correlated with them and may be correlated with condom use. For example, younger women hold lower bargaining power on some measures, but may also have had greater exposure

[42]Similar results to all of those results presented in this section are obtained for use in the last thirty days and are available on request. We restrict attention to current use and use in the last 30 days, since ever use does not proxy sustained adoption.

[43]Facilitator fixed effects are not included, since we lose power due to the low number of observations per cell if the sample is first split along seventeen facilitators and treatment status, then interacted with bargaining power.

Table 8: Likelihood of sex acts: Interaction of treatment and baseline MC use

	(1) **Sex act per week full endline period**	(2) **Sex act per week last 30 days**	(3) **Sex act per week last 14 days**
Treat × endline	0.150***	0.200***	0.197**
	(0.051)	(0.063)	(0.098)
Treat × endline × MCbaseline	-0.130*	-0.174**	-0.078
	(0.078)	(0.071)	(0.102)
Facilitator × endline f.e.'s	✓	✓	✓
Observations	806	536	367
Control mean	0.469	0.471	0.491

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are binary indicators for whether a respondent had a sex act in a particular week. MCbaseline is a binary indicator for whether a respondent currently uses MC at baseline. "Treat × endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat × endline" is the intent-to-treat effect. All regressions are respondent-level LPM fixed effects models with the respondent-week as the unit of observation. All specifications include individual fixed effects (N=56), and facilitator × endline fixed effects (N=17) for inference since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation.

to sex education at school and so may use condoms for that reason.[44] The results are striking: for almost every measure of bargaining power, we observe a *negative* interaction between baseline bargaining power and treatment, showing that women with the lowest bargaining power in our sample are the most likely to adopt female condoms as a result of the intervention.[45]

The descriptive evidence in Table 3 demonstrates that women with the lowest bargaining power in the population do not select into the condom programme. For Prediction 3a. this suggests that heterogeneity by bargaining power is unlikely to be observed among those women not using condoms at baseline, even though Section 5.1 showed evidence of strong adoption of female condoms among women who were not currently using male condoms at baseline. Indeed, Table A.19 shows that (conditional on the same set of controls) there is a lack of evidence of a heterogenous treatment effect between adoption and baseline bargaining power among these women.

Table A.20 in Appendix A.4 reports the correlation (conditional on the same set of controls) between the different measures of bargaining power and current use of female condoms at endline, among those women who were using male condoms at baseline.[46] We see consistent evidence of the negative correlation predicted by the model. On the other hand, as discussed in Section 5.1, we do not see evidence of a large degree of substitution away from male condoms. A possible explanation is that women with higher bargaining power who take up female condoms also intersperse their use with the use of male condoms. Indeed, 81% of women who are currently using female condoms at endline also report currently using male condoms. This "double protection" is a typical pattern of adoption observed in the medical literature, and is found to be associated with a large increase in the number of protected sex acts (Vijayakumar et al., 2006).

[44]Results are virtually unchanged if these controls are excluded. These results are available on request.

[45]We find similar results for the relationship between endline current use of female condoms and each of the baseline principal component measures of bargaining power and their square (see online Appendix). The dominant relationship is linear and negative; although the squared term carries a positive but small coefficient for some of the asset measures, which have a very long right tail. Interactions with the treatment are not included, since we lose considerable power when interacting both the linear and the squared terms. Similar results also hold if instead of using the principal components we use a score for each module, summing an individual's responses. Results are available on request. Negative effects are also estimated for many of the individual questions about bargaining power taken separately, although only some are significant. However, neither of these approaches take into account the fact that different questions may be almost a duplicate measure of the same construct.

[46]We omit the interaction terms between the measures of bargaining power and treatment, because power is severely weakened by the loss in sample size when we restrict attention to only those women who were or were not using condoms at baseline.

Table 9: Impacts on current use of female condoms – heterogeneity by bargaining power

	(1)	(2)	(3)	(4)	(5)
	β / (s.e.)	β / (s.e.)	β / (s.e.)	β / (s.e.)	β / (s.e.)
Treatment	0.128***	0.111***	0.159***	0.223**	0.193*
	(0.045)	(0.039)	(0.051)	(0.093)	(0.114)
Assets 1	-0.002				
	(0.010)				
Treat*Assets1	-0.044**				
	(0.017)				
Assets 2		0.004			
		(0.010)			
Treat*Assets2		-0.031**			
		(0.015)			
Assets 3			-0.001		
			(0.005)		
Treat*Assets3			-0.054***		
			(0.018)		
Decision-making				-0.015	
				(0.022)	
Treat*Decision				-0.079**	
				(0.038)	
Power dynamics					0.014
					(0.019)
Treat*Power dynamics					-0.037
					(0.039)
Controls	✓	✓	✓	✓	✓
Observations	198	198	198	180	180
Control mean endline	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Excluding attritters, N=206 women are in a stable relationship at baseline. N=198 are women who are in a stable relationship and have no missing values on the control variables. N=180 have non-missing values for all of the decision-making, power dynamics, and control variables. “Treatment” is a dummy for assignment to treatment thus the coefficient on “Treatment” is the intent-to-treat effect. “Assets 1”, “Assets 2” and “Assets 3” are the first three principal components from the assets module, as identified in Table A.11. “Decision-making” and “Power dynamics” are the first two principal components from all the survey questions referring to these two modules, as identified in Table A.12. Dependent variables are binary indicators for current use of female condoms. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Regressions do not include facilitator fixed effects due to loss of sample size where baseline use perfectly predicts endline use conditional on a given facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. Controls are: respondent’s age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

5.4 Heterogeneity by HIV status and risk perceptions

Women who are HIV-positive, and those who believe they are at high risk of being infected with HIV or other STIs, are likely to have a higher demand for health.[47] It follows straightforwardly from the model that these groups of women should therefore be more likely to use male condoms at baseline, and more likely to switch from unprotected sex to female condoms when female condoms are introduced.[48] To test for this, we collect self-reported data on HIV status as part of the baseline and endline surveys. As a proxy of beliefs about risk, respondents are asked to respond on a five-point scale to the question "what is the risk to you personally of contracting HIV/AIDS in a case of unprotected sex?".[49]

The data accord well with the model's predictions. In terms of HIV status, columns (1) and (2) of Table 10 show that women who report being HIV-positive at baseline indeed have a significantly larger treatment effect for "ever use of female condoms". Columns (3)-(6) of Table 10 show that HIV-positive women also have larger point estimates for the treatment effect on last thirty days and current female condom use, although neither are significantly different from the treatment effect for HIV-negative women. We do not observe significant treatment effects on male condom use either among HIV-positive or HIV-negative women (see online appendix).

In terms of risk beliefs, we first verify that a one point increase in the "risk perception for self" scale at baseline is correlated with an increase of 8.3 percentage points in the probability of ever having used male condoms at baseline (p-value 0.035).[50] To test whether the treatment effects differ by risk beliefs, we code a dummy variable "believes high risk to self" equal to one if the respondent's answer was above the median at baseline. Table A.22 in Appendix A.4 shows the results of re-estimating Equation 21 including this dummy and its interaction with treatment status. Columns (5) and (6) show that women with below-median perceptions of HIV risk at baseline increase their current use of both male and female condoms as a result of the treatment. Meanwhile women with above-median perceptions of HIV risk at baseline experience no treatment effect, as the negative interaction term completely offsets the main treatment effect for them.

[47] In principle, individuals who are already HIV-positive could either increase or decrease their preference for protection, depending on their level of altruism, their expectations of guilt or punishment from infecting another person, and the status of their stable partner if they have one. However, existing empirical evidence points towards decreased risk-taking after learning a HIV-positive status (di Paula et al., 2014).

[48] This prediction is reinforced if lower values of β^f are correlated with lower values of β^m, for example if the male partners of HIV-positive women also care more about protection. However, we do not observe proxies of β^m.

[49] As a framing question, respondents are first asked about the probability of a woman in general contracting HIV from unprotected sex. The pattern of results is similar when using this variable. These results are available on request.

[50] The correlation with current use of male condoms at baseline is 4.7 percentage points, but this is marginally insignificant (p-value 0.146).

Table 10: Treatment effects by baseline HIV status

	(1) **Ever FC** **HIV positive** **at baseline**	(2) **Ever FC** **HIV negative** **at baseline**	(3) **Last 30 days FC** **HIV positive** **at baseline**	(4) **Last 30 days FC** **HIV negative** **at baseline**	(5) **Current FC** **HIV positive** **at baseline**	(6) **Current FC** **HIV negative** **at baseline**
Treatment	0.392***	0.144***	0.120	0.047	0.200**	0.058
	(0.107)	(0.045)	(0.075)	(0.031)	(0.095)	(0.036)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	59	138	59	138	59	138
Control mean endline	0.133	0.067	0.000	0.017	0.033	0.017
$\chi^2(1) : (a) = (b)$		4.58		0.79		1.95
$\Pr > \chi^2$		0.037		0.358		0.162

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms. Columns 1-2 refer to whether the respondent has used ever used female condoms, columns 3-4 to whether she has used them in the last 30 days, and columns 5-6 to whether she is currently them. Odd-numbered columns present results for the subsample of individuals who report being HIV-positive at baseline; even-numbered columns present results for the subsample who report being HIV-negative at baseline. The bottom two rows present chi-squared statistics and their p-values for the test that the treatment effect is the same across the two subsamples. These are obtained from seemingly unrelated estimations. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

A possible explanation for these findings is that women with low priors about the risk of unprotected sex are the ones who update their beliefs as a result of the programme, and thus change their behaviour accordingly. Column (7) of Table A.22 in Appendix A.4 provides suggestive evidence that this is the case. The treatment has a positive and significant effect on risk perceptions for women with below-median baseline risk perceptions. This is again completely offset by a negative interaction term for women whose baseline risk beliefs were above the median, suggesting the programme did not shift their beliefs at all. However, we caution against over-interpretation of this result, since the median baseline risk perception was "risky" (four) and so women who had above-median perceptions already answered the maximum of "very risky" (five). Thus even if the programme did increase these women's perceptions of risk, we would not capture this in our measure.[51]

Finally, column (1) of Table A.30 in Appendix A.4 shows that we observe no evidence that treated respondents increase their knowledge of HIV. In fact, treated respondents score slightly lower on a set of six questions about how HIV can be transmitted, although this is only marginally significant (p=0.095) and baseline scores are already high.

5.5 Effects on bargaining power, STIs, violence, and well-being

As described in Section 5.2, we observe a significant increase in the number of sex acts per week and a significant decrease in the number of discussions or arguments. Both of these results suggest positive impacts on couples' utility. Columns (2)-(5) of Table A.30 in Appendix A.4 show that we observe almost no treatment impact on a variety of other outcome measures associated with welfare. The treatment group are no less likely to report being HIV-positive or having an STI. However, we do not place too much weight on this result, since it is unlikely that many respondents got newly tested for HIV or STIs in the time window between the start of the programme and the endline survey (four months). There is also no reduction in the likelihood of being pregnant, although we observe only a handful of pregnancies at baseline and endline, given that not being pregnant was a restriction for being accepted into the study sample.

Table A.29 in Appendix A.4 shows no substantial treatment impacts on the measures of bargaining power, as measured by indicators of decision-making and power dynamics.[52] The treatment significantly reduces the likelihood that the respondent decides over her health expenses and beliefs that her partner is involved with others, and increases the likelihood that she has more say and more power and that her

[51] We also see a significant treatment effect of 14.9 percentage points (p-value 0.022) on the proportion of women agreeing or strongly agreeing with the statement "my partner is involved with other people apart from me". This may also proxy increased risk perceptions. When we analyse the continuous version of this variable, we cannot reject that women with high and low baseline beliefs increase their belief equally in response to treatment.

[52] Assets at marriage are fixed at the time of marriage and can not be impacted by treatment

partner always wants to know where she is.

Reassuringly, we see no negative effects of the treatment on self-reported well-being, as measured through a survey module with twelve separate indicators. This is in contrast to the findings of Ashraf et al. (2014b) who see a decrease in women's reported well-being after being assigned to receive concealable contraceptives. The difference may reflect the fact that condom use is ultimately something agreed on and observed by both partners, rather than something hidden and conducive to moral hazard. Nor do we observe any adverse effects on reports of physical or emotional violence, as measured through five separate indicators. This may reflect the careful design of the programme, to mitigate the possibility that women suggesting new contraceptive use might face backlash from their partner.

5.6 Alternative explanations and robustness

Access: An alternative possible explanation for the negative interaction terms could be if women with lower bargaining power are less able than women with higher bargaining power to access male condoms (or other contraceptives) through the market or at health clinics. For example, women with lower bargaining power may be less confident to do so, or their partners may place restrictions on their movements. If so, the heterogeneity would be driven by heterogeneity in the extent to which the intervention alleviates the market access constraint, rather than male partners' direct unwillingness to use male condoms. However, if this was the case then we would expect to see a similar pattern of heterogeneity in current use of male condoms, since male condoms are also freely available from the health workers. Yet Table A.21 in Appendix A.4 shows that this is not the case: women with lower bargaining power are not consistently more likely than women with higher bargaining power to take up male condoms. If anything the interaction between treatment and bargaining power is positive, although in most cases it is not significant. Moreover, the same pattern of interactions between treatment and bargaining power is observed when we restrict the sample to women who have the easiest pre-treatment access to male condoms, as proxied by living below the median walking distance to the health centre (results available on request).[53] Thus even among these women, it is still the case that women with lower bargaining power adopt female condoms. Furthermore, as a kind of placebo test we also interact treatment with walking distance to the health centre itself (results available on request). This term is not significant for female (or male) condom adoption, and thus it does not appear that women with the lowest access are driving the treatment effects. Thus the results are more consistent with the idea formalised in the model that women with low bargaining power are constrained by their partners' preferences rather, than by alternative barriers to access.

[53]Distance to the health centre is negatively correlated with ever having used male condoms or other forms of contraception, and thus appears to be a good proxy of access.

Use of other contraceptive methods: Table A.23 in Appendix A.4 shows that the interaction between bargaining power and treatment is also not proxying a differential effect of treatment depending on whether the respondent is using other methods of contraception (i.e. the pill or injectables) at baseline. Our measures of bargaining power are positively correlated with current use of the pill at baseline, and negatively correlated with use of injectables at baseline, consistent with the arguments of Ashraf et al. (2014b) that women with low bargaining power may use concealables as a way to hide contraceptive use from their partner. However, when baseline use of other forms of contraception and its interaction with treatment is included into the regressions, the interactions between treatment and bargaining power remain negative and highly significant.

HIV status: Table A.24 in Appendix A.4 shows that heterogeneity by bargaining power is not proxying the observed heterogeneity by HIV status. This could have been the case since women with lower bargaining power are more likely to be HIV-positive. However, the interaction of the bargaining power measures with treatment remain negative and significant when controlling for HIV status and its interaction with treatment.

Risk beliefs: Table A.25 in Appendix A.4 also shows that heterogeneity by bargaining power is not proxying the heterogeneity by risk beliefs. This could have been the case if women with lower bargaining power revised their risk perceptions upwards more strongly as a result of the intervention. However, again the interaction of the bargaining power measures with treatment remain negative and significant when controlling for baseline risk beliefs and their interaction with treatment.

5.6.1 Bounding treatment effects for attrition

As mentioned in Section 4.1, attrition is slightly higher in the control group. For robustness we therefore conduct a bounds analysis on the main treatment effects following Lee (2009). We do not include facilitator fixed effects, since whether attrition is higher in the treatment or control group varies by facilitator. Inclusion of facilitator fixed effects would then imply that the monotonicity assumption required for Lee bounds would fail. Table A.26 in Appendix A.4 details the results of the bounds analysis on the main treatment effects without facilitator fixed effects. Whilst the lower bounds cannot rule out a treatment effect of zero for the various measures of female condom use when facilitator fixed effects are not included, we are able to rule out any sizeable negative effects, and the upper bounds are large and highly significant.

5.6.2 Diary robustness checks

As robustness checks, we also re-estimate the individual-level impacts using the diary data. First, to check that the diary respondents are representative of the full sample in terms of the impacts estimated from the survey data, we repeat estimation of Equation 21 with the survey data but only for diary respondents. Table A.27 in Appendix A.4 confirms that the results are not different from the main sample, although the result for female condom use in the last 30 days loses significance due to the loss of sample size.

Next, we re-estimate impacts using the main outcome variables measured from the diary data. We first use the linear probability fixed effects panel model from Equation 10.[54] Table A.27 in Appendix A.4 show the results. Again, we see significant impacts for the use of female condoms during the full endline period and the last 30 days, while we see no significant impacts on male condom use and on female condom use in the last 14 days. We also estimate ANCOVA specifications, for comparability with the main results (see online appendix).[55] The results are available in the online appendix. Consistent with the survey data, we see significant and positive impacts on ever use and use in the last 30 days of female condoms. The impact on use of female condoms in the last 14 days is no longer significant due to the loss of sample size. We see no significant impacts for the use of male condoms.

6 Cost-Benefit and Cost-Effectiveness Analysis

To further explore the policy implications of our results, we estimate the effects on the entire population of South Mozambique of scaling up the intervention to cover all women in the sexually active age-group (15-49 years) for the years 2015-30, excluding high-risk groups.[56] We first estimate the number of HIV

[54]Again, similar results are obtained with a logit specification (see A.4).

[55]Since the nature of the diary data is such that baseline observations are missing for some respondents, we follow McK and estimate:

$$Pr\left[Y_{if1}=1|Y_{if0},treat_{if},\eta_f\right]=\alpha+\delta Y_{if0}+\beta_1 treat_{if}+\gamma\text{missbase}_{if}+\sigma\text{missbase}_{if}\times Y_{if0}+\eta_f \tag{11}$$

where Y_{if1} is the outcome variable of interest for the endline period. Y_{if0} is the value of the outcome variable for the baseline period. $missbase_{if}$ is a dummy equal to one if the respondent is missing the value of the outcome variable during the baseline period. $missbase_{if}\times Y_{if0}$ then sets the baseline value to zero in the case that it is missing. Inclusion of this dummy means that δ is estimated only for respondents whose baseline data is not missing. $treat_{if}$ is again a dummy for being in the treatment group, and η_f is again a facilitator fixed effect. β_1 represents the intent-to-treat effect, this time as estimated on the diary subsample.

[56]In the epidemiological model that we use, adults above the median age of first sex are allocated into one of five risk categories, identified for males and females separately. These are: stable couples (men and women reporting a single partner in the last year); multiple partners (men and women with more than one partner in the last year); female sex workers and clients; men who have sex with men; and injecting drug users. Our intervention targets women in the first two categories, whose partners are estimated by the epidemiological model

infections and disability-adjusted life years (DALYs) that such a scale-up would help to avert. We then estimate the implied cost savings to the healthcare system of Mozambique, through reduced provision of ARTs and other treatments and through the reduction of productivity losses. Finally, we estimate the costs of the scale-up of our intervention. Comparing the programme costs to the healthcare system cost savings allows us to calculate the internal rate of return (IRR). This is an indicator of cost-benefit, which can be used to evaluate the policy as a financial investment. Meanwhile, comparing the programme costs to the DALYs averted allows us to calculate the incremental cost-effectiveness ratio (ICER). This measure is often used to compare the cost-effectiveness of policies across the public health spectrum, in terms of cost per DALY averted (Creese et al., 2002; Oster, 2005; Uthman et al., 2010; Bärnighausen et al., 2012; Walensky et al., 2013).

In light of the model and results presented in Section 5, we make intervention projections based on two different scenarios in terms of programme impact. In the first scenario we focus purely on the increase in condom coverage and marginal decrease in average condom efficacy when individuals adopt female condoms as a result of the intervention. In the second scenario we also take into account the estimated increase in the number of sex acts. This is especially important because an increase in the number of sex acts has the potential to offset partially or fully the benefits of the increased rate of condom coverage (Greenwood et al., 2017), especially if these sex acts are unprotected.

6.1 Modelling health impacts

To estimate the number of HIV infections and DALYs averted as a result of the scale-up of the intervention, we use the AIM module of the SPECTRUM suite of models, which is the software used by UNAIDS to make its national and global projections, and by governments to develop national strategic plans on HIV/AIDS (Stover; Stover et al., 2008, 2010, 2011, 2014, 2017).[57] The model combines epidemiological variables with behavioural factors based on gender and risk-group. This allows us to simulate the impact of an intervention such as ours, that targets partners' propensity to use condoms and the number of sex acts they engage in, on the entire population (Bärnighausen et al., 2012).[58]

Up to and including 2015 we take as given AIM's demographic estimations, including HIV transmission,

to be primarily in the second category. It does not target individuals in the last three, high-risk categories.

[57] The SPECTRUM suite is developed by Avenir Health, see http://www.avenirhealth.org/software-spectrum.php.

[58] Specifically, AIM estimates new HIV infections by gender and risk group, as a function of behavioural and epidemiological factors. These factors include: condom use, number of partners, number of sex acts, contacts per partner, ART use, voluntary medical male circumcision, and the prevalence of other STIs. The clinical progression of the HIV/AIDS epidemic is modelled as a function of CD4 count after HIV infection. The model has been fitted to antenatal clinic surveillance data, household and key risk group surveys, program statistics, and financial records for the North/Central/South regions of Mozambique since 1982 (Korenromp et al., 2015).

and its assumptions on the HIV/AIDS national strategic program in Mozambique, which influences factors such as ART coverage (Korenromp et al., 2015).[59] For the years after 2015, we update these with the most recent UNAIDS data. We then assume that all women in the sexually active age group receive the intervention in 2015, and from 2016-30 only 15 year old girls and female immigrants receive the intervention. We first simulate a control projection, where estimates from 2015-16 are taken and projections for 2017-30 are made with none of the epidemiological and behavioural parameters changed. We then simulate two intervention scenarios, where condom coverage, average condom efficacy, and (in scenario two) the number of sex acts are changed in line with the impacts of the intervention estimated from our experiment.

In both of the intervention scenarios, we assume that condom coverage increases by 10.5 percentage points. This aggregates the estimates from Table A.17 in Appendix A.4 of the treatment effects on the proportion of sex acts per respondent per week that are protected by female condoms during the full endline period in the diaries (10.5 percentage points) and male condoms (no significant effect). For the diaries our preferred estimate of adoption is the full endline period, as not all diary participants are interviewed each week and hence some of the variation from week-to-week may be driven by the sample composition. The full endline period captures all observations per week per respondent. We assume a conservative average condom effectiveness of 78.9%, based on the proportion of male and female condoms used by the treatment group in the endline data, and the effectiveness of male and female condoms calibrated to population health conditions in Mozambique.[60]

In the second scenario, we also adjust the average number of sex acts per respondent per year. Again, given that not all diary respondents participate in an interview every week, and some have vastly more sex acts than others, our most robust estimation approach is first to estimate the treatment effect on the probability that an individual has any sex acts in a given week in the full endline period (9.1 percentage points, significant at the 5% level). We then multiply this by the average number of sex acts per respondent per week in the control group in the full endline period (0.79 sex acts). We than add this to average number of sex acts per week in the AIM model for the relevant risk groups and multiply this by 52 to arrive at the additional sex acts per individual per year (3.5 sex acts). Figure A.2 in Appendix A.4 shows a graph of the HIV infections per year between 2015 and 2030 for the status quo and the two

[59] Mozambique's national HIV/AIDS program began in 2001 and includes condom promotion, community mobilisation, HIV counselling and testing, prevention targeted at female sex workers and clients, ART, prevention of mother-to-child transmission, and voluntary adult medical male circumcision.

[60] The general effectiveness of female and male condoms established in the medical literature is 79% and 85% respectively (Hatcher and Nelson, 2007). However, this can vary by factors such as underlying population health. The male condom effectiveness in AIM, based on careful calibration to data from Mozambique, is 80%. Our estimate of female condom effectiveness for Mozambique is therefore $\frac{79}{85} \times 80 = 74\%$. To establish aggregate condom effectiveness (which is the parameter required by AIM), we take the estimates from our endline sex-act level data that 38.3% of sex acts are protected by male condoms and 8.0% by female condoms. The mixture of condom use is therefore 17.3% female condoms and 82.7% male, leading to a weighted condom effectiveness of $0.173 \times 74 + 0.827 \times 80 = 78.9\%$.

intervention scenario's. In the control scenario, there are 388,437 new HIV infections and 59,743,636 new DALYS by 2030. In the first intervention scenario, the scaled-up programme averts 41,073 HIV infections and 66,514 DALYs. In the second intervention scenario, the programme averts 11,437 HIV infections and 25,875 DALYS.

6.2 Cost savings

To estimate the financial benefits of our intervention to the healthcare system, we focus on the reduction in the number of adults and children that require ART, and the number of mothers requiring Prevention of Mother-To-Child Transmission (PMTCT) for the period from 2015-2030. This is a very conservative estimate because there are other financial benefits to the health system, such as the reduced costs of unintended pregnancies and hospital admissions for AIDS-related diseases.

The tables in the online Cost Benefit Analysis Appendix show that in scenario one, the cumulative number of individuals that receive ART each year is reduced by 189,278 adults and 14,718 children, the cumulative number of children that receive cotrimoxazole[61] is reduced by 22,854, whilst the cumulative number of women who receive PMTCT is reduced by 17,024. The total implied discounted cost savings are 160,088,910 USD.[62] In scenario two, the cumulative number of individuals that receive ART each year is reduced by 39,148 adults and 3,135 children, the cumulative number of children that receive cotrimoxazole is reduced by 5,427 whilst the cumulative number of women who receive PMTCT is reduced by 4,051. The total implied discounted cost savings are 33,799,234 USD. [63]

To estimate the cost savings of our intervention in terms of productivity gains we focus on the reduction in productivity losses due to continued workforce participation of adults who did not get infected with HIV as a result of our intervention. To calculate this reduction in productivity losses we use each country's average gross national income (GNI) per working-age adult as a proxy for working-age adult economic productivity. It is assumed that individuals who fall ill will achieve 10% of the productivity of their HIV-negative counterparts, while individuals who receive ART will achieve 75% of the productivity of HIV-negative counterparts after the first year of falling ill Resch et al. (2011). This implies that the total cost savings from the reduction of productivity losses in scenario one are 72,388,635 USD and in scenario two are 17,803,780 USD.[64]

[61] Cotrimoxazole is most commonly used by people with HIV as prophylaxis against pneumonia

[62] This comprises 152,666,673 USD for adult ART; 3,449,062 USD for child ART; 850,692 USD for child cotrimoxazole, and 3,122,483 USD for PMTCT.

[63] This comprises 32,090,970 USD for adult ART; 751,265 USD for child ART; 204,644 USD for child cotrimoxazole, and 752,355 USD for PMTCT.

[64] We use the average GNI per capita of working age adults in 2014 (1,120 PPP USD) based on World Bank data. We discount the GNI per capita by 3% per year, in line with the discounting used for other costs. We use the PPP value as most of the costs we consider, such as the unit and distribution costs of condoms and the cost

6.3 Programme costs

We calculate an upper and a lower bound of the costs per participant. For the upper bound we use the full costs of the intervention as implemented, plus the full cost of acquiring and distributing the subsequent increase in the number of female condoms used between 2015 and 2030, assuming full subsidisation of female condom provision by the government. For the lower bound, we assume that the provision of information about female condoms is included into existing sex education programmes in schools and at health centres. This is a realistic add-on to such programmes, given that they already provide information about and practical demonstrations of male condoms, as well as information about HIV/AIDS and other STIs. The lower bound cost estimates therefore comprise just the costs of acquiring and distributing the additional number of female condoms when adoption subsequently increases, assuming that the government fully subsidises free provision of female condoms.

The total intervention cost per participant is 28.90 USD, including the costs of facilitator training, door-to-door recruitment, group organisation, purchase and distribution of condoms during the intervention itself (but not afterwards), administration, and monitoring and evaluation.[65] As stated above, we assume that all women aged 15-49 are treated in 2015, and then from 2016 onwards that only the new 15-year-olds and migrants are treated each year. All costs are in constant 2014 USD, and we apply a 3% discount rate as is common in this literature (Stover et al., 2017). This implies an initial intervention cost in 2015 of 47.8 million USD, and an average discounted cost of 3.6 million USD per year for the years 2016-2030 inclusive.

To estimate the cost of meeting the increased demand for female condoms after the intervention, we follow Dowdy et al. (2006) and assume a combined unit acquisition and distribution cost of 0.45 USD.[66] We estimate the average additional number of female condoms required per year to be 13.9 million in scenario one and 15.4 million in scenario two, with average discounted costs per year of 8.0 million USD and 8.9 million USD respectively.[67]

Taken altogether (see online Cost Benefit Analysis Appendix), for scenario one this implies an upper bound of 229,129,567 USD on the programme cost (i.e., the full intervention cost) and a lower bound

of ARTs are values at international prices.

[65] The total cost of the intervention for 298 women was 259,239 Mozambican Meticais, equating to 8,612 USD at an exchange rate of 1 MZN=0.03322 USD on 14 August 2014.

[66] This is based on the high-volume scenario of (Dowdy et al., 2006), in which the number of female condoms used is 30% of the number of male condoms used. In our setting the proportion of sex acts protected by female condoms as a proportion of the sex acts protected by male condoms in our treatment group's endline data is 27% (see Table A.17).

[67] Our estimate of the total number of female condoms comes from combining the proportion of the adult population in each risk group (estimated from AIM) with the estimated treatment effects on the percentage of acts protected by female condoms, and the estimated number of sex acts per individual per year. The latter is taken as constant in scenario one and increases with treatment in scenario two, in line with the discussion above.

of 127,791,649 USD (i.e., the cost of adding female condoms to existing sex education programmes). In scenario two, the upper bound is 243,416,199 USD and the lower bound is 142,078,704 USD.

6.4 Cost-benefit and cost-effectiveness results

In scenario one the ICER for the full intervention is -50 USD[68], a saving of 50 USD per DALY averted, meaning that scaling up the full intervention is therefore *very cost-effective* and offers a positive financial return (IRR=1.02). Meanwhile the ICER for the lower-cost, add-on intervention is -1,574 USD, a saving of 1,574 USD, meaning that adding female condom provision to existing sex education programs is also *very cost-effective* and in fact represents a substantial saving per DALY averted compared to the existing set of treatments, and offers a highly favourable return on investment of 1.82. Low-cost delivery mechanisms, such as adding female condoms to the curriculum of school sex education programmes, has the potential to be cost-effective and cost-saving.

In scenario two the ICER for the full intervention is 7,413 USD, meaning that a full scale up of the intervention is not cost-effective. However, the ICER for the lower bound is 3,497 USD, implying that adding female condom provision to existing sex education programs is *cost-effective*. The IRR for the upper-bound cost is 0.21 and for the lower-bound cost is 0.36. Despite being cost-effective in the lower bound scenario, the intervention does not offer a positive financial return on investment.

6.5 Discussion

In summary, only the full intervention in scenario two, taking account of the increase in risky sex acts, is *not* cost-effective. However, there are still several reasons to believe that our estimates of the IRR and ICER are conservative. First, we use an upper bound for the estimated costs of condoms, which is likely to be highly conservative given that the scale-up of the intervention to the entire female population of South Mozambique would lead to economies of scale in production or procurement. Second, as mentioned above, potentially sizeable benefits, such as reduction in unwanted pregnancies, indirect costs to the health system, costs for orphan care, are not included in our estimates. Third, as shown in Table 10, the impacts of the intervention on female condom use appear to be larger for women who are HIV-positive. Increased coverage among this group will have a disproportionately large impact on HIV transmission rates. However, this cannot be accounted for in the AIM projections since the coverage

[68] Following the recommendations of the Commission on Macroeconomics and Health, WHO-CHOICE deems interventions highly cost-effective if the ICER is less than GDP per capita, cost-effective if the ICER is between one and three times GDP per capita, or not cost-effective if the ICER is higher than three times the GDP per capita (Walensky et al., 2013) The GDP per capita of Mozambique was 511 USD in 2014.

Table 11: Cost-effectiveness and cost-benefit analysis

	Scenario 1		Scenario 2	
Year	Lower bound	Upper bound	Lower bound	Upper bound
HIV infections averted	41,073		11,437	
DALYs averted	66,514		25,875	
Total cost savings (USD)	232,477,545		51,603,014	
Annual GDP per capita (PPP USD)	1192		1192	
Programme cost (USD)	127,791,649	229,129,567	142,078,704	243,416,199
ICER (USD)	**-1,574**	**-50**	**3,497**	**7,413**
Internal rate of return (IRR)	**1.82**	**1.02**	**0.36**	**0.21**

Notes: In scenario 1 we focus purely on the increase in condom coverage and marginal decrease in average condom efficacy when individuals adopt female condoms as a result of the intervention. In the second scenario, we also take into account the estimated increase in the number of sex acts. We calculate an upper and a lower bound of the costs per participant. For the upper bound we use the full costs of the intervention as implemented by Pathfinder, plus the full cost of acquiring and distributing the subsequent increase in the number of female condoms used between 2015 and 2030, assuming full subsidisation of female condom provision by the government. For the lower bound, we assume that the provision of information about female condoms is included into existing sex education programmes in schools and at health centres. The GDP per capita in PPP USD for 2015 is used, as the majority of our scale-up costs are made in 2015. We then discount the 2015 figure at 3% per year.

rates for HIV-positive and HIV-negative individuals cannot be adjusted separately. Fourth, as shown in Table 3, the women in our sample are above-average in their bargaining power when compared to a representative sample from the DHS. Given that the estimated treatment effects on female condom use are *larger* for the women with *lower* bargaining power in our sample, it is therefore likely that the average impact of the intervention in the population would be larger than that observed in our sample. If so, and the intervention increased condom usage by 10.5 percentage points or more in the general population, then scaling up the full intervention (implying the upper bound on costs) would be cost-effective even in scenario two.

7 Conclusion

In terms of HIV/AIDS policy, our results imply that female condoms are largely taken up by women who are otherwise having unprotected sex, rather than by women who are otherwise using male condoms. This means that the correct cost comparison for free provision of female condoms is not to the free provision of male condoms, but rather to the costs of ARTs and other costs associated with unprotected sex. Given this, our simulations show that free provision of female condoms may be highly cost-effective, even implying a saving compared to the cost of treatment, if implemented via incorporation of female condoms into existing sex education programmes in schools and health centres.

Our results attribute this pattern of take-up to intra-household bargaining over contraceptive use. Female condoms appear to be the only STI protection method that some women with low bargaining power can convince their partners to use. Thus the finding that free provision of female condoms can be a cost-effective policy is particularly likely to hold in contexts where women have low bargaining power.

To aid more comprehensive welfare calculations and funding decisions, it would be useful for future studies to determine individual men and women's willingness-to-pay of individual men and women once they have tried both technologies. Evidence on longer-term adoption of female condoms is also crucial, given that our study and studies in the medical literature have tended to focus on the first three to six months.

More broadly, we have highlighted how low female bargaining power may lead to low adoption of technologies that improve household welfare, in cases where women have a stronger preference for adoption or face higher costs of non-adoption. There are many other examples of welfare-improving technologies where women may have a stronger willingness to adopt. For instance, women may have a higher demand for insurance, given evidence that they are more risk-averse. Women may also have a higher willingness-to-pay for household sanitation technologies such as private toilets, since women face larger stigma and risks to their personal safety from using facilities outside of the household. Furthermore, women may have a stronger preference for delaying the marriage and childbearing of adolescent girls, if women's preferences for lower fertility and higher birth spacing apply to their children. In each of these cases, information and social norm campaigns targeted specifically at men may be necessary to increase welfare-improving investments and adoption (Stopnitzky, 2017). If such campaigns still cannot resolve low adoption, then providing versions of the technology that are more acceptable to men, or bundling technologies with goods for which men have a high demand, may offer a second-best solution. These remain important topics for future research.

References

Another reason to prefer Ancova: dealing with changes in measurement between baseline and follow-up. URL https://blogs.worldbank.org/impactevaluations/another-reason-prefer-ancova-dealing-changes-measurement-between-baseline-and-follow.

The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. 1981.

A. Abadie, S. Athey, G. W. Imbens, and J. M. Wooldridge. When should you adjust standard errors for clustering? *NBER Working Paper No. 24003*, 2017.

J. R. Agnew, L. R. Anderson, J. R. Gerlach, and L. R. Szykman. Who chooses annuities? An experimental investigation of the role of gender, framing, and defaults. *American Economic Review*, 98(2): 418–422, 2008.

T. Amemiya. Some theorems in the linear probability model. *International Economic Review*, pages 645–650, 1977.

S. Anderson and J.-M. Baland. The Economics of ROSCAs and Intrahousehold Resource Allocation. *Quarterly Journal of Economics*, 117(3):963–995, 2002.

D. Ariely and G. Loewenstein. The heat of the moment: The effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, 19(2):87–98, 2006.

N. Ashraf, O. Bandiera, and B. K. Jack. No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics*, 120:1–17, 2014a.

N. Ashraf, E. Field, and J. Lee. Household bargaining and excess fertility: an experimental study in Zambia. *American Economic Review*, 104(7):2210–2237, 2014b.

O. P. Attanasio and V. Lechene. Efficient responses to targeted cash transfers. *Journal of Political Economy*, 122(1):178–222, 2014.

M. J. Bailey. More power to the pill: the impact of contraceptive freedom on women's life cycle labor supply. *Quarterly Journal of Economics*, 121(1):289–320, 2006.

S. Baird, E. Gong, C. McIntosh, and B. Özler. The heterogeneous effects of HIV testing. *Journal of Health Economics*, 37:98–112, 2014.

S. J. Baird, R. S. Garfein, C. T. McIntosh, and B. Özler. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *The Lancet*, 379(9823):1320–1329, 2012.

O. Bandiera, N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, M. Sulaiman, et al. Women's economic empowerment in action: Evidence from a randomized control trial in Africa. *Working paper No. 187*, 2015.

B. M. Barber and T. Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, 116(1):261–292, 2001.

T. Bärnighausen, D. E. Bloom, and S. Humair. Economics of antiretroviral treatment vs. circumcision for HIV prevention. *Proceedings of the National Academy of Sciences*, 109(52):21271–21276, 2012.

M. Beksinska, J. Smit, C. Joanis, and C. Hart. Practice makes perfect: reduction in female condom failures and user problems with short-term experience in a randomized trial. *Contraception*, 86(2): 127–131, 2012.

M. Bjorkman Nyqvist, L. Corno, D. de Walque, and J. Svensson. Using lotteries to incentivize safer sexual behavior: evidence from a randomized controlled trial on HIV prevention. *World Bank Policy Research Working Paper No. 7215*, 2015.

M. Browning and P.-A. Chiappori. Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, 66(6):1241–1278, 1998.

M. Bruhn and D. McKenzie. In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1(4):200–232, 2009.

V. Chernozhukov, I. Fernández-Val, J. Hahn, and W. Newey. Average and quantile effects in nonseparable panel models. *Econometrica*, 81(2):535–580, 2013.

P.-A. Chiappori. Collective labor supply and welfare. *Journal of Political Economy*, 100(3):437–467, 1992.

P.-A. Chiappori and S. Oreffice. Birth control and female empowerment: An equilibrium analysis. *Journal of Political Economy*, 116(1):113–140, 2008.

K.-H. Choi, J. Wojcicki, and D. Valencia-Garcia. Introducing and negotiating the use of female condoms in sexual relationships: Qualitative interviews with women attending a family planning clinic. *AIDS and Behavior*, 8(3):251–261, 2004.

E. N. Coman, M. R. Weeks, I. Yanovitzky, E. Iordache, R. Barbour, M. A. Coman, and T. B. Huedo-Medina. The impact of information about the female condom on female condom use among males and females from a US urban community. *AIDS and Behavior*, 17(6):2194–2201, 2013.

A. Creese, K. Floyd, A. Alban, and L. Guinness. Cost-effectiveness of HIV/AIDS interventions in Africa: a systematic review of the evidence. *The Lancet*, 359(9318):1635–1642, 2002.

J. Das, J. Hammer, and C. Sánchez-Paramo. The impact of recall periods on reported morbidity and health seeking behavior. *Journal of Development Economics*, 98(1):76–88, 2012.

D. De Walque, W. H. Dow, R. Nathan, R. Abdul, F. Abilahi, E. Gong, Z. Isdahl, J. Jamison, B. Jullu, S. Krishnan, et al. Incentivising safe sex: a randomised trial of conditional cash transfers for HIV and sexually transmitted infection prevention in rural Tanzania. *BMJ Open*, 2(1), 2012.

DHS. Mozambique Demographic and Health Survey 2011 – Final report. 2011. URL https://dhsprogram.com/pubs/pdf/FR266/FR266.pdf.

Á. D. di Paula, G. Shapira, and P. E. Todd. How beliefs about HIV status affect risky behaviors: Evidence from Malawi. *Journal of Applied Econometrics*, 29(6):944–964, 2014.

A. Donald, G. Koolwal, J. Annan, K. Falb, and M. Goldstein. Measuring women's agency. *World Bank Policy Research Working Paper No. 8148*, 2017.

D. W. Dowdy, M. D. Sweat, and D. R. Holtgrave. Country-wide distribution of the nitrile female condom (FC2) in Brazil and South Africa: a cost-effectiveness analysis. *AIDS*, 20(16):2091–2098, 2006.

E. Duflo. Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa. *The World Bank Economic Review*, 17(1):1–25, 2003.

E. Duflo and C. Udry. Intrahousehold resource allocation in Cote d'Ivoire: Social norms, separate accounts and consumption choices. *NBER Working Paper No. 10498*, 2004.

E. Duflo, P. Dupas, and M. Kremer. Education, HIV, and early fertility: Experimental evidence from Kenya. *American Economic Review*, 105(9):2757–2797, 2015.

P. Dupas. Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. *American Economic Journal: Applied Economics*, 3(1):1–34, 2011.

C. C. Eckel and P. J. Grossman. *Men, women and risk aversion: Experimental evidence*, volume 1. Elsevier, 2008.

M. Eswaran. The empowerment of women, fertility, and child mortality: towards a theoretical analysis. *Journal of Population Economics*, 15(3):433–454, 2002.

G. Farr, H. Gabelnick, K. Sturgen, and L. Dorflinger. Contraceptive efficacy and acceptability of the female condom. *American Journal of Public Health*, 84(12):1960–1964, 1994.

D. Firth. Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38, 1993.

A. L. Fontanet, J. Saba, V. Chandelying, C. Sakondhavat, P. Bhiraleus, S. Rugpao, C. Chongsomchai, O. Kiriwat, S. Tovanabutra, L. Dally, et al. Protection against sexually transmitted diseases by granting sex workers in Thailand the choice of using the male or female condom: results from a randomized controlled trial. *AIDS*, 12(14):1851–1859, 1998.

O. Galárraga, S. G. Sosa-Rubí, C. Infante, P. J. Gertler, and S. M. Bertozzi. Willingness-to-accept reductions in HIV risks: conditional economic incentives in Mexico. *The European Journal of Health Economics*, 15(1):41–55, 2014.

P. Gertler, M. Shah, and S. M. Bertozzi. Risky business: the market for unprotected commercial sex. *Journal of Political Economy*, 113(3):518–550, 2005.

S. Godlonton, A. Munthali, and R. Thornton. Responding to risk: Circumcision, information, and HIV prevention. *Review of Economics and Statistics*, 98(2):333–349, 2016.

C. Goldin and L. F. Katz. The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of Political Economy*, 110(4):730–770, 2002.

J. Greenwood, P. Kircher, C. Santos, and M. Tertilt. An equilibrium model of the African HIV/AIDS epidemic. *NBER Working Paper No. 18953*, 2017.

R. A. Hatcher and A. L. Nelson. *Contraceptive technology. Nineteenth Revised Edition.* Ardent Media, 2007.

G. Heinze and M. Schemper. A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21(16):2409–2419, 2002.

M. J. Hoppe, D. M. Morrison, M. R. Gillmore, B. Beadnell, D. H. Higa, and B. C. Leigh. Agreement of daily diary and retrospective measures of condom use. *AIDS and Behavior*, 12(1):113–117, 2008.

W. C. Horrace and R. L. Oaxaca. Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, 90(3):321–327, 2006.

G. King and L. Zeng. Logistic regression in rare events data. *Political Analysis*, 9(2):137–163, 2001a.

G. King and L. Zeng. Explaining rare events in international relations. *International Organization*, 55 (3):693–715, 2001b.

G. King and L. Zeng. Estimating risk and rate levels, ratios and differences in case-control studies. *Statistics in Medicine*, 21(10):1409–1427, 2002.

E. L. Korenromp, B. Gobet, E. Fazito, J. Lara, L. Bollinger, and J. Stover. Impact and Cost of the HIV/AIDS National Strategic Plan for Mozambique, 2015-2019? Projections with the Spectrum/Goals Model. *PloS one*, 10(11):e0142908, 2015.

W. Koster, M. G. Bruinderink, and W. Janssens. Empowering Women or Pleasing Men? Analyzing Male Views on Female Condom Use In Zimbabwe, Nigeria and Cameroon. *International Perspectives on Sexual and Reproductive Health*, 41(3):126–135, 2015.

T. Lancaster. Orthogonal parameters and panel data. *Review of Economic Studies*, 69(3):647–666, 2002.

D. S. Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102, 2009.

H. Leitgöb. The problem of modeling rare events in ML-based logistic regression.

M. Macaluso, R. Blackwell, D. J. Jamieson, A. Kulczycki, M. P. Chen, R. Akers, D.-j. Kim, and A. Duerr. Efficacy of the male latex condom and of the female polyurethane condom as barriers to semen during intercourse: a randomized clinical trial. *American Journal of Epidemiology*, 166(1):88–96, 2007.

M. Macia, P. Maharaj, and A. Gresh. Masculinity and male sexual behaviour in Mozambique. *Culture, Health & Sexuality*, 13(10):1181–1192, 2011.

J. E. Mantell, J. A. Smit, T. M. Exner, Z. Mabude, S. Hoffman, M. Beksinska, E. A. Kelvin, C. Ngoloyi, C.-S. Leu, and Z. A. Stein. Promoting female condom use among female university students in KwaZulu-Natal, South Africa: results of a randomized behavioral trial. *AIDS and Behavior*, 19(7): 1129–1140, 2015.

G. Miller and A. M. Mobarak. Gender differences in preferences, intra-household externalities, and low demand for improved cookstoves. *NBER Working Paper No. 18964*, 2013.

I. N. d. E. Ministério da Saúde. Inquérito de Indicadores de Imunização Malária e HIV/SIDA em Moçambique: Relatório de Indicadores Básicos de HIV. *DHS Publications*, 2015. URL https://dhsprogram.com/pubs/pdf/PR85/PR85.pdf.

E. Oster. Sexually transmitted infections, sexual behavior, and the HIV/AIDS epidemic. *Quarterly Journal of Economics*, 120(2):467–515, 2005.

A. Peters, W. Jansen, and F. van Driel. The female condom: the international denial of a strong potential. *Reproductive Health Matters*, 18(35):119–128, 2010.

A. Philpott, W. Knerr, and D. Maher. Promoting protection and pleasure: amplifying the effectiveness of barriers against sexually transmitted infections and pregnancy. *The Lancet*, 368(9551):2028–2031, 2006.

K. M. Pilz. Assessment of the female condom supply chain in Mozambique. *i+solutions Report*, 2014.

I. Rasul. Household bargaining over fertility: Theory and evidence from Malaysia. *Journal of Development Economics*, 86(2):215–241, 2008.

S. Resch, E. Korenromp, J. Stover, M. Blakley, C. Krubiner, K. Thorien, R. Hecht, and R. Atun. Economic returns to investment in AIDS treatment in low and middle income countries. *PloS one*, 6 (10):e25310, 2011.

J. P. Ruger. Combating HIV/AIDS in developing countries: Requires empowering people to act on their own terms. *BMJ: British Medical Journal*, 329(7458):121, 2004.

S. Schaner. Do opposites detract? Intrahousehold preference heterogeneity and inefficient strategic savings. *American Economic Journal: Applied Economics*, 7(2):135–174, 2015.

C. Stalgaitis and S. N. Glick. The use of web-based diaries in sexual risk behaviour research: a systematic review. *Sexually Transmitted Infections*, 90(5):374–381, 2014.

Y. Stopnitzky. No toilet no bride? Intrahousehold bargaining in male-skewed marriage markets in India. *Journal of Development Economics*, 127:269–282, 2017.

J. Stover. Projecting the demographic consequences of adult HIV prevalence trends: the Spectrum Projection Package. *Sexually Transmitted Infections*.

J. Stover, B. Fidzani, B. C. Molomo, T. Moeti, and G. Musuka. Estimated HIV trends and program effects in Botswana. *PLoS One*, 3(11):e3729, 2008.

J. Stover, L. Bollinger, and C. Avila. Estimating the impact and cost of the WHO 2010 recommendations for antiretroviral therapy. *AIDS Research and Treatment*, 2011, 2010.

J. Stover, E. L. Korenromp, M. Blakley, R. Komatsu, K. Viisainen, L. Bollinger, and R. Atun. Long-term costs and health impact of continued global fund support for antiretroviral therapy. *PLoS One*, 6(6): e21048, 2011.

J. Stover, T. B. Hallett, Z. Wu, M. Warren, C. Gopalappa, C. Pretorius, P. D. Ghys, J. Montaner, B. Schwartländer, N. P. T. S. Group, et al. How can we get close to zero? The potential contribution of biomedical prevention and the investment framework towards an effective response to HIV. *PloS One*, 9(11):e111956, 2014.

J. Stover, J. E. Rosen, M. N. Carvalho, E. L. Korenromp, H. S. Friedman, M. Cogan, and B. Deperthes. The case for investing in the male condom. *PloS one*, 12(5):e0177108, 2017.

H. Thirumurthy, S. H. Masters, S. Rao, M. A. Bronson, M. Lanham, E. Omanga, E. Evens, and K. Agot. Effect of providing conditional economic compensation on uptake of voluntary medical male circumcision in Kenya: a randomized clinical trial. *Journal of the American Medical Association*, 312(7): 703–711, 2014.

R. L. Thornton. The demand for, and impact of, learning HIV status. *American Economic Review*, 98 (5):1829–1863, 2008.

J. Trussell. Contraceptive failure in the United States. *Contraception*, 83:397–404, 2011.

UNAIDS. UNAIDS Prevention Gap Report. 2016a. URL http://www.unaids.org/sites/default/files/media_asset/2016-prevention-gap-report_en.pdf.

UNAIDS. HIV prevention among adolescent girls and young women. 2016b. URL http://www.unaids.org/sites/default/files/media_asset/UNAIDS_HIV_prevention_among_adolescent_girls_and_young_women.pdf.

O. A. Uthman, T. A. Popoola, M. M. Uthman, and O. Aremu. Economic evaluations of adult male circumcision for prevention of heterosexual acquisition of HIV in men in sub-Saharan Africa: a systematic review. *PloS one*, 5(3):e9628, 2010.

E. Van Tassel. Household bargaining and microfinance. *Journal of Development Economics*, 74(2):449–468, 2004.

G. Vijayakumar, Z. Mabude, J. Smit, M. Beksinska, and M. Lurie. A review of female-condom effectiveness: patterns of use and impact on protected sex acts and STI incidence. *International Journal of STD & AIDS*, 17(10):652–659, 2006.

R. P. Walensky, E. L. Ross, N. Kumarasamy, R. Wood, F. Noubary, A. D. Paltiel, Y. M. Nakamura, S. V. Godbole, R. Panchia, I. Sanne, et al. Cost-effectiveness of HIV treatment as prevention in serodiscordant couples. *New England Journal of Medicine*, 369(18):1715–1725, 2013.

R. K. Wanyenze, L. Atuyambe, V. Kibirige, S. Mbabazi, N. M. Tumwesigye, K. Djurhuus, and A. Namale. The new female condom (FC2) in Uganda: perceptions and experiences of users and their sexual partners. *African Journal of AIDS Research*, 10(3):219–224, 2011.

A Appendices

A.1 Theoretical Appendices

A.1.1 Proof that H* is increasing in female bargaining power

For convenience of notation, define

$$U_j(H) = u_j(P(H), H) \tag{1}$$

for partner $j = m, f$, where $P(H)$ describes the technological frontier. Equation 4 becomes

$$\max_H \left\{ \alpha U_f (H) + (1-\alpha) U_m (H) \right\}. \tag{2}$$

The first-order condition is

$$\alpha U_f' (H) + (1-\alpha) U_m' (H) = 0. \tag{3}$$

Note this implies that at the optimal choice $\widetilde{H}$, U_f' and U_m' must be of opposite signs. It follows from the single-crossing property in Equation 1 that at the optimum, $U_f' (H) > 0$ and $U_m' (H) < 0$.

The second-order condition is

$$\alpha U_f'' (H) + (1-\alpha) U_m'' (H) < 0. \tag{4}$$

Taking the first-order condition in Equation 3 as an implicit definition of $\widetilde{H} (\alpha)$, and differentiating with respect to α, we obtain

$$\left[\alpha U_f'' (H (\alpha)) + (1-\alpha) U_m'' (H (\alpha)) \right] \widetilde{H}' (\alpha) + U_f' (H) - U_m' (H) = 0, \tag{5}$$

which yields

$$\widetilde{H}' (\alpha) = -\frac{U_f' (H) - U_m' (H)}{\alpha U_f'' (H (\alpha)) + (1-\alpha) U_m'' (H (\alpha))}. \tag{6}$$

To determine the sign of the numerator, note that from the first-order condition we have that

$$-U_m' (H) = \frac{\alpha}{1-\alpha} U_f' (H), \tag{7}$$

and thus that

$$sgn \left[\widetilde{H}' (\alpha) \right] = sgn \left[U_f' (H) - U_m' (H) \right] = sgn \left[U_f' (H) \left(1 + \frac{\alpha}{(1-\alpha)} \right) \right] = sgn \left[U_f' (H) \right]. \tag{8}$$

As reasoned above, at the optimum $U_f'(H) > 0$ because of the single-crossing property. Thus $\widetilde{H}'(\alpha) > 0$, *QED.*

A.1.2 Model with transfers

We can generalize the model to include transfers in the following way. Let q_i be an action that spouse i can take, for example housework, with marginal cost to spouse i of unity and marginal benefit to the other spouse of $\phi(q_i)$. This nests the no-transfer case if $\phi(q) = 0$. Let $\phi(0) = 0$, and assume that $\phi'(q) \in [0,1]$ and $\phi''(q) < 0$, implying that transfers involve some friction. We normalise such that at no sex, $s = 0$, both transfers are equal to zero.

The individual utility functions with sex and transfers become

$$v_i(P, H, q_i, q_{-i}) = u_i(P, H) - q_i + \phi(q_{-i}) \tag{9}$$

All other aspects of the model are kept intact.

Extensive Margin: The couple will choose $s = 1$ if and only if there exists some $(P, H, q_m, q_f) \in \{US, FC, MC\} \times \times R_+^2$ such that $v_i(P, H, q_i, q_{-i}) \geq u_i^0 \quad \forall i = m, f$. It follows that the possibility of transfers increases the likelihood that $s = 1$ compared to the no-transfer case, insofar as there are cases where $s = 1$ occurs with transfers but would not if transfers were not possible. Note that it is still the case that the choice of $s = 0$ or $s = 1$ does not depend on α.

Intensive Margin: Suppose that the above condition is satisfied and thus that $s = 1$. The unconstrained household utility maximisation problem generalises to

$$\max_{H, q_m, q_f} \{(1-\alpha)[u_m(P(H), H) - q_m + \phi(q_f)] + \alpha[u_f(P(H), H) - q_f + \phi(q_m)]\} \tag{10}$$

Due to the separable form, the first-order condition with respect to H is the same for the model without transfers, namely

$$\alpha u_{fH}'(P(H), H) + (1-\alpha) u_{mH}'(P(H), H) = 0. \tag{11}$$

Thus the unconstrained function $\widetilde{H}(\alpha)$ is preserved. In addition we now have the complementary slackness conditions

$$(1-\alpha) \geq \alpha\phi'(q_m), \tag{12}$$

and

$$(1-\alpha)\,\phi'(q_f) \leq \alpha, \tag{13}$$

implying a solution $\widetilde{q}_j(\alpha)$ for $j = m, f$. Note that $\phi'(q) \leq 1)$ implies that only one of the complementary slackness conditions can hold with equality — i.e. q_f and q_m cannot be positive at the same time — and thus transfers will only occur in one direction. Intuitively, if α is low then $q_f > 0$, and vice versa if α is high. Taken together, this gives rise to implied utilities

$$\widetilde{V}_i(\alpha) = u_i\left(P(\widetilde{H}(\alpha)), \widetilde{H}(\alpha)\right) - \widetilde{q}_i(\alpha) + \phi(\widetilde{q}_{-i}(\alpha))\; i = m, f \tag{14}$$

with $\widetilde{V}_f'(\alpha) > 0$ and $\widetilde{V}_m'(\alpha) < 0$.

However, as before, if α is low enough such that $\widetilde{V}_f(\alpha) < u_f^0$ then the female's participation constraint binds. The couple instead choose an allocation that just satisfies her participation constraint, solving

$$\max_{H,q_m,q_f}\left\{U_m(P(H),H) - q_m + \phi(q_f)\,|U_f(P(H),H) - q_f + \phi(q_m) \geq u_f^0\right\}, \tag{15}$$

with the following Lagrangean

$$L = U_m(P(H),H) - q_m + \phi(q_f) + \mu_f\left\{U_f(P(H),H) - q_f + \phi(q_m) - u_f^0\right\}. \tag{16}$$

Since the female's participation constraint failed at the unconstrained solution, it follows that the constrained solution involves a larger implicit relative weight to the woman: $\mu_f^* \geq \alpha/(1-\alpha)$. The reverse logic applies if his participation constraint fails.

Taken together, this implies that $H^*(\alpha)$ is weakly increasing in α as in the no-transfer case, but that the range of values for which it is strictly increasing (i.e. in which an interior solution $\widetilde{H}$ is chosen) is smaller than in the no-transfer case. In terms of Figure 2, as transfers become less costly, the horizontal segments of the line move closer to one another vertically, and thus the range $\alpha_H - \alpha_L$ becomes smaller.

A.1.3 The limiting case of frictionless transfers

Consider the limiting case where transfers are frictionless, such that $\phi'(\cdot)$ is constant and equal to unity. In this case we can simply refer to q as the net transfer from her to him, which is negative if on net he transfers to her. Hence the household's unconstrained optimisation problem collapses to

$$\max_{H,q}\left\{(1-\alpha)\left[u_m(P(H),H) + q\right] + \alpha\left[u_f(P(H),H) - q\right]\right\}. \tag{17}$$

It is straightforward to see that this problem has no solution, except in the knife-edge case where $\alpha = 1/2$. Taking the first-order condition with respect to q, we obtain

$$1 - \alpha - \alpha = 0. \tag{18}$$

Since generically $\alpha \neq 1/2$, the solution will involve infinite transfers in one of the two possible directions. However, this then trivially leads to the failure of the donor's participation constraint. Suppose that $\alpha < 1/2$ whereby she is the donor. In that case the couple instead solves

$$\max_{H,q} \left\{ u_m \left(P(H), H \right) + q | u_f \left(P(H), H \right) - q \geq u_f^0 \right\}, \tag{19}$$

with Lagrangean

$$L = u_m \left(P(H), H \right) + q + \mu_f^* \left[u_f \left(P(H), H \right) - q - u_f^0 \right] \tag{20}$$

Note that the first-order condition with respect to q is $1 - \mu_f^* = 0$, implying $\mu_f^* = 1$. The first-order condition with respect to H therefore implies $u'_{fH} \left(P(H), H \right) = u'_{mH} \left(P(H), H \right)$. By a corresponding analysis of the case where $\alpha < 1/2$, we obtain that, with frictionless transfers, $u'_m \left(H \right) = u'_f \left(H \right)$ characterizes the couple's choice of H for any α. That is, the choice of contraceptive technology is independent of the bargaining weight. In terms of Figure 2, we reach the limiting case where the horizontal segments of the line become completely aligned vertically, and $\widetilde{H}$ is just a constant for an value of α.

A.2 Linear versus nonlinear regression specification

The regression equations in the main body of the paper are ANCOVA (Analysis of Covariance) linear probability models (LPM) of the following form:

$$Pr \left[Y_{if1} = 1 | Y_{if0}, treat_{if}, \eta_f \right] = \alpha + \delta Y_{if0} + \beta treat_{if} + \eta_f, \tag{21}$$

where Y_{if1} is the outcome variable of interest at endline, and Y_{if0} is its value at baseline. The variable $treat_{if}$ is a dummy for being assigned to the treatment group, i.e. to receiving the programme in the first rather than the second phase. The parameter β represents the intent-to-treat effect. The parameter η_f is a facilitator fixed effect, which is included for inference since randomisation was blocked on the seventeen facilitators (Bruhn and McKenzie, 2009). Standard errors are robust to individual-level heteroskedasticity, as this was the level of randomisation (Abadie et al., 2017).

Tables 5 and 7 display the main treatment impacts of the intervention. Since the majority of outcome variables are binary indicators for whether a respondent has used a certain STI protection technology or had a sex act in a particular week, the preferred specification could have been a logit. Two reasons in favour of logit are the fact that LPM tends to produce consistently biased and inconsistent estimates of structural parameters and marginal effects.[69] For maximum likelihood estimation of the logit model both the marginal effect and the standard deviation decrease as sample sizes increase (Amemiya, 1977; Horrace and Oaxaca, 2006).[70] However, estimating marginal effects with the wrong nonlinear model can also produce inconsistent estimates. This is especially true in our case because maximum likelihood estimators (MLEs) such as logit are well-known to suffer from small-sample bias and underestimation of the true probability of rare events (Leitgöb). The degree of bias is strongly dependent on the number of cases in the less frequent of the two categories in the dependent variable. Even with a large sample, if there are a few cells with very few observations this is problematic (King and Zeng, 2001a,b, 2002). Only 2% of women in our sample report they are currently using female condoms at baseline. With intent-to-treat effects of 5-20 percentage points female condom use remains a relatively rare event. This problem is exacerbated by the fact that we require fixed effects for facilitators in the estimating equation, as these were used for stratification (Bruhn and McKenzie, 2009). If the stratification dummy is used, the number of observations per cell is substantially reduced, and we run into the problem that facilitators perfectly predict outcome variables, and estimating a treatment effect becomes impossible because there is no variation between base and endline and treatment and control. This is true regardless of the specification (LPM, logit, or probit) but LPM doesn't drop these observations, while logit does and correctly so. The resulting variation in sample sizes for different dependent variables, while using the same regression specification, creates uncertainty about the comparison of effects.

One way to correct for this bias is to use the Bias Correction Method proposed by King and Zeng (2001a) through re-estimating our results with their 'rare event logit' (relogit) estimator. Alternatively, penalized maximum likelihood estimators can be used, such as firthlogit or exact logistic regression because they deal specifically with concerns about bias due to small samples and 'separation' in logistic regression (Firth, 1993; Heinze and Schemper, 2002).[71] These estimators have the attraction of producing finite, consistent estimates of regression parameters when the maximum likelihood estimates do not even exist because of complete or quasi-complete separation.

There is a trade-off between using the correction method and the penalized estimators. The correction method by (King and Zeng, 2001a) overcorrects bias in MLEs as the sample size gets small but it does allow for clustering of standard errors (Leitgöb). Firthlogit and exact logistic regression seem unbiased and converge, even in the case of low numbers of observations when stratification dummies are used (Heinze and Schemper, 2002). A complication, however, is that it does not allow clustering of standard errors that are robust to individual level heteroskedasticity. This is required because this was the level

[69]The marginal effect implied by LPM remains consistently biased even if sample sizes increase, while the inconsistency in LPM is driven by asymptotic bias

[70]http://davegiles.blogspot.co.uk/2012/06/another-gripe-about-linear-probability.html

[71]A condition in the data in which maximum likelihood estimates become inestimable because they tend to infinity

of randomization (Abadie et al., 2017).

To demonstrate robustness of our results to the use of nonlinear specifications, the results are re-estimated with nonlinear models. Tables 5, A.1, and A.2 demonstrate the estimates from the LPM, logit, Firthlogit, and relogit estimators for the main treatment effects in the survey, respectively. The sample sizes in the logit specification vary across specifications, but are consistent for the LPM, Firthlogit, and relogit. Qualitatively the estimates for the LPM specification, and the estimated marginal effects for the logit and Firthlogit are the same, although the LPM and Firthlogit estimates appear more consistent with each other. The predicted marginal effects for the treatment effect on ever use, last 30 day use, and current use are 26.3, 11.7, and 6.7 percentage points respectively, and are significant at the one percent level. Tables 7 and A.4 show the panel fixed effects specifications for the LPM and logit, respectively. These estimates are also consistent. The weekly nature of the diary data implies that we have multiple observations per respondent, and don't face the prefect predictor problem when we re-estimate the panel fixed effects LPM model with a logit model. Nonlinear panel models do suffer from the incidental parameter problem which causes parameters to not be point-identified (Hec, 1981; Lancaster, 2002; Chernozhukov et al., 2013).

Small sample sizes do become a problem when we estimate treatment impacts conditional on the level of baseline bargaining power. Due to the perfect predictor problem MLE logit estimations don't converge. Firthlogit and relogit overcome these problems. Relogit does so by estimating the same model as standard logistic regression, but correcting the estimates for the fact that in rare events data values of one are more statistically informative than values of zero. A correction of the weighting of these values can be used to reduce the variance and make the parameter estimates more informative (King and Zeng, 2001a). Firthlogit, rather then correcting the estimate ex-post, makes a systematic correction in the mechanism that produces the maximum likelihood estimate, namely the score equation, thereby not necessarily requiring the existence of a finite estimate, making the approach especially suitable for small samples (Firth, 1993). Firthlogit, however, doesn't allow for clustering of standard errors at the level of randomisation. The lack of clustering produces insignificant estimates, although still of the same sign as the LPM. If we re-estimate the LPM without clustering of standard errors we produce qualitatively similar results as with the Firthlogit. Fortunately relogit does allow for clustering and the results are qualitatively similar to the estimates from the LPM in terms of direction and significance. For ease of interpretation the predicted marginal heterogenous treatment effect on the current use of female condoms by the five bargaining power indices is presented in Figure A.1. These predicted marginal effects are based on the estimates from the rare events logit (see Table A.5). The dashed red line is the average "intent-to-treat" effect. The blue line represents the predicted margin of the treatment effect, estimated from the interaction of treatment with the bargaining power indices, with 95% confidence intervals. For all bargaining power principal components the treatment is significantly higher for those respondents with low female bargaining power, as predicted by our model.

Table A.1: Main treatment outcomes – Logit specification

	(1) **Ever FC** Mfx / (s.e.)	(2) **Ever MC** Mfx / (s.e.)	(3) **Ever other** Mfx / (s.e.)	(4) **Last 30 days FC** Mfx / (s.e.)	(5) **Last 30 days MC** Mfx / (s.e.)	(6) **Current FC** Mfx / (s.e.)	(7) **Current MC** Mfx / (s.e.)	(8) **Current other** Mfx / (s.e.)
Treatment	0.275***	0.003	0.011	0.129	-0.050	0.165**	0.064	0.030
	(0.068)	(0.045)	(0.045)	(0.081)	(0.057)	(0.077)	(0.059)	(0.054)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓	✓	✓
Observations	172	193	218	112	227	141	227	227
Control mean endline	0.088	0.824	0.735	0.010	0.363	0.020	0.353	0.412

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms, male condoms and other modern contraceptive methods. Columns 1-3 refer to whether the respondent has ever used the method, columns 4 and 5 to whether she has used it in the last 30 days (this was only asked for condoms, not for other contraceptive methods), and columns 6 to 8 to whether she is currently using it. All regressions are logit ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table A.2: Main treatment outcomes – Firthlogit specification

	(1) **Ever FC** Mfx / (s.e.)	(2) **Ever MC** Mfx / (s.e.)	(3) **Ever other** Mfx / (s.e.)	(4) **Last 30 days FC** Mfx / (s.e.)	(5) **Last 30 days MC** Mfx / (s.e.)	(6) **Current FC** Mfx / (s.e.)	(7) **Current MC** Mfx / (s.e.)	(8) **Current other** Mfx / (s.e.)
Treatment	0.198***	-0.001	0.012	0.074*	-0.049	0.105***	0.060	0.029
	(0.043)	(0.047)	(0.048)	(0.040)	(0.058)	(0.040)	(0.058)	(0.054)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓	✓	✓
Observations	227	227	227	227	227	227	227	227
Control mean endline	0.088	0.824	0.735	0.010	0.363	0.020	0.353	0.412

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms, male condoms and other modern contraceptive methods. Columns 1-3 refer to whether the respondent has ever used the method, columns 4 and 5 to whether she has used it in the last 30 days (this was only asked for condoms, not for other contraceptive methods), and columns 6 to 8 to whether she is currently using it. All regressions are firthlogit ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table A.3: Main treatment outcomes – relogit specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ever FC	**Ever MC**	**Ever other**	**Last 30 days FC**	**Last 30 days MC**	**Current FC**	**Current MC**	**Current other**
	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)
Treatment	1.565***	-0.182	0.128	1.496	-0.355	1.616**	0.250	0.126
	(0.465)	(0.400)	(0.408)	(1.062)	(0.297)	(0.762)	(0.289)	(0.340)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓	✓	✓
Observations	227	227	227	227	227	227	227	227
Control mean endline	0.088	0.824	0.735	0.412	0.363	0.020	0.353	0.412

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms, male condoms and other modern contraceptive methods. Columns 1-3 refer to whether the respondent has ever used the method, columns 4 and 5 to whether she has used it in the last 30 days (this was only asked for condoms, not for other contraceptive methods), and columns 6 to 8 to whether she is currently using it. All regressions are relogit ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table A.4: Impacts on likelihood of sex acts per respondent week – Logit

	(1) **# Sex acts per week full endline period** β / (s.e.)	(2) **# Sex acts per week last 30 days** β / (s.e.)	(3) **# Sex acts per week last 14 days** β / (s.e.)
Treat×endline	0.131**	0.168**	0.142*
	(0.054)	(0.068)	(0.084)
Observations	863	536	367
Control mean	0.469	0.471	0.491

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are binary indicators for whether a respondent had a sex act in a particular week. "Treat×endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat×endline" is the intent-to-treat effect. All regressions are respondent-level logit fixed effects models with the respondent-week as the unit of observation. All specifications include individual fixed effects (N=56), and facilitator×endline fixed effects (N=17) for inference since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation.

Table A.5: Impacts on current use of female condoms – heterogeneity by bargaining power – Relogit

	(1)	(2)	(3)	(4)	(5)
	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)
Treatment	28471757.599***	5.063***	21.857***	1.540	6.300**
	(0.763)	(0.850)	(0.983)	(1.218)	(2.859)
Assets 1	70020910.740***				
	(0.791)				
Treat×Assets1	-7.002e+07				
	(0.000)				
Assets 2		23.603***			
		(1.523)			
Treat×Assets2		-20.773***			
		(1.753)			
Assets 3			21.840***		
			(0.404)		
Treat×Assets3			-22.168***		
			(0.593)		
Decision-making				0.452	
				(0.519)	
Treat×Decision				-1.500**	
				(0.669)	
Power dynamics					1.308*
					(0.675)
Treat×Power dynamics					-1.590**
					(0.765)
Controls	✓	✓	✓	✓	✓
Observations	201	201	201	182	182
Control mean endline	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Excluding attritters, N=206 women are in a stable relationship at baseline. N=198 are women who are in a stable relationship and have no missing values on the control variables. N=180 have non-missing values for all of the decision-making, power dynamics, and control variables. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table A.11. "Decision-making" and "Power dynamics" are the first two principal component from all the survey questions referring to these two modules, as identified in Table A.12. Dependent variables are binary indicators for current use of female condoms. All regressions are relogit ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Regressions do not include facilitator fixed effects due to loss of sample size where baseline use perfectly predicts endline use conditional on a given facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table A.6: Impacts on current use of female condoms – heterogeneity by bargaining power – Firthlogit

	(1) Mfx / (s.e.)	(2) Mfx / (s.e.)	(3) Mfx / (s.e.)	(4) Mfx / (s.e.)	(5) Mfx / (s.e.)
Treatment	0.132**	0.114**	0.172**	0.096	0.452**
	(0.067)	(0.046)	(0.083)	(0.076)	(0.116)
Assets 1	0.014				
	(0.071)				
Treat×Assets1	-0.098				
	(0.146)				
Assets 2		0.045			
		(0.048)			
Treat×Assets2		-0.056			
		(0.065)			
Assets 3			0.030		
			(0.027)		
Treat×Assets3			-0.113		
			(0.128)		
Decision-making				-0.087	
				(0.064)	
Treat×Decision				-0.009	
				(0.104)	
Power dynamics					0.162**
					(0.071)
Treat×Power dynamics					-0.177***
					(0.059)
Controls	✓	✓	✓	✓	✓
Observations	198	198	198	182	182
Control mean endline	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Excluding attritters, N=206 women are in a stable relationship at baseline. N=198 are women who are in a stable relationship and have no missing values on the control variables. N=180 have non-missing values for all of the decision-making, power dynamics, and control variables. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table A.11. "Decision-making" and "Power dynamics" are the first two principal component from all the survey questions referring to these two modules, as identified in Table A.12. Dependent variables are binary indicators for current use of female condoms. All regressions are Firthlogit ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Regressions do not include facilitator fixed effects due to loss of sample size where baseline use perfectly predicts endline use conditional on a given facilitator. Standard errors are not clustered since this is not an option with Firthlogit. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

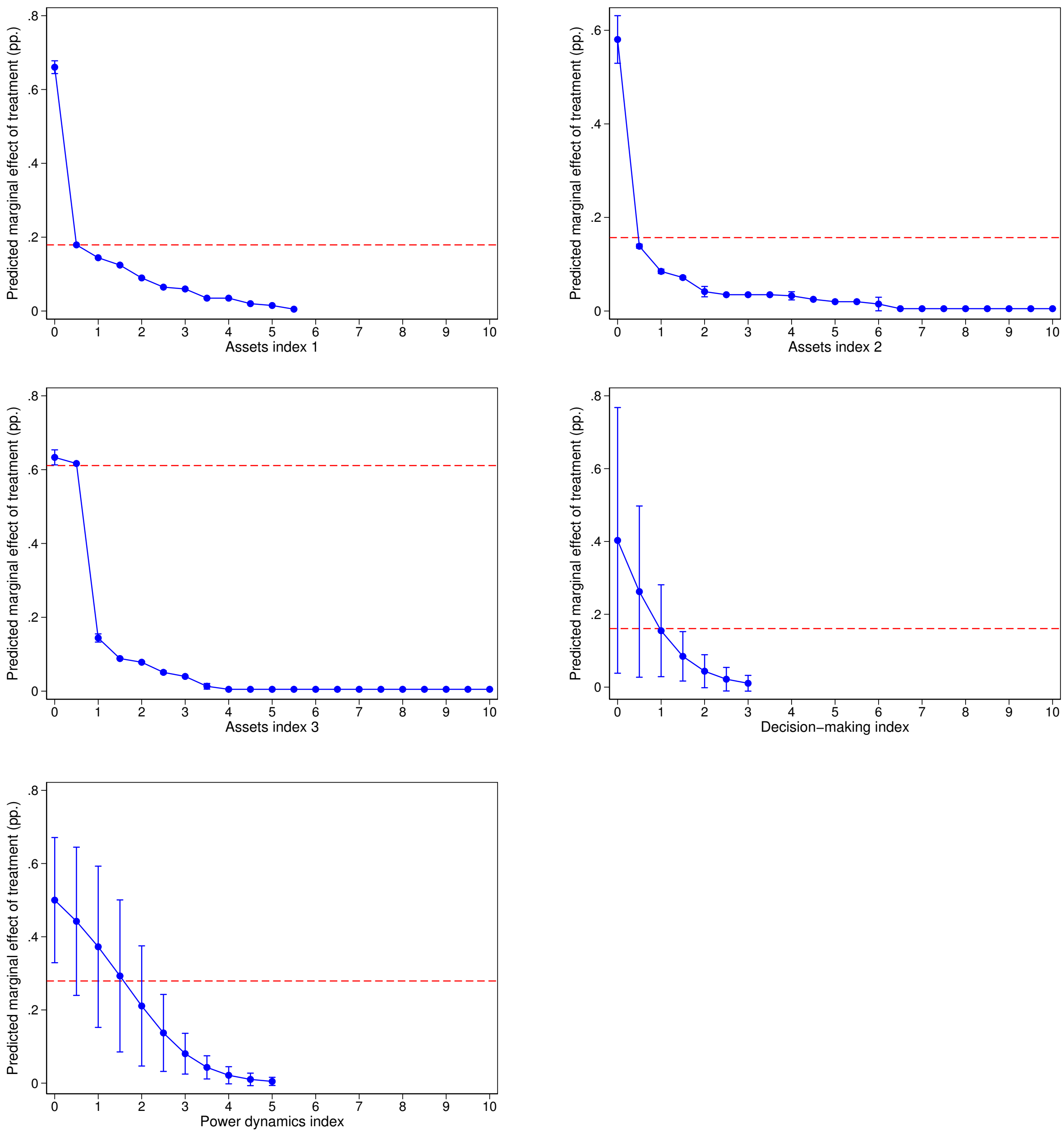
A.3 Additional Descriptive Data

Table A.7: Baseline balance on covariates – diary subsample

	Mean	Control Mean	Treatment Mean	t-test	Total N	Control N	Treatment N
Demographics							
Age	31.40	31.93	30.86	0.48	57	29	28
Education	5.91	5.45	6.39	-1.22	57	29	28
Literate	0.84	0.86	0.82	0.36	56	28	28
HH head	0.30	0.21	0.39	-1.53	57	29	28
Income							
Has job	0.37	0.38	0.36	0.17	57	29	28
Personal income last 30 days	987.72	927.59	1050.00	-0.30	57	29	28
Relationships							
In a stable relationship (incl. married)	0.82	0.83	0.82	0.06	57	29	28
Married (officially or unofficially)	0.53	0.59	0.46	0.91	57	29	28
Years relation	11.78	12.82	10.58	0.86	41	22	19
# Partners last 12 months	0.89	0.86	0.93	-0.69	57	29	28
Sexual behaviour							
Pregnant	0.00	0.00	0.00	0.00	57	29	28
HIV positive (self-report)	0.33	0.38	0.28	0.70	49	24	25
STD last 3 months (self-report)	0.12	0.12	0.12	0.05	49	24	25
Names FC as contraceptive	0.27	0.31	0.22	0.74	56	29	27

Notes: N=57 in the balanced panel. Lower sample sizes reflect observations that are missing or not applicable. “Treatment Mean” contains all individuals assigned to the treatment group, whether or not they attended the sessions. Column 4 presents the t-test statistic for the null hypothesis that the mean in the treatment group is equal to the mean in the control group.

Figure A.1: Treatment effect on current use of FC – Heterogeneity by bargaining power indices



Notes: Predicted marginal effects of treatment on the outcome variable "respondent is currently using female condoms at endline", estimated with rare events logit (see Table A.5), with 95% confidence intervals. Effects are estimated from the interaction of treatment with the bargaining power indices. Indices are constructed via PCA of the relevant survey modules, and are re-centered such that zero is the minimum and one unit represents one standard deviation. The treatment indicator is equal to one if the respondent was assigned to the treatment group, thus the estimates are "intent-to-treat" effects. The dashed red line is the average intent-to-treat effect. "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table A.11. "Decision-making" and "Power dynamics" are the first two principal component from all the survey questions referring to these two modules, as identified in Table A.12. All regressions are rare events logit ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Regressions do not include facilitator fixed effects due to loss of sample size where baseline use perfectly predicts endline use conditional on a given facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table A.8: Baseline balance on use – diary subsample

	Mean	Control Mean	Treatment Mean	t-test	Total N	Control N	Treatment N
Ever use survey							
Ever used FC	0.05	0.07	0.04	0.56	57	29	28
Ever used MC	0.77	0.79	0.75	0.38	57	29	28
Ever used other	0.84	0.86	0.82	0.41	57	29	28
Use last 30 days survey							
Used FC last 30 days	0.04	0.00	0.07	-1.44	57	29	28
Used MC last 30 days	0.35	0.28	0.43	-1.20	57	29	28
Current use survey							
Current use FC	0.02	0.00	0.04	-1.00	57	29	28
Current use MC	0.40	0.38	0.43	-0.37	57	29	28
Current use other	0.47	0.41	0.54	-0.91	57	29	28
Current use none	0.26	0.34	0.18	1.43	57	29	28
Ever use diaries							
Used FC in baseline weeks	0.04	0.03	0.04	-0.02	57	29	28
Used MC in baseline weeks	0.44	0.41	0.46	-0.38	57	29	28
Use last 30 days							
Used FC in baseline last 30 days	0.04	0.03	0.04	-0.02	57	29	28
Used MC in baseline last 30 days	0.53	0.52	0.54	-0.14	57	29	28
Use last 14 days							
Used FC in baseline last 14 days	0.04	0.03	0.04	-0.02	57	29	28
Used MC in baseline last 14 days	0.39	0.34	0.43	-0.64	57	29	28
Impact on discussions and sex acts							
% sex acts with discussion about condom use	0.16	0.13	0.18	-0.57	57	29	28
% sex acts with female-initiated discussion about condom use	0.12	0.07	0.18	-1.50	57	29	28
# sex acts per week	0.89	0.89	0.88	0.05	56	28	28

Notes: N=57 in the balanced panel. Lower sample sizes reflect observations that are missing or not applicable. "Treatment Mean" contains all individuals assigned to the treatment group, whether or not they attended the sessions. Column 4 presents the test statistic for the null hypothesis that the mean in the treatment group is equal to the mean in the control group. "Ever used other" and "Current use other" refer to use of any other modern contraceptive method apart from condoms, e.g. the pill, injectables, or an IUD. Percentage use represents the percentage of sex acts per individual which were protected by female condoms, male condoms and other contraceptives respectively.

Table A.9: Diary sample representativeness of full sample – covariates

	Mean	Survey Mean	Diary subsample Mean	t-test	Survey N	Diary subsample N
Demographics						
Age	30.95	30.74	31.92	-0.90	231	50
Years of education	6.21	6.30	5.76	1.17	230	50
Literate	0.85	0.86	0.82	0.73	229	49
HH head	0.25	0.23	0.34	-1.45	231	50
Income						
Has job	0.38	0.37	0.42	-0.61	228	50
Personal income last 30 days	871.71	816.67	1126.00	-1.27	231	50
Relationships						
In a stable relationship (incl. married)	0.85	0.85	0.84	0.22	231	50
Married (officially or unofficially)	0.60	0.62	0.50	1.50	230	50
Years relation	9.70	9.20	12.19	-1.97*	183	37
# Partners last 12 months	0.93	0.93	0.92	0.20	231	50
Sexual behaviour						
Pregnant	0.05	0.06	0.00	3.85***	230	50
HIV positive (self-report)	0.31	0.30	0.35	-0.56	201	43
STI last 3 months (self-report)	0.13	0.14	0.12	0.38	204	43
Names FC as contraceptive	0.41	0.43	0.31	1.64	229	49
Bargaining power (principle components)						
Assets 1	0.79	0.80	0.78	0.09	204	43
Assets 2	0.48	0.48	0.47	0.05	204	43
Assets 3	1.21	1.23	1.14	0.70	204	43
Decision-making	1.74	1.69	1.98	-1.58	160	34
Power dynamics	2.92	2.93	2.87	0.36	160	34

Notes: N=232 im the baseline sample of which N=57 are in the subsample who respond to the diaries. Lower sample sizes reflect observations that are missing or not applicable. “Survey Mean” contains all individuals in the balanced panel, whether or not they participated in the diaries. “Diary subsample Mean” contains just those individuals who responded to the diaries. Column 4 presents the t-test statistic for the null hypothesis that the mean in the diary subsample is equal to the mean in the full sample.

Table A.10: Diary sample representativeness of full sample – baseline use

	Mean	Survey Mean	Diaries Mean	t-test	Survey N	Diaries N
Ever use survey						
Ever used FC	0.08	0.09	0.05	0.94	297	55
Ever used MC	0.75	0.75	0.78	-0.56	297	55
Ever used other	0.74	0.72	0.85	-2.45**	297	55
Use last 30 days survey						
Used FC last 30 days	0.03	0.03	0.04	-0.35	297	55
Used MC last 30 days	0.32	0.32	0.36	-0.67	297	55
Current use survey						
Current use FC	0.02	0.02	0.02	0.27	297	55
Current use MC	0.39	0.39	0.40	-0.08	297	55
Current use other	0.41	0.39	0.49	-1.36	297	55

Notes: N=57 are in the subsample of survey respondents who respond to the diaries. N=298 are in the full baseline study sample. Lower sample sizes reflect observations that are missing. All values taken from the baseline survey. "Ever used other" and "Current use other" refer to use of any other modern contraceptive method apart from condoms, e.g. the pill, injectables, or an IUD. Column 4 presents the t-test statistic for the null hypothesis that the mean in the diary subsample is equal to the mean in the full sample.

Table A.11: Assets brought to the relationship – principal component analysis

	(PC1)	(PC2)	(PC3)	Unexplained
Assets brought to the relationship				
Jewellery			0.4091	0.2922
Livestock			0.7958	0.2609
Land		0.5964		0.4774
Electronic appliances		0.7373		0.3262
Money	0.6782			0.2543
Mobile phone	0.6587			0.2568
Kitchenware			0.4131	0.3372

Factor loadings from a PCA of all asset questions simultaneously. Only loadings greater than or equal to 0.25 are displayed. All variables are coded such that a higher value proxies greater bargaining power for the respondent. As shown, the asset questions load three separate components, all of which are used in the heterogeneity analysis.

Table A.12: Bargaining power – principal component analysis

	(PC1)	(PC2)	Unexplained
Decision-making (1= she is involved)			
Clothes	0.2728		0.6316
Phone credit	0.2678		0.6491
Children's education	0.3583		0.3933
Her health	0.3468		0.4190
Children's health	0.3288		0.4757
Her employment	0.3390		0.4288
Spending earnings	0.3216		0.5115
Visiting friends	0.3586		0.3648
Visiting family	0.3436		0.4152
In general, more influence			0.8677
In general, more power			0.9097
Power dynamics (1=completely disagree)			
We do what he wants		0.3039	0.7457
He won't let me wear certain things			0.8812
I'm quiet around him		0.3645	0.6395
He has more say about joint decisions		0.3081	0.7497
He controls who I spend time with		0.2915	0.7666
I feel trapped or stuck		0.2926	0.7293
He does what he wants		0.3173	0.7402
I'm more committed		0.2984	0.7438
He sees other people			0.8898
He wants to know where I am		0.2775	0.8013
He gets his way when we disagree			0.9082

Notes: Factor loadings from a PCA of 'decision-making' and 'power dynamics' bargaining power variables simultaneously. Only loadings greater than or equal to 0.25 are displayed. All variables are coded such that a higher value proxies greater bargaining power for the respondent.

Table A.13: Balance – principal components of bargaining power

	Mean	Control Mean	Treatment Mean	t-test	Total N	Control N	Treatment N
Assets 1	0.76	0.81	0.71	0.82	263	128	135
Assets 2	0.43	0.41	0.45	-0.33	263	128	135
Assets 3	1.18	1.24	1.13	0.94	263	128	135
Decision-making	1.79	1.83	1.75	0.60	235	114	121
Power dynamics	2.84	2.82	2.85	-0.17	235	114	121

Notes: All values taken from the baseline survey. "Assets 1", "Assets 2" and "Assets 3" represent the three principal components loaded by the assets module as identified in Table A.11. "Decision-making" and "Power dynamics" represent the principal component loaded by each of these survey modules as identified in Table A.12. All variables are coded such that a higher value proxies greater bargaining power for the respondent. All components are scaled such that the least empowered woman on that component has a score of zero. They are also normalised such that a one point increase in each component represents an increase of one standard deviation. "Treatment Mean" contains all individuals assigned to the treatment group, whether or not they attended the sessions. Column 4 presents the test statistic for the null hypothesis that the mean in the treatment group is equal to the mean in the control group. All values taken from the baseline survey (N=298). The assets module was enumerated to all women who lived with their partner at baseline, including a few who did not claim to be in a stable relationship (N=264). The decision-making module was enumerated to all respondents (N=298), except the questions "who has more influence" and "who has more power" which were asked only of women in a stable relationship at baseline (N=250). Power dynamics questions were only asked to women who were in a stable relationship at baseline (N=250). Any lower sample sizes reflect values missing or unwillingness to answer.

A.4 Additional Analyses

Table A.14: Treatment effects on female condom use, by baseline contraceptive use

	(1) **Ever FC** **No method** **at baseline**	(2) **Ever FC** **Some method** **at baseline**	(3) **Last 30 days FC** **No method** **at baseline**	(4) **Last 30 days FC** **Some method** **at baseline**	(5) **Current FC** **No method** **at baseline**	(6) **Current FC** **Some method** **at baseline**
Treatment	0.216***	0.168***	0.094**	0.023	0.125**	0.044
	(0.066)	(0.051)	(0.042)	(0.025)	(0.050)	(0.037)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	84	143	84	143	84	143
Control mean endline	0.100	0.081	0.000	0.016	0.000	0.032
$\chi^2(1) : (a) = (b)$		0.32		2.10		1.64
Pr $> \chi^2$		0.570		0.147		0.200

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms. Columns 1-2 refer to whether the respondent has ever used that method, columns 3-4 to whether she has used it in the last 30 days, and columns 5-6 to whether she is currently using it. Odd-numbered columns present results for the subsample of individuals not currently using any contraceptive at baseline; even-numbered columns present results for the subsample of individuals who are currently using some contraceptive method at baseline. The bottom two rows present chi-squared statistics and their p-values for the test that the treatment effect is the same across the two subsamples. These are obtained from seemingly unrelated estimations. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table A.15: Treatment effects – heterogeneity by relationship status

	(1) **Ever FC**	(2) **Ever MC**	(3) **Last 30 days FC**	(4) **Last 30 days MC**	(5) **Current FC**	(6) **Current MC**
Treatment	0.358***	-0.089	0.040	0.061	0.165*	0.179
	(0.103)	(0.112)	(0.054)	(0.153)	(0.088)	(0.150)
Stable relationship	0.030	-0.038	0.007	-0.052	0.024	-0.064
	(0.051)	(0.078)	(0.020)	(0.120)	(0.024)	(0.109)
Treat×Stable relationship	-0.202*	0.090	0.009	-0.132	-0.102	-0.141
	(0.109)	(0.121)	(0.064)	(0.166)	(0.093)	(0.162)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	227	227	220	221	227	227
Control mean endline	0.088	0.824	0.010	0.366	0.020	0.353

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. "Stable relationship" is a dummy equal to one if the respondent reports being in a stable relationship at baseline. Dependent variables are binary indicators for the current use of female and male condoms. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table A.16: Impacts on proportion of all sex acts protected – diary subsample

	(1) **% sex acts with FC full endline period** β / (s.e.)	(2) **% sex acts with MC full endline period** β / (s.e.)	(3) **% sex acts with FC last 30 days** β / (s.e.)	(4) **% sex acts with MC last 30 days** β / (s.e.)	(5) **% sex acts with FC last 14 days** β / (s.e.)	(6) **% sex acts with MC last 14 days** β / (s.e.)
Treat×endline	0.051**	-0.005	0.085**	0.015	0.067**	-0.004
	(0.023)	(0.087)	(0.034)	(0.119)	(0.031)	(0.135)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	349	349	204	204	143	143
Control mean	0.010	0.330	0.009	0.374	0.013	0.387

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are the number of sex acts. "Treat×endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat×endline" is the intent-to-treat effect. All regressions are linear probability models with the sex act as the unit of observation. Given that by definition there are no repeated baseline and endline observations at the level of the sex act, baseline values do not exist and so are not included as regressors. All specifications include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation.

Table A.17: Impacts on proportion of sex acts protected per respondent per week – diary subsample

	(1) **% fc-protected full endline period** β / (s.e.)	(2) **% mc-protected full endline period** β / (s.e.)	(3) **% fc-protected last 30 days** β / (s.e.)	(4) **% mc-protected last 30 days** β / (s.e.)	(5) **% fc-protected last 14 days** β / (s.e.)	(6) **% mc-protected last 14 days** β / (s.e.)
Treat×endline	0.105***	-0.065	0.105***	-0.047	0.094***	-0.222
	(0.029)	(0.104)	(0.034)	(0.122)	(0.035)	(0.144)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	398	398	259	259	179	179
Control mean	0.010	0.392	0.010	0.419	0.012	0.443

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are the percentage of sex acts protected by female or male condoms per week per respondent. "Treat×endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat×endline" is the intent-to-treat effect. All regressions are respondent-level OLS fixed effects models with the respondent-week as the unit of observation. All specifications include individual fixed effects (N=56), and facilitator×endline fixed effects (N=17) for inference since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation (Abadie et al., 2017).

Table A.18: Impacts on discussions per respondent per week – diary subsample

	(1) **% sex acts with discussion full endline period** β / (s.e.)	(2) **% sex acts with discussion last 30 days** β / (s.e.)	(3) **% sex acts with discussion last 14 days** β / (s.e.)
Treat×endline	-0.100	-0.194*	-0.356***
	(0.108)	(0.109)	(0.120)
Individual f.e.'s	✓	✓	✓
Facilitator×endline f.e.'s	✓	✓	✓
Observations	398	259	179
Control mean	0.227	0.275	0.411

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are binary indicators for discussions and female-initiated discussions about condom use at the sex act level, observed per respondent per week. "Treat×endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat×endline" is the intent-to-treat effect. All regressions are respondent level OLS fixed effects models with the respondent-week as the unit of observation. All specifications include individual fixed effects (N=56), and facilitator×endline fixed effects (N=17) for inference since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation.

Table A.19: Current use of female condoms at endline – women not using male condoms at baseline

	(1)	(2)	(3)	(4)	(5)
	β / (s.e.)	β / (s.e.)	β / (s.e.)	β / (s.e.)	β / (s.e.)
Assets 1	-0.009				
	(0.008)				
Assets 2		-0.017			
		(0.014)			
Assets 3			-0.019**		
			(0.009)		
Decision-making				-0.033	
				(0.024)	
Power dynamics					0.010
					(0.021)
Controls	✓	✓	✓	✓	✓
Observations	123	123	123	105	105
Control mean endline	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. The sample is restricted to women who report currently using male condoms at baseline, and who answer the assets module (columns 1-3) or all two survey modules on bargaining power (columns 4-5). “Assets 1”, “Assets 2” and “Assets 3” are the first three principal components from the assets module, as identified in Table A.11. “Decision-making” and “Power dynamics” are the first two principal components, as identified in Table A.12. Dependent variables are binary indicators for whether the respondent is currently using a female condom at endline. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. Controls are: respondent’s age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table A.20: Current use of female condoms at endline – women using male condoms at baseline

	(1) β / (s.e.)	(2) β / (s.e.)	(3) β / (s.e.)	(4) β / (s.e.)	(5) β / (s.e.)
Assets 1	-0.047**				
	(0.019)				
Assets 2		-0.021**			
		(0.010)			
Assets 3			-0.060**		
			(0.027)		
Decision-making				-0.117***	
				(0.044)	
Power dynamics					-0.040
					(0.043)
Controls	✓	✓	✓	✓	✓
Observations	75	75	75	75	75
Control mean endline	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. The sample is restricted to women who report currently using male condoms at baseline, and who answer the assets module (columns 1-3) or all two survey modules on bargaining power (columns 4-5). "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table A.11. "Decision-making" and"Power dynamics" are the first two principal components, as identified in Table A.12. Dependent variables are binary indicators for whether the respondent is currently using a female condom at endline. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table A.21: Impacts on current use of male condoms – heterogeneity by bargaining power

	(1) β / (s.e.)	(2) β / (s.e.)	(3) β / (s.e.)	(4) β / (s.e.)	(5) β / (s.e.)
Treatment	0.017	0.056	-0.009	0.040	0.138
	(0.076)	(0.065)	(0.106)	(0.142)	(0.181)
Assets 1	-0.060*				
	(0.031)				
Treat×Assets1	0.092*				
	(0.055)				
Assets 2		-0.045			
		(0.031)			
Treat×Assets2		0.076*			
		(0.040)			
Assets 3			-0.045*		
			(0.027)		
Treat×Assets3			0.082		
			(0.077)		
Decision-making				0.010	
				(0.049)	
Treat×Decision				0.008	
				(0.064)	
Power dynamics					0.043
					(0.043)
Treat×Power dynamics					-0.029
					(0.060)
Controls	✓	✓	✓	✓	✓
Observations	198	198	198	180	180
Control mean endline	0.353	0.353	0.353	0.353	0.353

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Excluding attritters, N=198 women are in a stable relationship at baseline. N=206 answer the assets module, including some women who live with their partner but do not report being in a stable relationship. N=160 have non-missing values for all of the decision-making and power dynamics questions (which is required for the principal components to be calculated) and all of the controls. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table A.11. "Decision-making" and "Power dynamics" are the first two principal component from each of these modules, as identified in Table A.12. Dependent variables are binary indicators for current use of male condoms. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Regressions do not include facilitator fixed effects due to loss of sample size where baseline use perfectly predicts endline use conditional on a given facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table A.22: Treatment effects – heterogeneity by baseline HIV risk beliefs

	(1) Ever FC	(2) Ever MC	(3) Last 30 days FC	(4) Last 30 days MC	(5) Current FC	(6) Current MC	(7) Risk perception for self Score 1-5
Treatment	0.208**	-0.002	0.072	0.078	0.159***	0.392***	0.427**
	(0.085)	(0.079)	(0.065)	(0.112)	(0.061)	(0.108)	(0.212)
Believes high risk to self	-0.085	0.019	-0.045	0.067	0.010	0.208**	
	(0.060)	(0.063)	(0.039)	(0.086)	(0.024)	(0.085)	
Treatment×Believes high risk to self	-0.032	-0.014	-0.034	-0.177	-0.111	-0.454***	-0.483**
	(0.097)	(0.094)	(0.068)	(0.129)	(0.068)	(0.126)	(0.236)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓	✓
Observations	227	227	227	227	227	227	226
Control mean endline	0.088	0.824	0.010	0.363	0.020	0.353	4.471

Notes: Significance levels $p < 0.10^*$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female and male condoms. "Believes high risk to self" is an indicator equal to one if the respondent's answer to the question "what is your risk of contracting HIV/AIDS in a case of unprotected sex?" was above the median on a 1-5 scale. In practice this corresponds to an answer of "5, very risky" since the median response was "4, risky". All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17), since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table A.23: Treatment effects – heterogeneity by baseline bargaining power and baseline use of other contraceptives

	(1) β / (s.e.)	(2) β / (s.e.)	(3) β / (s.e.)	(4) β / (s.e.)	(5) β / (s.e.)
Treatment	0.123**	0.109**	0.153***	0.207**	0.173
	(0.050)	(0.047)	(0.055)	(0.099)	(0.108)
Current use other (baseline)	-0.001	-0.002	-0.003	-0.026	-0.031
	(0.037)	(0.038)	(0.036)	(0.027)	(0.029)
Treat×Current use other (baseline)	0.014	0.005	0.016	0.036	0.058
	(0.077)	(0.078)	(0.077)	(0.075)	(0.078)
Treat×Assets1	-0.045**				
	(0.018)				
Assets 1					
Treat×Assets2		-0.031**			
		(0.016)			
Assets 2					
Treat×Assets3			-0.055***		
			(0.019)		
Assets 3					
Treat×Decision				-0.078**	
				(0.039)	
Decision-making					
Treat×Power dynamics					-0.038
					(0.040)
Power dynamics					
Controls	✓	✓	✓	✓	✓
Observations	198	198	198	180	180
Control mean endline	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. N=198 women are in a stable relationship at baseline. N=160 have non-missing values for all of the decision-making and power dynamics questions (required for the principal components to be calculated) and all of the controls. N=206 answer the assets module, including some women who live with their partner but do not report being in a stable relationship. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. "Decision-making" and "Power dynamics" are the first two principal component from each of these modules, as identified in Table A.12."Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table A.11. Dependent variables are binary indicators for the current use of female condoms. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Regressions do not include facilitator fixed effects due to loss of sample size where baseline use perfectly predicts endline use conditional on a given facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. "Current use other (baseline)" is a dummy equal to one if the respondent reports using any non-condom forms of modern contraception at baseline, which mainly comprises either the pill or injectables. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table A.24: Treatment effects – heterogeneity by baseline bargaining power and HIV status

	(1) β / (s.e.)	(2) β / (s.e.)	(3) β / (s.e.)	(4) β / (s.e.)	(5) β / (s.e.)
Treatment	0.089*	0.076*	0.120**	0.224**	0.205
	(0.051)	(0.043)	(0.058)	(0.089)	(0.136)
HIV positive (self-report)	0.017	0.020	0.020	0.046	0.046
	(0.042)	(0.041)	(0.042)	(0.045)	(0.046)
Treatment×HIV positive	0.084	0.086	0.083	0.066	0.034
	(0.098)	(0.098)	(0.099)	(0.101)	(0.104)
Treat×Assets1	-0.034*				
	(0.019)				
Assets 1					
Treat×Assets2		-0.024			
		(0.015)			
Assets 2					
Treat×Assets3			-0.048**		
			(0.019)		
Assets 3					
Treat×Decision				-0.087**	
				(0.042)	
Decision-making					
Treat×Power dynamics					-0.042
					(0.046)
Power dynamics					
Controls	✓	✓	✓	✓	✓
Observations	170	170	170	156	156
Control mean endline	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. N=198 women are in a stable relationship at baseline. N=160 have non-missing values for all of the decision-making and power dynamics questions (required for the principal components to be calculated) and all of the controls. N=206 answer the assets module, including some women who live with their partner but do not report being in a stable relationship. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. "Decision-making" and "Power dynamics" are the first two principal component from each of these modules, as identified in Table A.12. "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table A.11. Dependent variables are binary indicators for the current use of female condoms. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Regressions do not include facilitator fixed effects due to loss of sample size where baseline use perfectly predicts endline use conditional on a given facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. "HIV positive (self-report)" is a dummy equal to one if the respondent reports being HIV-positive at basline. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table A.25: Treatment effects – heterogeneity by baseline bargaining power and risk beliefs

	(1)	(2)	(3)	(4)	(5)
	β / (s.e.)	β / (s.e.)	β / (s.e.)	β / (s.e.)	β / (s.e.)
Treatment	0.249***	0.223***	0.279***	0.341***	0.268**
	(0.086)	(0.081)	(0.090)	(0.120)	(0.125)
Believes high risk to self	-0.029	-0.030	-0.035	-0.002	-0.007
	(0.042)	(0.043)	(0.042)	(0.035)	(0.038)
Treatment×Believes high risk to self	-0.151*	-0.145*	-0.147*	-0.137*	-0.101
	(0.081)	(0.081)	(0.081)	(0.075)	(0.079)
Treat×Assets1	-0.050**				
	(0.020)				
Assets 1	0.000				
	(0.011)				
Treat×Assets2		-0.033**			
		(0.016)			
Assets 2		0.008			
		(0.011)			
Treat×Assets3			-0.061***		
			(0.021)		
Assets 3			0.001		
			(0.005)		
Treat×Decision				-0.090**	
				(0.038)	
Decision-making				-0.015	
				(0.021)	
Treat×Power dynamics					-0.037
					(0.041)
Power dynamics					0.014
					(0.020)
Controls	✓	✓	✓	✓	✓
Observations	197	197	197	179	179
Control mean endline	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. N=198 women are in a stable relationship at baseline. N=160 have non-missing values for all of the decision-making and power dynamics questions (required for the principal components to be calculated) and all of the controls. N=206 answer the assets module, including some women who live with their partner but do not report being in a stable relationship. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. "Decision-making" and "Power dynamics" are the first two principal component from each of these modules, as identified in Table A.12. "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table A.11. Dependent variables are binary indicators for the current use of female condoms. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Regressions do not include facilitator fixed effects due to loss of sample size where baseline use perfectly predicts endline use conditional on a given facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. "Believes high risk to self" is an indicator equal to one if the respondent's answer to the question "what is your risk of contracting HIV/AIDS in a case of unprotected sex?" was above the median on a 1-5 scale. In practice this corresponds to an answer of "5, very risky" since the median response was "4, risky". Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table A.26: Lee bounds – primary outcome variables

	(1) **Ever FC**	(2) **Ever MC**	(3) **Ever other**	(4) **Last 30 days FC**	(5) **Last 30 days MC**	(6) **Current FC**	(7) **Current MC**	(8) **Current other**
Upper	0.223***	0.090	0.169*	0.066**	0.005	0.112***	0.137*	0.106
	(0.058)	(0.084)	(0.086)	(0.028)	(0.073)	(0.037)	(0.077)	(0.079)
Lower	0.045	-0.087	-0.008	-0.010	-0.173**	-0.020	-0.040	-0.071
	(0.079)	(0.061)	(0.065)	(0.010)	(0.087)	(0.014)	(0.084)	(0.084)
95% C.I. Upper bound	0.318	0.228	0.311	0.112	0.124	0.173	0.264	0.237
95% C.I. Lower bound	-0.084	-0.187	-0.115	-0.026	-0.315	-0.042	-0.178	-0.209
Proportion trimmed	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150
Observations	298	298	298	298	298	298	298	298
Control mean endline	0.088	0.824	0.735	0.010	0.363	0.020	0.353	0.412

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms, male condoms and other modern contraceptive methods. Columns 1-3 refer to whether the respondent has ever used the method, columns 4 and 5 to whether she has used it in the last 30 days (this was only asked for condoms, not for other contraceptive methods), and columns 6-8 to whether she is currently using it. Bounds do not include facilitator fixed effects, as attrition is not monotonic on treatment status conditional on facilitator fixed effects.

Table A.27: Treatment effects – survey variables, diary subsample

	(1) **Ever FC**	(2) **Ever MC**	(3) **Ever other**	(4) **Last 30 days FC**	(5) **Last 30 days MC**	(6) **Current FC**	(7) **Current MC**	(8) **Current other**
Treatment	0.249***	-0.040	0.040	0.087	-0.107	0.125*	0.069	0.068
	(0.094)	(0.093)	(0.066)	(0.059)	(0.132)	(0.068)	(0.133)	(0.108)
Observations	50	50	50	50	50	50	50	50
Control mean endline	0.088	0.824	0.735	0.010	0.363	0.020	0.353	0.412

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are binary indicators for the use of contraceptives at the level of the respondent. "Treatment" is an indicator for observations in the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. All regressions are linear probability models with the respondent as the unit of observation. Facilitator fixed effects are dropped because facilitator perfectly predicts outcomes for many observations in this subsample. Standard errors are clustered at the individual level, since this was the level of randomisation.

Table A.28: Impacts on condom use – fixed effects panel estimator, diary subsample

	(1) **FC** **full endline period** β / (s.e.)	(2) **MC** **full endline period** β / (s.e.)	(3) **FC** **last 30 days** β / (s.e.)	(4) **MC** **last 30 days** β / (s.e.)	(5) **FC** **last 14 days** β / (s.e.)	(6) **MC** **last 14 days** β / (s.e.)
Treat×endline	0.120**	-0.148	0.123**	-0.033	0.054	-0.099
	(0.054)	(0.093)	(0.050)	(0.103)	(0.045)	(0.113)
Individual f.e.'s	✓	✓	✓	✓	✓	✓
Facilitator×endline f.e.'s	✓	✓	✓	✓	✓	✓
Observations	383	383	252	252	175	175
Control mean	0.020	0.350	0.015	0.374	0.021	0.412

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are binary indicators for male and female condom use per week per respondent. "Treat×endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat×endline" is the intent-to-treat effect. All regressions are respondent level OLS fixed effects models with the respondent-week as the unit of observation. All specifications include individual fixed effects (N=57) and facilitator×endline fixed effects (N=17) for inference since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation.

Table A.29: Treatment effects – bargaining power

	Mfx	sd	p-val	N
Who decides about...				
...buying clothes for you?	-0.03	0.04	0.46	227
...buying phone credit?	0.03	0.04	0.52	227
...education for the children?	-0.03	0.04	0.46	226
...health expenses for you?	-0.10	0.04	0.01	227
...health expenses for the children?	-0.06	0.04	0.13	225
...if you are allowed to work?	-0.06	0.04	0.16	227
...how earnings are used?	-0.01	0.04	0.74	227
...visits to friends?	-0.00	0.04	1.00	226
...visits to family?	-0.01	0.05	0.80	226
Who usually has more say when you talk about serious things	0.11	0.05	0.03	177
In general, who do you think has more power in your relationship	0.11	0.05	0.02	177
Power dynamics				
Most of the time, we do what my partner wants to do	-0.03	0.05	0.45	193
My partner won't let me wear certain things	-0.01	0.05	0.82	193
When my partner and I are together, I'm pretty quiet	-0.04	0.05	0.37	193
My partner has more say about important decisions that affect us	-0.03	0.05	0.51	193
My partner tells me who I can spend time with	-0.03	0.05	0.52	193
I feel trapped or stuck in our relationship	-0.00	0.05	0.99	193
My partner does what he wants, even if I do not want him to	-0.05	0.05	0.27	193
I am more committed to our relationship than my partner is	0.04	0.05	0.34	193
My partner is involved with other people apart from me	-0.15	0.05	0.00	193
My partner always wants to know where I am	0.13	0.04	0.00	193
When my partner and I disagree, he gets his way most of the time	0.07	0.05	0.12	193

Notes: "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are bargaining power indicators measured at endline. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17), since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. N=232 for the endline survey excluding attritters.

Figure A.2: Simulation of annual number of HIV infections

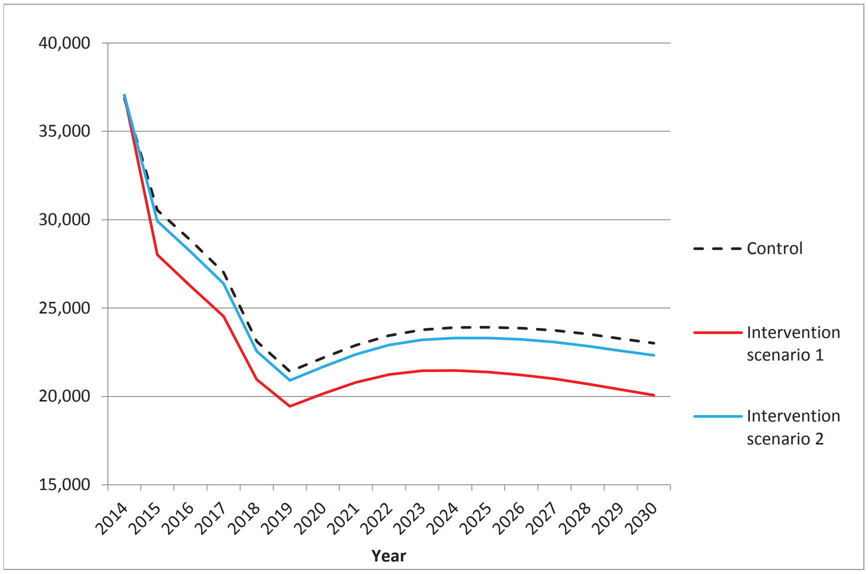


Table A.30: Treatment effects – other outcome variables

	(1) **HIV knowledge (score 0-6)**	(2) **HIV positive**	(3) **STI last 3 months**	(4) **Well-being (score 0-12)**	(5) **Violence (score 0-5)**
Treatment	-0.136*	-0.021	0.021	0.171	0.076
	(0.081)	(0.042)	(0.033)	(0.260)	(0.195)
Facilitator f.e.'s	✓	✓	✓	✓	✓
Observations	219	196	185	212	162
Control mean endline	5.758	0.313	0.054	8.135	1.149

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are as follows, all measured at endline: column 1, a score from six questions testing knowledge about how HIV can and cannot be transmitted; column 2, a self-reported dummy for HIV-positive status; column 3, a self-reported dummy for having had an STI in the last three months; column 4, a score from twelve questions on well-being (higher scores indicate greater well-being); column 5 a score from five questions about emotional and physical violence (a higher score indicates greater violence). N=232 for the endline survey excluding attritters, except the violence module where N=162 since these questions were only enumerated to women in a stable relationship. Missing observations reflect not applicable, does not want to answer, and cases where the facilitator fixed effect perfectly predicts the outcome variable. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17), since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.