Title

Evaluation of the association between the concentrations of key vaginal bacteria and the increased risk of HIV acquisition in African women from five cohorts: a nested case-control study

Journal

The Lancet Infectious Diseases, 18(5)

ISSN

1473-3099

Authors

McClelland, R Scott
Lingappa, Jairam R
Srinivasan, Sujatha
et al.

Publication Date

2018-05-01

DOI

10.1016/s1473-3099(18)30058-6

Peer reviewed

HHS Public Access

Author manuscript

Lancet Infect Dis. Author manuscript; available in PMC 2019 May 01.

Published in final edited form as:

Lancet Infect Dis. 2018 May ; 18(5): 554–564. doi:10.1016/S1473-3099(18)30058-6.

Key Vaginal Bacteria Associated with Increased Risk of HIV Acquisition in African Women: A Nested Case-Control Study

Prof R. Scott McClelland, MD[1,2,3,6,*], **Prof Jairam R. Lingappa, MD**[1,3,4,*], **Sujatha Srinivasan, PhD**[8], **John Kinuthia, MBChB**[11], **Prof Grace C. John-Stewart, MD**[2,3], **Prof Walter Jaoko, MBChB**[7], **Prof Barbra A. Richardson, PhD**[5,10], **Krista Yuhas, MSc**[3], **Tina L. Fiedler, BS**[8], **Prof Kishorchandra N. Mandaliya**[3], **Matthew M. Munch, BS**[8], **Nelly R. Mugo, MBChB**[12], **Prof Craig R. Cohen, MD**[13], **Prof Jared M. Baeten, MD**[1,2,3], **Prof Connie Celum, MD**[1,2,3], **Prof Julie Overbaugh, PhD**[9,10], and **Prof David N. Fredricks, MD**[1,8]

[1]Department of Medicine, University of Washington, Seattle, WA, USA;

[2]Department of Epidemiology, University of Washington, Seattle, WA, USA;

[3]Department of Global Health, University of Washington, Seattle, WA, USA;

[4]Department of Pediatrics, University of Washington, Seattle, WA, USA;

[5]Department of Biostatistics, University of Washington, Seattle, WA, USA;

[6]Institute of Tropical and Infectious Diseases, University of Nairobi, Nairobi, Kenya;

[7]Department of Medical Microbiology, University of Nairobi, Nairobi, Kenya;

[8]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;

[9]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;

[10]Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;

[11]Kenyatta National Hospital, Nairobi, Kenya, University of California at San Francisco, San Francisco, CA, USA

Corresponding Author: R. Scott McClelland, MD, MPH, Professor of Medicine, Epidemiology, and Global Health, University of Washington, Box 359909; 325 9th Avenue, Seattle, WA 98104, USA, Phone: 206-543-4278, Fax: 206-543-4818, mcclell@uw.edu.

CONTRIBUTORS

RSM, JRL, and DNF developed the study concept. RSM, JRL, GJS, JMB, CC, JO, and DNF secured funding for the parent P01 grant and for the individual cohorts that contributed samples and data. RSM, JRL, JK, GJS, WJ, KNM, NRM, CRC, JMB, and CC oversaw clinical field work. DNF oversaw the molecular microbiological work. SS and TLF developed the qPCR assays. TLF and MMM applied the qPCR assays to the samples. performed by TLF. SS and MMM generated and curated the pyrosequencing data. RSM, JRL, BAR, and KY developed the statistical analysis plan, and KY performed the statistical analysis. RSM and JRL developed the initial draft of the manuscript. All authors contributed to editing of the manuscript and approved the final draft for submission.

*RSM and JRL contributed equally to this work

CONFLICTS OF INTEREST

RSM receives research funding, paid to the University of Washington, from Hologic/Gen-Probe for a study of human papilloma virus screening. JRL is on the Scientific Advisory Board for Prosetta Biosciences, Inc. All other authors report nothing to disclose.

[12]Kenya Medical Research Institute, Nairobi, Kenya, University of California at San Francisco, San Francisco, CA, USA

[13]Department of Obstetrics, Gynecology & Reproductive Sciences, University of California at San Francisco, San Francisco, CA, USA

SUMMARY

Background: Disruptions of vaginal microbiota may increase women's susceptibility to HIV infection. Advances in molecular microbiology have enabled detailed examination of associations between vaginal bacteria and HIV acquisition. This prospective study utilized molecular characterization of vaginal microbiota to examine the link between vaginal bacteria and risk of HIV acquisition.

Methods: Data from five cohorts of African women including sex workers, pregnant/post-partum women, and women in serodiscordant relationships were utilized to conduct a nested case-control analysis comparing vaginal microbiota between women who acquired HIV (cases, N=87) versus women who remained seronegative (controls, N=262). First, deep sequence analysis of broad-range 16S rRNA gene polymerase chain reaction (PCR) products was applied to a subset of 55 cases and 55 controls. From these data, 20 taxa were selected for bacterium-specific quantitative PCR (qPCR) assays, which were examined in the full cohort.

Findings: Vaginal bacterial community diversity, measured by the Shannon Diversity Index, was higher in women who acquired HIV (median 1.3, interquartile range [IQR] 0.4–2.3) compared to seronegative controls (median 0.7, IQR 0.1–1.5), p=0.03. Seven taxa, *Parvimonas* species Types 1 and 2, *Gemella asaccharolytica, Mycoplasma hominis, Leptotrichia/Sneathia, Eggerthella* species Type 1, and vaginal *Megasphaera* species showed significant concentration-dependent associations with up to >4.5-fold increased odds of HIV acquisition after adjusting for potential confounding factors.

Interpretation: Vaginal microbial diversity and concentrations of key bacteria were associated with women's risk of HIV acquisition. Defining vaginal bacterial taxa associated with HIV risk could point to mechanisms influencing HIV susceptibility and provide important targets for future prevention research.

INTRODUCTION

In contrast to other parts of the world, where men account for most new infections, 56% of new human immunodeficiency virus (HIV) infections in Africa in 2015 were in women.[1] Bacterial vaginosis (BV), a condition characterized by the presence of complex anaerobic vaginal bacterial communities, may contribute to HIV transmission and the disproportionate burden of HIV in African women.[2] The specific bacteria underlying the association between BV and HIV remain poorly understood.

Advances in molecular microbiology have enhanced our understanding of normal and dysbiotic human microbiota.[3] These approaches have facilitated identification of distinct vaginal bacterial community types ranging from low-diversity, *Lactobacillus*-dominated bacterial communities to heterogeneous and highly diverse BV-associated communities

characterized by an abundance of anaerobic species.[3–5] Bacterial species vary in terms of their associations with BV,[6] with particular symptoms,[7] and with vaginal inflammation.[8] A recent study from South Africa demonstrated that young women with high-diversity vaginal bacterial communities had higher numbers of activated genital mucosal $CD4^+$ T-cell numbers and a four-fold increased risk of HIV acquisition compared to women with low-diversity *L. crispatus*-dominated communities.[9] Higher relative abundance of several bacterial taxa (*Prevotella melaninogenica, Veillonella montpellierensis, Mycoplasma, Prevotella bivia*, and *Sneathia sanguinegens*) were also associated with increased risk. One limitation of relative abundance data is that they do not provide absolute concentrations of bacteria, which can vary widely in women with the same relative abundance. The primary objective of this study was to test the hypothesis that concentrations of specific vaginal bacteria are associated with increased risk of HIV acquisition in women.

METHODS

Participants and Procedures

Participants from five cohorts in eastern and southern Africa were included in this nested case-control study. In the Mombasa Cohort,[10] women were ≥16 years old, HIV-seronegative, and self-identified as exchanging sex for cash or in-kind payment. In the Mama Salama Study,[11] women were ≥14 years old, pregnant, and HIV-seronegative either at enrollment or documented during routine antenatal care within the past three months. In three cohorts of HIV-serodiscordant heterosexual couples, the Partners in Prevention Herpes Simplex Virus (HSV)/HIV Transmission Study,[12] the Couples Observational Study,[13] and the Partners Pre-Exposure Prophylaxis (PrEP) Study,[14] couples included in the present analysis had an HIV-seropositive male partner, while the female partner was ≥18 years old and HIV-seronegative. Detailed procedures for each cohort have been published.[10–14] Each protocol received approval from country-specific and investigator-affiliated ethical review boards. Participants provided written informed consent.

Common procedures across all cohorts included enrollment with collection of demographic, medical, and sexual history data, as well as collection of vaginal samples for microbiota analyses. Sexually transmitted infections (STI) were evaluated and treated at baseline. Participants were asked to return every one to three months for HIV testing. Women who acquired HIV were defined as cases, and were compared to controls from the same cohort. Women who seroconverted for antibodies to HIV infection had plasma samples from preseroconversion visits evaluated for HIV RNA using nucleic acid amplification tests (NAAT). These data, together with HIV serology results, were used to identify the first visit with evidence of HIV infection (serum anti-HIV antibody, plasma HIV RNA, or both). Pre-HIV-infection genital samples were selected from visits at which participants were both HIV antibody negative and HIV RNA negative. All participants received risk-reduction education and free condoms. Additional details regarding procedures for each cohort are provided in Table 1.[10–14]

Laboratory

Vaginal swabs for DNA extraction and bacterial PCR were stored at −80°C and transported on dry ice to the Fred Hutchinson Cancer Research Center in Seattle, WA for analysis. See Appendix, page 1, for detailed laboratory methods.

A sequential approach was used in these experiments. First, broad-range 16S rRNA gene PCR with pyrosequencing was performed for a subset of all cases and one randomly selected control per case, available in Seattle for analysis in mid-2014. These data illustrated the overall distribution of bacterial taxa in cases versus controls. Second, relative abundance data were used to identify key bacteria to analyze using qPCR in all cases and controls.

Sample Size and Statistical Analyses

This study targeted at least 80 cases and 240 controls. Assuming α=0.05, this sample size provides >90% power to detect a ≥2.6-fold difference in the odds of detecting a vaginal bacterial taxon in cases versus controls, assuming ≥20% prevalence of the organism in controls.

The first step of the analysis, utilizing deep sequencing data from 55 cases and 55 matched controls, generated a large number of tests of association, so this step was considered to be hypothesis generating. Based on these initial comparisons, 20 bacteria were selected for directed hypothesis testing using qPCR in all cases and controls.

Pyrosequencing data were used to calculate two measures of bacterial community structure. The Chao1 Index provides an estimate of community richness, reflecting the number of different taxa.[15] The Shannon Index is a measure of diversity accounting for both the number of different taxa and the evenness of their distribution.[16] Statistics of ecological diversity and richness were calculated separately for each sample using read numbers classified to their most specific taxonomic rank, using implementations of R microbiome package (http://microbiome.github.io/microbiome/). Index values in cases and controls were compared using Wilcoxon rank-sum tests.

To identify potentially important species for quantitative analysis, unadjusted logistic regression was applied to the relative abundance data, with case status as the outcome and relative abundance percentage for each taxon separately as the exposure. To select the subset for qPCR testing, bacteria identified through pyrosequencing were ranked in descending order by score statistic, with logistic models run on each taxon in rank order of score statistic until a p-value of 0.2 was reached in univariate logistic regressions. Sixteen bacteria were identified for further study using qPCR based on the magnitude of odds ratios in logistic regression. The *Parvimonas* taxon was represented by qPCR assays to detect species Types 1 and 2, both of which were linked to HIV risk. The *Megasphaera* qPCR was a combined assay detecting vaginal *Megasphaera* Types 1 and 2, both of which have been linked to BV. The *Prevotella* qPCR was a genus-directed assay. This approach was chosen based on the pyrosequencing data, in which higher relative abundance of *P. timonensis, P. bivia*, and additional undifferentiated *Prevotella* taxa were all associated with similarly increased odds of HIV acquisition. Four additional bacteria were tested using qPCR despite lower score statistics. *Gardnerella vaginalis* was included because of its longstanding association with

BV and role in biofilm formation.[17] *Lactobacillus crispatus* and *L. jensenii* were included because of their well-recognized association with vaginal health.[4, 6] Finally, the relative abundance of *Atopobium vaginae* was significantly higher in cases versus controls in serodiscordant couples, so qPCR for this species was performed in all cohorts.

In the qPCR analyses, each bacterial taxon was analyzed as a four-category exposure including undetectable (reference category), first tertile, second tertile, and third tertile of concentrations. The one exception was the *Prevotella* genus, which had few undetectable samples, so was modeled in four quartiles. Conditional logistic regression was used to generate odds ratios (ORs) and 95% confidence intervals (CIs) testing the hypothesis that increasing quantities of the targeted bacteria were associated with increased or decreased risk for HIV acquisition. Regression models were stratified by cohort to address clustering of individuals within cohorts. The modeling approach assumed a baseline odds for each cohort-cluster. This was treated as a nuisance parameter and conditioned out of the likelihood in the conditional logistic regression models.[18] For each of 20 bacterial taxa in this primary analysis, a single joint p-value was used to assess the statistical significance of the overall association between bacterial quantity and HIV acquisition.

Potential confounders of the association between vaginal bacterial concentrations and HIV acquisition were selected *a priori* based on biologically plausible confounding effects. These included age (continuous), pregnancy and contraceptive status (categorical), number of sex partners in the past month (continuous), frequency of sex in the past month (continuous), and recent self-reported unprotected sex (binary). All potential confounders were included in a multivariable model stratified by cohort. Spearman correlation values were calculated for pairwise comparisons of bacterial concentrations.

To illustrate the difference in HIV risk associated with individual bacterial taxa compared to microscopic criteria for BV used in earlier studies, analyses were repeated using Nugent scores,[19] comparing normal microbiota (scores 0–3) to intermediate microbiota (scores 4–6), and BV (scores 7–10).

To examine the effect of the two-step experimental approach using analysis of pyrosequencing data to generate hypotheses for further investigation using qPCR assays, the association between detection of individual taxa using qPCR assays and HIV acquisition was evaluated in a validation subset of 128 women not included in the relative abundance analysis. Because the validation sample was independent of that used in the pyrosequencing analysis, it was possible to apply a Benjamini-Hochman false discovery rate of 0.20 to the 20 bacterial taxa examined. Three adjustments were made to the analytical approach because of the smaller dataset. First, this analysis used detection, rather than quantiles of each bacterial taxon. Second, confounders in the multivariable model were restricted to age, pregnancy/contraceptive status, and recent unprotected intercourse. Third, Firth logistic regression was used to generate 95%CIs and p-values if data were sparse (expected cell count <5 in a cross-tabulation of exposure and outcome).

Sensitivity analyses were performed, focusing on bacteria significantly associated with HIV acquisition in the primary analysis. First, analyses were repeated with the sex worker cohort,

pregnant/postpartum cohort, and HIV-serodiscordant couples cohorts separately to assess whether results were similar in each population. Second, since vaginal microbiota changes over time, the primary analysis was repeated after excluding cases and their matched controls where the case sample was collected >90 days prior to the first visit where HIV infection was identified. Third, the analysis was repeated after excluding cases sampled during acute or early HIV infection and their controls. Fourth, because inflammation caused by classical STIs might abrogate an effect of vaginal microbiota on HIV susceptibility mediated through an inflammatory mechanism, analyses were repeated in the subset of women without *Neisseria gonorrhoeae, Chlamydia trachomatis*, or *Trichomonas vaginalis* detected at the study visit. Fifth, because vaginal washing is a suspected risk factor for both HIV acquisition and BV,[20, 21] an analysis adjusting for vaginal washing status was performed in the subset of women in which these data were captured. Finally, analyses were repeated after stratifying by sample collection method (vaginal versus cervical swab).

Analyses were performed using IBM SPSS Statistics Version 23 (IBM, Armonk, NY), Stata version 13 (StataCorp, College Station TX), and R version 3.3.2 using the ggplot2 and RColorBrewer packages.

Role of the Funding Source

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

RESULTS

We identified 87 women who acquired HIV (cases), and 262 controls who did not (detailed enumeration in Table 1). For 72 (82.8%) cases, vaginal samples were collected a median of 141 days (interquartile range [IQR 84–250]) prior to HIV detection. For the remaining 15 cases, samples were collected during acute or early HIV infection. Baseline characteristics are presented in Table 2.

Overall vaginal bacterial community diversity in 55 cases versus 55 controls evaluated by pyrosequencing is presented in Figure S1 (Appendix, page 23). The Shannon Diversity Index was significantly higher in cases (median 1.3, IQR 0.4–2.3) compared to controls (median 0.7, IQR 0.1–1.5), p=0.03. Community richness, using the Chao1 Richness Estimator, was also higher in cases (median 38.5, IQR 14.0–59.0) compared to controls (median 25.0, IQR 10.0–54.2), though not significantly, p=0.17. Histograms illustrating Shannon and Chao1 distributions are shown in Figure S2 (Appendix, page 24).

Relative abundance of individual bacterial taxa was compared between cases and controls. Taxa that showed a statistical trend towards association with HIV acquisition (p<0.150) are shown in Table 3. Higher relative abundance of *Dialister* genus (OR 2.17, 95%CI 1.04–4.53), *Dialister* species Type 2 (OR 1.85, 95%CI 1.11–3.08), *D. microaerophillus* (OR 1.62, 95%CI 1.03–2.53) , *Gemella asaccharolytica* (OR 12.01, 95%CI 2.26–63.78) , *Eggerthella* species Type 1 (OR 2.06, 95%CI 1.23–3.45), *P. micra* (OR 3.26, 95%CI 1.17–9.07), and *Leptotrichia amnionii* (OR 2.67, 95%CI 1.26–5.65) were associated with significantly higher

odds of HIV acquisition. In contrast, higher relative abundance of *L. iners* (OR 0.54, 95%CI 0.36–0.80) was associated with significantly lower odds of acquiring HIV.

In univariate analyses of the 20 taxa selected for qPCR testing, *Parvimonas* species Type 1, *G. asaccharolytica*, *M. hominis*, *Leptotrichia/Sneathia*, *Porphyromonas* species Type 1, *Parvimonas* species Type 2, *G. vaginalis, Eggerthella* species Type 1, and *Megasphaera* showed significant associations with HIV acquisition (Table 4). Results were similar after adjustment for potential confounders, although *G. vaginalis* and *Porphyromonas* species Type 1 were no longer significantly associated with HIV acquisition (joint test $\alpha>0.05$). Four taxa, *Parvimonas* species Type 1 (tertile 1 adjusted OR [aOR] 1.67, 95%CI 0.61–4.57; tertile 2 aOR 3.01, 95%CI 1.13–7.99; tertile 3 aOR 4.64, 95%CI 1.73–12.46; joint test p=0.005), *G. asaccharolytica* (tertile 1 adjusted aOR 2.09, 95%CI 1.01–4.36; tertile 2 aOR 2.02, 95%CI 0.98–4.17; tertile 3 aOR 3.03, 95%CI, 1.46–6.30; joint test p=0.010), *M. hominis* (tertile 1 aOR 1.46, 95%CI 0.69–3.11; tertile 2 aOR 1.40, 95%CI 0.66–2.98; tertile 3 aOR 2.76, 95%CI 1.36–5.63; joint test p=0.048), and *Leptotrichia/Sneathia* (tertile 1 aOR 2.04, 95%CI 1.02–4.10; tertile 2 aOR 1.45, 95%CI 0.70–3.00; tertile 3 aOR 2.59, 95%CI 1.26–5.34; joint test p=0.046), showed associations that were strongest at the highest concentrations. A forest plot of adjusted odds ratios for the highest quantile of each taxon (Figure 1), illustrates how some species typically associated with vaginal dysbiosis and BV were strongly associated with HIV acquisition, while others were not. Importantly, this analysis identified individual bacterium-specific associations despite strong correlations between many bacteria included in the qPCR analyses (Figure 2).

In the validation subset, after adjustment for potential confounding factors and a 20% false discovery rate, detection by qPCR was associated with higher odds of HIV acquisition for five of the seven bacteria that demonstrated concentration-related associations in the primary analysis (Table S3, Appendix, page 11): *M. hominis* (aOR 2.71 95%CI 1.13–6.49), *Eggerthella* species Type 1 (aOR 2.50, 95%CI 1.07–5.85), *Leptotrichia/Sneathia* (aOR 2.47, 95%CI 0.98–6.22), *G. asaccharolytica* (aOR 2.45, 95%CI 1.04–5.78), and *Parvimonas* species Type 2 (aOR 2.43, 95%CI 1.03–5.70).

To compare the risk of HIV acquisition associated with the concentrations of individual bacterial taxa to the risk of HIV acquisition associated with microscopically identified vaginal dysbiosis (normal, intermediate, or BV),[19] the Nugent scoring system was applied to 76 cases and 230 controls with vaginal Gram stains available. Compared to women with normal microbiota, those with intermediate microbiota (aOR 2.50, 95%CI 1.15–5.40) and BV (aOR 2.10, 95%CI 1.14–3.88) had increased risk for HIV acquisition (joint test p=0.018).

Several sensitivity analyses were applied to the seven bacterial taxa associated with HIV acquisition in our primary multivariable analysis. Associations were similar in sex workers, pregnant and postpartum women, and serodiscordant couples (Table S4, Appendix, page 13), despite the differences in demographics, risk factors, and incidence observed in these cohorts (Table 1). In addition, analyses limited to cases where samples were collected <90 days prior to identification of HIV infection, or excluding cases with samples from acute HIV infection, or excluding cases with STIs (when available), or incorporating adjustment

for vaginal washing, or stratifying by swab collection method (vaginal versus cervical), had similar point estimates compared to the primary analyses (Tables S5-S9, Appendix, pages 14–21). Of note, many sensitivity analyses had much lower statistical power compared to the primary analyses, because they utilized smaller subsets of the data.

DISCUSSION

Using two sophisticated bacterial PCR approaches and data from five cohorts spanning six sub-Saharan African countries, this study was the first to demonstrate significant associations between the quantity of specific vaginal bacteria and women's risk of HIV acquisition. Concentrations of *Parvimonas* species Types 1 and 2, *G. asaccharolytica, M. hominis, Leptotrichia/Sneathia, Eggerthella* species Type 1, and *Megasphaera* were significantly associated with HIV risk. There were strong correlations between concentrations of many of the 20 bacteria evaluated with qPCR, suggesting they may frequently be found together, establishing high-risk bacterial communities. These findings were remarkably consistent across three distinct risk groups including sex workers, pregnant and post-partum women, and women in serodiscordant relationships.

Vaginal microbiota could influence women's risk of HIV acquisition at multiple levels.[22] First, genital inflammation, mediated by the presence of particular bacterial taxa or communities, is likely to influence HIV susceptibility.[9, 23] A recent study identified six vaginal bacterial genera independently associated with pro-inflammatory cytokines.[5] Two of these, *Sneathia* and *Gemella*, correspond to vaginal bacteria showing concentration-dependent associations with HIV risk in the present analysis. Second, vaginal dysbiosis has been associated with HIV inducing factors in vaginal fluid.[24] Third, many bacteria associated with BV produce sialidases and mucinases that disrupt the protective cervicovaginal mucus layer.[7]

In one earlier study, the presence of cultivable *Lactobacillus* species was associated with lower risk of acquiring HIV.[25] In addition, a recent study using molecular characterization of vaginal microbiota found that women with vaginal bacterial communities deficient in non-*iners* species of *Lactobacillus* were at increased risk for HIV infection.[9] The present analysis demonstrated an association between lower relative abundance of *L. iners* and HIV acquisition. However, the primary analysis using qPCR assays in the full dataset demonstrated no significant associations between concentrations of *Lactobacillus* species (*L. iners, L. crispatus,* and *L. jensenii*) and HIV acquisition. Given the negative correlations between concentrations of lactobacilli and BV-associated species, low relative abundance of *Lactobacillus* species may simply reflect the presence of high concentrations of BV-associated bacteria that impact HIV susceptibility.

This study used a novel sequential experimental approach that employed complementary methods to evaluate the vaginal microbiome. In the first step, relative abundance data were utilized to demonstrate the association between vaginal bacterial diversity and HIV acquisition, and to guide selection of a restricted set of bacteria for further investigation. In the second step, highly sensitive taxon-directed qPCR assays were used to test the hypothesis that concentrations of 20 key bacteria would be associated with HIV risk. The

two steps captured related but distinct exposures. Specifically, relative abundance is not the same as absolute quantity of a bacterial taxon. In addition, the qPCR assays are more sensitive, but may be less specific for detection of individual bacteria compared to broad range PCR with pyrosequencing.

This study included a validation analysis in a subset of samples that were not included in the pyrosequencing step, facilitating independent testing of hypotheses generated using pyrosequencing data. Two important points should be considered in comparing the primary analysis to the validation analysis. First, because of the smaller sample size in the validation, bacterial taxa were modeled in a binary fashion, rather than as four quantiles. Second, the validation subset was not a randomly generated group of cases and controls, so distribution across the five cohorts diverged from that of the full dataset. Despite these caveats, the validation confirmed an association between bacterial taxa and HIV acquisition for five of the seven bacteria identified in the primary analysis.

An important strength of this study was the large and geographically diverse sample, with individuals representing three distinct risk groups. A further strength in the study design was the collection of vaginal microbiota samples prior to HIV acquisition in >80% of cases, and shortly after HIV acquisition in the remainder. In addition, the analyses were robust in multiple sensitivity analyses testing assumptions in the experimental approach.

This study also had limitations. First, as an observational study, these analyses do not provide definitive evidence that the associations detected are caused by bacteria increasing HIV susceptibility. Second, these analyses did not explore mechanisms through which individual bacterial taxa may increase HIV risk. Such mechanistic data will help to further evaluate the likelihood of a causal link between vaginal bacteria and HIV susceptibility, and will be the focus of future studies. Third, despite adjustment for potential confounding factors, residual confounding is possible due to measurement error or unmeasured confounding factors. Fourth, longer intervals between sample collection and HIV acquisition could attenuate the observed associations,[26] although this source of variability was minimized by avoiding sample collection during menses. Fifth, sampling methods and laboratory procedures varied across the five cohorts included in this analysis. Related to this point, laboratory data on STIs and vaginal yeast were not available at all analysis visits, so the primary analysis does not include adjustment for these conditions. Finally, it should be noted that while multiple risk groups were included, all participants were from eastern and southern Africa. While this represents the region hardest hit by the HIV epidemic, the findings may not be generalizable to all geographic regions.

Higher diversity vaginal bacterial communities not dominated by lactobacilli are more common in African and Hispanic women compared to Caucasians and Asians,[4, 27] leading to the hypothesis that racial differences in vaginal microbiota may contribute to population-level differences in HIV transmission and prevalence.[28] Underscoring this point, recent studies suggest that vaginal dysbiosis accounts for 20–30% of the population attributable risk percent of HIV acquisition in African women.[29, 30] Because BV is an extremely heterogeneous condition,[3, 4] defining individual vaginal bacteria that are associated with

HIV risk in women could provide more specific targets and inform future strategies for HIV prevention research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of our clinical, laboratory, and administrative staff in Kenya, Uganda, Tanzania, Zambia, Botswana, South Africa, and the US, whose outstanding work has contributed to the success of this collaborative effort. We also acknowledge the time, effort, and commitment of the women in the cohorts that contributed to this analysis, whose participation made this research possible.

Funding: This work was supported by the National Institute of Child Health and Human Development of the National Institutes of Health (NIH; grant P01-HD64915). Support for the Mombasa Cohort was through the National Institute of Allergy and Infectious Diseases of the NIH (grant R37-AI38518). Support for the clinical trials in HIV-serodiscordant couples was through the Bill & Melinda Gates Foundation (grants OPP47674, 26469, and 41185). Infrastructure and logistical support was received from the University of Washington & Fred Hutchinson Cancer Research Center's Center for AIDS Research (NIH grant P30-AI27757). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. UNAIDS. Global AIDS Update. Geneva, Switzerland, 2016.
2. Low N, Chersich MF, Schmidlin K, et al. Intravaginal practices, bacterial vaginosis, and HIV infection in women: individual participant data meta-analysis. PLoS Med 2011; 8(2): e1000416. [PubMed: 21358808]
3. Fredricks DN, Fiedler TN, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. N Engl J Med 2005; 353: 1899–911. [PubMed: 16267321]
4. Ravel J, Gajer P, Abdo Z, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A 2011; 108 Suppl 1: 4680–7. [PubMed: 20534435]
5. Anahtar MN, Byrne EH, Doherty KE, et al. Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. Immunity 2015; 42(5): 965–76. [PubMed: 25992865]
6. Fredricks DN, Fiedler TL, Thomas KK, Oakley BB, Marrazzo JM. Targeted PCR for detection of vaginal bacteria associated with bacterial vaginosis. J Clin Microbiol 2007; 45(10): 3270–6. [PubMed: 17687006]
7. Srinivasan S, Hoffman NG, Morgan MT, et al. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. PLoS One 2012; 7(6): e37818. [PubMed: 22719852]
8. Srinivasan S, Morgan MT, Fiedler TL, et al. Metabolic signatures of bacterial vaginosis. mBio 2015; 6(2).
9. Gosmann C, Anahtar MN, Handley SA, et al. Lactobacillus-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young South African women. Immunity 2017; 46: 1–9. [PubMed: 28099858]
10. McClelland RS, Richardson BA, Cherutich P, et al. A 15-year study of the impact of community antiretroviral therapy coverage on HIV incidence in Kenyan female sex workers. AIDS 2015; 29(17): 2279–86. [PubMed: 26237099]
11. Kinuthia J, Drake AL, Matemo D, et al. HIV acquisition during pregnancy and postpartum is associated with genital infections and partnership characteristics. AIDS 2015; 29(15): 2025–33. [PubMed: 26352880]
12. Celum C, Wald A, Lingappa JR, et al. Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. N Engl J Med 2010; 362(5): 427–39. [PubMed: 20089951]

13. Lingappa JR, Petrovski S, Kahle E, et al. Genomewide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure. PLoS One 2011; 6(12): e28632. [PubMed: 22174851]
14. Baeten JM, Donnell D, Ndase P, et al. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N Engl J Med 2012; 367(5): 399–410. [PubMed: 22784037]
15. Chao A Nonparametric estimation of the number of classes in a population. Scand J Stat 1984; 11(4): 265–70.
16. Shannon CE. A mathematical theoury of communication. Bell Syst Tech J 1948; 27: 379–423, 623–56.
17. Swidsinski A, Mendling W, Loening-Baucke V, et al. An adherent Gardnerella vaginalis biofilm persists on the vaginal epithelium after standard therapy with oral metronidazole. Am J Obstet Gynecol 2008; 198(1): 97 e1–6. [PubMed: 18005928]
18. Langholz B, Goldstein L. Conditional logistic analysis of case-control studies with complex sampling. Biostatistics 2001; 2(1): 63–84. [PubMed: 12933557]
19. Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. J Clin Microbiol 1991; 29(2): 297–301. [PubMed: 1706728]
20. McClelland RS, Lavreys L, Hassan WM, Mandaliya K, Ndinya-Achola JO, Baeten JM. Vaginal washing and increased risk of HIV-1 acquisition among African women: a 10-year prospective study. AIDS 2006; 20(2): 269–73. [PubMed: 16511421]
21. McClelland RS, Richardson BA, Graham SM, et al. A Prospective Study of Risk Factors for Bacterial Vaginosis in HIV-1-Seronegative African Women. Sex Transm Dis 2008.
22. Schellenberg JJ, Plummer FA. The Microbiological Context of HIV Resistance: Vaginal Microbiota and Mucosal Inflammation at the Viral Point of Entry. International journal of inflammation 2012; 2012: 131243. [PubMed: 22506135]
23. Passmore JA, Jaspan HB, Masson L. Genital inflammation, immune activation and risk of sexual HIV acquisition. Curr Opin HIV AIDS 2016; 11(2): 156–62. [PubMed: 26628324]
24. Cohn JA, Hashemi FB, Camarca M, et al. HIV-inducing factor in cervicovaginal secretions is associated with bacterial vaginosis in HIV-1-infected women. J Acquir Immune Defic Syndr 2005; 39(3): 340–6. [PubMed: 15980696]
25. Martin HL, Richardson BA, Nyange PM, et al. Vaginal lactobacilli, microbial flora, and risk of human immunodeficiency virus type 1 and sexually transmitted disease acquisition. J Infect Dis 1999; 180(6): 1863–8. [PubMed: 10558942]
26. Srinivasan S, Liu C, Mitchell CM, et al. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. PLoS One 2010; 5(4): e10197. [PubMed: 20419168]
27. Zhou X, Brown CJ, Abdo Z, et al. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. The ISME journal 2007; 1(2): 121–33. [PubMed: 18043622]
28. Buve A, Jespers V, Crucitti T, Fichorova RN. The vaginal microbiota and susceptibility to HIV. AIDS 2014; 28(16): 2333–44. [PubMed: 25389548]
29. van de Wijgert JH, Morrison CS, Brown J, et al. Disentangling Contributions of Reproductive Tract Infections to HIV Acquisition in African Women. Sex Transm Dis 2009; 36(6): 357–64. [PubMed: 19434010]
30. Masese L, Baeten JM, Richardson BA, et al. Changes in the contribution of genital tract infections to HIV acquisition among Kenyan high-risk women from 1993 to 2012. AIDS 2015; 29(9): 1077–85. [PubMed: 26125141]
31. Gorgos LM, Sycuro LK, Srinivasan S, et al. Relationship of Specific Bacteria in the Cervical and Vaginal Microbiotas With Cervicitis. Sex Transm Dis 2015; 42(9): 475–81. [PubMed: 26267872]

RESEARCH IN CONTEXT

Evidence before this study

A PubMed search was performed on November 22, 2017, using the search terms ((((vagina*) AND bacteria*) AND HIV) AND acquisition) NOT review, without any date or language restrictions. The search returned 80 articles, of which nine addressed the hypothesis that vaginal microbiota may influence women's risk of acquiring HIV. All five prospective cohort studies in this group found that bacterial vaginosis (BV) was associated with increased risk of HIV acquisition. A single recent cohort study from South Africa used broad range bacterial PCR with deep sequencing to characterize the vaginal microbiome. In this cohort, young women with high-diversity vaginal bacterial communities had increased risk of HIV acquisition compared to women with low-diversity *Lactobacillus crispatus*-dominated communities. Higher relative abundances of *Prevotella melaninogenica, Veillonella montpellierensis, Mycoplasma, P. bivia,* and *Sneathia sanguinegens* were also associated with increased risk. Individual bacterial species were not measured using quantitative PCR methods that are more sensitive compared to deep sequencing approaches, and allow for assessment of absolute concentrations. Further research is needed to clarify the relationship between individual bacterial quantities and women's risk of HIV acquisition.

Added value of this study

This was a large nested case-control study of women from diverse regions within Africa, and representing three important risk groups; female sex workers, HIV-negative women in serodiscordant couples, and pregnant and postpartum women. State-of-the-art molecular techniques including broad-range 16S ribosomal RNA (rRNA) gene polymerase chain reaction (PCR) with pyrosequencing and taxon-directed quantitative PCR (qPCR) assays were applied sequentially to test the hypothesis that concentrations of specific vaginal bacteria are associated with increased risk of HIV acquisition in African women. This study is the first to demonstrate significant associations between the concentrations of specific vaginal bacteria and women's risk of HIV acquisition. *Parvimonas* species Types 1 and 2, *Gemella asaccharolytica, Mycoplasma hominis, Leptotrichia/Sneathia, Eggerthella* species Type 1, and vaginal *Megasphaera* species were significantly associated with up to >4.5-fold increases in HIV risk. Ten other vaginal bacterial taxa typically associated with vaginal dysbiosis, including *Gardnerella vaginalis* and *Atopobium vaginae,* showed weaker and non-significant associations with HIV acquisition. Three *Lactobacillus* species were associated with small and non-significant reductions in HIV risk. The use of broad-range PCR with deep sequencing to identify bacterial taxa for further study, followed by taxon-directed qPCR to test hypotheses related to individual bacterial taxa, illustrates the value of this sequential experimental approach in microbiome studies.

Implications of all the available evidence

Concentrations of key vaginal bacteria were strongly associated with women's risk of acquiring HIV. High concentrations of some bacteria were substantially more predictive of HIV risk compared to a microscopic diagnosis of BV. Key bacteria could increase HIV

susceptibility through multiple potential pathways including inflammation, production of HIV inducing factors, and disruption of physical and chemical barriers to infection. Defining vaginal bacterial taxa associated with HIV risk could point to mechanisms influencing HIV susceptibility and provide important targets for future prevention research.

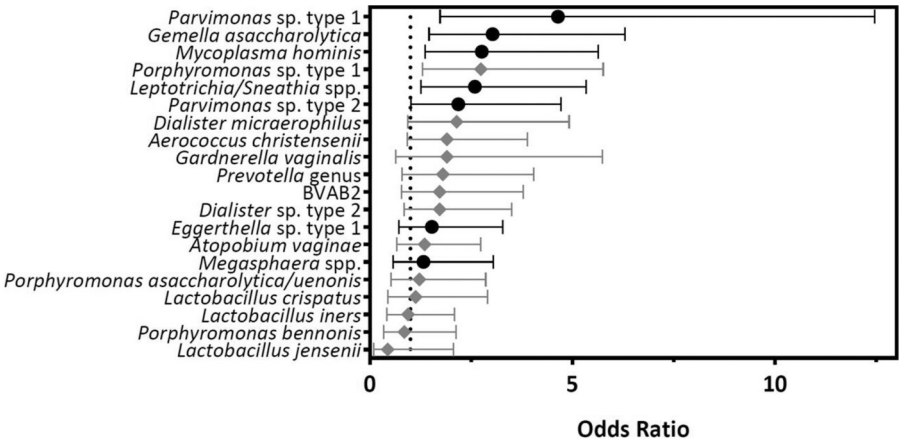

Figure 1: Forest plot showing point estimate and 95% confidence interval for the association between the highest bacterial quantile and HIV acquisition for each of 20 bacterial taxa

BVAB2, bacterial vaginosis associated bacterium 2; spp., species.

Each bacterial taxon was analyzed using conditional logistic regression in a multivariable model as a four category variable including undetectable, first tertile, second tertile and third tertile with one exception. *Prevotella* genus was modeled as four quartiles as it had few samples with bacterial DNA levels that were not detected. This plot shows the association between the highest quantile of bacterial concentration and HIV acquisition for each taxon. Adjusted odds ratios with 95% confidence intervals are shown. Bacteria that showed a significant concentration-dependent association with HIV acquisition after adjustment for potential confounding factors are shown in black, with point estimates illustrated by circles. Those bacteria that did not show an overall concentration-dependent association with HIV acquisition across the four quantiles are shown in grey, with point estimates shown as diamonds.

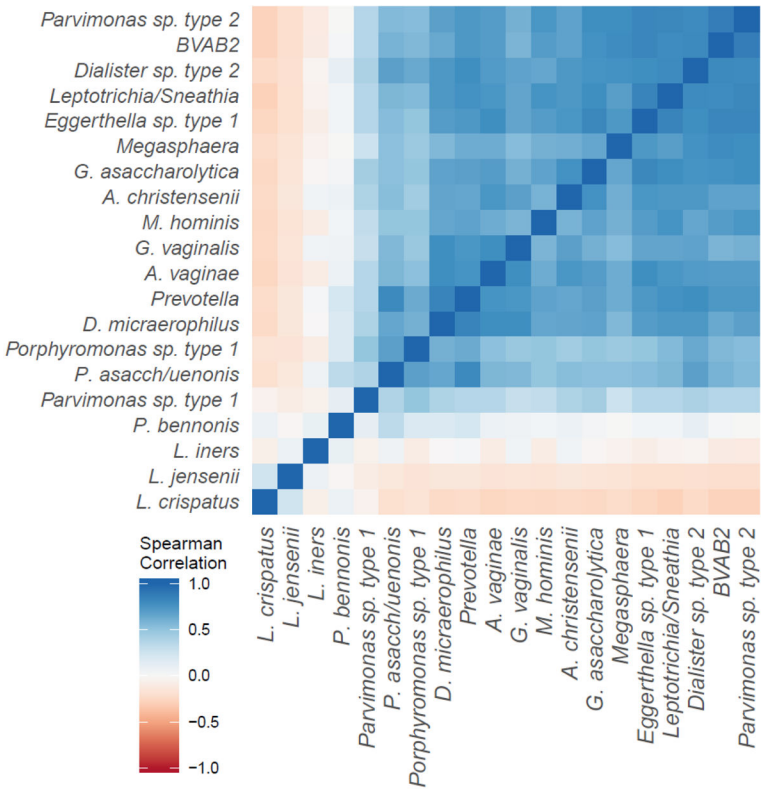

Figure 2: Heat map illustrating Spearman's correlation for quantity of 20 bacterial taxa in the cohort of 349 women from eastern and southern Africa

BVAB2, bacterial vaginosis associated bacterium 2; sp., species.

Table 1:

Summary of procedures and selection of cases and controls for five cohorts of African women

Cohort	Study Dates	Countries	Risk Group	Study Design	Visit Interval	Sample Timing and Type	Case Identification	Control Selection	Cases (n)	Controls (n)
Mombasa Cohort[10]	May 2010 to August 2014	Kenya	Female sex workers	Prospective cohort study	Monthly	Enrollment & monthly vaginal dry polyester swab *	Screening ELISA with positive results confirmed by second ELISA †	Incidence density sampling‡	10	30
Mama Salama Study[11]	May 2011 to August 2014	Kenya	Pregnant & postpartum women	Prospective cohort study	One to three months§	Enrollments & 1–3 monthly vaginal dry polyester swab *	Transcription mediated amplification‖	Incidence density sampling‡	28	84
Partners in Prevention HSV/HIV Transmission Study[12]	November 2004 to October 2008	Botswana Kenya South Africa Tanzania Uganda Zambia	HIV-serodiscordant couples	Phase III clinical trial¶	Three months	Enrollment cervical swab in media **	HIV rapid assay with positive results confirmed by ELISA, and in batch by HIV Western blot ††	Frequency matched by cohort‡‡	13§§	39
Couples Observational Study[13]	August 2007 to January 2010	Uganda South Africa	HIV-serodiscordant couples	Prospective cohort study	Three months	Enrollment cervical swab in media **	HIV rapid assay with positive results confirmed by ELISA, and in batch by HIV Western blot ††	Frequency matched by cohort‡‡	8§§	24
Partners PrEP Study[14]	July 2008 to December 2012	Kenya Uganda	HIV-serodiscordant couples	Phase III clinical trial¶¶	Monthly	Enrollment and annual vaginal dry polyester swab ¶¶	HIV rapid assay with positive results confirmed by ELISA, and in batch by HIV Western blot ††	Frequency matched by cohort‡‡	28§§***	85***

ELISA, enzyme linked immunoassay; HIV, human immunodeficiency virus; PrEP, pre-exposure prophylaxis. In all cohorts, the exposure variable for the primary analyses was the quantity of individual bacterial taxa detected using quantitative polymerase chain reaction assays.

*Dry, large-bulb polyester (Dacron) tipped push-off swab (Spin-Eze, Fitzco, Spring Park, Minnesota). Swabs were placed on ice (Mombasa Cohort) or in liquid nitrogen dry shippers (Mama Salama Study) immediately after collection, and transported to a central laboratory for inventory and storage at −80°C.

†Pishtaz HIV1,2 ELISA (Pishtaz Teb Diagnostics, Tehran, Iran) was used as the initial screening test, and all positives were confirmed with a the Vironostika HIV Uni-Form II Ag/Ab ELISA (bioMerieux, Marcy l'Etoile, France).

‡For each case, we selected three controls from the same cohort who remained HIV-negative, matched in calendar time to the cases, by randomly selecting from among all HIV-negative participants who contributed a follow-up visit with a vaginal swab for microbiota analyses within +/− 28 days of the case's exposure swab date. In addition to cases in women who were HIV-uninfected at enrollment, the Mombasa Cohort (N=2) and the Mama Salama Study (N=3) included women defined as having acute HIV infection (positive NAAT but a negative plasma antibody or rapid HIV test at enrollment). The Mama Salama Study also included enrollment seroconverters with a documented negative HIV rapid test <3 months prior to enrollment, and both a positive rapid test and positive NAAT for HIV at enrollment (N=10). Because all of these women were newly diagnosed, none were on antibiotic prophylaxis. Sensitivity analyses were performed in which these cases were excluded, as detailed in the methods and results sections of the manuscript.

§Pregnant women were enrolled at any stage of gestation. During pregnancy, they were asked to return at 20, 24, 32, and 36 weeks gestation. Following delivery, they were asked to return at 2, 6, 10, and 14 weeks, and at 6 and 9 months post-partum.

‖At each visit, testing for HIV was performed using the first-generation Gen-Probe HIV viral load assay (Hologic/Gen-Probe, San Diego, CA).

[¶] Herpes suppression with acyclovir did not significantly reduce HIV transmission risk compared to placebo, so women in both trial arms were eligible for the present analysis.

[**] Gen-Probe cervical collection swab sample collected in Gen-Probe medium (Hologic/Gen-Probe, San Diego, California). Female genital microbiome studies provide similar results with cervical and vaginal swabs.[31] Swabs were placed on ice after collection, and transported to a central laboratory for inventory and storage at −80°C.

[††] Dual rapid HIV antibody tests were performed at the clinic and confirmed with HIV EIA. HIV serostatus at enrollment visits for all participants and at follow-up visits for HIV seroconverters was evaluated using Western blot (Genetics Systems™ HIV, Bio-Rad laboratories, Hercules, California) at the University of Washington in batch at the end of each study.

[‡‡] For each HIV-serodiscordant couples cohort, a pool of non-seroconverting controls was identified that was frequency-matched by gender and site to represent the distribution of the enrolled study cohort. Controls for this study were then selected at random from each cohort's set of controls to match a 3:1 ratio of controls to cases.

[§§] For the Partners in Prevention HSV/HIV Study, Couples Observational Study, and Partners PrEP study, where vaginal samples were collected less frequently, cases were restricted to women who had a swab collected within 12 study months prior to seroconversion.

[‖‖] Participants randomized to PrEP exhibited a 67–75% reduction in HIV acquisition risk compared to those randomized to placebo, so only women in the placebo arm were eligible for the present analysis.

[¶¶] Copan Flocked swab (FloqSwabs™, Copan Diagnostics, Murrieta, CA). Swabs were placed on ice after collection, and transported to a central laboratory for inventory and storage at −80°C.

[***] From the initial set of 29 cases and 87 controls for the Partners PrEP Study, one participant initially identified to be a case was not confirmed on HIV testing performed at the conclusion of the trial.[14] This putative case was removed from the analysis. Because the HIV-serodiscordant couples cohorts used frequency matching of controls, no controls were removed after dropping that case. However, two controls were dropped for lack of a swab at the relevant visit.

Table 2:

Baseline characteristics of 349 women from eastern and southern Africa

		Full Cohort		Deep Sequencing Subset	
	Combined (N=349) Median (IQR) or n(%)	**Cases (N=87) Median (IQR) or n(%)**	**Controls (N=262) Median (IQR) or n(%)**	**Cases (N=55) Median (IQR) or n(%)**	**Controls (N=55) Median (IQR) or n(%)**
Age (years)	28 (22, 35)	26 (22, 30)	29 (23, 36)	26 (22, 30)	29 (20, 35)
Nationality					
Kenya	210 (60.2)	53 (60.9)	157 (59.9)	35 (63.6)	35 (63.6)
Uganda	112 (32.1)	27 (31.0)	85 (32.4)	20 (36.4)	19 (34.6)
South Africa	13 (3.7)	3 (3.5)	10 (3.8)	---	---
Tanzania	6 (1.7)	---	6 (2.3)	---	1 (1.8)
Botswana	5 (1.4)	1 (1.2)	4 (1.5)	---	---
Zambia	3 (0.9)	3 (3.5)	---	---	---
Education (years)	8 (6, 10)	8 (7, 10)	8 (6, 10)	8 (6, 10)	7 (5, 9)
Married	265 (75.9)	66 (75.9)	199 (76.0)	44 (80.0)	43 (78.2)
Enrollment cohort					
Mombasa Cohort	40 (11.5)	10 (11.5)	30 (11.5)	7 (12.7)	7 (12.7)
Mama Salama Study	112 (32.1)	28 (32.2)	84 (32.1)	20 (36.4)	20 (36.4)
Partners in Prevention HSV/HIV Transmission Study	52 (14.9)	13 (14.9)	39 (14.9)	---	---
Couples Observational Study	32 (9.2)	8 (9.2)	24 (9.2)	---	---
Partners PrEP Study	113 (32.4)	28 (32.2)	85 (32.4)	28 (50.9)	28 (50.9)
Pregnancy and contraception status					
Not pregnant, no modern contraception	173 (49.6)	35 (40.2)	138 (52.7)	17 (30.9)	31 (56.4)
Not pregnant, oral contraceptive	18 (5.2)	7 (8.1)	11 (4.2)	5 (9.1)	2 (3.6)
Not pregnant, DMPA	55 (15.8)	18 (20.7)	37 (14.1)	13 (23.6)	8 (14.6)
Not pregnant, IUD	8 (2.3)	2 (2.3)	6 (2.3)	2 (3.6)	2 (3.6)
Not pregnant, implant	18 (5.2)	5 (5.8)	13 (5.0)	5 (9.1)	2 (3.6)
Pregnant	77 (22.1)	20 (23.0)	57 (21.8)	13 (23.6)	10 (18.2)
Number of recent sex partners *					
0	70 (20.1)	14 (16.1)	56 (21.4)	10 (18.2)	14 (25.5)
1	268 (76.8)	71 (81.6)	197 (75.2)	44 (80.0)	38 (69.1)
>1	11 (3.2)	2 (2.3)	9 (3.4)	1 (1.8)	3 (5.5)

	Full Cohort			Deep Sequencing Subset	
	Combined (N=349) Median (IQR) or n(%)	**Cases (N=87) Median (IQR) or n(%)**	**Controls (N=262) Median (IQR) or n(%)**	**Cases (N=55) Median (IQR) or n(%)**	**Controls (N=55) Median (IQR) or n(%)**
Frequency of sex in the past month[†]	3 (1, 5)	3 (1, 6)	3 (1, 5)	3 (1, 7)	2 (0, 5)
Any recent unprotected sex[‡]	121 (34.7)	36 (41.4)	85 (32.4)	22 (40.0)	18 (32.7)
On examination[§]					
Abnormal vaginal discharge[‖]	41 (13.9)	12 (16.2)	29 (13.2)	9 (19.1)	8 (17.8)
Genital ulceration[¶]	6 (2.0)	2 (2.7)	4 (1.8)	0 (0.0)	0 (0.0)
Cervical mucopusn[¶]	4 (1.4)	3 (4.1)	1 (0.5)	3 (6.4)	0 (0.0)
Vaginal gram stain Nugent score[**]					
Normal (0–3)	161 (52.6)	28 (36.8)	133 (57.8)	21 (38.9)	36 (69.2)
Intermediate (4–6)	45 (14.7)	16 (21.1)	29 (12.6)	12 (22.2)	9 (17.3)
BV (7–10)	100 (32.7)	32 (42.1)	68 (29.6)	21 (38.9)	7 (13.5)
Laboratory-confirmed STIs at Analysis Visits					
Neisseria gonorrhoeae[††]	12 (4.6)	2 (2.9)	10 (5.2)	2 (4.4)	2 (4.9)
Chlamydia trachomatis[‡‡]	9 (3.8)	3 (5.0)	6 (3.4)	2 (5.0)	0 (0.0)
Trichomonas vaginalis[§§]	27 (8.0)	11 (12.9)	16 (6.3)	4 (7.3)	1 (1.8)
Syphilis seropositive[‖‖]	6 (3.1)	2 (4.4)	4 (2.6)	0 (0)	1 (3.7)
HSV-2 seropositive[¶¶]	165 (83.3)	45 (90.0)	120 (81.1)	32 (91.4)	27 (87.1)
Vaginal yeast on wet mount[***]	29 (19.1)	9 (23.7)	20 (17.5)	8 (29.6)	5 (18.5)
Any antibiotic use past 90 days[†††]	22 (11.5)	7 (14.3)	15 (10.5)	4 (10.5)	4 (10.5)

BV, bacterial vaginosis; DMPA, depo medroxyprogesterone acetate; HIV, human immunodeficiency virus; HSV, herpes simplex virus; IQR, interquartile range; IUD intrauterine contraceptive device; PrEP, pre-exposure prophylaxis

[*] Past week for Mombasa Cohort women, past month otherwise.

[†] Imputed for Mombasa Cohort women as 4 times the past week frequency.

[‡] Past week for Mombasa Cohort, past month otherwise.

[§] A total of 55 women in the Mama Salama Study were not examined at the swab collection date (per study schedule). One woman in Partners in Prevention HSV/HIV was not examined for genital ulceration and cervical mucopus.

[‖] For full cohort, N=294; for deep sequencing subset, N=92.

¶For full cohort, N=293, for deep sequencing subset, N=92.

**Not available for Couples Observational Study women and missing for 11 women from other cohorts. For full cohort, N=306, for deep sequencing subset, N=106.

††For full cohort, N=262; for deep sequencing subset, N=86. A total of 74 women in the Mama Salama Study were not assessed for NG at the swab collection date (per study schedule), and NG is missing for 13 women from other cohorts.

‡‡For full cohort, N=239; for deep sequencing subset, N=77. A total of 74 women in the Mama Salama Study and 23 women in the Mombasa Cohort were not assessed for CT at the swab collection date (per study schedules), and CT is missing for 13 women from other cohorts.

§§For full cohort, N=33. Missing for 10 women from Partners in Prevention HSV/HIV, COS, and PrEP cohorts.

‖‖For full cohort N=196; for deep sequencing subset, N=51. Syphilis status at the study visit was not available for the Mama Salama Study, and was not available at the study visit for 41 women from other cohorts.

¶¶For full cohort N=198; for deep sequencing subset N=59. HSV-2 serostatus was not available for the Mama Salama Study, and was missing for 39 women from other cohorts.

***For full cohort N=152; for deep sequencing subset, N=54. Yeast data from were not available for the Partners in Prevention HSV/HIV Study, Couples Observational Study, or Partners PrEP study.

†††For full cohort N=192; for deep sequencing cohort N=76. Antibiotic use data were not available for the Partners in Prevention HSV/HIV Study or the Couples Observational Study, both of which used only enrollment samples for the present analysis. Data on antibiotic use in the past 90 days were not available for the study visit for 73 women from other cohorts. In the Partners PrEP study, only antibiotics provided for STIs were captured.

Table 3:

Univariate logistic regression comparing relative abundance of vaginal bacteria identified by pyrosequencing in 55 women who acquired HIV and 55 women who remained uninfected

	Percent Relative Abundance				
Bacterial taxon*	**Median (Range) Cases (n=55)**	**Median (Range) Controls (n=55)**	**SD**	**OR (95% CI) per 1-SD change**	**p value**
Lactobacillus iners	8.36 (0, 99.85)	76.46 (0, 99.94)	42.02	0.54 (0.36, 0.80)	0.002
Gemella asaccharolytica	0.01 (0, 5.70)	0 (0, 0.82)	0.94	12.01 (2.26, 63.78)	0.004
Eggerthella species Type 1	0 (0, 2.47)	0 (0, 1.70)	0.51	2.06 (1.23,3.45)	0.006
Leptotrichia amnionii	0.01 (0, 26.40)	0 (0, 13.71)	5.40	2.67 (1.26, 5.65)	0.01
Dialister species Type 2	0 (0, 7.61)	0 (0, 5.14)	1.70	1.85 (1.11, 3.08)	0.02
Parvimonas micra	0 (0, 9.98)	0 (0, 2.31)	1.12	3.26 (1.17, 9.07)	0.02
Dialister micraerophilus	0.08 (0, 1.60)	0 (0, 1.43)	0.33	1.62 (1.03, 2.53)	0.04
Dialister†	0 (0, 0.29)	0 (0, 0.10)	0.04	2.17 (1.04, 4.53)	0.04
Mycoplasma†	0 (0, 1.53)	0 (0, 0.02)	0.17	1.09E+08 (0.67, 1.79E+16)‡	0.06
Lactobacillus reuteri/vaginalis	0 (0, 2.76)	0 (0, 15.15)	1.90	0.33 (0.10, 1.07)	0.07
Prevotella timonensis	0.10 (0, 16.68)	0 (0, 19.11)	3.96	1.52 (0.97, 2.37)	0.07
BVAB2	0 (0, 8.66)	0 (0, 5.10)	1.31	1.65 (0.96, 2.82)	0.07
Mycoplasma hominis	0.03 (0, 3.59)	0 (0, 1.24)	0.61	2.36 (0.92, 6.08)	0.08
Corynebacterium amycolatum	0 (0, 0.06)	0 (0, 0.19)	0.03	0.51 (0.24, 1.08)	0.08
Aerococcus†	0 (0, 2.66)	0 (0, 0.85)	0.31	2.01 (0.89, 4.58)	0.09
Porphyromonas	0 (0, 4.46)	0 (0, 0.80)	0.49	2.24 (0.86, 5.83)	0.10
Megasphaera species Type 1	0 (0, 17.76)	0 (0, 4.39)	2.45	2.57 (0.82, 8.07)	0.11
Prevotella†	0.04 (0, 15.12)	0 (0, 3.32)	1.71	1.97 (0.86, 4.52)	0.11
Coriobacteriaceae†	0 (0, 0.01)	0 (0, 0.08)	0.01	0.35 (0.09, 1.33)	0.12
Prevotella bivia	0 (0, 68.15)	0 (0, 32.23)	9.54	1.75 (0.86, 3.59)	0.13
Megasphaera†	0 (0, 0.81)	0 (0, 0.17)	0.10	2.05 (0.80, 5.24)	0.13

CI, confidence interval; OR, odds ratio; SD, standard deviation

*The table includes only bacterial taxa that were associated with HIV acquisition with a p-value ≤0.15.

†Some taxa were identified only to the genus level, while others were identified to the species level. Where this occurred, the genus level group excluded the individual species that were identified separately.

‡OR and CI are highly sensitive to two outlier *Mycoplasma* values. When these two values are removed, the OR (95% CI) is 30.23 (0.93, 984.53), p=0.06.

Table 4:

Unadjusted and odds ratios showing the association between vaginal bacterial quantity by qPCR and HIV acquisition in 87 women who became HIV infected versus 262 women who remained HIV uninfected

Bacterial Taxon (Quantiles show log_{10} genome copies/ swab)	Combined (n=349) n (%)	Controls (n=262) n (%)	Cases (n=87) n (%)	OR (95% CI)	p value	aOR* (95% CI)	p value
Aerococcus christensenii					0.257		0.350
Undetectable	170 (48.7)	135 (51.5)	35 (40.2)	1		1	
Tertile 1 (2.2–5.2)	60 (17.2)	44 (16.8)	16 (18.4)	1.36 (0.67, 2.73)		1.34 (0.64, 2.80)	
Tertile 2 (5.3–6.9)	60 (17.2)	43 (16.4)	17 (19.5)	1.58 (0.79, 3.14)		1.42 (0.70, 2.91)	
Tertile 3 (6.9–8.8)	59 (16.9)	40 (15.3)	19 (21.8)	1.91 (0.96, 3.80)		1.90 (0.92, 3.89)	
Atopobium vaginae					0.529		0.777
Undetectable	134 (38.4)	106 (40.5)	28 (32.2)	1		1	
Tertile 1 (2.0–6.1)	71 (20.3)	53 (20.2)	18 (20.7)	1.28 (0.65, 2.51)		1.24 (0.61, 2.49)	
Tertile 2 (6.1–7.6)	72 (20.6)	51 (19.5)	21 (24.1)	1.56 (0.81, 3.00)		1.37 (0.69, 2.70)	
Tertile 3 (7.6–9.1)	72 (20.6)	52 (19.8)	20 (23.0)	1.48 (0.75, 2.92)		1.35 (0.66, 2.74)	
BVAB-2					0.063		0.127
Undetectable	226 (64.8)	180 (68.7)	46 (52.9)	1		1	
Tertile 1 (2.0–5.1)	41 (11.7)	28 (10.7)	13 (14.9)	1.79 (0.86, 3.70)		1.78 (0.83, 3.81)	
Tertile 2 (5.2–6.6)	41 (11.7)	26 (9.9)	15 (17.2)	2.26 (1.11, 4.62)		2.10 (1.00, 4.44)	
Tertile 3 (6.6–8.1)	41 (11.7)	28 (10.7)	13 (14.9)	1.85 (0.87, 3.91)		1.72 (0.78, 3.78)	
Dialister micraerophilus					0.086		0.305
Undetectable	76 (21.8)	63 (24.0)	13 (14.9)	1		1	
Tertile 1 (2.3–5.0)	92 (26.4)	72 (27.5)	20 (23.0)	1.41 (0.64, 3.08)		1.32 (0.58, 2.98)	
Tertile 2 (5.1–6.6)	90 (25.8)	65 (24.8)	25 (28.7)	1.98 (0.91, 4.32)		1.55 (0.69, 3.50)	
Tertile 3 (6.6–8.5)	91 (26.1)	62 (23.7)	29 (33.3)	2.56 (1.16, 5.66)		2.14 (0.94, 4.92)	
Dialister species Type 2					0.061		0.165
Undetectable	162 (46.4)	127 (48.5)	35 (40.2)	1		1	
Tertile 1 (2.0–5.2)	62 (17.8)	51 (19.5)	11 (12.6)	0.78 (0.37, 1.66)		0.67 (0.30, 1.48)	
Tertile 2 (5.3–7.3)	63 (18.1)	44 (16.8)	19 (21.8)	1.52 (0.78, 2.96)		1.32 (0.66, 2.64)	
Tertile 3 (7.3–8.8)	62 (17.8)	40 (15.3)	22 (25.3)	2.09 (1.07, 4.07)		1.72 (0.85, 3.50)	
Eggerthella species Type 1					0.020		0.041
Undetectable	186 (53.3)	151 (57.6)	35 (40.2)	1		1	

Bacterial Taxon (Quantiles show log_{10} genome copies/ swab)	Combined (n=349) n (%)	Controls (n=262) n (%)	Cases (n=87) n (%)	OR (95% CI)	p value	aOR* (95% CI)	p value
Tertile 1 (2.3–6.0)	55 (15.8)	39 (14.9)	16 (18.4)	1.77 (0.89, 3.50)		1.79 (0.88, 3.64)	
Tertile 2 (6.0–7.5)	54 (15.5)	33 (12.6)	21 (24.1)	2.77 (1.43, 5.36)		2.62 (1.31, 5.22)	
Tertile 3 (7.5–9.1)	54 (15.5)	39 (14.9)	15 (17.2)	1.72 (0.84, 3.55)		1.53 (0.72, 3.28)	
Gemella asaccharolytica					0.003		0.010
Undetectable	196 (56.2)	161 (61.5)	35 (40.2)	1		1	
Tertile 1 (2.0–5.2)	51 (14.6)	36 (13.7)	15 (17.2)	1.91 (0.95, 3.87)		2.09 (1.01, 4.36)	
Tertile 2 (5.3–6.7)	51 (14.6)	34 (13.0)	17 (19.5)	2.34 (1.17, 4.69)		2.02 (0.98, 4.17)	
Tertile 3 (6.7–8.8)	51 (14.6)	31 (11.8)	20 (23.0)	3.22 (1.59, 6.49)		3.03 (1.46, 6.30)	
Gardnerella vaginalis					0.024		0.077
Undetectable	38 (10.9)	33 (12.6)	5 (5.7)	1		1	
Tertile 1 (2.3–6.6)	104 (29.8)	83 (31.7)	21 (24.1)	1.68 (0.59, 4.80)		1.97 (0.67, 5.80)	
Tertile 2 (6.6–8.7)	104 (29.8)	68 (26.0)	36 (41.4)	3.52 (1.26, 9.80)		3.24 (1.13, 9.27)	
Tertile 3 (8.8–10.3)	103 (29.5)	78 (29.8)	25 (28.7)	2.18 (0.75, 6.35)		1.90 (0.63, 5.74)	
Lactobacillus crispatus					0.625		0.469
Undetectable	268 (76.8)	199 (76.0)	69 (79.3)	1		1	
Tertile 1 (2.1–5.3)	27 (7.7)	23 (8.8)	4 (4.6)	0.51 (0.17, 1.51)		0.44 (0.14, 1.37)	
Tertile 2 (5.4–7.5)	27 (7.7)	20 (7.6)	7 (8.0)	1.01 (0.41, 2.49)		1.07 (0.42, 2.75)	
Tertile 3 (7.6–9.1)	27 (7.7)	20 (7.6)	7 (8.0)	1.01 (0.41, 2.51)		1.13 (0.44, 2.91)	
Lactobacillus iners					0.896		0.946
Undetectable	61 (17.5)	44 (16.8)	17 (19.5)	1		1	
Tertile 1 (1.9–7.3)	97 (27.8)	72 (27.5)	25 (28.7)	0.92 (0.44, 1.92)		1.10 (0.51, 2.41)	
Tertile 2 (7.3–8.7)	96 (27.5)	74 (28.2)	22 (25.3)	0.76 (0.36, 1.60)		0.88 (0.40, 1.93)	
Tertile 3 (8.7–9.9)	95 (27.2)	72 (27.5)	23 (26.4)	0.81 (0.38, 1.73)		0.94 (0.42, 2.09)	
Lactobacillus jensenii					0.564		0.589
Undetectable	301 (86.2)	223 (85.1)	78 (89.7)	1		1	
Tertile 1 (2.0–4.4)	16 (4.6)	12 (4.6)	4 (4.6)	0.96 (0.30, 3.07)		1.21 (0.36, 4.05)	
Tertile 2 (4.5–6.2)	16 (4.6)	13 (5.0)	3 (3.4)	0.66 (0.18, 2.36)		0.61 (0.16, 2.28)	
Tertile 3 (6.3–8.2)	16 (4.6)	14 (5.3)	2 (2.3)	0.41 (0.09, 1.83)		0.44 (0.09, 2.06)	
Leptotrichia/Sneathia					0.025		0.046
Undetectable	143 (41.0)	118 (45.0)	25 (28.7)	1		1	
Tertile 1 (1.7–4.7)	69 (19.8)	49 (18.7)	20 (23.0)	1.98 (1.01, 3.91)		2.04 (1.02, 4.10)	

Lancet Infect Dis. Author manuscript; available in PMC 2019 May 01.

Bacterial Taxon (Quantiles show log_{10} genome copies/ swab)	Combined (n=349) n (%)	Controls (n=262) n (%)	Cases (n=87) n (%)	OR (95% CI)	p value	aOR* (95% CI)	p value
Tertile 2 (4.7–7.5)	68 (19.5)	50 (19.1)	18 (20.7)	1.73 (0.86, 3.46)		1.45 (0.70, 3.00)	
Tertile 3 (7.5–9.6)	69 (19.8)	45 (17.2)	24 (27.6)	2.78 (1.39, 5.57)		2.59 (1.26, 5.34)	
Megasphaera					0.024		0.038
Undetectable	233 (66.8)	185 (70.6)	48 (55.2)	1		1	
Tertile 1 (2.1–6.3)	39 (11.2)	22 (8.4)	17 (19.5)	3.00 (1.47, 6.14)		3.15 (1.45, 6.81)	
Tertile 2 (6.3–7.7)	39 (11.2)	28 (10.7)	11 (12.6)	1.50 (0.70, 3.22)		1.43 (0.65, 3.14)	
Tertile 3 (7.7–9.0)	38 (10.9)	27 (10.3)	11 (12.6)	1.57 (0.71, 3.50)		1.32 (0.57, 3.05)	
Mycoplasma hominis					0.018†		0.048
Undetectable	206 (59.0)	165 (63.0)	41 (47.1)	1		1	
Tertile 1 (2.0–5.1)	48 (13.8)	35 (13.4)	13 (14.9)	1.48 (0.72, 3.03)		1.46 (0.69, 3.11)	
Tertile 2 (5.1–6.5)	47 (13.5)	34 (13.0)	13 (14.9)	1.58 (0.76, 3.28)		1.40 (0.66, 2.98)	
Tertile 3 (6.5–8.4)	48 (13.8)	28 (10.7)	20 (23.0)	3.02 (1.52, 6.01)		2.76 (1.36, 5.63)	
Parvimonas species Type 1					0.002		0.005
Undetectable	287 (82.2)	227 (86.6)	60 (69.0)	1		1	
Tertile 1 (2.4–3.1)	21 (6.0)	14 (5.3)	7 (8.0)	1.99 (0.77, 5.18)		1.67 (0.61, 4.57)	
Tertile 2 (3.1–4.2)	21 (6.0)	12 (4.6)	9 (10.3)	3.02 (1.20, 7.60)		3.01 (1.13, 7.99)	
Tertile 3 (4.2–7.0)	20 (5.7)	9 (3.4)	11 (12.6)	4.93 (1.92, 12.63)		4.64 (1.73, 12.46)	
Parvimonas species Type 2					0.001		0.004
Undetectable	226 (64.8)	181 (69.1)	45 (51.7)	1		1	
Tertile 1 (2.4–5.4)	41 (11.7)	22 (8.4)	19 (21.8)	3.82 (1.82, 7.98)		3.52 (1.63, 7.61)	
Tertile 2 (5.4–7.0)	41 (11.7)	33 (12.6)	8 (9.2)	0.95 (0.41, 2.20)		0.85 (0.36, 2.02)	
Tertile 3 (7.0–8.2)	41 (11.7)	26 (9.9)	15 (17.2)	2.30 (1.10, 4.80)		2.18 (1.01, 4.72)	
Porphyromonas asaccharolytica/uenonis					0.673		0.740
Undetectable	76 (21.8)	60 (22.9)	16 (18.4)	1		1	
Tertile 1 (2.0–3.4)	92 (26.4)	66 (25.2)	26 (29.9)	1.55 (0.73, 3.28)		1.40 (0.65, 3.05)	
Tertile 2 (3.5–5.2)	90 (25.8)	69 (26.3)	21 (24.1)	1.21 (0.55, 2.67)		1.00 (0.44, 2.26)	
Tertile 3 (5.2–8.6)	91 (26.1)	67 (25.6)	24 (27.6)	1.44 (0.64, 3.23)		1.22 (0.52, 2.86)	
Porphyromonas species Type 1					0.029		0.068
Undetectable	204 (58.5)	161 (61.5)	43 (49.4)	1		1	
Tertile 1 (2.0–3.3)	49 (14.0)	38 (14.5)	11 (12.6)	1.18 (0.54, 2.55)		1.14 (0.52, 2.52)	
Tertile 2 (3.4–4.9)	48 (13.8)	35 (13.4)	13 (14.9)	1.48 (0.71, 3.07)		1.22 (0.57, 2.62)	

Bacterial Taxon (Quantiles show log_{10} genome copies/ swab)	Combined (n=349) n (%)	Controls (n=262) n (%)	Cases (n=87) n (%)	OR (95% CI)	p value	aOR* (95% CI)	p value
Tertile 3 (4.9–8.0)	48 (13.8)	28 (10.7)	20 (23.0)	2.93 (1.45, 5.94)		2.74 (1.30, 5.76)	
Porphyromonas bennonis					0.907		0.951
Undetectable	245 (70.2)	183 (69.8)	62 (71.3)	1		1	
Tertile 1 (1.9–2.6)	35 (10.0)	28 (10.7)	7 (8.0)	0.73 (0.30, 1.79)		0.79 (0.31, 1.97)	
Tertile 2 (2.6–3.6)	34 (9.7)	25 (9.5)	9 (10.3)	1.06 (0.46, 2.43)		0.99 (0.42, 2.33)	
Tertile 3 (3.6–6.4)	35 (10.0)	26 (9.9)	9 (10.3)	1.01 (0.42, 2.44)		0.85 (0.34, 2.13)	
Prevotella					0.343		0.543
Quartile 1 (≤4.3)	88 (25.2)	70 (26.7)	18 (20.7)	1.00		1	
Quartile 2 (4.4–5.9)	87 (24.9)	67 (25.6)	20 (23.0)	1.26 (0.59, 2.69)		1.30 (0.60, 2.83)	
Quartile 3 (5.9–7.9)	87 (24.9)	65 (24.8)	22 (25.3)	1.40 (0.67, 2.91)		1.26 (0.59, 2.68)	
Quartile 4 (8.0–9.8)	87 (24.9)	60 (22.9)	27 (31.0)	1.98 (0.92, 4.26)		1.80 (0.80, 4.04)	

aOR, adjusted odds ratio; CI, confidence interval; OR, odds ratio

*Adjustment was made in a multivariable model stratified by cohort for the following confounding variables: age (continuous), pregnancy and contraceptive status (six categories: 1) pregnant, 2) non-pregnant and no hormonal contraception, 3) combination oral contraceptive, 4) depot medroxyprogesterone acetate, 5) intra uterine contraceptive device, 6) progesterone-only implant), number of sex partners in the past month (continuous), frequency of sex in the past month (continuous), and any recent self-reported unprotected sex (binary).