Telling truth from Ys: Accuracy of self-reported semen exposure assessed by a semen Y-chromosome biomarker

Janet Rosenbaum, PhD* [1,2] Johan Melendez, MS[2,3] Eve Rose, MPH[4]
Gina Wingood, ScD[4] Ralph DiClemente, PhD[4] Jonathan Zenilman, MD[2,3]

March 16, 2012

[1] Maryland Population Research Center, University of Maryland, College Park.[1]
[2] Sexually Transmitted Disease Center, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland
[3] Department of Infectious Diseases, Johns Hopkins Medical Institutions, Baltimore, Maryland
[4] Behavioral Sciences and Health Education, Rollins School of Public Health, Emory University, Atlanta, Georgia

Acknowledgements: The authors thank Jennifer Brown, Shai Davis, Maria Gallo, Fran Goldscheider, Joshua Green, Margaret Kaner, and Jessica Sales for helpful conversations, as well as participants at seminars at George Mason University Center for Social Science Research, the Office of Survey Methods Research at the Bureau of Labor Statistics, and the American Public Health Association Statistics Section.

[1]* Corresponding author: Janet Rosenbaum, Maryland Population Research Center, Cole Student Center 0124N, University of Maryland, College Park, MD, 20742. janet@post.harvard.edu.
This research was funded by T-32 AI050056 from the Sexually Transmitted Diseases Division of the Centers for Disease Control and Prevention (Dr. Zenilman) and the Eunice Kennedy Shriver National Institute for Child Health and Human Development grant R24-HD041041 to the Maryland Population Research Center. The data collection was funded by R01 MH061210 from the Center for Mental Health Research on AIDS, National Institute of Mental Health, Bethesda, Maryland (Drs DiClemente and Wingood).

Short summary: This study compares adolescent women's reports of condom use with the results of a biomarker for semen exposure. Adolescent women who reported perfect condom use but tested positive on the semen exposure biomarker were more likely than other participants to report pregnancy 6 months later.

Abstract

Background: Adolescents may use condoms inconsistently or incorrectly, or may over-report condom use. This study used a semen exposure biomarker to evaluate the accuracy of female adolescents' reports of condom use.
Methods: The sample comprised 715 sexually active African-American female adolescents, ages 15–21. At baseline, 6, and 12 months, participants completed a 40 minute interview and were tested for semen Y-chromosome with polymerase chain reaction from a self-administered vaginal swab. Data analysis predicted pregnancy using a Poisson working model, controlling for oral contraception, reported condom use, and coital frequency.
Results: At the 3 waves, 30%, 20% and 15% of adolescents who reported always using condoms tested positive for semen exposure. Six months later, 20.4% and 16.2% of the adolescents who under-reported semen exposure reported pregnancy, a higher pregnancy rate than among other adolescents, even those who never used condoms (14.2% and 11.8%). Suspected over-reporters had respectively 3.23 (95% confidence interval (1.61, 6.45)) and 2.21 (0.94, 5.20) greater likelihood of reporting pregnancy than other participants.
Conclusions: Respondents who reported perfect condom use but tested positive for semen Y-chromosome were more likely to become pregnant than other women, even women who reported never using condoms, adjusting for reported frequency of sex. Women who reported perfect condom use but tested positive for semen Y-chromosome may have over-reported their condom use and under-reported coital frequency.

Introduction

Adolescents may use condoms inconsistently or incorrectly, and may over-report their condom use. Biomarkers for semen exposure can identify apparent over-reporters, allowing interventions to target populations at risk from inconsistent condom use. This study uses a biomarker for semen exposure to evaluate the accuracy of adolescent women's reports of condom use.

Only 62% of US adolescents report having used condoms at last intercourse (1), and many adolescents never use condoms at all. Survey research in adolescent populations is also subject to reporting validity issues because adolescents may give inaccurate responses on surveys, especially for sensitive topics such as sexual behavior (2–5). Inconsistent reporting has been used to identify inaccurate self-reports of sexual behavior (3–5) and abortions (6), in addition to other sensitive behavior including cigarette smoking (7–10), alcohol, and illegal drug use (8, 9, 11–15). Respondents who give two logically contradictory reports about their sexual behavior within the same wave or across waves must have given at least one inaccurate response (16), but inconsistency is limited because it's not known which report is accurate.

Researchers have also identified biomarkers for semen exposure to decrease reliance on self-report, including sexually transmitted infection (STI) diagnoses and detection of substances present in semen. Respondents who report 100% condom use but were subsequently diagnosed with a sexually transmitted infection (STI) may have over-reported their condom use, but this measure has low sensitivity even in high STI prevalence populations (17–19). Condom use data can also be validated using biomarkers for semen exposure (20), including substances in the seminal plasma, including prostate specific antigen (PSA) (21, 22), semenogelins, and acid phospotase; and spermatazoa and other cells present in the semen, such as Y-chromosome (Yc) DNA (23–28). PSA has been

studied the most extensively of the semen biomarkers (20) and PSA concentration predicts pregnancy in condom efficacy trials (29), but PSA is only detectable for 24–48 hours post-coitus (22, 30). The short time-frame for the PSA is useful for condom effectiveness tests because swabs can be done immediately after condom use, but the time-frame is a limitation for studies with other purposes, such as to measure semen exposure over a longer time period.

Tests for Y-chromosome seem to be more sensitive than tests for PSA; Yc is detectable for 14 days post-coitus even during menses (26, 27, 31). One study that used this biomarker on archived samples from Baltimore STI clinic patients found that among those who reported 100% condom use in the previous two weeks (n=141), 55% tested positive for Yc, suggesting substantial over-reporting (28). Women with a history of STIs were more likely to be suspected over-reporters of perfect condom use in the past 2 weeks, suggesting self-presentation bias (32).

We are also interested in whether smoking mediates results from the Y-chromosome test because of growing evidence of the impact of smoking and nicotine on gynecological health from both randomized experiments and observational studies. Nicotine reduced physiological sexual arousal in non-smoking women in a randomized experiment (33), paralleling results in males (34). Smoking initiation is followed by changes in the cervical epithelium predictive of subsequent cervical cancer, reinforcing evidence that smoking could be causal rather than confounding (35). Smokers are more likely than non-smokers to test positive for mycoplasma genitalium (36), genital herpes (37), HPV (38), and bacterial vaginosis (39). Smokers are more likely to have HPV infection persistence and cervical abnormalities, proportional to smoking intensity and younger of smoking initiation (40). Smokers have lower vulvic vascularization and report lower orgasm frequency, proportional to smoking intensity (41), although smokers have higher self-reported sexual functioning (42). Smokers also have lower serum oestradiol levels and more abnormal vaginal flora (43). Although smokers also have less ideal health behaviors, such as lower likelihood of adhering recommended cervical cancer screening guidelines (44, 45), these experiments and prospective studies suggest that smoking causes changes to the vaginal environment, and is not just a marker for other risky health behavior.

Researchers have speculated that respondents who give inaccurate survey responses do not misreport their behavior intentionally, and in fact they may believe that their survey responses accurately portray their own behavior. Respondents giving inaccurate survey responses may thus be more likely to take health risks because they do not acknowledge their own risk behavior and may underestimate their STI risk (3). Few studies have tested whether under-reporting of health risk behavior predicts health risks. This study uses a biomarker for semen exposure to measure under-report of semen exposure among women adolescents in an urban Atlanta safe sex intervention. We test the hypothesis that under-reporters of semen exposure are more likely to become pregnant than women who accurately report their semen exposure, controlling for contraception and sexual frequency. The proposed mechanism is that under-reporters of semen exposure may either be unwilling or unable to recall their true behavior, or may engage in more risky behavior than accurate reporters. Following research finding that people with strong norms are more likely to misreport behavior that violates those norms (2, 3, 46–53), we also hypothesize that women who were randomized to the intervention would be more likely to under-report semen exposure after participating in the intervention.

Under-report of semen exposure was determined by comparing adolescent women's reports of condom use frequency with results of a biomarker test for semen exposure in the last two weeks that detects semen on a vaginal swab. This study is, to our knowledge, the first longitudinal use of the Yc biomarker to predict subsequent outcomes.

Method

Data

Participants were 715 African-American females, 15 to 21 years old (mean age 17.6), who reported sexual activity in the past 60 days and were not trying to get pregnant. Participants were recruited for a randomized trial of an HIV prevention program that enrolled African-American females at a publicly funded STI clinic, a teen clinic based in a large public hospital, and a family planning clinic, all in Atlanta in 2002–04: 847 participants were eligible, of whom 84% agreed to participate (32). This study uses both the experimental and control groups without evaluating the intervention's effectiveness.

Participants completed a 40 minute interview administered via audio computer-assisted self-interviewing (ACASI) and tested for Y-chromosome DNA. After the interview, trained monitors instructed participants how to collect vaginal fluid using a life-like model of a vagina. Participants performed a 10–15 second vaginal sweep using the Becton Dickinson 'swube applicator,' and the swabs were frozen and shipped to the Johns Hopkins Division of Infectious Disease Laboratory, where they were tested using the Yc-PCR test. All samples were processed by a female technician to avoid technician Yc contamination. The Y-chromosome biomarker was evaluated with a polymerase chain reaction (Yc-PCR). The Yc-PCR assay is sensitive to 5 copies of Yc for up to 14 days after coitus. The estimated specificity is 92% (95% CI (80, 98)): i.e., 92% of women in the calibration trial who had protected sex tested negative for Yc-PCR, and the 8% of women had each had digital or oral genital contact by their male partner, so the false positives could be explained by epithelial cells (25–27).

Follow-up interviews were administered at 6 and 12 months using similar methods. Follow-up proportion, as a proportion of wave 1 participants, was 84.8% at wave 2 and 84.5% at wave 3, and 78.3% participated in all 3 waves. The Institutional Review Board at Emory University approved the study protocol prior to implementation. Participants were paid $50 upon completion of each wave.

Variables

Condom use is measured in the past 60 days, past 14 days, and at last sex. Condom use in the past 60 days is a proportion measured by a 2 question sequence: how many times the respondent had vaginal sex in the past 60 days and "How many of the X [number specified in previous question] times you've had sex in the past 60 days did you use a condom?" Most respondents reported either never or always using condoms. Condom use was categorized in 6 categories: condoms used in 0% of vaginal sex acts in the past 60 days, 1–50% of vaginal sex acts, 51–99% of vaginal sex acts, 100% of the vaginal sex acts, respondent reported no sex in the past 60 days, and respondent did not answer the question of how many times they had vaginal sex in the past 60 days. Condom use in the past 14 days was measured identically.

Condom use at last sex is measured by the question, "The very last time you had sex, what type(s) of protection did you use?" with instruction to choose as many of the 6 options as applicable: male condoms, female condoms,spermicide, foam, withdrawal, and hormonal contraception via injection, patch, or implant. A separate question assessed oral contraceptive use.

We used two indicators of suspected over-report: the Yc-PCR test and inconsistent report within the same wave. Respondents who reported 100% condom use in the past 14 days but who tested positive for Yc are called suspected over-reporters. Respondents who report 100% condom use in the past 14 days but report not using a condom at last sex are called inconsistent reporters.

At each of the 3 waves, more respondents skipped the question about number of episodes of coitus in the past

60 days than in the past 14 days, which may indicate that they had trouble enumerating each episode of coitus from the past 60 days. Pregnancy is self-reported. Abuse and forced sex by boyfriend and casual partners were asked only to respondents over age 18 (n=385).

Analysis

Horizons intervention participants reported serially monogamous relationships: over 80% had steady boyfriends and only 22–31% reported casual partners in each of the three waves, although only about 30% of respondents reported the same boyfriend as in the previous wave. Due to the relatively small number of partners, this study used self-reported pregnancy as the outcome of interest rather than STI diagnosis because pregnancy can occur even in monogamous relationships with an STI-negative partner, whereas STI diagnosis requires an STI positive partner.

Data analysis was performed in Stata SE 11.1.

Women who reported using condoms for all coitus in the past 14 days but who tested positive for Yc are labeled suspected over-reporters.

Differences between groups were evaluated with the chi-squared test, Fisher's exact test, and Cuzick's non-parametric test for trend, a generalization of Wilcoxon's sign rank test.

We estimated relative risks using a Poisson working model because estimators from logistic regression are inconsistent when the outcome is not rare — as is true in this case because pregnancy was reported by 10% of the sample at each wave. Poisson regression yields consistent and unbiased estimators (**? ? ?**). Control variables were oral contraceptive use, number of coital episodes in the past 60 days, condom use in the past 60 days, and a binary variable for non-report of condom use in the past 60 days.

Regressions were limited to participants in all 3 waves to ensure comparable groups (n=560). Fisher's exact test evaluated whether suspected over-reporters differed in their likelihood of being lost to follow-up.

Results

Suspected over-report of condom use

At wave 1, 29% of the sample reported having used condoms for all coitus in the past 14 days. Among the women who reported perfect condom use, 30% tested positive for Y-chromosome (Yc) , suggesting over-report of condom use in the past 14 days (Figure 1). At wave 2, 45% of women reported having used condoms for all coitus in the past 14 days, but 20% tested positive for Yc. At wave 3, 44% of women reported having used condoms for all coitus in the past 14 days, but 15% tested positive for Yc.

Suspected over-reporters of condom use were more likely to report being pregnant at the following wave than any other category of condom users, including never users: 20% of wave 1 suspected over-reporters reported at wave 2 being pregnant, and 16% of wave 2 suspected over-reporters reported at wave 3 being pregnant (Table 1). Suspected over-reporters were also more likely to report pregnancy than any other category of contraceptive users at last sex, including those using no contraception (Table 2). Suspected over-reporters were 2–3 times more likely to report pregnancy in the next wave than other respondents, adjusting for contraception and frequency of sex (Table 3), with risk ratios of 3.23(1.61, 6.45) for wave 2 pregnancy and 2.21 (0.94, 5.20) for wave 3 pregnancy.

At all 3 waves, suspected over-reporters reported fewer episodes of coitus in the past 60 days than all other categories of condom users (Table 1). Few respondents who reported having never used condoms in the past 60 days also reported using other contraception at last sex (Table 4).

Suspected over-reporters did not differ at median from other respondents in lifetime and past 60 day alcohol and marijuana use, number of alcoholic drinks when drinking, beliefs about the safety of sex while drinking alcohol, or reported sex while high on drugs or alcohol. They also did not differ in age, number of days between surveys, having been emotionally or physically abused in lifetime or in last 60 days by boyfriend or casual sex partners, and having ever been forced to have vaginal or anal sex.

At wave 3, there were more suspected over-reporters in the intervention group than in the control group (6.57% vs. 3.16%, Wilcoxon p=0.05), but there were no significant differences in proportion of suspected over-reporters in intervention than control group at wave 2.

Suspected over-reporters were not more likely to drop out of the study at waves 2 or 3 (Fisher's exact p=0.9 at wave 2, p=0.2 at wave 3).

Smoking and the semen exposure biomarker

Suspected over-reporters of condom use were more likely to report being cigarette smokers than any other category of condom users: about 35% of suspected over-reporters at waves 1 and 2 and 31% of suspected over-reporters at wave 3. The proportion of smokers was inversely proportional to condom use category, with suspected over-reporters having the highest smoking prevalence (test for trend p=0.01, 0.04, 0.01) (Table 1).

Smoking more cigarettes per day predicted a greater chance of testing positive on the Yc-PCR among all respondents (Cuzick's test for trend p=0.02, 0.000, 0.000 at waves 1, 2, and 3, respectively) but not if analysis was restricted to current smokers. Neither smoking status (Fisher's exact p=0.3, 0.2) nor number of cigarettes per day (Test for trend p=0.9, 0.9) predicted pregnancy in the subsequent wave.

Inconsistent report of condom use

Respondents also reported condom use inconsistently within the same wave: reporting 100% condom use within the last 14 days, but saying that they did not use a condom at last sex, which was done by 24%, 22%, and 19% of those reporting 100% condom use in the past 14 days in the respective 3 waves (Figure 1). Inconsistent reporters of condom use were not more likely to get pregnant, using Poisson regression with the same control variables as above (RRs 0.55 (0.14, 2.01), 1.87 (0.80, 3.96)).

Discussion

A significant number of intervention participants seem to have over-estimated the correctness and/or consistency of their condom use. This overestimation of condom use may be unintentional but is nonetheless associated with greater pregnancy rates in the next 6 months. Suspected over-reporters of condom use — who report using condoms in 100% of coital acts during the past 2 weeks but who test positive for Y-chromosome — were 2–3 times more likely to become pregnant, controlling for contraception use and frequency of sex. All women participating in the study stated that they did not want to become pregnant, but about 10% of the sample became pregnant at each wave of the study, implying that these pregnancies were unplanned.

Apparent condom overreport may arise from intentionally concealing condom non-use, from unintentionally imperfect use of condoms that caused condoms to leak or break, or from partner birth control sabotage, such as deception about whether condoms were used or surreptitious removal of condom mid-coitus, but these factors would not explain why suspected over-reporters may have a higher pregnancy rate than respondents reporting no

condom use in the past 60 days. The higher pregnancy rate among suspected over-reporters than condom non-users is not attributable to never-users of condoms using other common contraception methods such as withdrawal because virtually no never-users of condoms reported using other contraceptive methods at last sex. Suspected over-reporters also report having had sex fewer times in the last 60 days than do never-users.

The higher pregnancy rate among suspected over-reporters could be explained if over-reporters also under-reported vaginal intercourse. This explanation is consistent with past evidence that suspected over-reporting of condom use is attributable to self-presentation bias (32), as well as with the larger number of sexual partners among suspected-over-reporters. Respondents often misreport related issues (55), so suspected over-reporters may over-report condom use and under-report frequency of coitus. The data does not yield potential explanations for why suspected over-reporters would under-report coitus, such as substance use. Compared with other respondents, suspected over-reporters do not report more frequent or greater drug or alcohol use, use during sex, or greater acceptance of sex under the influence.

Smokers were more likely to be suspected over-reporters of condom use and to test positive on the Yc-PCR regardless of their reported condom use, and likelihood of testing positive increased in proportion to the number of cigarettes smoked. This difference is consistent with the hypothesis that the Yc-PCR test is more sensitive for smokers than non-smokers; the mechanism could be that nicotine causes lower physiological arousal (33), which could increase the length of time that Y-chromosome is detectable in the vagina.

Suspected overreport decreased over the course of the study — from 30% at baseline to 20% at 6 months and 15% at 12 months — despite greater proportions of women reporting having used condoms consistently over the course of the study. More intervention participants were suspected of condom use over-report, which may suggest that the intervention induced self-presentation bias similar to that observed among respondents with previous STI diagnoses (32). The intervention did not appear responsible for the lower prevalence of suspected over-reporting of condom use. Taking the survey may have made respondents feel more conscious of their condom use frequency and correctness, which may explain the decrease in suspected over-report, but the decrease is not attributable to suspected over-reporters leaving the study at greater rates.

The Yc-PCR test identifies a group at distinctly higher risk of unplanned pregnancy than respondents who gave inconsistent reports of their condom use in the same wave. Suspected over-reporters are more likely to become pregnant, but respondents who give inconsistent reports of their condom use in the same wave — reporting 100% condom use in the past 2 weeks, but reporting no condom used at last sex — are not more likely to become pregnant.

Limitations

Respondents were safe sex intervention participants, so these results may not generalize to all adolescents.

The Horizons intervention was designed to prevent STIs rather than pregnancy, and pregnancy was not a main outcome of the Horizons evaluation. Pregnancy was self-reported and may be under-reported, so this study may under-estimate the relationship between positive Yc-PCR tests and pregnancy. If semen exposure under-reporting is attributable to reporting bias rather than imperfect condom use, we would expect women suspected of over-reporting condom use to under-report pregnancy as well. Pregnancy may be even higher among women suspected of over-reporting condom use than found in this study's results, so this study may under-state the effect.

Conclusions

In this sample, 15–30% of women who reported perfect condom use tested positive for semen Y-chromosome, suggesting that they over-reported condom use. The Yc-PCR identifies women at high risk of pregnancy who portray themselves as perfect condom users. The Yc-PCR is not currently suitable for a clinical setting because of the long processing time. If adapted to clinical setting, the Yc-PCR could be a valuable tool for preventing unplanned pregnancy among adolescent women by alerting women that they have not been using condoms correctly or consistently.

References

[1] Danice K. Eaton, Laura Kann, Steve Kinchen, Shari Shanklin, James Ross, Joseph Hawkins, William A. Harris, Richard Lowry, Tim McManus, David Chyen, Connie Lim, Nancy D. Brener, and Howell Wechsler. Youth risk behavior surveillance — united states, 2007. *Mortality and Morbidity Weekly Report*, 57(SS04):1–131, June 6, 2008 2008.

[2] ND Brener, JOG Billy, and WR Grady. Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: evidence from the scientific literature. *Journal of Adolescent Health*, 33:436–457, 2003.

[3] Janet E Rosenbaum. Reborn a virgin: adolescents' retracting of virginity pledges and sexual histories. *American Journal of Public Health*, 96(6):1098–1103, 2006.

[4] CS Alexander, MR Somerfield, ME Ensminger, KE Johnson, and YJ Kim. Consistency of adolescents' self-report of sexual behavior in a longitudinal study. *Journal of Youth and Adolescence*, 22:455–71, 1993.

[5] DM Upchurch, LA Lillard, CS Aneshensel, and NF Li. Inconsistencies in reporting the occurrence and timing of first intercourse among adolescents. *Journal of Sex Research*, 39:197–206, 2002.

[6] Haishan Fu, Jacqueline E. Darroch, Stanley K. Henshaw, and Elizabeth Kolb. Measuring the extent of abortion underreporting in the 1995 national survey of family growth. *Family Planning Perspectives*, 30(3):128–138, 1998.

[7] RCME Engels, RA Knibbe, and MJ Drop. Inconsistencies in adoelscents' self-reports of initiation of alcohol and tobacco use. *Addictive Behaviors*, 22:613–623, 1997.

[8] AM Shillington and JD Clapp. Self-report stability of adolescent substance use: are there differences for gender, ethnicity and age? *Drug and Alcohol Dependence*, 60:19–27, 2000.

[9] W Pedersen. Reliability of drug use responses in a longitudinal study. *Scandinavian Journal of Psychology*, 31:28–33, 1990.

[10] WR Stanton, M McClelland, C Elwood, D Ferry, and PA Silva. Prevalence, reliability and bias of adolescents' reports of smoking and quitting. *Addiction*, 91:1705–1714, 1996.

[11] SL Bailey, RL Flewelling, and JV Rachal. Characterization of inconsistencies in self-reports of alcohol and marijuana use in a longitudinal study of adolescents. *Journal of Studies on Alcohol*, 53:636–47, 1992.

[12] M Fendrich and DP Rosenbaum. Recanting of substance use reports in a longitudinal prevention study. *Drug and Alcohol Dependence*, 70:241–253, 2003.

[13] M Fendrich and JYS Kim. Multiwave analysis of retest artifact in the national longitudinal survey of youth drug use. *Drug and Alcohol Dependence*, 62:239–253, 2001.

[14] M Fendrich and CM Vaughn. Diminished lifetime substance use over time: an inquiry into differential underreporting. *Public Opinion Quarterly*, 58:96–123, 1994.

[15] BS Mensch and DB Kandel. Underreporting of substance use in a national longitudinal youth cohort: individual and interviewer effects. *Public Opinion Quarterly*, 52:100–124, 1988.

[16] EJ Reinisch, RM Bell, and PL Ellickson. How accurate are adolescent reports of drug use? Technical Report N-3189 CHF, RAND, 1991.

[17] ML Shew, GJ Remafedi, LH Bearinger, PL Faulkner, BA Taylor, SJ Potthoff, and MD Resnick. The validity of self-reported condom use among adolescents. *Sexually Transmitted Diseases*, 24(9):503–510, October 1997.

[18] DP Orr, DJ Fortenberry, and MJ Blythe. Validity of self-reported sexual behaviors in adolescent women using biomarker outcomes. *Sexually Transmitted Diseases*, 24(5):261–266, 1997.

[19] JM Zenilman, CS Weisman, AM Rompalo, N Ellish, D Upchurch, E Hook, and DD Celentano. Condom use to prevent incident stds: The validity of self-reported condom use. *Sexually Transmitted Diseases*, 22(1):15–21, 1995.

[20] Christine K Mauck. Biomarkers of semen exposure. *Sexually Transmitted Diseases*, 36(3, supplement):S81–S83, March 2009.

[21] M Macaluso, L Lawson, and G et al Horton. Efficacy of the female condom as a barrier to semen during intercourse. *American Journal of Epidemiology*, 157(4):289–297, 2003.

[22] M Macaluso, L Lawson, and R et al Akers. Prostate-specific antigen in vaginal fluid as a biologic marker of condom failure. *Contraception*, 59(3):195–201, 1999.

[23] N Chomont, G Grésenguet, M Lévy, H Hocini, P Becquart, M Matta, J Tranchot-Diallo, L Andreoletti, MP Carreno, MD Kazatchkine, and L Bélec. Detection of y chromosome dna as evidence of semen in cervicovaginal secretions of sexually active women. *Clinical and Diagnostic Laboratory Immunology*, 8(5):955–958, 2001.

[24] Nicolas Chomont, Gérard Grésenguet, Hakim Hocini, Pierre Becquart, Mathieu Matta, Laurent Andreoletti, Ali Si-Mohamed, Marie-Paule Carreno, Michel Kazatchkine, and Laurent Bélec. Polymerase chain reaction for y chromosome to detect semen in cervicovaginal fluid: a prerequisite to assess hiv-specific vaginal immunity and hiv genital shedding. *AIDS*, 15(6):801–802, April 2001.

[25] Jonathan M. Zenilman, Jeffrey Yuenger, Noya Galai, Charles F Turner, and Susan M Rogers. Polymerase chain reaction detection of y chromosome sequences in vaginal fluid: Preliminary studies of a potential biomarker for sexual behavior. *Sexually Transmitted Diseases*, 32(2):90–94, February 2005.

[26] JH Melendez, JA Giles, JD Yuenger, TD Smith, KG Ghanem, K Reich, and JM Zenilman. Detection and quantification of y-chromosomal sequences by real-time pcr using the lightcycler system. *Sexually Transmitted Diseases*, 34(8):617–619, August 2007.

[27] KG Ghanem, JH Melendez, C McNeil-Solis, JA Giles, J Yuenger, TD Smith, and J Zenilman. Condom use and vaginal y-chromosome detection: The specificity of a potential biomarker. *Sexually Transmitted Diseases*, 34(8):620–623, August 2007.

[28] RA Jadack, J Yuenger, KG Ghanem, and JM Zenilman. Polymerase chain reaction detection of y-chromosome sequences in vaginal fluid of women accessing a sexually transmitted disease clinic. *Sexually Transmitted Diseases*, 33(1):22–25, 2006.

[29] Terri L. Walsh, Ron G. Frezieres, Karen Peacock, Anita L. Nelson, Virginia A. Clark, Leslie Bernstein, and Brian G.D. Wraxall. Use of prostate-specific antigen (psa) to measure semen exposure resulting from male condom failures: implications for contraceptive efficacy and the prevention of sexually transmitted disease. *Contraception*, 67(2):139–50, 2003.

[30] TL Walsh, RG Frezieres, and et al. Nelson, AL. Evaluation of prostate-specific antigen as a quantifiable indicator of condom failure in clinical trials. *Contraception*, 60:289–298, 1999.

[31] Rebecca M. Brotman, Johan H. Melendez, Tukisa D. Smith, Noya Galai, and Jonathan M. Zenilman. Effect of menses on clearance of y-chromosome in vaginal fluid: Implications for a biomarker of recent sexual activity. *Sexually Transmitted Diseases*, 37(1):1–4, 2010.

[32] E Rose, RJ Diclemente, GM Wingood, JM Sales, TP Latham, RA Crosby, J Zenilman, J Melendez, and J Hardin. The validity of teens' and young adults' self-reported condom use. *Arch Pedatri Adolesc. Med.*, 163(1):61–64, January 2009.

[33] Christopher B Harte and Cindy M Meston. The inhibitory effects of nicotine on physiological sexual arousal in nonsmoking women: results from a randomized, double-blind, placebo-controlled, cross-over trial. *The Journal of Sexual Medicine*, 5(5):1184–1197, May 2008. PMID: 18331269.

[34] Christopher B Harte and Cindy M Meston. Acute effects of nicotine on physiological and subjective sexual arousal in nonsmoking men: a randomized, double-blind, placebo-controlled trial. *The Journal of Sexual Medicine*, 5(1):110–121, January 2008. PMID: 17971108.

[35] Y T Ma, S I Collins, L S Young, P G Murray, and C B J Woodman. Smoking initiation is followed by the early acquisition of epigenetic change in cervical epithelium: a longitudinal study. *British Journal of Cancer*, 104(9):1500–1504, April 2011. PMID: 21487403.

[36] V L Short, P A Totten, R B Ness, S G Astete, S F Kelsey, P Murray, and C L Haggerty. The demographic, sexual health and behavioural correlates of mycoplasma genitalium infection among women with clinically suspected pelvic inflammatory disease. *Sexually Transmitted Infections*, 86(1):29–31, February 2010. PMID: 19703841.

[37] Thomas L Cherpes, Leslie A Meyn, Marijane A Krohn, and Sharon L Hillier. Risk factors for infection with herpes simplex virus type 2: role of smoking, douching, uncircumcised males, and vaginal flora. *Sexually Transmitted Diseases*, 30(5):405–410, May 2003. PMID: 12916131.

[38] Thomas Iftner, Sonja Eberle, Angelika Iftner, Barbara Holz, Norbert Banik, Wim Quint, and Anja-Natascha Straube. Prevalence of low-risk and high-risk types of human papillomavirus and other risk factors for HPV infection in germany within different age groups in women up to 30 years of age: an epidemiological observational study. *Journal of Medical Virology*, 82(11):1928–1939, November 2010. PMID: 20872721.

[39] L Uscher-Pines, AL Hanlon, and DB Nelson. Racial differences in bacterial vaginosis among pregnant women: the relationship between demographic and behavioral predictors and individual bv-related microorganism levels. *Matern Child Health J*, 13(4):512–9, 2009.

[40] Koji Matsumoto, Akinori Oki, Reiko Furuta, Hiroo Maeda, Toshiharu Yasugi, Naoyoshi Takatsuka, Yasuo Hirai, Akira Mitsuhashi, Takuma Fujii, Tsuyoshi Iwasaka, Nobuo Yaegashi, Yoh Watanabe, Yutaka Nagai, Tomoyuki Kitagawa, and Hiroyuki Yoshikawa. Tobacco smoking and regression of low-grade cervical abnormalities. *Cancer Science*, 101(9):2065–2073, September 2010. PMID: 20626752.

[41] Cesare Battaglia, Bruno Battaglia, Fulvia Mancini, Nicola Persico, Rossella E Nappi, Roberto Paradisi, and Stefano Venturoli. Cigarette smoking decreases the genital vascularization in young healthy, eumenorrheic women. *The Journal of Sexual Medicine*, 8(6):1717–1725, June 2011. PMID: 21477023.

[42] Christian W Wallwiener, Lisa-Maria Wallwiener, Harald Seeger, Alfred O Mück, Johannes Bitzer, and Markus Wallwiener. Prevalence of sexual dysfunction and impact of contraception in female german medical students. *The Journal of Sexual Medicine*, 7(6):2139–2148, June 2010. PMID: 20487241.

[43] J D Wilson, R A Lee, A H Balen, and A J Rutherford. Bacterial vaginal flora in relation to changing oestrogen levels. *International Journal of STD & AIDS*, 18(5):308–311, May 2007. PMID: 17524189.

[44] Shannon D MacLaughlan, Jason A Lachance, and Annie Gjelsvik. Correlation between smoking status and cervical cancer screening: a cross-sectional study. *Journal of Lower Genital Tract Disease*, 15(2):114–119, April 2011. PMID: 21478697.

[45] Anthony M A Smith, Wendy Heywood, Richard Ryall, Julia M Shelley, Marian K Pitts, Juliet Richters, Judy M Simpson, and Kent Patrick. Association between sexual behavior and cervical cancer screening. *Journal of Women's Health (2002)*, 20(7):1091–1096, July 2011. PMID: 21682554.

[46] L Midanik. Perspectives on the validity of self-report alcohol use. *British Journal of Addictions*, 84:1419–1423, 1989.

[47] RW Pearson, M Ross, and RM Dawes. *Questions about Questions*, chapter Personal recall and the limits of retrospective questions in surveys, pages 65–94. Russell Sage Foundation, New York, NY, 1992.

[48] TJ DeMaio. *Surveying subjective phenomena, vol. 2*, chapter Social desirability and survey measurement: A review, pages 257–282. Russell Sage Foundation, 1984.

[49] BD Silver, BA Anderson, and PR Abramson. Who overreports voting? *Am Political Sci Rev*, 80:613–624, 1986.

[50] S Presser. Is inaccuracy on factual survey items item-specific or respondent-specific? *Public Opinion Quarterly*, 48:344–355, 1984.

[51] BD Silver, PR Abramson, and BA Anderson. The presence of others and overreporting of voting in american national elections. *Public Opinion Quarterly*, 50:228–239, 1986.

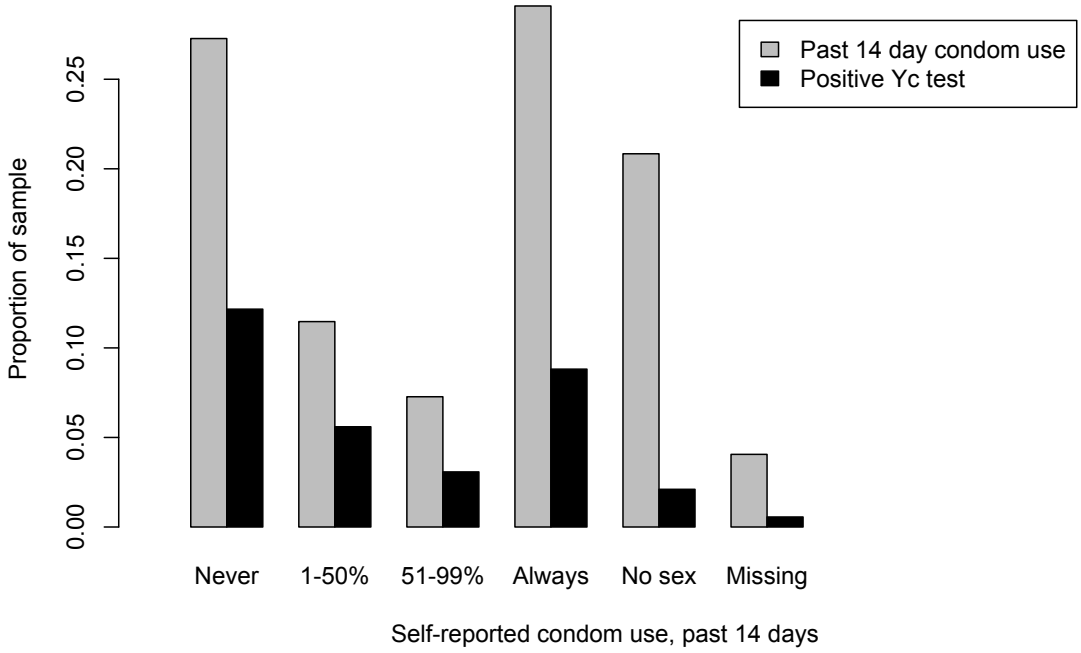
[52] RA Bernstein, A Chadha, and R Montjoy. Overreporting voting: why it happens and why it matters. *Public Opinion Quarterly*, 65:22–44, 2001.

[53] WB Locander, S Sudman, and NM Bradburn. An investigation of interview method, threat, and response distortion. *Journal of the American Statistical Association*, 71:269–275, 1976.

[54] Jun Zhang and Kai F Yu. What's the relative risk? a method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*, 280(19):1690–1, 1998.

[55] L Hamilton. Self-reports of academic performance: Response errors are not well behaved. *Sociological methods and research*, 10:165–185, 1981.

Figure 1: Condom use in the past 14 days, compared with results from (1) the Yc-PCR biomarker for semen exposure and (2) reporting of not using a condom at last sex, all measured at baseline. Suspected over-reporters report having had sex in the past 14 days, during which they used condoms 100% of the time, but test positive on the Yc-PCR. Inconsistent reporters report having had sex in the past 14 days, during which they used condoms 100% of the time, but report no condom at last sex. Missing are those who participated in the wave but did not answer the question how many times they had sex in the past 14 days. Suspected over-report and inconsistent report are correlated with Pearson correlations of r=0.28 ($p < 0.0001$), r=0.21 ($p < 0.0001$), and r=0.03 ($p = 0.4$) at waves 1, 2, and 3, respectively.



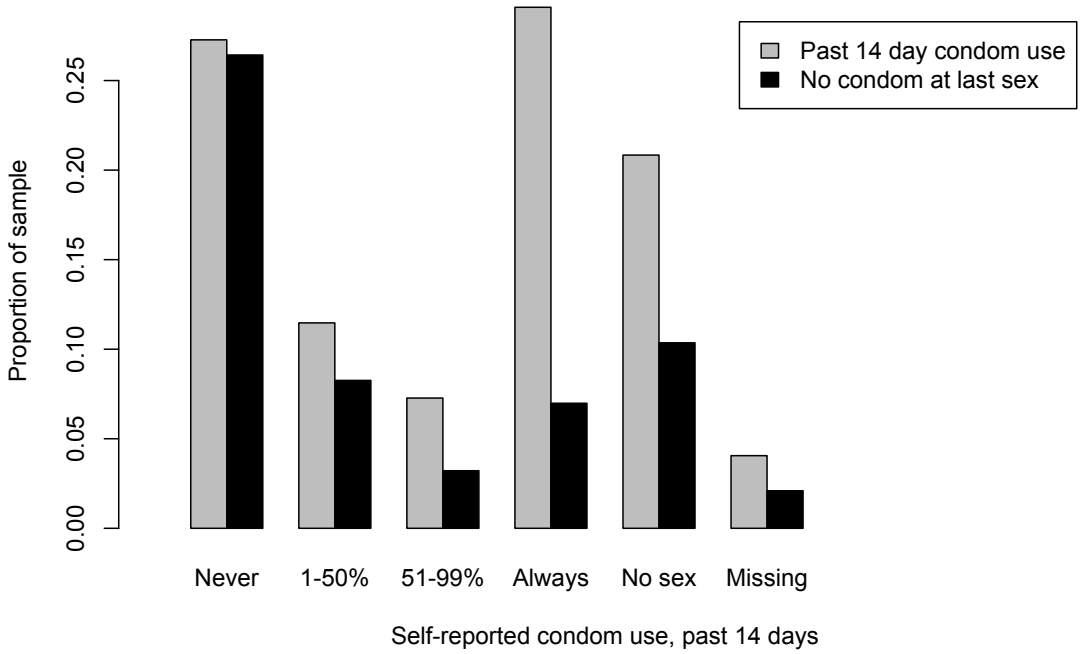


Table 1: Prevalence of smoking and pregnancy and coital frequency, by suspected over-report condom use in past 60 days. Missing are those who participated in the wave but did not answer the question how many times they had sex in the past 60 days. P is from Cuzick's non-parametric test for trend evaluated across the first six columns from suspected over-report to missing, which evaluates whether suspected over-reporters have greater prevalence/frequency of the behaviors than never-condom users, and whether the behaviors are less common among those with greater condom use, and whether the behavior is least common among respondents who skipped the question about frequency of sex in the past 60 days. Analysis was limited to those who participated in all 3 waves of the survey.

	Report 100% use	Condom use in past 60 days						
	positive biomarker	0%	1–50%	51–99%	100%	Missing	Total	P
Wave 1								
Number	49	113	104	96	81	117	560	
% Smoke, wave 1	34.7	24.8	26.0	16.7	17.3	17.1	21.8	0.01
% Pregnant, wave 2	20.4	14.2	6.7	11.5	2.5	2.6	8.8	0.002
Times sex 60 days	7.3	13.3	13.0	14.5	6.7		12.5	0.9
Wave 2								
Number	37	110	81	86	117	129	560	
% Smoke, wave 2	35.1	29.1	23.5	23.3	20.5	10.9	21.8	0.04
% Pregnant, wave 3	16.2	11.8	12.4	8.1	6.0	6.2	9.1	0.03
Times sex 60 days	11.0	17.3	16.4	14.6	7.6		13.6	0.01
Wave 3								
Number	29	129	74	81	143	104	560	
% Smoke, wave 3	31.0	26.4	29.7	18.5	16.1	18.3	21.8	0.01
Times sex 60 days	17.7	16.5	13.9	18.8	6.0		12.5	0.000

Table 2: Proportion reporting pregnancy by contraception used at last sex. Condoms include both male and female condoms. Hormonal contraception includes oral contraception as well as hormonal implants, injections, and patches. Suspected over-report is defined as those who reported 100% condom use in the past 14 days but tested positive on the Yc-PCR test, suggesting semen exposure in the past 14 days. P is from Fisher's exact test.

	Number, wave 1	% pregnant, wave 2	Number, wave 2	% pregnant, wave 3
Contraception at last sex				
None	316	9.8	259	10.8
Hormonal only	52	1.9	30	3.3
Male/female condoms only	190	5.3	210	6.2
Hormonal and condoms	71	7.0	52	1.9
Suspected over-reporters	64	17.2	41	14.6
Total	693	8.4	592	8.3
Fisher's exact test		**p=0.02**		**p=0.05**

Table 3: Prediction of pregnancy with a Poisson working model. Analysis was limited to those who participated in all 3 waves of the survey (n=560). Item non-response on sex past 60 days are those who participated in the wave but did not answer the question how many times they had sex in the past 60 days; for these item non-responders, condom use in the past 60 days was imputed from condom use in the past 14 days for those who answered that question. Predictors of pregnancy are measured in the wave prior to pregnancy.

Variable	Risk ratio (95% CI)	p
Pregnancy at wave 2		
Suspected over-report	3.23 (1.61, 6.45)	***
Birth control pill	0.45 (0.14, 1.43)	
Times sex past 60 days	1.02 (1.01, 1.04)	**
Proportion condom use past 60 days	0.54 (0.25, 1.16)	
Item non-response: sex in past 60 days	0.55 (0.15, 1.99)	
Pregnancy at wave 3		
Suspected over-report	2.21 (0.94, 5.20)	+
Birth control pill	0.17 (0.02, 1.17)	+
Times sex past 60 days	1.03 (1.01, 1.04)	****
Proportion condom use past 60 days	0.60 (0.28, 1.29)	
Item non-response: sex in past 60 days	1.38 (0.58, 3.26)	

Table 4: Number of respondents reporting each type of contraception at last sex, comparing suspected over-reporters with respondents reporting never using condoms in the past 60 days. Respondents can endorse more than one option. Zeroes are not listed

	Wave 1		Wave 2		Wave 3	
	Never condoms (n=150)	Suspected Over-report (n=63)	Never condoms (n=123)	Suspected Over-report (n=40)	Never condoms (n=161)	Suspected Over-report (n=29)
Female condom		2				
Foam		2				
Withdrawal	3					2
Spermicide	1				1	
Condoms	1	44	2	29	1	24
Depo/Norplant/patch	1	4	1	5		3
Oral contraception	18	6	13	3	17	3

Oral contraception is the response to the question of whether they take the birth control pill, asked separately from question about contraception at last sex.