Telling truth from Ys: An evaluation of whether the accuracy of self-reported semen exposure assessed by a semen Y-chromosome biomarker predicts pregnancy in a longitudinal cohort study of pregnancy

Janet E. Rosenbaum, Ph.D.¹,*, Jonathan Zenilman, MD², Johan Melendez, MS², Eve Rose, MPH³, Gina Wingood, ScD³, and Ralph DiClemente, PhD³

1Department of Epidemiology and Biostatistics, School of Public Health, SUNY Downstate Medical Center, Brooklyn, NY, USA
2Department of Infectious Diseases, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA
3Behavioral Sciences and Health Education, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

Abstract

Objectives—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 and predict subsequent pregnancy.

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. We predicted pregnancy from semen exposure under-report using multivariate regression controlling for oral contraception, reported condom use, and coital frequency.

Results—At the 3 surveys, 30%, 20% and 15% of adolescents who reported always using condoms tested positive for semen exposure. At 6 month follow-up, 20.4% and 16.2% of the adolescents who under-reported semen exposure reported pregnancy, a higher pregnancy rate than accurate reporters of semen exposure, even accurate reporters who reported never using condoms (14.2% and 11.8%). Under-reporters of semen exposure were 3.23 (95% confidence interval (1.61,
6.45) times as likely to become pregnant at 6-month follow-up and 2.21 (0.94, 5.20) times as likely to become pregnant at 12-month follow-up as accurate reporters who reported not using contraception, adjusting for self-reported coital frequency.

**Conclusions**—Adolescents who under-report semen exposure may be at uniquely high risk for unplanned pregnancy and STIs, and may also under-report coital frequency. Condom efficacy trials that rely on self-report may yield inaccurate results. Adapted to a clinical setting, the Yc-PCR could alert women to incorrect or inconsistent condom use.

**Keywords**
Adolescent; pregnancy; contraception; DNA amplification

---

**INTRODUCTION**

**Background/rationale**

Adolescents may use condoms inconsistently or incorrectly, and may under-report semen exposure. 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 and examines whether the biomarker predicts subsequent pregnancy.

Only 60% 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 behaviour (2). Inconsistent reporting has been used to identify inaccurate self-reports of sexual behaviour (3), abortions (4), cigarette smoking (5, 6), alcohol, and illegal drug use (5, 7, 8).

Researchers have identified biomarkers for semen exposure to decrease reliance on self-report, including STI diagnoses and detection of substances in semen. Respondents who report 100% condom use but were subsequently diagnosed with a STI may have over-reported their condom use, but this measure has low sensitivity even in high-prevalence populations (9–11). Condom use can be validated using semen exposure biomarkers in the seminal plasma (12), such as prostate specific antigen (PSA) (13, 14), semenogelins, and acid phosphotase, and spermatazoa and other cells present in the semen, such as Y-chromosome (Yc) DNA (15–17). PSA has been studied extensively and can predict pregnancy in condom efficacy trials (18), but PSA is only detectable for 24–48 hours post-coitus (19). The short time frame for 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 measuring semen exposure over longer time periods. Y-chromosome tests seem more sensitive than PSA; Yc is detectable for 14 days post-coitus, even during menses (20–22). One study used the Yc biomarker on archived samples from Baltimore STI clinic patients and found that, among women who reported 100% condom use in the previous two weeks (n=141), 55% tested positive for Yc, suggesting substantial under-reporting of semen exposure (23). Women with a history of STIs were more likely to under-report semen exposure, suggesting self-presentation bias (24).
Objectives

Biomarkers with prognostic value are more useful than biomarkers that only reveal past behaviour. This study is the first to examine whether semen exposure under-reporters were more likely to become pregnant than women who reported their semen exposure accurately, controlling for contraception and coital frequency. We predicted pregnancy instead of STIs from the Yc biomarker because 69–78% of the sample had steady partners with no casual partners in each of the three surveys, so semen exposure is unlikely to predict greater STI risk.

Associations between Yc-PCR and pregnancy could be confounded because participants with other risk factors --- including risky sexual behaviour, smoking, and substance use -- may be more likely to under-report semen exposure and get pregnant. This study identifies potential confounders between Yc-PCR and pregnancy.

Past evaluations of accurate report of sensitive behaviours have used inconsistent survey responses to identify potentially inaccurate responses (2–8). We evaluate whether logically inconsistent reporting of condom use predicts subsequent pregnancy as well as the Y-chromosome test.

METHODS

Study Design

We tested these hypotheses in longitudinal data from a randomized trial of an HIV prevention program (25); we used both experimental and control groups without evaluating the intervention. We predicted pregnancy 6 months after semen exposure under-report, controlling for contraception and coital frequency.

Participants

The 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 (84% of eligible participants.) Among the 715 baseline participants, 607 (84.8%) participated at 6-months, 606 (84.5%) participated at 12-months, and 560 (78.3%) participated at both follow-ups.

Setting

The study enrolled African-American females at a publicly funded STI clinic, a teen clinic based in a large public hospital, and a family planning clinic in Atlanta, Georgia, USA in 2002–2004.

Participants completed a 40 minute interview administered via audio computer-assisted self-interviewing 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.’ The swabs were frozen and shipped to the Johns Hopkins Division of Infectious Disease Laboratory, where they were tested using the Yc-PCR test. Female technicians...
processed all samples to avoid Yc contamination. The Y-chromosome biomarker was evaluated with a polymerase chain reaction (Yc-PCR), which is sensitive to 5 copies of Yc for up to 14 days after coitus. The estimated specificity is 92% (95% CI (80, 98)): 92% of women in the calibration trial who had protected sex tested negative for Yc-PCR, and the remaining 8% had digital or oral genital contact with their male partner, so false positives could be explained by epithelial cells (17, 20, 21).

Follow-up interviews were administered at 6 and 12 months using similar methods. The Institutional Review Board at Emory University approved the entire study protocol prior to implementation. Participants were told that a vaginal swab would be tested “for the presence of sperm to see if you have had sex,” but they were not given results. Participants were paid $50 upon completion of each survey.

Variables

Condom use was measured in the past 60 days, past 14 days, and at last sex. Condom use in the past 60 days was measured by a 2 question sequence: how many times the respondent had vaginal sex in the past 60 days and “How many of the N times you’ve had sex in the past 60 days did you use a condom?” Condom use was categorized in 6 categories: 0%, 1–50%, 51–99%, 100%, no sex in the past 60 days, and missing coital frequency due to item non-response.

The Y-chromosome test is sensitive for up to 14 days, so condom use in the past 14 days was necessary to identify under-reporters of semen exposure. Condom use in the past 14 days was measured identically to condom use in the past 60 days, but it had more missing data because some respondents reported no coitus in the past 14 days. Primary analyses used condom use in the past 60 days to maximize power, which were then repeated with 14 day condom use.

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.

Respondents who reported 100% condom use in the past 14 days but who tested positive for Yc are called under-reporters of semen exposure. We compare under-reporters of semen exposure with accurate reporters of semen exposure.

Respondents who reported 100% condom use in the past 14 days but report not using a condom at last sex are called inconsistent reporters.

Most participants reported serially monogamous relationships: over 80% had steady boyfriends, and only 22–31% reported casual partners in each of the three surveys. Due to the small number of partners, this study focused on pregnancy rather than STI diagnosis because pregnancy can occur even in monogamous relationships with a STI-negative partner.
At 6-month and 12-month specimen collections, participants were asked verbally whether they were currently pregnant; pregnant participants were tested for gonorrhoea and chlamydia via urine test rather than vaginal swab. Current pregnancy was coded as yes, no, or missing from self-reported pregnancy at 6 and 12 months.

Due to the small number of pregnancies (59 and 56 at 6-months and 12-months), biological predictors of pregnancy were considered as potential confounders between semen under-report and pregnancy: oral contraception, contraception at last sex, coital frequency, and non-response to coital frequency.

Additional potential confounders between under-report of semen exposure and pregnancy included smoking; assignment to intervention; age; number of days between surveys; substance use (lifetime and past 60 day alcohol and marijuana use, number of alcoholic drinks when drinking); substance use and sex (beliefs about the safety of sex while drinking alcohol, sex while high on drugs or alcohol); and abuse (emotional and physical abuse in lifetime, last 60 days, by boyfriend or casual sex partners, history of forced vaginal or anal sex.)

**Statistical methods**

Data were analysed using Stata SE 11.1—We chose to use non-parametric statistical tests because they require fewer distributional assumptions than the standard parametric tests. We tested differences between non-ordered groups using the Pearson chi-squared test. We compared groups using the Mann-Whitney-Wilcoxon rank-sum test. We tested differences between ordered groups using Cuzick’s test for trend, a generalization of Wilcoxon’s signed rank test (26).

The outcome of pregnancy was reported by 10% of the sample at 6- and 12-month follow-ups. Estimators from logistic regression are inconsistent for non-rare outcomes, so a logistic regression with the outcome of pregnancy would yield inconsistent estimators. We estimated relative risks using a Poisson working model, which yields consistent and unbiased estimators (27, 28). Control variables were oral contraceptive use, number of coital episodes, condom use, and a binary variable for non-report of condom use, all in the past 60 days. Regressions were limited to participants in both follow-ups to ensure comparable groups (n=560).

Missing Data—Missing data can bias results if data is differentially missing. To evaluate whether semen exposure under-reporters differed in loss to follow-up, we used the Pearson chi-squared test. We used an indicator for missing coital frequency in multivariate regressions.

**RESULTS**

Main results

Under-report of semen exposure—At baseline, 29% of participants who reported coitus in the past 14 days (n=537) reported having used condoms for all coitus in the past 14 days; among these women, 30% tested positive for Y-chromosome (Yc), suggesting under-
report of semen exposure in the past 14 days. At 6-months, 45% of participants who reported coitus in the past 14 days (n=445) reported having used condoms for all coitus in the past 14 days, but 20% tested positive for Yc. At 12-months, 44% of participants who reported coitus in the past 14 days (n=434) reported having used condoms for all coitus in the past 14 days, but 15% tested positive for Yc.

Intervention participants were not more likely than control to report consistent condom use at baseline (p=0.8) and 6-months (p=0.8), but were more likely at 12-months (50.0% versus 39.0%, p=0.02).

We stratified accurate reporters of semen exposure by reported condom use in the past 60 days. Under-reporters of semen exposure were more likely to report subsequent pregnancy than all strata of accurate reporters, including women who reported never using condoms (Figure 1). Among baseline under-reporters, 20% reported pregnancy at 6-months; among under-reporters at 6-months, 16% reported pregnancy at 12-months.

We stratified accurate reporters of semen exposure by reported contraceptive use at last sex. Under-reporters of semen exposure were more likely to report pregnancy than all strata of accurate reporters, including those reporting no contraception (Figure 2). After adjusting for contraception and frequency of sex, under-reporters of semen exposure had a risk ratio for pregnancy of 3.23 (1.61, 6.45) at 6-months and 2.31 (0.94, 5.20) at 12-months (Table 1).

To examine whether greater prevalence of pregnancy among under-reporters of semen exposure could be attributed to greater coital frequency, we compared self-reported coital frequency of under-reporters with accurate-reporters, stratified by condom use in the past 60 days. Under-reporters of semen exposure reported fewer episodes of coitus in the past 60 days than four of five strata of accurate reporters at baseline and 6-months (not shown).

To examine whether greater prevalence of pregnancy among under-reporters of semen exposure could be attributed to contraception other than male condoms and oral contraception, we enumerated contraception used at last sex. Few respondents who reported not using condoms reported using other forms of contraception at last sex (Table 2).

Other risk factors and the semen exposure biomarker

Using non-parametric statistical test, under-reporters of semen exposure did not differ at median from accurate reporters of semen exposure in age, number of days between surveys, substance use, substance use in sex, and past or current abuse (not shown).

Under-reporters of semen exposure were more likely to report being cigarette smokers than accurate reporters of semen exposure. The proportion of smokers was inversely proportional to condom use category, with semen exposure under-reporters having the highest smoking prevalences (Cuzick’s test for trend p=0.01, p=0.05, p=0.01 in the respective time periods). Smoking more cigarettes per day predicted a greater chance of positive Yc- PCR among all respondents (Cuzick’s test for trend p=0.02, 0.000, 0.000 at baseline, 6-months, and 12-months) 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 6 months later.
Semen exposure under-reporters were over-represented in the treatment versus the control group at 12-months (6.6% vs. 3.2%, Wilcoxon p=0.05), but not at 6-months.

Under-reporters of semen exposure were not more likely to drop out of the study at 6-months and 12-months (Chi-squared p=0.9 and p=0.2).

**Inconsistent reporting of condom use**

Some respondents reported condom use inconsistently within the same survey: reporting sex within the last 14 days with 100% condom use, but no condom at last sex. At baseline, 6-months, and 12-months, inconsistent reporters of condom use comprised 24%, 22%, and 19% of those reporting 100% condom use in the past 14 days. Inconsistent reporting of condom use was not significantly correlated with under-reporting of semen exposure (Pearson’s correlations 0.09, 0.06, and −0.03; p=0.2, 0.4, 0.6). Inconsistent reporters of condom use were not more likely to get pregnant, controlling for oral contraception, coital frequency, and condom use (RRs 0.55 (0.14, 2.01), 1.87 (0.80, 3.96)).

**DISCUSSION**

**Key results**

Fifteen to 30% of adolescents who reported condom use during every vaginal sex act in the past 14 days tested positive for semen Y-chromosome, suggesting under-report of semen exposure. Adolescents who under-reported semen exposure were 2–3 times more likely to be pregnant 6 months later than adolescents who reported semen exposure accurately. All participants stated that they did not want to become pregnant, implying that these pregnancies were unplanned.

**Interpretation**

Under-report of semen exposure could be due to intentional or unintentional incorrect or inconsistent condom use. Incorrect or inconsistent use alone cannot explain why adolescents who under-report their semen exposure have a higher pregnancy rate than accurate reporters who report no condom use. The higher pregnancy rate among under-reporters than accurate reporters who do not use condoms is also not attributable to other contraception use because few condom non-users reported other contraceptive methods. Under-reporters of semen exposure also reported having had sex fewer times in the last 60 days than accurate reporters who reported never using condoms.

The higher pregnancy rate among under-reporters could be explained if under-reporters of semen exposure also under-reported vaginal intercourse. This explanation is consistent with past evidence that under-reporting of semen exposure is attributable to self-presentation bias (24). The data do not explain why under-reporters of semen exposure would under-report coitus; compared with accurate reporters of semen exposure, under-reporters do not report more frequent or greater drug or alcohol use, use during sex, or greater acceptance of sex under the influence.

This finding may have implications for studies of condom effectiveness that do not measure semen exposure. Women who become pregnant during FDA-mandated studies of condom
effectiveness may have used condoms inconsistently and under-reported their semen exposure. Condom trials that exclude biomarkers could cause condoms to appear to fail and suggest a serious methodological problem with condom trials that rely on self-reported condom use and coital frequency.

Smokers were more likely to under-report semen exposure and to test positive on the Yc-PCR regardless of reported condom use, and their 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 smoking predicts lower physiological arousal (29, 30), which could increase the length of time that Y-chromosome is detectable in the vagina. Alternatively, smokers could be more likely to under-report both semen exposure and coital frequency and to have unprotected sex.

Under-report of semen exposure 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. Taking the survey may have made respondents attend to condom use consistency and correctness, which may explain the decrease in under-reported semen exposure, but the decrease is not attributable to under-reporters leaving the study at greater rates. More intervention participants under-reported semen exposure at 12-months, which may suggest that the intervention induced self-presentation bias similar to that among respondents with previous STI diagnoses (24).

Inconsistent reporting of risk behaviours has been a widespread marker of invalid survey responses (2–8), but this analysis demonstrated that the Yc-PCR has greater prognostic value: inconsistent reporting of condom use does not predict later pregnancy, whereas under-report of semen exposure does.

**Limitations**

The research measured under-report of semen exposure but not breakage, slippage, late application, early removal, or leakage, so it could not discriminate between inconsistent and imperfect condom use.

Pregnancy may have been under-reported because respondents may have terminated pregnancies between survey waves. Under-reporting of known current pregnancies is likely negligible because participants were told that pregnant women would have chlamydia and gonorrhoea tests via urine test rather than vaginal swab to protect fetal health. If pregnancy were under-reported, we would expect under-reporters of semen exposure also to under-report pregnancy, and the true association between the Yc-PCR and pregnancy would be greater than found in this study.

**Generalisability**

These results may generalise to similar urban African-American female adolescents who seek care at sexual health clinics.
Conclusions

In this sample, 15–30% of women who reported consistent condom use tested positive for semen Y-chromosome, suggesting under-report of semen exposure. Under-reporters of semen exposure were 2–3 times more likely to become pregnant than accurate reporters who reported never using condoms; this excess risk could be partially due to under-reporting of coital frequency.

Although not currently suitable for clinical settings due to long processing times, future technological improvements could allow the Yc-PCR to be used in clinical settings to alert women to incorrect or inconsistent condom use and help prevent STIs.

Acknowledgments

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, the American Public Health Association Statistics Section, and the working group on semen exposure biomarkers.

FUNDING: This research was funded by the Sexually Transmitted Diseases Division of the Centers for Disease Control and Prevention (T-32 AI050056, Dr. Zenilman), National Institute for Child Health and Human Development (R24-HD041041, Maryland Population Research Center), and the SUNY Downstate School of Public Health. 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)

References


Key messages

• 15–30% of adolescents who reported consistent condom use tested positive for semen Y-chromosome from a vaginal swab, suggesting under-report of semen exposure.

• Adolescents who under-reported semen exposure were 2–3 times more likely than other adolescents to become pregnant 6 months later, controlling for coital frequency and use of other contraception.

• Under-reporters of semen exposure may also under-report coital frequency, suggesting that condom efficacy studies that use semen exposure biomarkers are more accurate.
Figure 1. Percent currently pregnant by condom use in the past 60 days (n=560 respondents at both follow-ups)

“Positive Yc test” reported 100% condom use and tested positive for semen Y-chromosome, suggesting semen exposure in the past 14 days. “100% condom use” reported 100% condom use and tested negative for semen Y-chromosome. Missing participated in the survey but did not report coital frequency. P is from Cuzick’s non-parametric test for trend. “n” are the number who endorsed each item at baseline and 6-month follow-up, respectively.
Figure 2. Percent currently pregnant, by contraception use at last sex

Hormonal contraception includes oral contraception as well as hormonal implants, injections, and patches. “Positive Yc test” reported 100% condom use and tested positive for semen Y-chromosome, suggesting semen exposure in the past 14 days. P is from Fisher’s exact test.
Table 1
Prediction of pregnancy at 6-months and 12-months with multivariate regression (n=560).

<table>
<thead>
<tr>
<th>Outcome: Pregnancy at 6- months</th>
<th>Risk Ratio (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under-report semen exposure</td>
<td>3.23 (1.61, 6.45)</td>
<td>***</td>
</tr>
<tr>
<td>Oral contraception</td>
<td>0.45 (0.14, 1.43)</td>
<td></td>
</tr>
<tr>
<td>Coital frequency past 60 days</td>
<td>1.02 (1.01, 1.04)</td>
<td>**</td>
</tr>
<tr>
<td>Proportion condom use past 60 days</td>
<td>0.54 (0.25, 1.16)</td>
<td></td>
</tr>
<tr>
<td>Item non-response: coital frequency</td>
<td>0.55 (0.15, 1.99)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome: Pregnancy at 12- months</th>
<th>Risk Ratio (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under-report semen exposure</td>
<td>2.21 (0.94, 5.20)</td>
<td>+</td>
</tr>
<tr>
<td>Oral contraception</td>
<td>0.17 (0.02, 1.17)</td>
<td>+</td>
</tr>
<tr>
<td>Coital frequency past 60 days</td>
<td>1.03 (1.01, 1.04)</td>
<td>****</td>
</tr>
<tr>
<td>Proportion condom use past 60 days</td>
<td>0.60 (0.28, 1.29)</td>
<td></td>
</tr>
<tr>
<td>Item non-response: coital frequency</td>
<td>1.38 (0.58, 3.26)</td>
<td></td>
</tr>
</tbody>
</table>

* p ≤0.1,
* p ≤0.05,
** p ≤0.01,
*** p ≤0.001,
**** p ≤0.0001

Item non-response are respondents who did not report coital frequency 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 were measured in the survey prior to pregnancy.
Table 2

Number of respondents reporting each type of contraception at last sex, comparing under-reporters of semen exposure with respondents reporting no condom use in the past 60 days.

<table>
<thead>
<tr>
<th></th>
<th>Baseline (n=63)</th>
<th>6-months (n=150)</th>
<th>12-months (n=123)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Semen exposure under-report</td>
<td>No condoms</td>
<td>Semen exposure under-report</td>
</tr>
<tr>
<td>Female condom</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Foam</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Withdrawal</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Spermicide</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Male condom</td>
<td>44</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>Depo/Norplant/Patch</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Oral contraception</td>
<td>6</td>
<td>18</td>
<td>3</td>
</tr>
</tbody>
</table>

Oral contraception is measured by a separate item from contraception at last sex.

Respondents can endorse more than one option. Zeroes are not listed.