Published in final edited form as:

Ann Epidemiol. 2016 September ; 26(9): 612–618.e2. doi:10.1016/j.annepidem.2016.07.011.

The Validity of Self-Reported Behaviors: Methods for Estimating Underreporting of Risk Behaviors

Marlena S. Norwood¹, James P. Hughes^{1,2}, and K. Rivet Amico³

¹University of Washington, Department of Biostatistics, Seattle, WA

²Fred Hutchinson Cancer Research Center, Seattle, WA

³University of Michigan, Department of Health Behavior and Health Education, School of Public Health, Ann Arbor, MI

Abstract

Purpose—When individuals underreport risk behaviors, data gathered from public health research and practice will underestimate risk. To date, there is little guidance on if or how reports can be adjusted to better reflect true levels of a risk behavior in a given cohort, sample or, by extension, population.

Methods—We develop the underreporting correction factor (UCF), which can be used to correct estimates of the prevalence of a risk behavior using self-report of the behavior and a specific (but not necessarily sensitive) biomarker. The UCF rests on three assumptions: (1) there is no overreporting of the behavior (2) the biomarker can only be acquired if the person engages in the behavior, and (3) the presence of the biomarker does not affect reporting of the behavior. We investigate the sensitivity of the UCF to violation of these assumptions and develop confidence intervals for the UCF and the corrected prevalence of the behavior.

Results—The UCF is most sensitive to the second assumption (biomarker perfectly specific). We apply the UCF to estimates of sexual risk behaviors in various settings using a variety of biomarkers.

Conclusions—Implementation of the UCF corrects for underreporting and more accurately quantifies risk in cohorts.

BACKGROUND

Behavioral and epidemiological research often rely on self-reports of risk behaviors (e.g. condomless sex, injection needle sharing). However, people may underreport these behaviors due to their stigmatizing or “undesirable” aspects.^{1, 2} In HIV research, for

Contact Information for Corresponding Author: James P. Hughes, PhD, Department of Biostatistics, University of Washington 357232, Seattle, WA 98195, T: 206-616-2721, F: 206-616-2724, jphughes@uw.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflicts of Interest The authors declare no conflict of interest.

example, self-reported condomless sex is often underreported^{3, 4} due to social desirability bias or individuals being unaware of condom malfunction.⁵ Underreporting of such behaviors may lead to bias (typically, attenuation) of the relative risk associated with that behavior.⁶

Attempts to assess the validity of self-reported risk sexual behaviors have included interviewing sexual partners for discordant responses,⁷ interviewing the same individual at multiple time points^{8–11} and using specific data collection strategies, such as computer administered surveys^{12, 13} neutral interviewing strategies,^{14, 15} and randomized response techniques¹⁶. Increasingly, researchers are using biomarkers, such as prostate-specific antigen (PSA), to assess the validity of self-reported condomless sex.¹⁷ The proportion of individuals who do not report condomless sex but who test PSA-positive is used as a measure of the validity of self-report.⁶ The main drawback of this approach is that it relies heavily on the sensitivity of the biomarker. For PSA, sensitivity declines quickly over the 48 hours following intercourse, making it impossible to identify underreporting over longer intervals.¹⁸

Statistical approaches to adjusting for measurement error using a gold standard (e.g. regression calibration^{19, 20}) have been used to adjust risk-outcome associations for measurement bias, but also require a biomarker with high sensitivity and specificity. STIs, for example, are imperfectly sensitive as markers for condomless sex since one may engage in the behavior but not become infected with an STI. Our approach relaxes the assumption of perfect sensitivity in favor of alternative assumptions (see below) that are particularly relevant to the setting of self-reports of stigmatizing behaviors.

We propose an alternative method for quantifying underreporting by comparing biomarker status in those who report the risk behavior versus those who do not report the risk behavior. First, we develop an index of underreporting, the “underreporting correction factor” (UCF), that can be used to quantify underreporting and correct estimates of the risk behavior prevalence at the group or cohort level. Next, we detail the underlying assumptions of the UCF, test its sensitivity to those assumptions, and evaluate its finite sample size bias. Finally, we apply the UCF to data from completed studies. We conclude with recommendations for the application of the UCF in research and practice.

METHODS

The Underreporting Correction Factor

The Underreporting Correction Factor (UCF) is calculated with reference to a reported behavior (**R**), true behavior (**T**), and biomarker (**B**: e.g., pregnancy, PSA, HSV-2). Each measure is characterized as present (1) or absent (0) for each participant. The true behavior is, of course, unobserved. The observed data on the biomarker and the behavior can be represented as a 2×2 table:

		<i>Biomarker</i>	
		present	absent
<i>Reported</i>	present (yes)	a	b

Behavior	absent (no)	<hr/>	
		c	d

Underreporting can be quantified in terms of the probability that an individual truly engaged in a behavior given that they report *not* having engaged in that behavior, i.e. $P(T=1|R=0)$. In the context of condomless sex, $P(T=1|R=0)$ would be the probability that an individual had condomless sex given he/she reports no condomless sex.

Given certain assumptions (see below) $P(T=1|R=0)$ can be expressed in terms of the *observed* measures R and B (see appendix A, [section S1 for proof](#)):

$$P(T=1|R=0) = \frac{P(B=1|R=0)}{P(B=1|R=1)}. \quad (1)$$

We refer to $\frac{P(B=1|R=0)}{P(B=1|R=1)}$ as the 'underreporting correction factor' (UCF). The UCF may be estimated as:

$$\widehat{UCF} = \frac{\frac{c}{c+d}}{\frac{a}{a+b}} \quad (2)$$

The validity of equation (1) depends on three assumptions:

1. There is no overreporting of the behavior: all people that report the behavior truly engaged in the behavior. For example, if a participant reports condomless sex, he or she did indeed have condomless sex.
2. The biomarker is perfectly specific for the behavior: the biomarker can only be acquired if the person engages in the behavior. For example, the probability of pregnancy in the absence of condomless sex is 0.
3. The presence of the biomarker does not affect reporting of the behavior. For example, early pregnancy or having an STI does not increase the likelihood of reporting condomless sex.

$$P(T=0|R=1)=0$$

$$P(B=1|T=0)=0$$

$$P(R=0|T=1, B=1) = P(R=0|T=1, B=0)$$

Note that, if these three assumptions hold, equation (1) is valid even if the biomarker is not 100% sensitive (i.e. if $P(B=1|T=1) < 1$). For large samples, a 95% confidence interval (CI) for the UCF can be computed by exponentiating:²¹

$$\log \widehat{UCF} \pm 1.96 * \sqrt{\frac{1}{c} + \frac{1}{a} - \frac{1}{c+d} - \frac{1}{a+b}} \quad (3)$$

(where the upper bound of the CI is truncated at 1.0, if necessary).

Importantly, the UCF can be used to provide a corrected estimate of the true prevalence of the risk behavior. The reported prevalence of the risk behavior is $P(R=1)$, and the true prevalence of the risk behavior is $P(T=1)$. Then (see appendix A, [section S2](#))

$$P(T=1) = P(R=1) + P(R=0) * UCF \quad (4)$$

and $P(T=1)$ can be estimated as

$$\hat{P}(T=1) = \frac{(a+b)}{N} + \frac{(c+d) * \widehat{UCF}}{N} = \frac{a+b}{N} \left(1 + \frac{c}{a}\right) \quad (5a)$$

where $N = a + b + c + d$. In other words, the UCF can be used to correct the reported risk prevalence and estimate the true level of risk in the population. A 95% confidence interval (based on the delta method²²) for the true prevalence of the risk behavior is given by:

$$\hat{P}(T=1) \pm 1.96 * \sqrt{\frac{(a+b)(a+c)(bc(N-a) + a^2d)}{(aN)^3}} \quad (5b)$$

Note that equation (5a) could be applied in a setting where only the reported behavior (with no biomarker) is available if an estimate of the UCF that is appropriate to that population is available from a previous study.

Use of the UCF is most straightforward when assumption 2 (perfectly specific biomarker) is met. If the specificity of the biomarker is less than 100% but known, an adjusted UCF may be obtained using methods described in appendix B, section S1. Further, if the lack of specificity is caused by a lag (of known duration) between performance of the behavior and detectability of the biomarker, the specificity can be estimated using methods described in appendix B, section S2.

Evaluating Sensitivity to Assumptions

We investigated the impact of violations of assumptions 1 – 3 by computing the UCF for fixed values of $P(T=1|R=0)$ and known levels of violation of assumptions 1–3. The difference between the UCF and the true value of $P(T=1|R=0)$ is the bias introduced by violation of a given assumption.

Evaluating Finite Sample Bias and Calculating Confidence Interval Probabilities

With an infinite sample size, $\widehat{UCF}$ is an unbiased estimate of the UCF. However, when applied to smaller sample sizes $\widehat{UCF}$ may be biased even if assumptions 1 – 3 are met. We evaluated this finite sample bias for $\widehat{UCF}$ and the corrected prevalence of the behavior (equation 5a) by simulating datasets (a dataset is a table of $R \times B \times T$) with all combinations of the following characteristics: sample size ($N = 600$ and 150), biomarker sensitivity ($P(B = 1|T = 1) = 0.3$ and 0.075), and behavior prevalence (common or uncommon). For each of these eight combinations, we simulated three different true values of $P(T = 1|R = 0)$ (0.1 , 0.3 , and 0.7) (i.e. different amounts of underreporting), resulting in a total of 24 scenarios. Each scenario was simulated 10,000 times. Bias is computed as the difference between the average of $\widehat{UCF}$ (over the 10,000 simulations) and the true UCF.

Using the simulated datasets, we also evaluate two methods of calculating confidence intervals for the UCF - the log relative risk confidence interval (equation 2) and the Koopman asymptotic score²¹ to assess the relative accuracy of the two methods under the different scenarios. Additionally, we evaluated the confidence interval for the corrected estimate of risk (equation 5b). We compare the observed coverage of each approach to the nominal level (95%).

Sources of Data

HIV Prevention Trials Network (HPTN) 068.²³—Study participants were young South African women in high school who agreed to participate in a randomized clinical trial of conditional cash transfer for HIV prevention. Sexual behavior was measured using ACASI. At enrollment, participants were asked if they ever had (vaginal or anal) sex and were screened for HSV-2, a biomarker for unprotected sex.

HPTN 039.²⁴—Study participants consisted of HIV-uninfected, HSV-2 positive women from Zambia, Zimbabwe, and South Africa. In quarterly face-to-face interviews over eighteen months, participants were asked about condom use and sexual behavior over the last three months and were tested for pregnancy. The data presented here are from the cross-sectional time points at three month intervals. Women were classified as having an incident pregnancy if they tested positive at the current quarterly visit and tested negative for pregnancy their previous quarterly visit.

Mombasa Cohort/Lifecourse Study.²⁵—Study participants were female sex workers enrolled in a prospective cohort study in Mombasa, Kenya. At enrollment, participants were asked about unprotected sex in the past week and were tested for PSA, a biomarker for unprotected sex.

Other sources of data are from articles about the validity of self-reported behaviors published in peer-reviewed journals found through a search of the *PubMed* database. We selected studies for analyses if they cross-tabulated sexual behavior (or quantitative measurements that could be split into yes/no categories, e.g. number of condomless sexual encounters), and positive/negative biomarker status.

All data analyses and simulations were conducted using R (version 3.1.3).

RESULTS

Sensitivity to Assumptions

Assumption 1—In assumption 1 we assume no overreporting of the risk behavior. In many, but not all, areas of interest the reported behavior is unidirectional in departure from the “true” behavior. For example, when the behavior has negative desirability (e.g. condomless sex among people living with HIV or individuals subject to frequent risk reduction counselling in a trial), the true behavior is likely to be underreported. Exceptions to this are possible, e.g. individuals believing that unprotected sex is required for study participation.¹²

Figure 1A shows the effect of violations of assumption 1. As overreporting increases, the UCF becomes slightly biased. However, the bias is small for low amounts of overreporting. The bias from overreporting does not depend on the sensitivity of the biomarker or the underlying prevalence of the behavior (not shown in figure), but it does depend on the true value of $P(T=1|R=0)$. As $P(T=1|R=0)$ increases, overreporting introduces more bias; for example, when overreporting is equal to 0.5 ($P(T=0|R=1)=0.5$), the bias in the UCF varies from 0.1 (when $P(T=1|R=0)=0.1$) to 0.4 (when $P(T=1|R=0)=0.5$).

Assumption 2—Assumption 2 states that the biological marker is perfectly specific. However, some biomarkers are intrinsically non-specific with regard to the behavior of interest (e.g. HIV is not a perfectly specific biomarker for unprotected sex since it can also be acquired by injection drug use, or perinatally). Even biomarkers that appear perfectly specific (e.g. pregnancy as a biomarker for unprotected sex) may violate assumption 2 due to a lag between the time of the risk behavior and the time of a positive biomarker test (e.g. the lag between condomless sex and positivity of a pregnancy test). In this case the biomarker (e.g. positive pregnancy test in the past 3 months) may not precisely reflect the behavior (e.g. unprotected sex in the past 3 months) (see Fig. A.1 and appendix B, section S2).

Figure 1B simulates violations of the second assumption. Bias due to an imperfectly specific biomarker depends on the sensitivity of the biomarker. If the sensitivity of the biomarker is low, then the bias in the UCF is very large even with small deviations from perfect specificity. For example, at a low sensitivity ($P(B=1|T=1)=0.25$), and a small deviation from perfect specificity ($P(B=0|T=0)=0.9$), the UCF would equal approximately 0.6 when the true value is 0.1, a bias of 0.5. As the sensitivity of the biomarker increases, the bias declines. Additionally, for a given specificity, the bias of the UCF decreases as the true underreporting, $P(T=1|R=0)$, increases.

Assumption 3—The third assumption – that the presence of the biomarker does not affect reporting of the behavior – is reasonable for biomarkers like PSA whose presence is unknown to the individual. However, the assumption may be violated for biomarkers that are visible or affect the health of the individual (e.g. pregnancy or symptomatic STI) and may deter underreporting of the behavior. For example, a visibly pregnant woman is probably less likely to report that she has not had condomless sex.

Figure 1C simulates violations of the third assumption. In this figure Φ is a quantity between 0 and 1 that represents the differential reporting among participants with and without the biomarker. Specifically,

$$\Phi = \frac{P(R=0|T=1, B=1)}{P(R=0|T=1, B=0)}$$

If the third assumption is met, $\Phi = 1$. A smaller Φ represents less underreporting among participants with the biomarker. As Φ declines from 1, the UCF is negatively biased. Additionally, as $P(T=1|R=0)$ increases, the UCF is more negatively biased at a given value of Φ . Overall, violation of assumption 3 (when $\Phi < 1$) leads to partial, rather than complete, adjustment for underreporting – the adjusted true prevalence of the risk behavior will be greater than $P(R=1)$ but less than $P(T=1)$ (in contrast, violations of assumptions 1 and 2 introduce bias that is unbounded and over-estimates $P(T=1|R=0)$).

Finite Sample Bias and Confidence Interval Coverage

Table 1 presents finite sample bias and confidence interval coverage probabilities for the $\widehat{UCF}$ and corrected prevalence of the behavior, $\hat{P}(T=1)$ for 8 scenarios ($N = 100$ or 600 , behavior prevalence common or uncommon, and biomarker sensitivity .075 or .3) and three values of $P(T=1 | R=0)$ (0.1, 0.3, 0.7). For “big N,” ($N = 600$), the bias in $\widehat{UCF}$ is very small, especially for low to moderate values of $P(T=1|R=0)$ (0.1 and 0.3). The bias increases, but is still minimal, for sample sizes of about 150. The bias in the corrected prevalence of behavior mirrors the pattern in the bias of $\widehat{UCF}$ – minimal even with small sample sizes.

The log CI coverage met or exceeded the nominal level of 95% coverage in all simulations with the exception of dataset 2 and 6: when the sample size of the group reporting the behavior is larger than the sample size of the group not reporting the behavior, and the biomarker is insensitive. The Koopman CI performed well in all simulations, giving a nominal or conservative CI. We recommend the log interval in most cases since it is simple to compute and gives at least nominal coverage. For datasets with similar qualities to dataset 2 and 6, we recommend the Koopman CI. For extremely small sample sizes (not modeled here), the asymptotic methods that we have studied may not be appropriate. Instead, we recommend an exact confidence interval for the ratio of independent binomial proportions (see, for example, the Chan-Zhang CI²¹).

The CI coverage for the corrected prevalence of the behavior (equation 5b) varied based on the parameters of the simulated dataset and the true value of $P(T=1|R=0)$. In some cases, the CI had 100% coverage, and in many it was nominal or conservative. However, in datasets 2, 4, 6, and 8 – all scenarios where the sensitivity of the biomarker was very low (0.075) – the CI did not achieve 95%. Therefore, we advise caution in the interpretation of the CI for the corrected prevalence of the behavior when the biomarker has extremely low sensitivity.

Examples Using Completed Trials Data

Table 2 presents the UCF (equation 2) and the true prevalence of the risk behavior (equation 5a) for several datasets.

We find varying levels of underreporting. In the HPTN 068 population of high school girls from South Africa the $\widehat{UCF}$ for “ever had sex” was about 14% (95% CI: 10% – 22%) using HSV-2 as a (specific, but <100% sensitive) biomarker. The estimated prevalence of “ever had sex” increased from 27% to 37% (95% CI: 33% – 42%) after adjustment for underreporting. Using the same biomarker in a study of young women in Kampala, Uganda, the $\widehat{UCF}$ for “ever had sex” was 48% (95% CI: 33% – 69%) and the estimated prevalence of “ever had sex” rose from 90% to 95% (95% CI: 93% – 97%). Underreporting of unprotected sex in the last 48 hours using PSA as a biomarker for female sex workers in Guinea, Mombasa, and Madagascar ranged from about 40% to almost 60% and the estimated true prevalence of unprotected sex increased notably in all cases after correction for underreporting. Using the Y-chromosome biomarker in a sample of women in the Baltimore Health Department STD clinic, the UCF for “unprotected sex in the past 2 weeks” was 79% and the estimated prevalence of unprotected sex in the past two weeks increased from (reported) 65% to (true) 93% (95% CI: 83% – 100%).

Table 3 presents an analysis of cross-sectional data collected at multiple time points in HPTN 039, with pregnancy as the biomarker. Because the women in this study were tested for pregnancy every three months, most of the pregnancies were sub-clinical. As a result, many of the women may not have known they were pregnant, so the third assumption (that the presence of the biomarker does not affect reporting of the behavior) may be reasonable in this context. To the extent that assumption 3 is violated, the UCF and the prevalence of unprotected sex are likely underestimated. We applied the UCF to these data with adjustment for lags (between the time of sex leading to pregnancy and first positive pregnancy test) of various durations, as described in appendix B, sections S1, S2.

Over the eighteen-months, assuming no lag between unprotected sex and a positive pregnancy test, the $\widehat{UCF}$ stayed relatively constant, varying from 20% to about 30%. When we accounted for a lag time of 10 days and 15 days between unprotected sex and a positive pregnancy test the $\widehat{UCF}$ decreased to about 12–29% and 8–25%, respectively. The reported prevalence of unprotected sex over the eighteen-month study ranged from 75% to 81 %, which was only slightly increased with application of the UCF: 77% to 84% over the eighteen months (using a lag of 15 days).

DISCUSSION

The UCF provides an index of underreporting of self-reported risk behaviors and can be used to generate a corrected estimate that better approximates the true prevalence of the risk behavior for a sample. When applied to existing datasets, the UCF was often large, suggesting that the actual degree of risk behavior may be substantially underestimated.

The UCF depends on three assumptions: no overreporting of the behavior, perfect specificity of the biomarker, and self-report not dependent on biomarker. If these assumptions hold,

then the UCF does not depend on the sensitivity of the biomarker. Although overreporting can induce significant bias, we believe that in practice the degree of overreporting will be small (and primarily due to imperfect recall) so that violations of the first assumption will not likely lead to significant bias. Similarly, violations of the third assumption (self-report not dependent on biomarker) are not as problematic because $\hat{P}(T=1)$ will be bounded between $P(R=1)$ and $P(T=1)$. However, low sensitivity biomarkers depend heavily on the perfect specificity assumption and large bias can result if this assumption is violated. If the specificity of the biomarker can be estimated, however, an adjusted estimate of the UCF is possible.

The UCF can, potentially, be extrapolated from one sample (used to generate the UCF) to a similar sample that collects behavioral data but lacks an objective biomarker. Clearly, such extrapolation must be done with caution, but may be warranted given similar populations and factors leading to underreporting of risk behaviors.

The UCF may be useful in other areas of research that utilize self-reports subject to underreporting and for which highly specific binary biomarkers are available, such as medication adherence²⁶ or prenatal smoking exposure.²⁷ Our approach may also be used to correct for overreporting by simply reversing the outcome (i.e. instead of considering overreporting of protected sex, we evaluate underreporting of unprotected sex).

Other statistical approaches to adjusting for measurement error using a gold standard (e.g. regression calibration^{19, 20}) similarly require a biomarker, although both sensitivity and specificity need to be high. The approach we propose relaxes the assumption of perfect sensitivity, and, as we demonstrate, can be implemented on sample data where self-report of stigmatizing behaviors and biological (observed) outcome are collected concurrently.

Several limitations in our work should be noted. We have discussed the dependence of the UCF on the three assumptions mentioned above. Although we can adjust the UCF for imperfectly specific biomarkers and adjust our interpretation of the UCF for violations of the third assumption, we are unable to adjust the UCF for violations of the first assumption. However, as shown in the analysis of assumption one (see Results: Sensitivity to Assumptions), a notable proportion (>10%) of the population would have to overreport their behavior for this to introduce a meaningful bias into the UCF.

Use of the UCF to estimate the corrected prevalence of risk could improve the accuracy of estimates of behavior based on self-reports. When risk behavior assessments from behavioral surveys are used to allocate resources to promote safe/safer behaviors, application of the UCF may more accurately approximate true rates of risk in the population and thus better guide community, agency and public health responses.

Acknowledgments

Source of Funding: Supported by the National Institute for Allergy and Infectious Disease (2R01AI029168) and the National Institute of Child Health and Human Development (R01HD072617).

APPENDIX A

S1: Derivation of the Underreporting Correction Factor

Suppose

R = individual's reported behavior

T = individual's true behavior

B = presence of biomarker

Assume

1. $P(R = 1|T = 0) = 0$ which implies $P(T = 1|R = 1) = 1$
2. $P(B = 0|T = 0) = 1$
3. $P(R = 0|T = 1, B = 0) = P(R = 0|T = 1, B = 1)$

Under assumptions 1 – 3,

$$P(T=1|R=0) = \frac{P(B=1|R=0)}{P(B=1|R=1)}$$

Proof:

Using the basic rules of probabilities,

$$P(T=1|R=0) = \frac{P(B=1, T=1|R=0)}{P(B=1|T=1, R=0)} = \frac{P(B=1|R=0) * P(T=1|B=1, R=0)}{P(B=1|T=1, R=0)}.$$

However, by Bayes' Rule,

$$P(T=1|B=1, R=0) = \frac{P(B=1|T=1, R=0) * P(T=1|R=0)}{P(B=1|T=1, R=0) * P(T=1|R=0) + P(B=1|T=0, R=0) * P(T=0|R=0)}$$

and assuming that the specificity of the biomarker is perfect (assumption 2), $P(B = 1|T = 0, R = 0) = 0$ so $P(T = 1|B = 1, R = 0) = 1$. Thus,

$$P(T=1|R=0) = \frac{P(B=1|R=0)}{P(B=1|T=1, R=0)}. \quad (\text{A.1})$$

Note: if the specificity of the biomarker is not perfect, the above equation does not hold - see appendix B, section 1.

Further,

$$P(B=1|R=1) = P(B=1|T=1, R=1) P(T=1|R=1) + P(B=1|T=0, R=1) P(T=0|R=1)$$

but since, by assumption 1, $P(T=1|R=1) = 1$, we have

$$P(B=1|R=1) = P(B=1|T=1, R=1). \quad (\text{A.2})$$

However, using assumption 3

$$\begin{aligned} P(R=0|T=1, B=0) &= P(R=0|T=1, B=1) \\ \frac{P(R=0|T=1, B=0)}{P(R=1|T=1, B=0)} &= \frac{P(R=0|T=1, B=1)}{P(R=1|T=1, B=1)} \\ \frac{P(B=0|T=1, R=0) P(R=0|T=1)}{P(B=0|T=1, R=1) P(R=1|T=1)} &= \frac{P(B=1|T=1, R=0) P(R=0|T=1)}{P(B=1|T=1, R=1) P(R=1|T=1)} \\ \frac{P(B=0|T=1, R=0)}{P(B=0|T=1, R=1)} &= \frac{P(B=1|T=1, R=0)}{P(B=1|T=1, R=1)} \\ P(B=1|T=1, R=0) &= P(B=1|T=1, R=1) \end{aligned} \quad (\text{A.3})$$

Combining A.1, A.2 and A.3

$$P(T=1|R=0) = \frac{P(B=1|R=0)}{P(B=1|R=1)}. \quad (\text{A.4})$$

We define the underreporting correction factor (UCF) as

$$UCF = \frac{P(B=1|R=0)}{P(B=1|R=1)} \quad (\text{A.5})$$

Note that if assumption 3 is violated, we expect $P(R=0|T=1, B=0) > P(R=0|T=1, B=1)$ (underreporting is greater in participants without the biomarker compared with participants with the biomarker), or, equivalently, $P(B=1|T=1, R=0) < P(B=1|T=1, R=1)$. Then the UCF will underestimate $P(T=1|R=0)$. That is, if assumption 3 is violated,

$$UCF < P(T=1|R=0)$$

S2: Estimating the True Prevalence of Behavior

The results from section S1 can be used to estimate the “true” behavior rate $P(T=1)$ instead of the reported behavior rate $P(R=1)$.

By the total probability rule,

$$P(T=1) = P(T=1|R=1) * P(R=1) + P(T=1|R=0) * P(R=0) \quad (\text{A.6})$$

Then, since $P(T=1|R=1) = 1$ (assumption 1) and $P(T=1|R=0) = \frac{P(B=1|R=0)}{P(B=1|R=1)}$ (equation A.4),

$$P(T=1) = P(R=1) + UCF * P(R=0) \quad (\text{A.7})$$

FIGURE A.1.

Illustration of a lag between time of the risk behavior and the time of a positive test on the biomarker. Assume the behavior and the biomarker are measured at time t . The participant's report at time t reflects behavior over the intervals h and r (i.e. $T_t = T_h \cup T_r$). However, if there is a lag of length l , the value of the biomarker at time t (B_t) reflects the behavior over the intervals l and r (i.e. $B_t = B_r \cup B_l$). Note that the length of l and h are equal.

APPENDIX B

S1: Adjustment of the Probability of Underreporting for Imperfect Specificity

If the biomarker is not perfectly specific (violation of assumption 2) we can adjust the estimate of $P(T=1|R=0)$ by incorporating a known specificity into the equation.

Using only assumptions 1 and 3, one can show

$$UCF = P(T=1|R=0) + \frac{(1 - P(B=0|T=0)) * (1 - P(T=1|R=0))}{P(B=1|T=1)}$$

Using algebra to solve for the true probability of underreporting, we get

$$P(T=1|R=0) = \frac{UCF * P(B=1|T=1) + P(B=0|T=0) - 1}{P(B=1|T=1) + P(B=0|T=0) - 1} \quad (\text{B.1})$$

In order to adjust the UCF closer to the true probability of underreporting, $P(T=1|R=0)$, we must have an estimate of the specificity $P(B=0|T=0)$ and the sensitivity $P(B=1|T=1)$ of the biomarker. If less than perfect specificity is because of a lag, see appendix B, section S2 for estimating specificity. If the sensitivity of the biomarker is unknown, it may be estimated as follows.

By the total probability rule, we know that

$$P(B=1) = P(B=1|T=1) * P(T=1) + P(B=1|T=0) * P(T=0)$$

Therefore, the sensitivity of the biomarker is equal to

$$P(B=1|T=1) = \frac{P(B=1) - (1 - P(B=0|T=0)) * (1 - P(T=1))}{P(T=1)}$$

$P(B=1)$ is observed and $P(B=0|T=0)$ is the biomarker specificity which must be known or estimated to apply the correction. $P(T=1)$ is estimated in appendix 1, section S2, but there is a circular relationship by which $P(T=1)$ depends on $P(T=1|R=0)$ (equation A.6) which depends on $P(B=1|T=1)$ (equation B.1), which depends back on $P(T=1)$. In order to address this, we created an iterative process of adjustment.

1. Assume perfect specificity, $P(B=0|T=0)$, and calculate UCF using the basic method (equation A.5).
2. Calculate $P(T=1)$ using the basic method (equation A.7)
3. Find the specificity $P(B=0|T=0)$ of the biomarker (if lag, see appendix B, section S2).
4. Use the UCF from step 1, $P(T=1)$ from step 2, and $P(B=0|T=0)$ from step 3 to find $P(T=1|R=0)$ (equation B.1).
5. Iterate through steps 2 – 4 using the new $P(T=1|R=0)$ until $P(T=1|R=0)$ does not change.

The value of $P(T=1|R=0)$ when it no longer changes is the final adjusted UCF . This adjusted UCF can then be used in equation A.7 to estimate the corrected prevalence of the behavior.

Confidence intervals for this method may be computed using the percentile bootstrap method.

S2: Estimating Biomarker Specificity with a Lag

To estimate the apparent specificity of the biomarker when there is a lag, let B_t be the biomarker result measured at time t and T_t be the true (unobserved) behavior at time t (see Fig. A.1).

Then, the apparent specificity of the biomarker is $1 - P(B_t=1|T_t=0)$ and

$$P(B_t=1|T_t=0) = P(B_r=1 \text{ or } B_l=1|T_t=0) = P(B_r=1|T_t=0) + P(B_l=1|T_t=0).$$

Assuming the biomarker is completely specific when measured over the same interval as the behavior, $P(B_r=1|T_t=0) = 0$. Further, T_t tells us nothing about the interval l (see Fig. A.1) so we assume $P(B_l=1|T_t=0) = P(B_l=1)$ giving

$$P(B_t=1|T_t=0)=P(B_t=1).$$

The incidence of the biomarker (as observed in the data) can be used to estimate this probability as

$$P(B_t=1|T_t=0)=P(B_t=1)=incidence * l$$

where l is the lag between the time of the risk behavior and the time of a positive biomarker test. Thus, the apparent specificity of the biomarker is

$$P(B_t=0|T_t=0)=1 - incidence * l.$$

REFERENCES

1. O'Sullivan LF. Challenging assumptions regarding the validity of self-report measures: the special case of sexual behavior. *J Adolesc Health*. United States. 2008;207–8.
2. Catania JA, Gibson DR, Chitwood DD, Coates TJ. Methodological problems in AIDS behavioral research: influences on measurement error and participation bias in studies of sexual behavior. *Psychol Bull*. 1990; 108(3):339–62. [PubMed: 2270232]
3. Gallo M, Behets F, Steiner M, Hobbs M, Hoke T, Van Damme K, et al. Prostate-specific antigen to ascertain reliability of self-reported coital exposure to semen. *Sexually Transmitted Diseases*. 2006; 33(8):476–479. [PubMed: 16865047]
4. Gallo M, Behets F, Steiner M, Thomsen S, Ombidi W, Luchters S, et al. Validity of self-reported 'safe sex' among female sex workers in Mombasa, Kenya--PSA analysis. *International Journal of STD & AIDS*. 2007; 18(1):33–38. [PubMed: 17326860]
5. Zenilman JM, Weisman CS, Rompalo AM, Elish N, Upchurch DM, Hook EW 3rd, et al. Condom use to prevent incident STDs: the validity of self-reported condom use. *Sex Transm Dis*. 1995; 22(1):15–21. [PubMed: 7709320]
6. Zimmerman R, Morisky D, Harrison L, Mark H. Validity of Behavioral Measures as Proxies for HIV-Related Outcomes. *Journal of Acquired Immune Deficiency Syndrome*. 2014; 66(Supplement 3):S285–S292.
7. Bell DC, Montoya ID, Atkinson JS. Partner concordance in reports of joint risk behaviors. *Journal of Acquired Immune Deficiency Syndromes*. 2000; 25(2):173–181. [PubMed: 11103048]
8. Beguy D, Kabiru C, Nderu E, Ngware M. Inconsistencies in self-reporting of sexual activity among young people in Nairobi, Kenya. *Journal of Adolescent Health*. 2009; 45(6):595–601. [PubMed: 19931832]
9. Palen LA, Smith EA, Caldwell LL, Flisher AJ, Wegner L, Vergnani T. Inconsistent reports of sexual intercourse among South African high school students. *J Adolesc Health*. 2008; 42(3):221–7. [PubMed: 18295129]
10. Rosenbaum JE. Reborn a virgin: adolescents' retracting of virginity pledges and sexual histories. *Am J Public Health*. 2006; 96(6):1098–103. [PubMed: 16670225]
11. Upchurch D, Lillard L, Aneshensel C, Fang Li N. Inconsistencies in reporting the occurrence and timing of first intercourse among adolescents. *Journal of Sex Research*. 2002; 39(3):197–206. [PubMed: 12476267]
12. Turner A, De Kock A, Meehan-Ritter A, Blanchad K, Sebola M, Hoosen A, et al. Many vaginal microbicide trial participants acknowledged they had misreported sensitive sexual behavior in face-to-face interviews. *Journal of Clinical Epidemiology*. 2009; 62(7):759–765. [PubMed: 19013762]

13. Morrison-Beedy D, Carey MP, Tu X. Accuracy of audio computer-assisted self-interviewing (ACASI) and self-administered questionnaires for the assessment of sexual behavior. *AIDS Behav.* 2006; 10(5):541–52. [PubMed: 16721506]
14. Kelly CA, Soler-Hampejsek E, Mensch BS, Hewett PC. Social desirability bias in sexual behavior reporting: evidence from an interview mode experiment in rural Malawi. *Int Perspect Sex Reprod Health.* 2013; 39(1):14–21. [PubMed: 23584464]
15. Mensch BS, Hewett PC, Gregory R, HELLERINGER S. Sexual behavior and STI/HIV status among adolescents in rural Malawi: an evaluation of the effect of interview mode on reporting. *Stud Fam Plann.* 2008; 39(4):321–34. [PubMed: 19248718]
16. Williams BL, Suen HK, Baffi CR. A controlled randomized response technique. *Eval Health Prof.* 1993; 16(2):225–38. [PubMed: 10125778]
17. Gallo M, Steiner M, Hobbs M, Warner L, Jamieson D, Macaluso M. Biological Markers of Sexual Activity: Tools for Improving Measurement in HIV/Sexually Transmitted Infection Prevention Research. *Sexually Transmitted Diseases.* 2013; 40(6):447–452. [PubMed: 23677018]
18. Macaluso M, Lawson L, Akers R, Valappil T, Hammond K, Blackwell R, et al. Prostate-specific antigen in vaginal fluid as a biologic marker of condom failure. *Contraception.* 1999; 59(3):195–201. [PubMed: 10382083]
19. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr.* 1997; 65(4 Suppl):1179s–1186s. [PubMed: 9094918]
20. Keogh RH, White IR. A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Stat Med.* 2014; 33(12):2137–55. [PubMed: 24497385]
21. Fagerland M, Lydersen S, Laake P. Recommended confidence intervals for two independent binomial proportions. *Statistical Methods in Medical Research.* 2011; 24(2):224–254. [PubMed: 21996567]
22. Oehlert GW. A note on the delta method. *American Statistician.* 1992; 46(1):27.
23. Network HPT. HPTN 068. 2014.
24. Celum C, Wald A, Hughes J, Sanchez J, Reid S, Delany-Moretlwe S, et al. Effect of aciclovir on HIV-1 acquisition in herpes simplex virus 2 seropositive women and men who have sex with men: a randomised, double-blind, placebo-controlled trial. *Lancet.* 2008; 371(9630):2109–19. [PubMed: 18572080]
25. Wilson, KS. Doctoral Dissertation. University of Washington; Seattle, Washington: 2015. The relationship between recent intimate partner violence, unprotected sex, and detectable plasma viral load in high-risk HIV-positive women in Kenya.
26. Liu J, Liu S, Yan J, Q, Huang H. Adherence to Immunosuppressive Medication in Renal Transplant Recipients From Follow-up Outpatient in China: Association of 2 Different Measurement Methods. *Clin Ther.* 2015; 37(11):2572–80. [PubMed: 26519232]
27. Searles Nielsen S, Dills RL, Glass M, Mueller BA. Accuracy of prenatal smoking data from Washington State birth certificates in a population-based sample with cotinine measurements. *Ann Epidemiol.* 2014; 24(3):236–9. [PubMed: 24461931]
28. Kelly C, Hewett P, Mensch B, Rankin J, Nsomba S, Kalibala S, et al. Using biomarkers to assess the validity of sexual behavior reporting across interview modes among young women in Kampala, Uganda. *Studies in Family Planning.* 2014; 45(1):43–58. [PubMed: 24615574]
29. Mose F, Newman L, Njunguna R, Tamoo H, John-Stewart G, Farquhar C, et al. Biomarker evaluation of self-reported condom use among women in HIV-discordant couples. *International Journal of STD & AIDS.* 2013; 24(7):537–540. [PubMed: 23970768]
30. Aho J, Koushik A, Diakite S, Loua K, Nguyen V, Rashed S. Biological validation of self-reported condom use among sex workers in Guinea. *AIDS and Behavior.* 2010; 14(6):1287–1293. [PubMed: 19680799]
31. Jadack RA, Yuenger J, Ghanem KG, Zenilman J. Polymerase chain reaction detection of Y-chromosome sequences in vaginal fluid of women accessing a sexually transmitted disease clinic. *Sex Transm Dis.* 2006; 33(1):22–5. [PubMed: 16385218]

FIGURE 1.

Effect of assumptions 1 – 3 on bias of UCF where $Bias = UCF - P(T = 1|R = 0)$. **A.**

Assumption 1. Effect of overreporting of the risk behavior on bias in the UCF. **B.**

Assumption 2. Effect of sensitivity and specificity of the biomarker on bias in the UCF. Bias is 0, regardless of sensitivity, when specificity is 1. **C. Assumption 3.** Bias in the UCF when the presence of the biomarker increases the likelihood of reporting the behavior. Φ represents the ratio of the probability of underreporting among those with and without the biomarker (see text). Assumption 3 requires that $\Phi = 1$. Violations of this assumption are expressed in terms of Φ from 0 to 1.

TABLE 1

Bias and coverage probabilities for 95% confidence intervals for the UCF and the corrected prevalence of the behavior based on 10,000 simulations as a function of dataset size (N), biomarker sensitivity $P(B=1|T=1)$, and true $P(T=1|R=0)$. The bias and the confidence interval may not be computable (NC) for some datasets. Specifically, the UCF cannot be computed when the estimate of $P(T=1|R=1)=0$. The log interval cannot be computed if the estimate of $P(T=1|R=0)=0$ or 1 and the Koopman interval cannot be calculated if the estimate of $P(T=1|R=0)=0$. Simulations where the proportion of simulations resulting in NC exceeded 5% are noted. NC's were not factored into the calculation of bias or coverage probability.

Parameters			UCF						P(Y=1)					
			Bias			Log CI Coverage			Koopman CI Coverage			Bias		
Data	R1 N	R0 N	True $P(T=1 R=0)$			0.1	0.3	0.7	0.1	0.3	0.7	0.1	0.3	0.7
			$P(B=1 T=1)$											
1	500	100	0.3	0.001	0.002	0.961	0.956	0.950	0.963	0.955	0.960	0.000	0.000	0.001
2	500	100	0.075	0.005	0.008	$a_{0.920}$	$a_{0.961}$	0.967	0.954	0.965	0.964	0.001	0.001	0.003
3	100	500	0.3	0.002	0.007	0.955	0.953	0.953	0.977	0.993	0.998	0.002	0.006	0.014
4	100	500	0.075	0.019	0.045	0.975	0.961	0.959	0.985	0.999	>0.999	0.016	0.045	0.105
5	100	50	0.3	0.004	0.009	$a_{0.950}$	0.961	0.955	0.961	0.964	0.970	0.001	0.003	0.006
6	100	50	0.075	0.012	0.018	$a_{0.876}$	$a_{0.960}$	$a_{0.972}$	0.952	0.965	0.980	0.006	0.015	0.038
7	50	100	0.3	0.005	0.017	0.972	0.962	0.956	0.974	0.981	0.991	0.003	0.011	0.025
8	50	100	0.075	0.035	0.071	$a_{0.962}$	$a_{0.967}$	0.982	0.969	0.997	0.999	0.018	0.062	0.143

^aProportion of NC intervals greater than 5%

^bLow coverage probability resulting from failure to cover a high proportion of simulations where $\widehat{UCF} = 0$, unrepresentative of typical use

TABLE 2

Application of UCF and corrected prevalence of behavior to multiple sources of data. 95% CIs computed with the log method (equation 2)

Data Source	Study Population	Study N	Biomarker (B)	Reported Behavior (R)	Biomarker Pos. B = 1	Biomarker Neg. B = 0	$\widehat{UCF}$	Uncorrected Prevalence of Behavior P(R=1)	Corrected Prevalence of Behavior
HPTN 068 ²³	High school girls in South Africa	2525	HSV-2	Ever had sex	81	610	0.14 (0.10–0.22)	0.27	0.37 (0.33–0.42)
				Never had sex	31	1803			
Mombasa Cohort ^{2,5}	Women in Mombasa, Kenya	389	PSA	Unprotected sex	11	19	0.33 (0.19–0.56)	0.08	0.38 (0.21–0.54)
				No unprotected sex	43	316			
Kelly, CA e.a. (2014) ²⁸	Young women in Kampala, Uganda	938	HSV-2	Ever had sex	420	425	0.48 (0.33–0.69)	0.90	0.95 (0.93–0.97)
				Never had sex	22	71			
Mose, F e.a. (2013) ²⁹	Women in HIV-discordant couples in Nairobi, Kenya	117	PSA	<100% condom use last month	3	16	0.65 (0.2–1.00)	0.16	0.70 (0.06–1.00)
				100% condom use last month	10	88			
Aho, J e.a. (2010) ³⁰	Female sex workers in Guinea	215	PSA	Unprotected intercourse last 48 hrs.	12	7	0.57 (0.38–0.84)	0.09	0.60 (0.40–0.81)
				No unprotected intercourse last 48 hrs.	70	126			
Gallo, MF e.a. (2007) ⁴	Female sex workers in Mombasa, Kenya	377	PSA	1 unprotected sex act last 48 hrs.	14	34	0.39 (0.23–0.66)	0.13	0.46 (0.28–0.65)
				Protected sex only or no sex last 48 hrs.	37	292			
Gallo, MF e.a. (2006) ³	Female sex workers in Madagascar	330	PSA	1 unprotected sex act last 48 hrs.	57	39	0.50 (0.38–0.64)	0.29	0.64 (0.55–0.74)
				Protected sex only or no sex last 48 hrs.	69	165			
Jadack, RA e.a. (2006) ³¹	Women in Baltimore Health Dept. STD Clinic	103	Y- Chromosome	No condom use past 2 weeks	47	20	0.79 (0.57–1.00)	0.65	0.93 (0.83–1.00)
				Condom use every time past 2 weeks	20	16			

TABLE 3

Application of specificity-adjusted UCF and corrected prevalence of behavior to female pregnancy data from HPTN 039 at quarterly visits. 95% CIs for unadjusted (no lag) $\widehat{UCF}$ computed with the Koopman method; 95% CIs for adjusted (10 day lag, 15 day lag) $\widehat{UCF}$ computed with a percentile bootstrap; 95% CIs for corrected prevalence of behavior computed with the delta method.

Mo.	N^a	R^b	$B = 1^c$	$B = 0$	$P(R=1)$	No lag, Unadjusted Estimates		10-day lag, Specificity-Adjusted Estimates		15-day lag, Specificity-Adjusted Estimates	
						$\widehat{UCF}$	Corrected Prevalence	$\widehat{UCF}$	Corrected Prevalence	$\widehat{UCF}$	Corrected Prevalence
3	1148	1	32	895	0.81	0.26 (0.07 – 1.00)	0.86 (0.78 – 0.93)	0.18 (0.00 – 0.71)	0.84 (0.79 – 0.94)	0.14 (0.00 – 0.69)	0.83 (0.78 – 0.94)
		0	2	219							
6	1101	1	41	823	0.79	0.36 (0.13 – 1.00)	0.86 (0.78 – 0.94)	0.29 (0.00 – 0.79)	0.85 (0.78 – 0.95)	0.25 (0.00 – 0.77)	0.84 (0.77 – 0.95)
		0	4	233							
9	1040	1	33	773	0.78	0.21 (0.06 – 0.88)	0.82 (0.75 – 0.89)	0.13 (0.00 – 0.55)	0.80 (0.75 – 0.90)	0.08 (0.00 – 0.53)	0.79 (0.74 – 0.89)
		0	2	232							
12	1007	1	21	749	0.77	0.31 (0.08 – 1.00)	0.84 (0.73 – 0.94)	0.24 (0.00 – 0.89)	0.82 (0.74 – 0.97)	0.20 (0.00 – 0.88)	0.81 (0.73 – 0.97)
		0	2	235							
15	977	1	21	735	0.77	0.33 (0.08 – 1.00)	0.85 (0.74 – 0.96)	0.26 (0.00 – 0.97)	0.83 (0.75 – 0.99)	0.21 (0.00 – 0.97)	0.82 (0.74 – 0.99)
		0	2	219							
18	966	1	30	695	0.75	0.20 (0.05 – 0.87)	0.80 (0.73 – 0.88)	0.12 (0.00 – 0.53)	0.78 (0.73 – 0.88)	0.08 (0.00 – 0.50)	0.77 (0.71 – 0.87)
		0	2	239							

^a N decreases as participants are lost to follow-up

^b R is the reported behavior; i.e. $R = 1$ indicates reported unprotected sex, $R = 0$ indicates no reported unprotected sex

^c B is the biological biomarker status; i.e. $B = 1$ indicates pregnant, $B = 0$ indicates not pregnant