Annals of Epidemiology 26 (2016) 612–618

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

The validity of self-reported behaviors: methods for estimating underreporting of risk behaviors

Marlena S. Norwood [a], James P. Hughes PhD [a,b,*], K. Rivet Amico PhD [c]

[a] Department of Biostatistics, University of Washington, Seattle
[b] Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, Seattle, WA
[c] Department of Health Behavior and Health Education, School of Public Health, University of Michigan, Ann Arbor

ARTICLE INFO

Article history:
Received 13 February 2016
Accepted 31 July 2016
Available online 8 August 2016

Keywords:
Self-reported behaviors
HIV
Sexual behaviors
Risk behaviors
Sexually transmitted infections

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 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 techniques [16]. 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. Sexually transmitted infections (STI), for example, are imperfectly sensitive as markers for condomless sex because 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 the following) 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

The authors declare no conflict of interest.

* Corresponding author. Department of Biostatistics, University of Washington, 357232, Seattle, WA 98195. Tel.: +1-206-616-2721; fax: +1-206-616-2724.
E-mail address: jphughes@uw.edu (J.P. Hughes).

http://dx.doi.org/10.1016/j.annepidem.2016.07.011
1047-2797/

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 UCF is calculated with reference to a reported behavior (R), true behavior (T), and biomarker (B: e.g., pregnancy, PSA, herpes simplex virus type 2 [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:

Reported behavior	Biomarker	
	Present	Absent
Present (yes)	a	b
Absent (no)	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, that is, $P(T = 1|R = 0)$. 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 the following), $P(T = 1|R = 0)$ 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 $P(B = 1|R = 0)/P(B = 1|R = 1)$ as the UCF. The UCF may be estimated as:

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

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.

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

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.

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

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.

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

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

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

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% CI (based on the delta method [22]) for the true prevalence of the risk behavior is given by:

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

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. Furthermore, 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 CI 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 data sets (a data set is a table of R × B × T) with all combinations of the following characteristics: sample size ($N = 600$ and 150), biomarker sensitivity [$P(B = 1|T = 1) = .3$ and .075], and behavior prevalence (common or uncommon). For each of these eight combinations, we simulated three different true values of $P(T = 1|R = 0)$ (.1, .3, and .7) (i.e., different amounts of