HPTN
HIV Prevention Trials Network

Methods for Estimating Underreporting of Risk Behaviors

Jim Hughes, SCHARP

HPTN STATISTICAL PLENARY 2016

National Institute of Allergy and Infectious Diseases
NIMH National Institute of Mental Health
National Institutes of Health
U.S. Department of Health and Human Services
National Institutes of Health
NIDA National Institute on Drug Abuse

Background

- Individuals often underreport risky behaviors
  - Stigma
  - Social desirability bias
- Biomarkers may be used to validate self-reports
  - e.g. PSA as marker of condomless sex in women
- Existing methods depend on biomarkers with high specificity and sensitivity

Goal

- Develop an Underreporting Correction Factor (UCF) appropriate for biomarkers with high specificity, imperfect sensitivity
- Use UCF to estimate true prevalence of risky behavior

Definitions

- **T** = True behavior
  - e.g. condomless sex
- **R** = Self-reported behavior
  - e.g. reported condomless sex
- **B** = Biomarker
  - 100% specific
  - <100% sensitive
  - e.g. pregnancy
- each coded as + or -

Definitions

Specificity = P(B- | T-)

Sensitivity = P(B+ | T+)

- **T** = True behavior
  - e.g. condomless sex
- **R** = Self-reported behavior
  - e.g. reported condomless sex
- **B** = Biomarker
  - 100% specific
  - <100% sensitive
  - e.g. pregnancy
- each coded as + or -

Underreporting Correction Factor

- Define the “Underreporting Correction Factor”

$$\mathbf{UCF} = \frac{\mathbf{P(B+|R-)}}{\mathbf{P(B+|R+)}}$$

- Under <u>certain assumptions</u>

$$\mathbf{P(T+) = P(R+) + P(R-)*UCF}$$

Example

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

$P(R+) = 100/500 = .2$

Example

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

$P(R+) = 100/500 = .2$

$UCF = P(B+ \mid R-) / P(B+ \mid R+)$

$= (5/400) / (10/100) = .125$

- 12.5% of those who reported no risky behavior actually had the behavior.

Example

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

$P(R+) = 100/500 = .2$

$P(T+) = P(R+) + P(R-)*UCF$

$= .2 + .8 *.125 = .3$

- True prevalence of the behavior is 50% larger than reported prevalence

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present				
	absent				
	Total	100		400	

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present				
	absent				
	Total	100		400	

1) Biomarker 100% specific

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present		0		0
	absent				
	Total	100		400	

1) Biomarker 100% specific

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present	10	0	5	0
	absent				
	Total	100		400	

1) Biomarker 100% specific

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present	10	0	5	0
	absent				
	Total	100		400	

1) Biomarker 100% specific
2) No overreporting

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present	10	0	5	0
	absent		0		
	Total	100		400	

1) Biomarker 100% specific
2) No overreporting

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present	10	0	5	0
	absent	90	0		
	Total	100		400	

1) Biomarker 100% specific
2) No overreporting

HPTN
HIV Prevention
Trials Network

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present	10	0	5	0
	absent	90	0		
	Total	100		400	

1) Biomarker 100% specific
2) No overreporting
3) Underreporting unrelated to biomarker (5/10 = ?/90)

HPTN
HIV Prevention
Trials Network

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present	10	0	5	0
	absent	90	0	45	
	Total	100		400	

1) Biomarker 100% specific
2) No overreporting
3) Underreporting unrelated to biomarker (5/10 = 45/90)

HPTN
HIV Prevention
Trials Network

Assumptions

		Reported Behavior	
		present	absent
Biomarker	present	10	5
	absent	90	395
	Total	100	400

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present	10	0	5	0
	absent	90	0	45	350
	Total	100		400	

1) Biomarker 100% specific
2) No overreporting
3) Underreporting unrelated to biomarker (5/10 = 45/90)

Assumptions

		Reported Behavior			
		present		absent	
True Behavior		present	absent	present	absent
Biomarker	present	10	0	5	0
	absent	90	0	45	350
	Total	100		400	

$$P(T+) = 150/500$$

$$= 100/500 + 50/500$$

$$= 100/500 + 400/500 * (5/400)/(10/100)$$

$$= P(R+) + P(R-) * \underbrace{P(B+ \mid R-)/P(B+ \mid R+)}_{\text{UCF}}$$

Assumptions

Assumption	If violated …
There is no overreporting of the risky behavior	• UCF biased ↑ • P(T+) biased ↑ • Bias unlikely to be large
Underreporting of the behavior is unrelated to biomarker status	• Expect less underreporting for B+ • UCF biased ↓ • P(T+) biased towards P(R+) (incomplete adjustment)
The biomarker is 100% specific for the behavior	• UCF biased ↑ • P(T+) biased ↑ • Severe bias possible for low sensitivity biomarker

Assumptions

- Consider biomarker specificity carefully
  - e.g. HIV is not 100% specific for unprotected sex
  - Imperfect specificity can result from lags between behavior and biomarker positivity e.g. pregnancy as a biomarker for unprotected sex in the last 3 months
    - Correction for lags, known specificity possible, but more complex

Example

HPTN 068

– Adolescent girls in South Africa @ enrollment

		Ever had sex	
		Yes	No
HSV-2	Positive	81	31
	Negative	610	1803
	Total	691	1834

- P(R+) = 0.27 (95% CI: 0.25 – 0.29)
- UCF = 0.14 (95% CI: 0.10 – 0.22)

P(T+) = 0.37 (95% CI: 0.33 – 0.42)

Summary

- UCF depends on highly specific biomarker
  - no overreporting
  - no differential underreporting
- Use UCF to improve estimates of risky/stigmatizing behavior based on self-reports
  - Provides population estimate, not individual correction

HPTN
HIV Prevention
Trials Network

ACKNOWLEDGEMENTS

The HIV Prevention Trials Network is sponsored by the National Institute of Allergy and Infectious Diseases, the National Institute of Mental Health, and the National Institute on Drug Abuse, all components of the U.S. National Institutes of Health.

Sponsored by NIAID, NIDA, NIMH, NICHHD under Cooperative Agreement # UM1 AI068619, and grants R01 AI029168, R01 HD072617

Norwood M, Hughes JP, Amico KR. The Validity of Self-Reported Behaviors: Methods for Estimating Underreporting of Risk Behaviors. Submitted to Annals of Epidemiology