Practice of Epidemiology

Test-Retest Reliability of a Sexual Behavior Interview for Men Residing in Brazil, Mexico, and the United States

The HPV in Men (HIM) Study

Alan G. Nyitray, Jongphil Kim, Chiu-Hsieh Hsu, Mary Papenfuss, Luisa Villa, Eduardo Lazcano-Ponce, and Anna R. Giuliano

Initially submitted March 27, 2009; accepted for publication July 1, 2009.

Understanding the natural history of sexually transmitted infections requires the collection of data on sexual behavior. However, there is concern that self-reported information on sexual behavior may not be valid, especially if study participants are culturally and linguistically distinct. The authors completed a test-retest reliability study of 1,069 men recruited in Brazil, Mexico, and the United States in 2005 and 2006. All of the men completed the same computer-assisted self-interview approximately 3 weeks apart. Refusal rates, kappa coefficients, and intraclass correlation coefficients were calculated for the full sample and by country, age, and lifetime number of female sex partners. Reliability coefficients for each study site and the combined population were high for almost all questions. With few exceptions, the authors found high test-retest reliability with a computer-assisted self-interview on sexual behavior used in 3 culturally and linguistically distinct countries.

data collection; internationality; men; questionnaires; reproducibility of results; sexual behavior

Abbreviations: CASI, computer-assisted self-interviewing; HPV, human papillomavirus; ICC, intraclass correlation coefficient.
reliability for a non-audio CASI that collected information on sexual health history and sexual behavior in 3 languages from men recruited in 3 different countries.

MATERIALS AND METHODS

Study population

Beginning in March 2005, men were recruited in Brazil (São Paulo), Mexico (Cuernavaca), and the United States (Tampa, Florida) for the HPV in Men (HIM) Study—a cohort study of the natural history of anogenital human papillomavirus (HPV). Men were enrolled if they were between the ages of 18 and 70 years; resided in one of the targeted recruitment areas; had no prior anal cancer, penile cancer, or genital warts; had no current diagnosis of a sexually transmitted disease, including human immunodeficiency virus; had no history of imprisonment, homelessness, or drug treatment in the prior 6 months; and were willing to engage in study visits every 6 months for 4 years. Additional details on the study design and population have been previously published (36, 37).

In Brazil, men were recruited from a large clinic in São Paulo that tests for human immunodeficiency virus and sexually transmitted diseases and from the general population through radio and print advertisements. In Mexico, men were recruited through a large health plan in the state of Morelos. In the United States, men were recruited from a large university campus and the general community in Tampa, Florida. Participants received a nominal monetary incentive for their participation. Men found to be illiterate or innumerate during the consenting or interview process were removed from analysis for the current CASI reliability study. All enrolled participants consented to the study protocol, which was approved by the human subjects protection committee of the Ludwig Institute for Cancer Research in Brazil; the ethical committee of the Center for Sexually Transmitted Diseases and AIDS in São Paulo; the National Institute of Public Health of Mexico; and the University of South Florida.

Recruited in 2005 and 2006, the first 1,069 men to complete their run-in (test) and baseline (retest) visits by CASI comprised the participants in the current reliability study. Age varied by study site, with the median age of participants in Brazil and Mexico (33 years in both countries) being higher than that of participants recruited in the United States (23 years). As expected, racial and ethnic characteristics also varied by study site. Overall, approximately one-half (50.8%) of participants reported a nonwhite race, while 41.4% reported a Hispanic ethnicity. Other population characteristics are provided in Table 1.

Procedure

Men expressing interest in the study came to the clinic for an initial visit. After consenting to the research and receiving instructions for using the CASI, participants completed the self-interview and then were sampled at anogenital sites for HPV. The CASI was written in the primary language of the region (Portuguese, Spanish, or English) and elicited information about participant demographic characteristics, substance use, sexual health history, and sexual behaviors implicated in the transmission of HPV. The men completed an identical CASI retest approximately 3 weeks later (the median test-retest interval was 21 days in Brazil, 25 days in Mexico, and 16 days in the United States). A Kruskal-Wallis test confirmed a statistically significant difference in test-retest interval by site (P < 0.0001). Per the protocol, men were not counseled or educated about HPV at either the test or the retest, although the informed consent form contained basic information on HPV and staff answered men’s impromptu questions. Men did not receive their first HPV test results until 6 months later, at a subsequent clinic visit.

Interview measures

The interview contained 88 items. The majority of the questions had previously been administered to US men in a paper-and-pencil format and generally were found to have excellent reliability (38).

Participants’ sexual health was assessed with 18 questions about past sexually transmitted infections, the existence of a current sex partner, circumcision status, and the sexual health histories of their partners (ever having a partner with a sexually transmitted disease, genital warts, or an abnormal Papanicolaou smear). In addition, 45 sexual behavior items assessed incidence and frequency of penetrative sexual behaviors (vaginal, anal, and oral) with women and men; age at first intercourse; number of female and male partners; frequency of condom use for vaginal and anal sex; incidence and frequency of sex with “steady” and casual partners; time since last vaginal sex and anal sex; and history of paying for sex. Participants were asked to recall their frequency of sexual intercourse and numbers of sex partners for varying periods of time, including the last month, the last 3 months, and over a lifetime. Participants could choose to refuse to respond to any question by clicking a “refuse” button. Participants could answer “Don’t know” or “Don’t remember” for some nominal items—for example, regarding past sexually transmitted disease diagnoses and use of a condom; however, participants were not given the option of answering “Don’t know” for interval and ordinal items—for example, regarding their lifetime number of sex partners.

A subset of interview items was selected for assessment of reliability, with preference being given to items for which reliability coefficients would be less affected by the test-retest interval; therefore, items with only a 1-month recall period were not assessed. A total of 38 variables were assessed, including 9 interval, 4 ordinal, and 25 nominal variables. Variables assessed included 14 sexual health history variables and 24 sexual behavior variables.

Data analysis

For each item, reliability coefficients were calculated for each of the 3 study sites. We calculated combined population coefficients by averaging study-site coefficients after weighting them by the inverse of their variances (39). Because age (40, 41) and number of sex partners (40, 42) are associated with increasing measurement error, reliability coefficients were stratified by age (<30 years vs. ≥30 years).
and lifetime number of female sex partners (≤7 partners (median) vs. >7 partners) reported at retest.

For nominal variables, the kappa (κ) statistic was calculated (43). Because the κ statistic can be unstable in situations where there are sparse data (44), κ was not computed for variables where the number of cases or noncases was less than 5 (7). For ordinal variables, a weighted κ statistic was calculated (45) to allow credit for partial agreement. Benchmarks for interpreting κ and weighted κ values followed those of Landis and Koch (46): poor reliability, κ < 0.00; slight reliability, κ = 0.00–0.20; fair reliability, κ = 0.21–0.40; moderate reliability, κ = 0.41–0.60; substantial reliability, κ = 0.61–0.80; and almost perfect reliability, κ ≥ 0.81.

Interval variables were assessed using the intraclass correlation coefficient (ICC) (47). All ICCs created using nonnormal variables were transformed using Fisher’s z transformation before calculation of confidence intervals (48). Confidence intervals were then transformed back to the original scale. ICCs approaching 1.0 indicate high test-retest reliability.

During exploratory analysis, extreme outliers were identified in 2 variables: number of different female sex partners in the past 3 months (a value of 11,111,109,632 on both test and retest) and age at first sexual intercourse with women (a value of 1,993 on test). Each observation was removed prior to analyses. Subsequent text regarding outliers identified in scatterplots does not include these 2 observations.

Refusal rates were assessed. Refusals were not included in reliability coefficient calculations.

RESULTS
With exceptions for skip patterns, participants at each study site answered virtually all of the 38 questions under study. The average refusal rate on retest for all questions was 1.0% in Mexico, 1.3% in the United States, and 2.5% in Brazil (data not shown).

For all nominal and ordinal questions, κ and weighted κ reliability coefficients for each study site and the combined population were almost perfect (κ ≥ 0.81) or substantial (κ = 0.61–0.80). Table 2 provides coefficients for 18 items for which reliability was less than 0.81.

Site-specific ICCs for all interval questions were 0.85 or more, with the exception of ICCs in Brazil and Mexico for 3 items asking men to report their numbers of sex partners. Scatterplots identified several extreme outliers in the bivariate distributions of all 3 items. Specifically, for the variable “number of sex partners other than a ‘steady’ partner in the past 3 months,” when 1 oullingy identified in the Mexico sample was removed, the Mexico ICC increased from 0.61 to 0.84. For the same item, when 2 oullingy participants in the Brazil sample were removed, the Brazil ICC increased from 0.10 to 0.79. For “lifetime number of male anal sex partners,” when 1 outlier identified in the Brazil scatterplot was removed from the data set, the ICC for Brazil increased from 0.50 to 0.99. For the variable “number of different female sex partners in the past 3 months,” when 1 outlier in the Brazil scatterplot was removed, the ICC for Brazil increased from 0.58 to 0.92.

After taking into account these outliers, test-retest reliability was generally high and consistent across sites: Reliability coefficients differed by no more than 17 percentage points among study sites.

All nominal and ordinal items had substantial or almost perfect reliability regardless of participant age. All interval variables had ICC scores greater than or equal to 0.85 for both age groups, with the exception of the older men’s answers to the same 3 interval variables as those discussed above: lifetime number of male anal sex partners, number of different female sex partners in the past 3 months, and number of sex partners other than a “steady” partner in the past 3 months (Table 3). After removing the outliers discussed above (all of which involved men over age 32 years), the ICCs for these 3 questions increased to 0.84 or more.

Reliability coefficients were also stratified by lifetime number of female sex partners. Whether men reported numbers of partners above or below the median number of 7, reliability coefficients for all nominal and ordinal variables were substantial or almost perfect, except for 2: ever having vaginal, anal, or oral sex (κ = 0.39 for men with >7 partners) and ever paying a man for sex (κ = 0.54 for men with ≤7 partners) (Table 3). Two interval variables also showed lower reliability: lifetime number of male anal sex partners (ICC = 0.50 for men with ≤7 partners) and number of sex partners other than a “steady” partner in the past 3 months (ICC = 0.29 for men with >7 partners). Removal of the outliers identified above increased the ICC for these 2 interval variables to 0.85 or more.

DISCUSSION
In this test-retest reliability study of a CASI instrument, 1,069 men in Brazil, Mexico, and the United States were asked the same questions on sexual health history and sexual behavior at test and retest. For each study site and the 3 study sites combined, κ and weighted κ reliability coefficients for nominal and ordinal questions, respectively, were substantial (0.61–0.80) or almost perfect (≥0.81). However, while the combined population ICC scores for all interval variables were greater than or equal to 0.85, the study site-specific reliability of 3 interval variables was of concern: lifetime number of male anal sex partners, number of different female sex partners in the past 3 months, and number of sex partners other than a “steady” partner in the past 3 months. The apparently poorer reliability of these variables was due to the presence of a small number of outliers identified in scatterplots. When these outliers were removed, study site-specific ICCs increased to 0.79 or more. These outlying observations also distorted the ICCs for several interval variables after results were stratified by age and median lifetime number of female sex partners.

The ability of a small number of observations to distort a reliability coefficient is discussed in the literature (7, 49); however, to our knowledge, the impact on reliability in practice has rarely been described (50, 51). In the current study, 1 or 2 participants with highly discrepant test and retest answers were able to increase the variance of an item by more than 2 magnitudes. For example, 1 Brazilian participant reported 2,000 lifetime male anal sex partners at test...
and only 20 such partners at retest. Removing this individual decreased the variance in the item from 0.00602 to 0.00002 and increased the ICC from 0.50 to 0.98. Such scenarios also underscore the importance of weighting by the inverse of the variance when combining coefficients in order to reduce the contribution of less reliable coefficients.

In total, outliers were observed in data reported by 3 out of 338 men in Brazil, 1 out of 327 men in Mexico, and no men out of 404 in the United States. The numbers of outliers in the 3 countries did not differ significantly ($P = 0.11$). Nevertheless, a higher number of outliers in Brazil may have occurred if the Brazilian men had less lifetime exposure to computer technology. Since a substantially higher percentage of Brazilian participants were aged 45 years or older as compared with Mexican and US participants, it is possible that these older men had less comfort with the technology and therefore had more reporting errors (17, 30, 32). It is also possible that a higher number of outliers for Brazil, in comparison with Mexico or the United States, may have occurred if this cross-national instrument was less culturally appropriate for Brazilian men. The questionnaire from which the current study’s CASI was created was developed in the southwestern United States near the Mexican border. Creation of the questionnaire in this region may have led to an instrument that was somewhat more culturally appropriate for US and Mexican participants and less so for Brazilian participants. If this was the case, this may also account for the somewhat higher rate of question refusal among Brazilians. A review of the 4 outlier participants’ responses for all interview items revealed that 2 of the Brazilian men also gave illogical answers to a number of other questions, while the remaining 2 outliers’ interviews were generally unremarkable.

Men found to be illiterate or innumerate during the consenting and interviewing procedures were not included in the current CASI study; however, staff may not have been able to determine if a man was illiterate or innumerate prior to the time of consent. Therefore, it was possible that some men were illiterate or innumerate, but were not identified and included in the current study. Table 1 presents the characteristics of participants at retest (baseline) in the HPV in Men (HIM) Study, Brazil, Mexico, and the United States, 2005–2006.

### Table 1. Characteristics of Participants at Retest (Baseline) in the HPV in Men (HIM) Study, Brazil, Mexico, and the United States, 2005–2006

| Variable                  | Brazil ($n = 338$) | Mexico ($n = 327$) | United States ($n = 404$) | Combined Population ($n = 1,069$) |
|---------------------------|--------------------|--------------------|---------------------------|----------------------------------|
|                           | No. | %    | No. | %    | No. | %    | No. | %    | No. | %    |
| Age, years                |      |      |      |      |      |      |      |      |      |      |
| 18–19                     | 11  | 3.3  | 14  | 4.3  | 93  | 23.0 | 118 | 11.0 |      |      |
| 20–24                     | 51  | 15.1 | 48  | 14.7 | 153 | 37.9 | 252 | 23.6 |      |      |
| 25–29                     | 63  | 18.6 | 70  | 21.4 | 41  | 10.2 | 174 | 16.3 |      |      |
| 30–34                     | 71  | 21.0 | 65  | 19.9 | 30  | 7.4  | 166 | 15.5 |      |      |
| 35–39                     | 44  | 13.0 | 56  | 17.1 | 34  | 8.4  | 134 | 12.5 |      |      |
| 40–44                     | 50  | 14.8 | 60  | 18.4 | 26  | 6.4  | 136 | 12.7 |      |      |
| 45–70                     | 48  | 14.2 | 14  | 4.3  | 27  | 6.7  | 89  | 8.3  |      |      |
| Refused to respond$^a$    | 0   | 0.0  | 0   | 0.0  | 0   | 0.0  | 0   | 0.0  |      |      |
| Median age, years         | 33  |      | 33  |      | 23  |      | 30  |      |      |      |
| Race                      |      |      |      |      |      |      |      |      |      |      |
| White                     | 207 | 61.2 | 15  | 4.6  | 296 | 73.3 | 518 | 48.5 |      |      |
| Black                     | 101 | 29.9 | 0   | 0.0  | 53  | 13.1 | 154 | 14.4 |      |      |
| Asian/Pacific Islander    | 3   | 0.9  | 0   | 0.0  | 22  | 5.5  | 25  | 2.3  |      |      |
| American Indian           | 20  | 5.9  | 0   | 0.0  | 0   | 0.0  | 20  | 1.9  |      |      |
| Mixed/other               | 0   | 0.0  | 311 | 95.1 | 33  | 8.2  | 344 | 32.2 |      |      |
| Refused to respond        | 7   | 2.1  | 1   | 0.3  | 0   | 0.0  | 8   | 0.8  |      |      |
| Ethnicity                 |      |      |      |      |      |      |      |      |      |      |
| Hispanic                  | 63  | 18.6 | 327 | 100.0 | 52 | 12.9 | 442 | 41.4 |      |      |
| Non-Hispanic              | 265 | 78.4 | 0   | 0.0  | 351 | 86.9 | 616 | 57.6 |      |      |
| Refused to respond        | 10  | 3.0  | 0   | 0.0  | 1   | 0.3  | 11  | 1.0  |      |      |
| Marital status            |      |      |      |      |      |      |      |      |      |      |
| Single, never married     | 148 | 43.8 | 87  | 26.6 | 294 | 72.8 | 529 | 49.5 |      |      |
| Married                   | 99  | 29.3 | 186 | 56.9 | 58  | 14.4 | 343 | 32.1 |      |      |
| Cohabitating              | 59  | 17.5 | 36  | 11.0 | 19  | 4.7  | 114 | 10.7 |      |      |
| Divorced/widowed          | 31  | 9.2  | 18  | 5.5  | 31  | 7.7  | 80  | 7.5  |      |      |
| Refused to respond        | 1   | 0.3  | 0   | 0.0  | 2   | 0.5  | 3   | 0.3  |      |      |

Table continues
able to identify all of the persons whose level of illiteracy increased their risk of providing unreliable responses. Reliability can also be affected by the number of days between test and retest (4). However, the outliers in Brazil had test-retest intervals of less than 22 days. In addition, reliability in this study was largely consistent across sites, even though median test-retest intervals varied from 16 days at the US study site to 25 days at the Mexican study site.

After stratification by lifetime number of female sex partners, lower reliability was also identified for 2 nominal variables: ever having vaginal, anal, or oral sex ($\kappa = 0.39$ for men with $>7$ partners) and ever paying a man for sex ($\kappa = 0.54$ for men with $\leq 7$ partners). For the variable “ever having vaginal, anal, or oral sex,” 7 out of 441 men who reported more than the median number of 7 sex partners at retest reported at test that they had never had anal, vaginal, or oral sex. In addition to inviting concern about validity, this item’s $\kappa$ coefficient may have been rendered unstable because of the small number of cases. It is also plausible that the multiple sexual behaviors addressed in the question confused some men. For the second variable, only 6 men with fewer than the median number of sex partners acknowledged ever paying a man for sex, possibly lending instability to the $\kappa$ coefficient. Variables with sparse data may simply reflect a lack of behavioral heterogeneity in the population or cultural stigmatization attached to certain behaviors (9).

During the design of this study, we decided not to report coefficients where the number of cases or noncases was less than 5, in an effort to report only stable coefficients. In future evaluations of test-retest reliability, investigators may wish to consider increasing this minimum requirement. However, absent the small number of extreme outliers and the presence of sparse data for 2 questions, virtually all coefficients indicated that the CASI interview items under study captured data on these men’s sexual behaviors in a highly reliable manner. This result may have been
Table 2. Selecta Kappa and Intraclass Correlation Coefficients for Computer-assisted Self-Interview Items, by Study Site, HPV in Men (HIM) Study, Brazil, Mexico, and the United States, 2005–2006

| Interview Item | Brazil | Mexico | United States | Combined Population |
|----------------|--------|--------|---------------|--------------------|
|                | No. c  | k or ICC 95% CI | No. c  | k or ICC 95% CI | No. c  | k or ICC 95% CI | No. c  | k or ICC 95% CI |
| Interval variables |        |        |               |                    |        |        |               |                    |        |        |               |                    |
| Lifetime no. of male anal sex partners | 94     | 0.50 0.33, 0.64 | 32     | 0.98 0.96, 0.99 | 35     | 0.99 0.98, 0.99 | 161    | 0.99 0.99, 0.99 |
| No. of different female sex partners in the past 3 months | 281    | 0.58 0.50, 0.65 | 292    | 0.88 0.85, 0.90 | 380    | 0.95 0.94, 0.96 | 953    | 0.94 0.93, 0.95 |
| No. of sex partners other than a “steady” partner in the past 3 months | 114    | 0.10 0.00, 0.28 | 86     | 0.61 0.46, 0.73 | 81     | 0.96 0.94, 0.97 | 281    | 0.95 0.94, 0.96 |
| Ordinal variables |        |        |               |                    |        |        |               |                    |        |        |               |                    |
| Frequency of condom use for anal sex in the past 3 months | 236    | 0.87 0.84, 0.90 | 178    | 0.83 0.78, 0.87 | 182    | 0.79 0.73, 0.84 | 596    | 0.84 0.81, 0.86 |
| Nominal variables |        |        |               |                    |        |        |               |                    |        |        |               |                    |
| Ever being diagnosed with a sexually transmitted disease | 336    | 0.84 0.81, 0.87 | 327    | 0.80 0.76, 0.84 | 404    | 0.89 0.87, 0.91 | 1,067   | 0.85 0.83, 0.87 |
| Ever being diagnosed with Chlamydia | 338    | 0.74 0.69, 0.78 | 327    | 0.79 0.75, 0.83 | —g     | —g     | 1,069   | 0.77 0.74, 0.79 |
| Ever being diagnosed with genital warts | 337    | 0.74 0.69, 0.78 | —g     | —g     | 404    | 0.81 0.77, 0.84 | 1,068   | 0.76 0.73, 0.78 |
| Ever being diagnosed with hepatitis B | 337    | 0.79 0.75, 0.83 | —g     | —g     | —g     | —g     | 1,068   | 0.76 0.73, 0.78 |
| Ever being diagnosed with hepatitis C | 337    | 0.71 0.65, 0.76 | 327    | 0.85 0.82, 0.88 | 404    | 0.85 0.82, 0.88 | 1,068   | 0.80 0.78, 0.82 |
| Ever being diagnosed with human immunodeficiency virus | 337    | 0.68 0.62, 0.73 | —g     | —g     | —g     | —g     | 1,068   | 0.68 0.65, 0.71 |
| Ever having a sex partner with a sexually transmitted disease | 338    | 0.73 0.68, 0.78 | 327    | 0.78 0.73, 0.82 | 404    | 0.80 0.76, 0.83 | 1,069   | 0.77 0.74, 0.79 |
| Ever having a sex partner with genital warts | 337    | 0.78 0.73, 0.82 | 327    | 0.78 0.73, 0.82 | 404    | 0.76 0.72, 0.80 | 1,068   | 0.77 0.74, 0.79 |
| Ever having a sex partner with an abnormal Papanicolaou test | 336    | 0.82 0.78, 0.85 | 327    | 0.75 0.70, 0.79 | 404    | 0.79 0.75, 0.82 | 1,067   | 0.79 0.77, 0.81 |
| Ever having vaginal, anal, or oral sex | 327    | 0.74 0.69, 0.79 | 326    | 0.80 0.76, 0.84 | 404    | 0.80 0.76, 0.83 | 1,057   | 0.78 0.76, 0.80 |
| Ever having oral sex | 309    | 0.92 0.90, 0.94 | 302    | 0.88 0.85, 0.90 | 390    | 0.78 0.74, 0.82 | 1,001   | 0.88 0.87, 0.89 |
| Ever exchanging sex for money/drugs | 315    | 0.90 0.88, 0.92 | 305    | 0.85 0.82, 0.88 | 390    | 0.78 0.74, 0.82 | 1,010   | 0.86 0.84, 0.88 |
| Ever paying a man for sex | 124    | 0.63 0.51, 0.73 | —g     | —g     | —g     | —g     | 218     | 0.68 0.60, 0.75 |
| Sex with someone other than a “steady” partner in the past 3 months | 257    | 0.93 0.91, 0.94 | 282    | 0.83 0.79, 0.86 | 283    | 0.96 0.95, 0.97 | 822     | 0.93 0.92, 0.94 |

Abbreviations: CI, confidence interval; HPV, human papillomavirus; ICC, intraclass correlation coefficient.

a Omitted are reliability coefficients for 20 items for which all study site coefficients were greater than or equal to 0.81.

b Unless otherwise noted, the total coefficient was derived by averaging the 3 study site coefficients after weighting them by the inverse of the variance of each site, as described by Fleiss (39).

c Number of participants, excluding refusals and missing observations.

Kappa and ICC coefficients were \( z \)-transformed for estimation of confidence intervals, as described by Rosner (48).

e One outlier (a value of 11,111,109,632 on both test and retest) was removed before calculation of ICCs.

f Weighted k, following the method of Cicchetti and Allison (45).

g The k value was unstable because of a low number of cases or noncases; likewise, a stable k value could not be calculated for any study site for “ever being diagnosed with syphilis” and “frequency of condom use for paid anal/vaginal sex.”

h Because of instability in the k value for at least 1 study site for this item, k was derived directly from the total sample, rather than from the average of the k coefficients from each study site.
obtained because participants in each country were primarily educated men living in an urban setting. In collecting data on abortions from Mexican women, Lara et al. (32) found audio-CASI more appropriate for urban and educated participants than for rural residents in Mexico. Success with audio-CASI has also been reported from Brazil, where it was not only acceptable to men recruited at a health clinic in Rio de Janeiro but also elicited more reports of sensitive sexual behaviors (24, 52).

To our knowledge, no studies have evaluated cross-national test-retest reliability for a sexual behavior survey administered by CASI; however, investigators in 3 reliability studies of Table 3. Select Kappa Statistics and Intraclass Correlation Coefficients for Computer-assisted Self-Interview Items, by Age and Lifetime Number of Female Sex Partners, HPV in Men (HIM) Study, Brazil, Mexico, and the United States, 2005–2006

| Interview Item | Age, years | Lifetime No. of Female Sex Partners | <30 | ≥30 | ≤7 | >7 |
|----------------|------------|-------------------------------------|-----|-----|----|----|
|                |            |                                     |     |     |    |    |
| Interval variables |            |                                     |     |     |    |    |
| Lifetime no. of male anal sex partners | 0.89 | 0.82, 0.93 | 0.50 | 0.34, 0.63 | 0.50 | 0.33, 0.64 | 0.62 | 0.45, 0.75 |
| No. of different female sex partners in the past 3 months | 0.91 | 0.89, 0.92 | 0.60 | 0.54, 0.65 | 0.88 | 0.86, 0.90 | 0.93 | 0.92, 0.94 |
| No. of sex partners other than a “steady” partner in the past 3 months | 0.74 | 0.65, 0.81 | 0.12 | 0.00, 0.28 | 0.86 | 0.80, 0.90 | 0.29 | 0.15, 0.42 |
| Ordinal variables |            |                                     |     |     |    |    |
| Frequency of condom use for anal sex in the past 3 months | 0.81 | 0.76, 0.85 | 0.88 | 0.85, 0.90 | 0.78 | 0.73, 0.82 | 0.85 | 0.82, 0.88 |
| Nominal variables |            |                                     |     |     |    |    |
| Ever being diagnosed with Chlamydia | 0.72 | 0.68, 0.76 | 0.80 | 0.77, 0.83 | 0.77 | 0.73, 0.80 | 0.80 | 0.76, 0.83 |
| Ever being diagnosed with genital warts | 0.81 | 0.78, 0.84 | 0.73 | 0.69, 0.77 | 0.70 | 0.66, 0.74 | 0.79 | 0.75, 0.82 |
| Ever being diagnosed with herpes | _ | 0.86 | 0.84, 0.88 | 0.77 | 0.73, 0.80 | 0.80 | 0.76, 0.83 |
| Ever being diagnosed with hepatitis B | _ | 0.76 | 0.72, 0.79 | 0.69 | 0.64, 0.73 | 0.78 | 0.74, 0.81 |
| Ever being diagnosed with hepatitis C | 0.79 | 0.76, 0.82 | 0.78 | 0.74, 0.81 | 0.72 | 0.68, 0.76 | 0.85 | 0.82, 0.87 |
| Ever being diagnosed with human immunodeficiency virus | _ | 0.72 | 0.68, 0.76 | 0.61 | 0.56, 0.66 | 0.76 | 0.72, 0.80 |
| Ever being diagnosed with nongonococcal urethritis | _ | 0.77 | 0.73, 0.80 | 0.79 | 0.76, 0.82 | _ | _ |
| Ever having a sex partner with a sexually transmitted disease | 0.73 | 0.68, 0.78 | 0.74 | 0.70, 0.78 | 0.76 | 0.72, 0.79 | 0.77 | 0.73, 0.81 |
| Ever having a sex partner with genital warts | 0.82 | 0.79, 0.85 | 0.74 | 0.70, 0.78 | 0.81 | 0.78, 0.84 | 0.73 | 0.68, 0.77 |
| Ever having a sex partner with an abnormal Papanicolaou test | 0.78 | 0.74, 0.81 | 0.79 | 0.76, 0.82 | 0.77 | 0.73, 0.80 | 0.81 | 0.78, 0.84 |
| Ever having vaginal, anal, or oral sex | 0.81 | 0.78, 0.84 | 0.73 | 0.69, 0.77 | 0.81 | 0.78, 0.84 | 0.39 | 0.31, 0.47 |
| Ever having oral sex | 0.89 | 0.87, 0.91 | 0.82 | 0.79, 0.85 | 0.88 | 0.86, 0.90 | 0.76 | 0.72, 0.80 |
| Ever paying a man for sex | 0.71 | 0.58, 0.80 | 0.67 | 0.56, 0.75 | 0.54 | 0.39, 0.66 | 0.75 | 0.65, 0.83 |

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient; HPV, human papillomavirus.

* Omitted are reliability coefficients for 21 items for which all study site coefficients were greater than or equal to 0.81.
* Kappa and ICC coefficients were z-transformed for estimation of confidence intervals, as described by Rosner (48).
* One outlier (a value of 11,111,109,632 on both test and retest) was removed before calculation of ICCs.
* Weighted k, following the method of Cicchetti and Allison (45).
* The k value was unstable because of a low number of cases or noncases; likewise, a stable k value could not be calculated for any category for “ever being diagnosed with syphilis” and “frequency of condom use for paid anal/vaginal sex.”

Am J Epidemiol 2009;170:965–974
adults have reported results for audio-CASI in more homogenous samples (34, 53, 54). These studies are difficult to compare with the current study because of different study populations, test-retest intervals, or reporting methods.

Schlecht et al. (40) assessed the reliability of sexual behavior data collected by face-to-face interview and self-administered questionnaire using pooled, multinational data from 6 studies. In that analysis, reliability suffered substantially when women in different countries reported their lifetime number of sex partners (ICC = 0.08–0.94). In contrast, the current study found almost perfect reliability when this question was asked of men in Brazil, Mexico, and the United States. These heterogeneous results may be due to the use of different survey methods or to the fact that Schlecht et al. assessed the reliability of separate and distinct studies that not only had different survey methods but also different and lengthy test-retest intervals (6 weeks to 5 years) (40). Therefore, the current study may have been more suited for a comparison of cross-national reliability, since identical protocols with relatively similar and short test-retest intervals were used at all study sites.

Refusal rates were generally low, which has been found previously with surveys of human sexuality (28, 55). However, the question on lifetime number of female sex partners was refused by 11.7% and 12.6% of Brazilian participants on test and retest, respectively. This rate of question refusal may be due to the targeted recruitment of Brazilian men in a clinic setting. Higher refusal rates from participants and less preference for using audio-CASI in a clinic setting, as compared with a face-to-face interview, have been reported previously (14, 52). Also noteworthy is that items requiring the participant to provide a numeral, as opposed to a nominal, answer were approximately twice as likely to be refused in Brazil and Mexico as in the United States, where participants refused numerical items and nominal items at about the same rates (data not shown).

This study had limitations. Reliability cannot be used as a surrogate measure for validity, since item reliability is not sufficient for validity. Additionally, because of the targeted recruitment, these results should not be generalized to the entire populations of the 3 study countries.

Comparisons of sexual behavior by country may be helpful in attempts to deliver large-scale programs for the prevention of sexually transmitted diseases; however, if sexual behavior measures are to be used cross-nationally, they should produce reliable data for each locale (8). With few exceptions, we found high reliability using a single CASE instrument in 3 culturally and linguistically distinct countries. While not guaranteeing validity, these results indicate that for the current instrument, use of CASI among men in diverse settings produces reliable data on sexual behavior.

ACKNOWLEDGMENTS

Author affiliations: H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida (Alan G. Nyitray, Jongphil Kim, Mary Papenfuss, Anna R. Giuliano); Mel and Enid Zuckerman College of Public Health, Tucson, Arizona (Alan G. Nyitray, Chiu-Hsieh Hsu); Ludwig Institute for Cancer Research, São Paulo, Brazil (Luiza Villa); and Instituto Nacional de Salud Pública, Cuernavaca, Mexico (Eduardo Lazcano-Ponce).

This work was supported by the National Cancer Institute, US National Institutes of Health (grant RO1CA098803 to A. R. G.).

The authors thank the following persons for their help with gathering the data required for this study: Lenice Galan and Maria Luiza Baggio in São Paulo, Brazil; Manuel Quiterio Trenado in Cuernavaca, Mexico; and Martha Abrahamsen in Tampa, Florida.

The contents of this report are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

Conflict of interest: none declared.

REFERENCES

1. Potterat JJ, Phillips L, Muth JB. Lying to military physicians about risk factors for HIV infections. JAMA. 1987;257(13):1727.
2. Stoneburner RL, Chiasson MA, Solomon K, et al. Risk factors in military recruits positive for HIV antibody [letter]. N Engl J Med. 1986;315(21):1355.
3. Franco EL. Statistical issues in studies of human papillomavirus infection and cervical cancer. In: Franco E, Monsonego J, eds. New Developments in Cervical Cancer Screening. Oxford, United Kingdom: Blackwell Science Ltd; 1997:39–50.
4. Catania JA, Gibson DR, Chitwood DD, et al. Methodological problems in AIDS behavioral research: influences on measurement error and participation bias in studies of sexual behavior. Psychol Bull. 1990;108(3):339–362.
5. Weinhardt LS, Forsyth AD, Carey MP, et al. Reliability and validity of self-report measures of HIV-related sexual behavior: progress since 1990 and recommendations for research and practice. Arch Sex Behav. 1998;27(2):155–180.
6. Downey L, Ryan R, Roffman R, et al. How could I forget—inaccurate memories of sexually intimate moments. J Sex Res. 1995;32:177–191.
7. Schroder KE, Carey MP, Vanable PA. Methodological challenges in research on sexual risk behavior: II. Accuracy of self-reports. Ann Behav Med. 2003;26(2):104–123.
8. Slaymaker E. A critique of international indicators of sexual risk behaviour. Sex Transm Infect. 2004;80(suppl 2):i13–i21.
9. Cleland J, Boerma JT, Carael M, et al. Monitoring sexual behaviour in general populations: a synthesis of lessons of the past decade. Sex Transm Infect. 2004;80(suppl 2):i1–i7.
10. Jobe JB, Pratt WF, Tourangeau R, et al. Effects of interview mode on sensitive questions in a fertility survey. In: Lyberg L, Biemer P, Collins M, et al. eds. Survey Measurement and Process Quality. New York, NY: John Wiley & Sons, Inc; 1997:311–329.
11. Wellings K, Collumbien M, Slaymaker E, et al. Sexual behaviour in context: a global perspective. Lancet. 2006;368(9548):1706–1728.
12. Schrimshaw EW, Rosario M, Meyer-Bahlburg HF, et al. Test-retest reliability of self-reported sexual behavior, sexual orientation, and psychosocial milestones among gay, lesbian, and bisexual youths. Arch Sex Behav. 2006;35(2):225–234.
13. Des Jarlais DC, Panee D, Milliken J, et al. Audio-computer interviewing to measure risk behaviour for HIV among injecting drug users: a quasi-randomised trial. *Lancet*. 1999; 353(9165):1657–1661.

14. Ghanem KG, Hutton HE, Zenilman JM, et al. Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients. *Sex Transm Infect*. 2005;81(5):421–425.

15. Gross M, Holte SE, Marmor M, et al. Anal sex among HIV-seronegative women at high risk of HIV exposure. The HIVNET Vaccine Preparedness Study 2 Protocol Team. *J Acquir Immune Defic Syndr*. 2000;24(4):393–398.

16. Hewett PC, Mensch BS, Ribeiro MC, et al. Using sexually transmitted infection biomarkers to validate reporting of sexual behavior within a randomized, experimental evaluation of interviewing methods. *Am J Epidemiol*. 2008;168(2):202–211.

17. Hewitt M. Attitudes toward interview mode and comparability of reporting sexual behavior by personal interview and audio computer-assisted self-interviewing: analyses of the 1995 National Survey of Family Growth. *Sociol Methods Res*. 2002;31(1):3–26.

18. Richman WL, Kiesler S, Weisband S, et al. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *J Appl Psychol*. 1999;84(5):754–775.

19. Romer D, Hornik R, Stanton B, et al. “Talking” computers: a reliable and private method to conduct interviews on sensitive topics with children. *J Sex Res*. 1997;34(1):3–9.

20. Gates NJ, Sonenstein FL. Heterosexual genital sexual activity among adolescent males: 1988 and 1995. *Fam Plann Perspect*. 2000;32(6):295–297, 304.

21. Hewett PC, Mensch BS, Erulkar AS. Consistency in the reporting of sexual behaviour by adolescent girls in Kenya: a comparison of interviewing methods. *Sex Transm Infect*. 2004;80(suppl 2):i43–i48.

22. Le LC, Blum RW, Magnani R, et al. A pilot of audio computer-assisted self-interview for youth reproductive health research in Vietnam. *J Adolesc Health*. 2006;38(6):740–747.

23. Macalino GE, Celentano DD, Latkin C, et al. Risk behaviors by audio computer-assisted self-interviews among HIV-seropositive and HIV-seronegative injection drug users. *AIDS Educ Prev*. 2002;14(5):367–378.

24. Simoes AA, Bastos FI, Moreira RI, et al. A randomized trial of audio computer and in-person interview to assess HIV risk of injecting drug users: a quasi-randomised trial. *Lancet*. 2005;81(5):421–425.

25. Turner CF, Ku L, Rogers SM, et al. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science*. 1998;280(5365):867–873.

26. Brown JL, Vanable PA, Eriksen MD. Computer-assisted self-interviews: a cost effectiveness analysis. *Behav Res Methods*. 2000;32(4):331–338.

27. Beebe TJ, Harrison PA, Park E, et al. The effects of data collection mode and disclosure on adolescent reporting of health behavior. *Soc Sci Comput Rev*. 2006;24(4):476–488.

28. NIMH Collaborative HIV/STD Prevention Trial Group. The feasibility of audio computer-assisted self-interviewing in international settings. *AIDS*. 2007;21(suppl 2):S49–S58.

29. Hewett PC, Erulkar AS, Mensch BS. The feasibility of computer-assisted survey interviewing in Africa—experience from two rural districts in Kenya. *Soc Sci Comput Rev*. 2004;22(3):319–334.

30. Jaya, Hindin MJ, Ahmed S. Differences in young people’s reports of sexual behaviors according to interview method-ology: a randomized trial in India. *Am J Public Health*. 2008; 98(1):169–174.

31. Jennings TE, Lucenko BA, Malow RM, et al. Audio-CASI vs interview method of administration of an HIV/STD risk of exposure screening instrument for teenagers. *Int J STD AIDS*. 2002;13(11):781–784.

32. Lara D, Strickler J, Díaz-Olavarrieta C, et al. Measuring induced abortion in Mexico: a comparison of four methodologies. *Sociol Methods Res*. 2004;32(4):529–558.

33. Potdar D, Koenig MA. Does Audio-CASI improve reports of risky behavior? Evidence from a randomized field trial among young urban men in India. *Stud Fam Plann*. 2005;36(2):107–116.

34. Williams ML, Freeman RC, Bowen AM, et al. A comparison of the reliability of self-reported drug use and sexual behaviors using computer-assisted versus face-to-face interviewing. *AIDS Educ Prev*. 2000;12(3):199–213.

35. 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):541–552.

36. Giuliano AR, Lazcano-Ponce E, Villa LL, et al. The Human Papillomavirus Infection in Men Study: human papillomavirus prevalence and type distribution among men residing in Brazil, Mexico, and the United States. *Cancer Epidemiol Biomarkers Prev*. 2008;17(8):2036–2043.

37. HPV Study Group in Men From Brazil, USA and Mexico. Human papillomavirus infection in men residing in Brazil, Mexico, and the USA. *Salud Publica Mex*. 2008;50(5):408–418.

38. Nyitray AG, Harris RB, Abalos AT, et al. Test–retest reliability and predictors of unreliable reporting for a sexual behavior questionnaire for U.S. men. *Arch Sex Behav*. 2009;Aug 25 [epub ahead of print]. (http://springerlink.com/content/4805j886x14287/).

39. Fleiss JL. *Statistical Methods for Rates and Proportions*. New York, NY: John Wiley & Sons, Inc; 1981.

40. Schlecht NF, Franco EL, Rohan TE, et al. Repeatability of sexual history in longitudinal studies on HPV infection and cervical neoplasia: determinants of reporting error at follow-up interviews. *J Epidemiol Biostat*. 2001;6(5):393–407.

41. Van Duynhoven YT, Nagelkerke NJ, Van De Laar MJ. Repeatability of the reliability of self-reported drug use and sexual behaviors using computer-assisted versus face-to-face interviewing. *AIDS Educ Prev*. 2000;12(3):199–213.

42. Durant LE, Carey MP. Reliability of retrospective self-reports of sexual and nonsexual health behaviors among women. *J Sex Marital Ther*. 2002;28(4):331–338.

43. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37–46.

44. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Psychol Meas*. 1977;33(1):159–174.

45. HPV Study Group in Men From Brazil, USA and Mexico. Human papillomavirus infection in men residing in Brazil, Mexico, and the USA. *Salud Publica Mex*. 2008;50(5):408–418.

46. Nyitray AG, Harris RB, Abalos AT, et al. Test–retest reliability and predictors of unreliable reporting for a sexual behavior questionnaire for U.S. men. *Arch Sex Behav*. 2009;Aug 25 [epub ahead of print]. (http://springerlink.com/content/4805j886x14287/).

47. Fleiss JL. *Statistical Methods for Rates and Proportions*. New York, NY: John Wiley & Sons, Inc; 1981.

48. Schlecht NF, Franco EL, Rohan TE, et al. Repeatability of sexual history in longitudinal studies on HPV infection and cervical neoplasia: determinants of reporting error at follow-up interviews. *J Epidemiol Biostat*. 2001;6(5):393–407.

49. Van Duynhoven YT, Nagelkerke NJ, Van De Laar MJ. Repeatability of the reliability of self-reported drug use and sexual behaviors using computer-assisted versus face-to-face interviewing. *AIDS Educ Prev*. 2000;12(3):199–213.

50. Durant LE, Carey MP. Reliability of retrospective self-reports of sexual and nonsexual health behaviors among women. *J Sex Marital Ther*. 2002;28(4):331–338.

51. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37–46.

52. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Psychol Meas*. 1977;33(1):159–174.

53. McGraw KO, Wong SP. Forming inferences about some in-ternal correlation coefficients. *Psychol Methods*. 1996;1(1): 30–46.

54. Rosner B. *Fundamentals of Biostatistics*. 5th ed. Pacific Grove, CA: Duxbury Press; 2000.

55. Schroder KE, Carey MP, Vanable PA. Methodological challenges in research on sexual risk behavior: I. Item content, scaling, and data analytical options. *Ann Behav Med*. 2003;26(2):76–103.
50. Carey MP, Carey KB, Maisto SA, et al. Assessing sexual risk behaviour with the Timeline Followback (TLFB) approach: continued development and psychometric evaluation with psychiatric outpatients. *Int J STD AIDS*. 2001;12(6): 365–375.

51. Jaccard J, McDonald R, Wan CK, et al. The accuracy of self-reports of condom use and sexual behavior. *J Appl Soc Psychol*. 2002;32(9):1863–1905.

52. Simões AA, Bastos FI, Moreira RI, et al. Acceptability of audio computer-assisted self-interview (ACASI) among substance abusers seeking treatment in Rio de Janeiro, Brazil. *Drug Alcohol Depend*. 2006;82(suppl 1):S103–S107.

53. Krawczyk CS, Gardner LI, Wang JC, et al. Test-retest reliability of a complex human immunodeficiency virus research questionnaire administered by an audio computer-assisted self-interviewing system. *Med Care*. 2003;41(7):853–858.

54. Wolford G, Rosenberg SD, Rosenberg HJ, et al. A clinical trial comparing interviewer and computer-assisted assessment among clients with severe mental illness. *Psychiatr Serv*. 2008; 59(7):769–775.

55. Johnson AM, Copas AJ, Erens B, et al. Effect of computer-assisted self-interviews on reporting of sexual HIV risk behaviours in a general population sample: a methodological experiment. *AIDS*. 2001;15(1):111–115.