Published in final edited form as:

J Acquir Immune Defic Syndr. 2014 August 15; 66(0 3): S285–S292. doi:10.1097/QAI.0000000000000241.

Validity of Behavioral Measures as Proxies for HIV-related Outcomes

Rick S. Zimmerman, PhD,

University of Missouri – St. Louis, College of Nursing

Donald E. Morisky, PhD,

University of California – Los Angeles Fielding School of Public Health

Lana Harrison, PhD, and

University of Delaware, Department of Sociology and Criminal Justice and Center for Drug and Alcohol Studies

Hayley Mark, PhD, MPH, RN

Johns Hopkins University School of Nursing

Abstract

Over the last 30 years, expectations for the quality, validity, and objectivity of the outcome measures used to assess the impact of behavior change interventions related to HIV have steadily increased. At this point (mid-2014 at this writing), biological evidence or biomarkers of the incidence of HIV and other sexually transmitted infections [STI] in a target population is clearly preferable to self-reports of behavior. This kind of evidence is, however, much more expensive to collect than participants' reports of behavior change (e.g., increased condom use, reduced substance use or abstinence from substance use, and high levels of medication adherence). In addition, while potentially less subject to reporting bias, biomarkers and biological outcomes have their own flaws.

In this paper we review the literature on the validity of self-reports of outcomes most relevant to HIV behavior change interventions, sexual behavior (ever having had sex and condom use), substance use, and medication adherence. We note the extent to which they may be adequate outcome measures without biological data, and the conditions under which they may be most likely to be sufficient. We also argue, like many others, that where possible, both self-report and biological measures should be collected.

Keywords

validity; measurement; behavior; biomarker; HIV

Correspondence should be addressed to Rick Zimmerman, University of Missouri – St. Louis College of Nursing, 220 NAB, 1 University Blvd., St. Louis, MO 63121-4400. zimmermanri@umsl.edu. Phone: +1-314-516-7685; Fax: +1-314-516-7082.

The preparation of this article was facilitated by United States Agency for International Development Cooperative Agreement #AID-OAA-A-12-00058 to the Johns Hopkins University, Bloomberg School of Public Health, Center for Communication Programs.

INTRODUCTION

With significant funding for studies of HIV behavior change interventions over the past 30 years, expectations for data quality have risen. In 1985 a study of increased knowledge about HIV might receive funding and was certainly publishable, but by the 1990s self-reports of behavior (e.g., condom use, frequency of unprotected sex) were minimum end points required of an intervention study. Also in the 1990s, the National Institute of Mental Health (NIMH, one of the primary institutes at NIH funding HIV prevention research) made a conscious decision to fund a large number of prevention trials with self-reported behavior change as the primary outcome rather than a small number of trials using more expensive biological outcomes. Despite this history, as this is being written (in 2014), biological end points such as the incidence of HIV and other STIs in the community are now often expected as indicators of the impact of an intervention for HIV prevention (including secondary prevention in those living with HIV). This paper summarizes current knowledge on the validity of self-reporting of behaviors, particularly those related to outcomes examined in studies of HIV behavior change interventions, such as sexual behaviors, substance use, and medication adherence. Thus, our aim is to provide a view of the extent to which and the conditions under which self-reports may be reasonable proxies for biological evidence or biomarkers of HIV-related outcomes, or valid indicators of those outcomes. Specifically, to better ascertain the reliability of reports of sexual behaviors, we review key studies that have conducted assessments in the following areas: (a) reports of ever having had sex, primarily examining the consistency of those reports over time, including reports at one time of having had sex and reports at a later time of never having had sex; (b) the validity of self-reports of condom use and protected/unprotected sex assessed by using other self-report measures or biological markers; and (c) ways to improve the validity of reporting of sexual behavior. We also review studies of self-reports of substance abuse and medication adherence.

FINDINGS

Sexual Behavior

Unprotected sex as a vector for sexually transmitted infections (STIs) is a major risk factor for morbidity and mortality worldwide. Traditionally, measurement of unprotected sex has relied on data collected through self-reporting. Unlike substance use, where biological testing (urinalysis, hair analysis, etc.) is the “gold standard” to which self-reports can be compared, there is no readily available gold standard for assessing whether or how frequently sexual behaviors (e.g., unprotected anal or vaginal intercourse) have occurred, although some biological measures are available and continue to be tested.^{1,2}

One way that many researchers have attempted to assess the validity of reports of “ever having had sex” is to determine whether those who report “ever having had sex” at one time point indicate at a later time point that they have never had sex. Table 1 presents data from all studies we found showing the proportion of individuals who indicated they had had sex at one time point and later contradicted those reports and said they had never had sex.^{3–7} Percentages ranged from about 4% (for a 2-week interval between measurements)⁷ to about 10–11% for a 1-year interval for US adolescents^{3–5} and up to 16% for a 4 to 6-month time

interval for South African adolescents.⁶ While some of this self-contradiction may be due to respondents reconsidering whether or not the event constituted sex or sexual intercourse (i.e., re-classification of the event as something other than sex), these percentages might also be considered a reasonable approximation of misreporting or lack of validity of responses about ever having had sex. Indeed, discrepancies in longitudinal reports, especially by adolescents, on a variety of variables that are quite consistent over time (e.g., race, age, gender, grade in school, number of siblings) typically range from 3% to 5%, suggesting that reports of ever having had sex are only somewhat more discrepant than reports for these much less sensitive variables. This suggests that at least a portion of the inconsistency observed over time is due to lack of attention to detail on surveys generally, rather than a calculated attempt to keep a secret from or not look bad to the researcher.

While other self-reports, such as self-reported honesty, and comparison of multiple measures of condom use to each other have often been used to assess the validity of self-reports of condom use,⁸ biological outcomes (e.g., positive results on STI tests) and biomarkers are increasingly being used for this purpose.^{1,2} As calls have grown for incorporating biomarkers into studies of sexual risk behavior, an accurate measurement of semen exposure has become an urgent priority for investigators seeking to measure the effect of interventions to reduce sexual exposure to HIV and STIs through various kinds of behavior change, including reducing the number and/or concurrency of partners and increasing male or female condom use.

Currently, biomarkers for unprotected sex are limited to markers of semen exposure in women. There are two broad categories of biomarkers: markers of seminal plasma, such as prostate-specific antigen (PSA),^{1,9} and markers of spermatozoa, such as Y-chromosome DNA (Yc DNA).^{2,10} Since PSA, a protein that occurs at high concentrations in seminal fluid, is expressed independently of spermatozoa, it can be a measure of disease risk even in men with low or no sperm count. Although PSA, which is the most frequently used marker for semen, can be found in women in very small amounts in various tissues and body fluids, it is not usually found in vaginal secretions. Yc DNA is contained in sperm cells, and fragments from spermatozoa can be detected in vaginal fluid. Polymerase chain reaction (PCR) and fluorescent in situ hybridization (FISH) are assays used to detect Y-bearing male cells. Both methods test vaginal secretions collected by swab.

PSA and Yc DNA differ in their sensitivities, rates of decay, and cost. In general, PSA is detected more often than Yc DNA immediately after exposure to large amounts of semen, but Yc DNA has a longer time to decay. Jamshidi and colleagues recently compared sensitivities and rates of decay of PSA and Yc DNA in vaginal fluid specimens. Sensitivities for PCA and Yc DNA immediately after women were inoculated with large volumes of semen (1000 microliters) were 0.96 and 0.72, respectively.¹¹ At 24 hours post-exposure, the sensitivity of being PCA- or Yc DNA-positive had dropped to .21 for PCA and to .49 for YcDNA, and at 48 hours, to 0.07 and .21. Both measures are very limited in their ability to detect exposures to tiny amounts of semen and thus may not be reliable indicators of condom failure. Finally, the assay for Yc-DNA is more expensive than that for PSA (approximately \$20-\$30 for Yc-DNA).¹²

Table 2 presents data from all studies which we could find, dating back as far as the 1960s, that have used positive STI tests or biomarkers to assess the accuracy of reporting consistent condom use.^{13–21} An early test (microscopic search for sperm)¹³ and positive STI tests (either at the time of the self-report or at a follow-up visit after treatment for STIs)^{14,15} have shown self-reports about consistent condom use to be problematic (i.e., associated with positive biological test results) from 10% to 19% of the time. The six published studies using PSA or Yc-DNA showed discrepant results between 13% and 56% of the time, with an average of 38% of individuals with discrepancies.^{16–21}

Biological indicators to assess sexual behavior outcomes of HIV behavior change interventions, while they may be desirable measures, are significantly more expensive and intrusive than self-report methods and thus are not always feasible to collect. A variety of methods have been used by researchers over the past several decades to improve the quality of self-reports about sensitive behaviors such as ever having had sex and condom use. Probably the best known method and the one that has had the most significant impact on the validity of self-reports of sensitive behaviors is the use of audio computer-assisted self-interviewing (ACASI).^{22–23} Individuals are given a laptop and headphones where they can see and hear questions and record their responses. Because this method affords greater privacy than face-to-face interviews and even self-administered questionnaires, where staff hover near individuals completing their surveys, ACASI has been found to yield prevalences of the most sensitive behaviors that can be as high as three times those of prevalences reported using self-administered surveys. Further, the method yields prevalences for a wide variety of behaviors that are 25% to 50% higher than those found using other data collection methods.²² Greater reporting of sensitive behaviors with ACASI than with other data collection methods has generally held for many kinds of studies, with two occasional (and not consistent) exceptions: some uses of ACASI in low income countries did not find that this improved the quality of self-reports, and no differences were found between ACASI and other methods in a few studies where statistical power was inadequate to detect any differences.

As we have noted elsewhere,²⁴ a wide variety of other methodological innovations have been developed to improve the validity of self-reports of sensitive behaviors. Two others that are potentially important include the appropriate choice of reporting intervals for self-reports of behavior (i.e., condom use over the last week, month, or year)⁸ and the use of daily diaries.²⁵ In general, “moderate” reporting periods (3 to 6 months for condom use, for example, for college students) appear to yield the best data, since they allow long enough recording to result in sufficient variance a behavior to detect effects and short enough to enable respondents to remember the behavior reasonably well. Daily (or some other short reporting period, e.g., weekly) diaries generally yield greater frequency of a variety of sensitive behaviors, and thus probably produce more valid results. However, respondent burden and cost can both be much greater than when retrospective reports are collected over longer time intervals, and diaries require that the frequency of the behavior is great enough to measure in this fine-grained way.

In sum, research that has assessed the validity of self-reports of sexual behavior indicates that reports of ever having had sex show relatively low rates of inconsistency (estimates

suggest they may be accurate 85–90% of the time), using relatively weak assessment methods that simply determine whether individuals contradict these reports over a period of time. Reports of condom use, when assessed in conjunction either with an early biomarker or with STI tests, have been found to be consistent with STI test results about 80–90% of the time. Comparison of reports of condom use with more “state-of-the-art” biomarkers (PSA and Yc-DNA) have found that an average of about 38% of respondents over-report consistent condom use (i.e., report consistent condom use though the biomarker suggests at least some intercourse experience(s) during which a condom was not used). Thus, among the three results reported here, consistency in reporting ever having sex, self-reports of condom use when compared to STI test results or an early biomarker, and self-reports of condom use when compared to more recently developed biomarkers, this last result is the outlier. Therefore, we tentatively conclude that self-reports of condom use appear to be quite problematic. However, more work may be necessary on the accuracy of these biomarkers to measure unprotected intercourse. Indeed, while the tests may lack specificity because individuals inaccurately report consistent condom use (e.g., in one initial test of the Yc-DNA biomarker, 11% of individuals who had agreed to use condoms consistently for sex with their partner for the duration of the study had positive test results),¹⁶ some of the error may be due to the test rather than to the person not being honest or accurate in self-reports. There is always more methodological work to be done and new potential methods that integrate more and more technology may yield increasingly valid self-reports of sensitive behaviors, including sexual behaviors related to HIV behavior change outcomes.

Substance Use

In addition to a variety of sexual behaviors such as early adolescents’ reports of ever having had sex and individuals’ reports of having engaged in sexual activity for pay that do not receive approval by society or parents, alcohol (for those under 21) and illicit drug use are viewed negatively by society or at least portions of it. Therefore people may deliberately respond untruthfully to survey questions about their use. As a result, there have been continuing efforts to refine testing for substance abuse. For many years, law enforcement officials used crude forms of drug testing, such as slurred speech tests to check for alcohol intoxication, and looking for pupil constriction or needle marks to assess narcotic use. Walking in a straight line, standing on one foot, or reciting the alphabet backwards are still used routinely in checks for alcohol intoxication. The first biochemical test for drugs was the Breathalyzer, created in the 1950s.²⁶ The military employed the first wide-scale drug testing with urine testing of returning Vietnam veterans in the 1970s. The Drug Use Forecasting (DUF) study of arrestees in major US cities, developed by the National Institute of Justice in the late 1980s, was the first large study to use urinalysis.²⁷ In 1986 the creation of the “Drug-Free Federal Workforce” paved the way for widespread drug testing of both federal and private employees. A monograph on the validity of self-reports of substance use, published by the National Institute on Drug Abuse in 1997, concluded that results for adolescents and the general population of adults were generally valid, but results were more problematic for high-risk and treatment populations.^{28,31}

Urine generally detects drug use in the past 1 to 3 days for most drugs, but not necessarily use in the past few hours (at a cost of between \$10 and \$50 per test or panel of tests). It is

not appropriate for alcohol, which is quickly metabolized.²⁹ Other tissues that can be tested for the presence of illicit drugs include sweat, hair, blood, and meconium from pregnant women. Blood offers a very narrow window of detectability but is preferred in medical settings with proper equipment to determine recent use and impairment. Hair testing generally detects drug use over the past 90 days. Sweat is usually collected during a period of several days to weeks by wearing a tamper-proof pad, although some new tests are being developed to detect recent use. Breath testing is available only for alcohol.²⁹

Most research on the validity of self-reported drug use had been conducted with criminal justice and treatment populations, who are much more likely to be heavily involved with drugs. However, a national study, known as the Validity Study, was conducted in 2000 and 2001 in conjunction with the National Survey of Drug Use and Health (NSDUH), the US's largest and oldest survey of drug use in the general population. This study collected urine and hair samples. The study limited respondents to those between the ages of 12 and 25, in the coterminous United States. All respondents were asked a series of questions about memory and confidentiality and were then asked a second set of questions about recent drug use. They were offered \$25 each for urine and hair samples. Approximately 90% of those interviewed agreed to provide either a hair or urine specimen, and 81% provided both samples.³²

The NSDUH has always been attentive to privacy and validity concerns. The study employed self-reported answer sheets through 1998 and then introduced the audio computer assisted self-interview (ACASI) method in 1999. The Validity Study was methodologically identical to the NSDUH, but a random half of the sample received a persuasion experiment designed to increase validity. The other half of participants were asked a few questions about what they thought of the study. About 3800 urine tests and 2000 hair tests were conducted in this study. Hair and urine specimens were analyzed with screening and confirmation tests, with levels of detection for screening tests set to lower-than-normal levels to ensure that all presumptive positives would be tested by gas chromatography/mass spectrometry (GC/MS) confirmation. Samples were screened and confirmed at actual metabolite levels.

The congruence between self-reported drug use and urine results was generally quite good, although it was dominated by self-reported non-users who tested negative.³² Tobacco and marijuana self-reports had greater congruence with urinalyses than those for cocaine, amphetamines, and opiates. Specifically, for 7-day tobacco use, there were 8.8% ($k=.65$) false negatives in self-reports for those who tested positive by urinalysis (this was the proportion of those with positive urinalysis tests who said they had not smoked). For 30-day marijuana use, the proportion of false negative self-reports was 3.2% ($k=.48$), but for 7-day and 3-day marijuana use, it was 4.5% ($k=.59$) and 5.2% ($k=.59$) respectively. It should also be pointed out that for 7-day and 3-day marijuana use, 3% and 1.5% respectively reported use and tested negative. For cocaine and amphetamines (both substances for which there were very small numbers of "true" positives according to urinalysis), the proportions of false negatives were 0.8% ($k=.28$) and .07% ($k=.10$), respectively. There were over-reporters for all the drugs, although under-reporters generally outnumbered over-reporters. Over-reporting may be related to the fact that the drug tests were not sensitive enough or the time

periods were not specific enough. There was little correspondence between hair and urine results and surprisingly few positive hair tests for any of these substances. Of course, very few respondents tested positive by hair or urine for any of the drugs—a fact of some note!

The Validity Study found discrepancies in both *self-report* and *urine* and *hair* test results that cannot be easily explained. Urine and hair testing technology are designed to err on the side of avoiding false positives, but there appeared to be many cases, particularly with tobacco and marijuana, where self-reported use patterns should have produced positive urine tests. The study concluded that the window of detectability for drugs and the cut-off levels used to assign positive status to a drug test should be considered guidelines at best. Hair testing is still not considered a valid and reliable way to screen for drug use in the community.³³

Although the majority of respondents had little difficulty understanding the drug-related questions, and felt very certain about the accuracy of their answers to these questions, they expressed much less faith in other people. Over half (58%) thought that most people would report using drugs less often than they did. Seventy-five percent said they were not embarrassed by answering the questions, but only 59% felt that most people would feel the same way. Twelve percent were concerned about the confidentiality of their own answers, but over one-quarter thought that most people would be very concerned that others might have access to their answers. Although 90% reported that they were completely truthful in answering the drug-related questions, only 16% thought most people would be completely truthful. The much higher percentages reported for “most people” make one wonder if respondents were projecting their own feelings onto others.

Statistical models found self-reports of perceived privacy and truthfulness of survey responses, as well as religiosity, to be positively associated with validity (i.e., consistency between self-reports and urinalysis results), while difficulty in understanding questions had a negative association with validity. Other predictors of consistency between self-reports and urinalysis were passive exposure and having drug-using friends.³² Both of these may actually have been indicators of passive contamination by marijuana smoked by others.

The Validity Study questionnaire repeated the drug questions at a later point in the same survey. Although there were no significant differences in the prevalence rates in responses to the two sets of questions, a surprising number of respondents gave inconsistent answers on the two sets. Since the second set was delivered after the persuasion experiment was given to half the respondents, it was hypothesized that the persuasion experiment would increase self-reporting rates. This was true, even in logistic regression models. However, some respondents who received the persuasion experiment did change their answers about drug use in the second set of questions, from “use” to “no use.”³²

The results from the Validity Study underscore the fact that despite assurances of confidentiality, under-reporting of use of illicit drugs, especially those with significant legal consequences, continues to be an issue for research. Clearly small proportions of respondents who have recently used a drug do not report that use. As noted above, however,

some of these respondents may be testing positive due to passive exposure to the drug through friends.

Although it is important to employ biological tests to measure licit and illicit drug use, the tests have their limitations. Research is needed to improve the validity of biological testing, as well as to improve methods for asking about sensitive subjects. The Validity Study findings indicate that it may be useful to ask drug-related questions twice, perhaps varying the format. The persuasion experiment increased the accuracy of self-reported drug use, suggesting that it helps to explain to individuals the necessity for accurate information about their drug use.

Medication Adherence

There are a multitude of methods for measuring adherence, each with its own distinctive advantages and disadvantages.³⁴ The simplest and most convenient way to measure this is through self-reports: the patient is asked, “Did you take your medication every day during the past two weeks”? The most likely response is “yes”. The assumption underlying self-report measures is that patients are honest in their reports and recall is perfect. The advantage of this method of assessment is that the process is convenient and inexpensive. However, patients are not always honest and recall is not always perfect. Also, a number of potential validity problems are associated with this method. Responses are personal and idiosyncratic and thus may bear little relationship to “reality” as seen by others. More importantly, people may respond in such a manner as to please the person asking the question. (This is often the case in the collection of data on condom use behavior among commercial sex workers.)

A second way of measuring adherence is to count the number of pills in the container when the patient makes a follow-up visit. This assessment is also simple and cheap; it is based on the assumption that the difference between the number of pills ideally taken and the number of pills remaining in the container equal the number of pills taken. However, patients may forget (purposely) to bring the container with them or they may remove pills from the container prior to the visit. Kalichman and colleagues examined the convergent validity of two self-report adherence measures administered by ACASI: (a) self-reported recall of missed doses (SR-recall) and (b) a single-item visual analogue rating scale (VAS).³⁵ Adherence was also monitored using unannounced phone-based pill counts that served as an objective benchmark. The VAS obtained adherence estimates that paralleled unannounced pill counts. In contrast, SR-recall of missed medications consistently overestimated adherence. The computer-administered VAS was less influenced by response biases than SR-recall of missed medication doses. Adherence self-efficacy has often been found to be a good predictor of adherence behavior.³⁶

Documenting prescription refills is another method of assessing adherence. That is, pharmacy refills are correlated with self-reports to obtain a measure of criterion-related validity. However, people do not always get their medications refilled at the same pharmacy because of discounts offered by some pharmacies. A fourth method of measuring adherence involves the use of a computer chip inserted into the cap of the pill container. Every time the container is opened, the chip will stamp the date and time of the opening. This method is

known as Medication Event Monitoring System or MEMS. The assumption behind MEMS is that every time the cap is opened, the patient takes a pill. This system is designed to provide insights into patterns of adherence. For example, if the medication is prescribed for 30 days and the patient opens the container 30 times, then there is perfect adherence. In a prospective cardiovascular study comparing MEMS Cap to pill counts, MEMS Cap results identified non-adherence 28% of the time compared to only 10% for pill counts.³⁷ However, many patients put their pills into a weekly or monthly container or open the container many times to show their friends. This results in a skewed distribution resulting in both over- and under-reporting. The "gold standard" for medication-taking is a chemical marker inserted into the pill/tablet that can be tested in the urine or finger stick. Unfortunately, this is invasive, and obtrusive and inefficient in a clinical setting. So the vast amount of literature indicates that self-reports are one of the most valid, efficient and a simple method to assess adherence to medical recommendations.

Mosca and colleagues conducted a 4-month prospective, non-randomized, controlled study of elderly patients followed by a community pharmacist.³⁸ Multi-compartment compliance aids (MCAs) (e.g., refrigerator magnets, stickers on mirrors, pill container for weekly/monthly use, and electronic reminder devices (timed alarms, watches, smart phones, and medication containers with chips) were used to assess self-reported adherence and clinical biomarkers of elderly patients followed in a community pharmacy. All received regular pharmacy counselling. Blood pressure (BP), lipid profile and blood sugar were assessed at baseline and monthly. The Morisky self-reported adherence scale was administered at baseline and at the end of the study.³⁹ Significant improvements in the intervention group, but not in a control group, were found for blood glucose levels ($p < 0.001$), total cholesterol ($p = 0.018$), and systolic ($p < 0.001$) and diastolic ($p = 0.012$) blood pressure levels.

Measurement options and relationships observed between adherence measures and biological outcomes in the HIV arena mirror those for adherence generally. Several articles have described the use of various methods of assessing ARV adherence, from self-reports to pill counts to MEMS, including a recent variation of MEMS in which various pills are organized in a pill-box and instances of opening of the lids for the trays in the organizer send data to the researcher via phone or internet transmission.⁴⁰ A recent publication has proposed "quality standards" for the various kinds of self-report measures used to assess adherence, based on best practices data for each.⁴¹ A recent meta-analysis which analyzed the correlations between adherence measures and various viral load measures⁴² found that correlations varied depending on the cutoff used for viral load (VL). When converted from correlations to an effect size d , the relationship between adherence results and viral load was weakest with a cutoff of $VL < 400$ ($d = .35$), moderate for $VL < 100$ ($d = .51$) and moderately large when VL was measured as a continuous variable ($d = .71$). These results show some congruence between VL and adherence measures, but also indicate a significant amount of discrepancy between them, especially for the less specific measures of VL. Another recent study has concluded that while single studies are often under-powered, researchers can achieve greater statistical power for comparisons of measurement methods by combining datasets across studies.⁴³

A final method of assessing adherence is the use of interventions that both measure adherence and remind individuals to take their medication. However, personal reminders can require an extensive time investment from healthcare providers. Electronic reminders (automatically sent reminders without personal contact between the healthcare provider and patient) are therefore being increasingly used to improve adherence. These reminders are automatically sent to patients at the appropriate time without involvement of a healthcare provider. Examples include reminder messages automatically sent to a patient's mobile phone by a short message service (SMS), an electronic reminder device (ERD) that provides patients with an audio and/or visual reminder at predetermined times, and text messages sent to patients' pagers to alert them of their medication.⁴⁴ Interventions using reminders are primarily based on the theory of behavioral learning.⁴⁵ According to this theory, behavior depends on stimuli or cues, either internal (thoughts) or external (environmental) cues, and non-adherent behavior can be modified after sufficient repetition of external stimuli or cues such as reminders.

An example of a simple intervention is a reminder to patients of their desired medication intake pattern. Reminders are especially useful for patients who are unintentionally non-adherent, i.e., patients who are willing to take their medication but forget it or are inaccurate in their timing. Forgetfulness is commonly reported as a barrier to adherence by various patient populations. Though the percentages of patients reporting this barrier range from 22% to 73% across studies, forgetting to take a dose is the most frequently cited reason for non-adherence. A recent review looked at ten studies of patients with HIV, hypertension, glaucoma or asthma, which all used electronic reminders for patients and compared the adherence of those receiving SMS reminders with the adherence of patients using a beeper (type of ERD) as a reminder.⁴⁶ For patients diagnosed with HIV, the review authors found a significant difference in favor of SMS reminders for short-term assessment. Stratified by the type of electronic reminder, this review shows that SMS reminders in particular but ERD as well can be effective strategies for improving patients' adherence in the short run. Interestingly, self-reports were found to be relatively accurate only for hypertension and glaucoma.

CONCLUSION

In this paper we have reviewed current knowledge concerning the validity of self-reports of three kinds of behaviors related to HIV—sexual behaviors, substance use, and medication adherence. In general, by most assessment methods, reports of sexual behaviors seem valid, except when compared with the biomarkers PSA and Pc-DNA. It may be that the greater sensitivity of these new measures puts into question the validity of reports of consistent condom use; or the lack of congruence may raise questions about the validity of these biomarkers to assess self-reports of condom use. For general population samples and for relatively widely used substances (including tobacco use by 12 to 18-year-olds and marijuana use by adolescents and adults), self-reports of substance use appear to be quite valid. Reports on less prevalent drugs and those for which legal penalties are greater (e.g., cocaine and amphetamines) and reports by treatment and high-risk samples appear to be significantly more problematic, with high levels of under-reporting occasionally found as assessed by urinalysis. Measures of medication adherence have improved over the decades.

Generally, self-reports are reasonably valid, although high-risk samples of people who may lose benefits or be expelled from a program if they are not taking their medications regularly may significantly under-report. Newer adherence measures (such as MEMS and EDR methods) appear to produce more valid data than most of the other reporting methods (e.g., clinical self-report, pill counts, prescription refill data).

In conclusion, consistent with recent statements from experts in the field,^{47–48} we propose that self-reports and biological measures (when available and shown to be highly specific and sensitive) should be jointly used for all three types of behaviors. We propose this strategy rather than the use of biological end points only, because there are some weaknesses and flaws in biological measures *as outcomes of a behavior change intervention*. Indeed there may be problems with the specificity of some biomarkers; that is, the biomarker may incorrectly suggest a poor behavioral outcome in general or for some people. Unfortunately, while collecting both self-report and biological data is considered the best approach, so far we have seen very little development of specific methods for combining the data from those two types of methods.

Measurement of key outcomes for HIV behavior change interventions should use methods likely to yield the highest levels of validity of self-reports for all three types of behaviors. These include using a reporting period that is moderate in length for sexual behavior and substance use, using the best reporting methods available (i.e., ACASI for all three behaviors and daily diaries under certain circumstances), and using continuing technological improvements such as video or phone-based evidence of medication-taking or electronic pill caps to assess medication adherence.

REFERENCES

1. Macaluso M, Lawson L, Akers R, Valappil T, Hammond K, Blackwell R, Hortin G. Prostate-specific antigen in vaginal fluid as a biologic marker of condom failure. *Contraception*. 1999 Mar; 59(3):195–201. PubMed PMID: 10382083. [PubMed: 10382083]
2. Zenilman JM, Yuenger J, Galai N, et al. Polymerase chain reaction detection of Y chromosome sequences in vaginal fluid: Preliminary studies of a potential biomarker for sexual behavior. *Sex Transm Dis*. 2005; 32(2):90–94. [PubMed: 15668614]
3. Zimmerman RS, Langer LM. Improving prevalence estimates of sensitive behaviors: The randomized lists technique and self-reported honesty. *J Sex Res*. 1995; 32:107–117.
4. Upchurch DM, Lillard LA, Aneshensel CS, et al. Inconsistencies in reporting the occurrence and timing of first intercourse among adolescents. *J Sex Res*. 2002; 39(3):197–206. [PubMed: 12476267]
5. Rosenbaum JE. Reborn a virgin: Adolescents' retracting of virginity pledges and sexual histories. *Am J Pub Health*. 2006; 96(6):1098–1103. [PubMed: 16670225]
6. Palen L-A, Smith EA, Caldwell LL, et al. Inconsistent reports of sexual intercourse among South African high school students. *J Adoles Health*. 2008; 42(3):221–227.
7. Rosenbaum JE. Truth or consequences: The inter-temporal consistency of adolescent self-report on the Youth Risk Behavior Survey. *Am J Epi*. 2009; 169(11):1388–1397.
8. 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.
9. Macaluso M, Lawson L, Akers R, Valappil T, Hammond K, Blackwell R, Hortin G. Prostate-specific antigen in vaginal fluid as a biologic marker of condom failure. *Contraception*. 1999 Mar; 59(3):195–201. PubMed PMID: 10382083. [PubMed: 10382083]

10. Ghanem KG, Melendez JH, McNeil-Solis C, et al. Condom use and vaginal Y-chromosome detection: The specificity of a potential biomarker. *Sex Transm Dis.* 2007; 34(8):620–623. [PubMed: 17308500]
11. Jamshidi R, Penman-Aguilar A, Wiener J, Gallo M, Zenilman JM, Melendez JH, Macaluso M. Detection of Two Biological Markers of Intercourse: Prostate-specific Antigen and Y-Chromosomal DNA. *Contracep.* 2013; 88:749–757.
12. Gallo MF, Steiner MJ, Hobbs MM, Warner L, Jamieson DJ, Macaluso M. Biological Markers of Sexual Activity: Tools for Improving Measurement in HIV/Sexually Transmitted Infection Prevention Research. *Sex Transm Dis.* 2013; 40(6):447–452. [PubMed: 23677018]
13. Udry JR, Morris NM. A method for validation of report sexual data. *J Marriage Fam.* 1967; 29(3): 443–446.
14. Zenilman JM, Weisman CS, Rompalo Am, et al. Condom use to prevent incident STDs: The validity of self-reported condom use. *Sex Transm Dis.* 1995; 22:15–21. [PubMed: 7709320]
15. Shew ML, Remafedi GJ, Bearinger LH, et al. The validity of self-reports condom use among adolescents. *Sex Transm Dis.* 1997; 24(9):503–510. [PubMed: 9339967]
16. Jadack RA, Yuenger J, Ghanem KG, et al. Polymerase chain reaction detecting of Y-chromosome sequences in vaginal fluid of women accessing a sexually transmitted disease clinic. *Sex Transm Dis.* 2006; 33:22–25. [PubMed: 16385218]
17. Gallo MF, Behets FM, Steiner MJ, et al. Prostate-specific antigen to ascertain reliability of self-reported coital exposure to semen. *Sex Trans Dis.* 2006; 33:376–479.
18. Gallo MF, Behets FM, Steiner MJ, et al. Validity of self-reported ‘safe sex’ among female sex workers in Mombasa, Kenya—PSA analysis. *Int J STD AIDS.* 2007; 18:33–38. [PubMed: 17326860]
19. Rose E, DiClemente RJ, Wingood GM, et al. The validity of teens’ and young adults’ self-reported condom use. *Arch Pediatr Adolesc Med.* 2009; 63:61–64. [PubMed: 19124705]
20. Minnis AM, Steiner MJ, Gallo MF, et al. Biomarker validation of reports of recent sexual activity: Results of a randomized controlled study in Zimbabwe. *Am J Epidemiol.* 2009; 170:918–924. [PubMed: 19741042]
21. Aho J, Koushik A, Diakite SI, et al. Biological validation of self-reported condom use among sex workers in Guinea. *AIDS Behav.* 2010; 14:1287–1293. [PubMed: 19680799]
22. 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. [PubMed: 9572724]
23. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Pub Health.* 2005; 27(3):281–291.
24. Zimmerman, R.; Atwood, K.; Cupp, P. Methods for collecting data about sensitive topics. Chapter in DiClemente. In: Crosby; Salazar, editors. *Methods for Health Promotion.* San Francisco, CA: Jossey-Bass; 2006.
25. Leigh BC, Gillmore MR, Morrison, et al. Comparison of diary and retrospective measures for recording alcohol consumption and sexual activity. *J Clin Epidemiol.* 1998; 51(2):119–127.
26. Borkenstein RF, Smith HW. The Breathalyzer and its applications. *J Med Sci Law.* 1961; 2:13–22.
27. Wish ED, O’Neil JA. Urine testing for drug use among male arrestees—United States 1989. *MMWR.* 1989; 38(45):780–783. [PubMed: 2509890]
28. Harrison, LD.; Hughes, A., editors. *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates.* NIDA Research Monograph 167. Washington, DC: Supt. of Docs, U.S. Govt. Print. Off.; 1997. <http://www.nida.nih.gov/pdf/monographs/monograph167/download167.html>
29. Harrison LD. The validity of self-reported data on drug use. *J Drug Issues.* 1995; 25(1):91–111.
30. Magura S, Kang S-Y. Validity of self-reported drug use in high risk populations: A meta-analytic review. *Subst Use & Misuse.* 1996; 31(0):1131–1151.
31. Gwet, KL. *Handbook of inter-rater reliability.* 3rd ed.. Gaithersburg, MD: Advanced Analytics, LLC; 2012.

32. Harrison, LD.; Martin, SS.; Enev, T., et al. Comparing Drug Testing and Self-Report of Drug Use among Youths and Young Adults in the General Population. DHHS Publication No. SMA 07-4249. Rockville, MD: SAMHSA; <http://www.oas.samhsa.gov/validity/drugTest.cfm>
33. Drug Testing Advisory Board. Minutes, September 11, 2013. Bethesda, MD: Center for Substance Abuse Prevention. SAMHSA; 2013.
34. Liu H, Golin CE, Miller LG, et al. How best to measure medication adherence? A comparison study of multiple measures of adherence to inhibitors of the HIV protease. *Ann Int Med*. 2006; 134:968–977. [PubMed: 11352698]
35. Kalichman SC, Amaral CM, Swetzes C, et al. A simple single item rating scale to measure adherence: Further evidence for convergent validity. *J Intl Assoc Phys AIDS Care*. 2009; 8(6): 367–374.
36. Gifford, AI; Bormann, JE.; Shively, MJ., et al. Predictors of self-reported adherence and plasma HIV concentrations in patients on multidrug antiretroviral regimens. *J Acquire Immune Defic Syndr*. 2000; 23(5):386–395.
37. Parker CS, Chen Z, Kimmel SE. Adherence to warfarin assessed by electronic pill caps, clinician assessment, and patient reports: Results from the IN-RANGE study. *J Gen Intern Med*. 2007; 22(9):1254–1259. [PubMed: 17587092]
38. Mosca C, Castel-Branco M, Ribeiro-Rama AC, et al. Assessing the impact of multi-compartment compliance aids on clinical outcomes in the elderly: a pilot study. *Int J Clin Pharm*. 2014; 36(1): 98–104. [PubMed: 24293332]
39. Morisky DE, DiMatteo MR. Improving the measurement of self-reported medication nonadherence: Final response. *J Clin Epidemio*. 2011; 64:258–263.
40. Bangsberg DR. Preventing HIV antiretroviral resistance through better monitoring of treatment adherence. *J Infect Dis*. 2008; 197:S272–S278. [PubMed: 18447613]
41. Williams AB, Amico KR, Bova C, Womack JA. A proposal for quality standards for measuring medication adherence in research. *AIDS Behav*. 2013; 17:284–297. [PubMed: 22407465]
42. Kahana SY, Rohan J, Allison S, et al. A meta-analysis of adherence to antiretroviral therapy and virologic responses in HIV-infected children, adolescents, and young adults. *AIDS Behav*. 2013; 17:41–60. [PubMed: 22411426]
43. Liu H, Wilson IB, Goggin K, et al. MACH14: A multi-site collaboration on ART adherence among 14 institutions. *AIDS Behav*. 2013; 17:127–141. [PubMed: 22864921]
44. Wise J, Operario D. Use of electronic reminder devices to improve adherence to antiretroviral therapy: A systematic review. 2008; 22(6):495–504.
45. Leventhal H, Cameron L. Behavioral theories and the problem of compliance. *Patient Educ Couns*. 1987; 10:117–138.
46. Vervloet M, Linn JA, van Weert DH, et al. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. *J Am Med Inform Assoc*. 2012; 19:696–704. [PubMed: 22534082]
47. Pequegnat W, Fishbein M, Celentano D, Ehrhardt A, Garnet G, Holtgrave D, Jaccard J, Schachter J, Zenilman J. NIMH/APPC Workgroup On Behavioral and Biological Outcomes in HIV/STD Prevention Studies: A position statement. *Sex Trans Dis*. 2000; 27(3):127–132.
48. Fishbein M, Pequegnat W. Evaluating AIDS prevention interventions using behavioral and biological outcome measures. *Sex Trans Dis*. 2000; 27(2):100–110.

Table 1

Discrepancy between 2 reports of ever having had sex in published studies

Reference	Population	Time Interval Between Measures	% of those who said they had sex at Time 1 who said they hadn't at Time 2	Sample Size	95% Confidence Interval
3	US 9th and 10th graders	1 year	9.7%	1886	8.4%, 11.0%
4	US 7th to 12th graders *	1 year	11.2%	3855	10.2%, 12.2%
5	US 7th to 12th graders *	1 year	10.7%	3855	9.7%, 11.7%
6	South African 8th to 10th graders	4-6 months	16% (average over 4 time intervals)	713	13.3%, 18.7%
7	US 9th to 12th graders	2 weeks	4.1%	4619	3.5%, 4.7%

* Same dataset, slightly different analyses

Table 2

Discrepancy Between Self-Reports of Condom Use and Biological Measures

Reference	Population	Biological Measure	% of Results Discrepant with Self-Reports of Consistent Condom Use	Sample Size	95% Confidence Interval
13	Low SES African American women	Microscopic search for sperm	5 of 36 (14%) positive lab results for 3 of 15 women	58	5.0%, 23.0%
14	Adult men and women recruited at Baltimore City Health Dept. STI clinics	Incident STIs at 3-month follow-up visit after treatment for any STIs	18.9%	598	15.8%, 22.0%
15	Adolescents recruited at school and community-based clinics	Diagnosed STIs	10.6%	540	8.0%, 13.2%
16	Adult men and women recruited at Baltimore City Health Dept. STI clinics	Yc-DNA	55.6%	141	47.4%, 63.8%
17	Female sex workers in Madagascar	PSA	39%	332	33.8%, 44.2%
18	Female sex workers in Kenya	PSA	13%	210	8.4%, 17.5%
19	African American female teens and young adults recruited at STI clinic and family planning clinic	Yc-DNA	33.9%	484	29.7%, 38.1%
20	Sexually active HIV-negative women in Zimbabwe	PSA	48%	910	45.8%, 51.2%
21	Female sex workers in Guinea	PSA	36%	223	29.7%, 42.3%