REVIEW PAPER

HIV Risk Behavior Self-Report Reliability at Different Recall Periods

Lucy E. Napper · Dennis G. Fisher ·
Grace L. Reynolds · Mark E. Johnson

Abstract Few studies have investigated the optimal length of recall period for self-report of sex and drug-use behaviors. This meta-analysis of 28 studies examined the test-retest reliability of three commonly used recall periods: 1, 3, and 6 months. All three recall periods demonstrated acceptable test-retest reliability, with the exception of recall of needle sharing behaviors and 6-months recall of some sex behaviors. For most sex behaviors, a recall period of 3 months was found to produce the most reliable data; however, 6 months was best for recalling number of sex partners. Overall, shorter periods were found to be more reliable for recall of drug-use behaviors, though the most reliable length of recall period varied for different types of drugs. Implications of the findings and future directions for research are discussed.

Keywords Reliability · Recall period · Sex · Drug use · Self-report

Introduction

The accurate assessment of health risk behaviors is essential for those wanting to describe and predict trends, identify populations at risk, and evaluate the effectiveness of interventions, as well as to advocate for support and to develop policies and programs (Brener et al. 2003; Falck et al. 1992; Kalichman et al. 1997; Kauth et al. 1991; Weinhardt et al. 1998). The assessment of Human Immunodeficiency Virus (HIV) risk behaviors is complex due to the inherently private, often stigmatized, and sometimes illegal nature of these drug-use and sexual risk behaviors. To assess such risk behaviors, researchers often rely on individuals' self-reports for both practical and ethical reasons.

A variety of approaches have been utilized to assess the reliability of people's self-reports of HIV risk behaviors. Research has demonstrated that data collected from high-risk populations, such as drug users, are, on the whole, reliable (Darke 1998; Dowling-Guyer et al. 1994; Goldstein et al. 1995; Johnson et al. 2000). However, there are a variety of factors likely to affect the reliability of data collected in this manner, such as individuals' motivation and ability to respond accurately. Fear of legal reprisal or self-presentation biases may lead participants to hide behaviors that they perceive to be undesirable or stigmatized (Catania et al. 1990a; Hser et al. 1992; Latkin et al. 1993; Weinhardt et al. 1998).

Reliability of measures of HIV risk behavior may also be impacted by factors associated with the measure itself (Blair and Burton 1987). For example, the length of time for which participants are asked to recall risk behaviors, or the recall period, is likely to affect the reliability of a measure (Blair and Burton 1987). It seems reasonable to expect people to find it easier to remember behaviors over recent short periods of time compared to longer periods of time. Easier-to-remember recall periods should be more reliable than longer more difficult-to-recall periods.

The length of the recall period used for self-report instruments has implications for the strategy used to recall behavioral frequency (Conrad et al. 1998; Jaccard and Wan

L. E. Napper · D. G. Fisher (✉) · G. L. Reynolds
Center for Behavioral Research and Services, California State University, Long Beach, 1090 Atlantic Avenue, Long Beach, CA 90813, USA
e-mail: dfisher@csulb.edu

M. E. Johnson
Behavioral Health Research and Services, University of Alaska Anchorage, Anchorage, AK, USA

1995; McFarlane and Lawrence 1999). This, in turn, affects the reliability of self-report data. Shorter recall periods are more likely to lead people to use episodic recall strategies, such as enumeration (Bogart et al. 2007), that are thought to be more reliable than other strategies (Conrad et al. 1998; Jaccard et al. 2002). Enumeration involves scanning a recall period for a particular behavior and counting all recalled instances of that behavior that occurred within the recall period (Jaccard and Wan 1995). Episodic enumeration may be common when behaviors are infrequent, irregular, or distinctive (Blair and Burton 1987; Conrad et al. 1998). However, if a behavior occurs frequently and episodes are indistinct, enumeration may become increasingly difficult, time consuming, and less likely to occur (Blair and Burton 1987).

Instead of enumeration, rate-based inferences (Conrad et al. 1998) may be used to recall how often an event occurs during a representative period (e.g., once a week) and multiply it by the length of the recall period (e.g., 12 times in a 3-months period). If the number of events or the rate of events is not retrievable, other strategies such as qualitative impressions, memory assessments, or normative expectations may be used (Conrad et al. 1998; Jaccard and Wan 1995). The use of such mental calculations and impressions can be imprecise and inconsistent (Bogart et al. 2007; Downey et al. 1995; Jaccard et al. 2002). As the recall period increases, so too may the use of these recall strategies, leading to the risk of reduced reliability of self-report data (Conrad et al. 1998).

Several researchers have examined the relationship between recall period and the test-retest reliability of self-report data, with some providing evidence to support the concept that shorter recall periods may be more reliable than longer recall periods. Kauth et al. (1991) compared sexual risk behavior self reports for 2-week, 3- and 12-months periods and argued that, as the length of recall period increased, inconsistency in responding increased. However, this study did not use a true test-retest methodology, and instead extrapolated data from 2 weeks and 3 months to be equivalent to 12 months. Catania (unpublished data cited in Catania et al. 1990b) assessed the test-retest reliability of college students' reports of frequency of vaginal intercourse using varying recall periods. Their results suggested that, as the recall interval increased from 1 month to 1 year, test-retest reliability of the measure decreased. In a study examining recall of substance use, Martin et al. (1998) found that shorter recall periods (30 and 90 days) were more reliable than longer recall periods (180 and 360 days).

On the other hand, some studies have failed to find differences in reliability between different recall periods, or have found inconsistent results. Using the Timeline Followback (TLFB) method, a variety of studies found no differences in reliability as the length of recall period increased (Carey et al. 2001; Ehrman and Robbins 1994; Levy et al. 2004; Sacks et al. 2003). Klinkenberg et al. (2002) compared recall periods of 3 and 6 months and found that recall of alcohol and drug use was more reliable at 3 months and recall of number of sexual partners more reliable at 6 months. Jaccard et al. (2002) examined self-reports of condom use and sexual behaviors, and found recall periods of 3 and 6 months to be more optimal than 1 and 12 months. Jaccard et al. (2004) compared recall periods of 1, 3, 6 and 12 months, and found that, for those with multiple sex partners, recall errors in self-report of number of sex partners increased as recall period increased. However, the correlation between self-report and behavior was highest at 6 months which Jaccard et al. attributed to restricted variability in responses for shorter recall periods. Graham et al. (2003) compared recall periods of one, two, and 3 months and found evidence that, for a high frequency behavior (i.e., heterosexual vaginal sex), accuracy decreased as recall period increased. However, findings indicated that, for infrequent behaviors, reliability did not decrease over time.

Given that research addressing the reliability of different recall periods has produced varying results and little consensus on what length of recall period is optimal (Jaccard et al. 2002), there have been a variety of calls for further research to examine this topic (Catania et al. 1990b, 1993; Downey et al. 1995; Noar et al. 2006; Schroder et al. 2003). The lack of agreement on the appropriate recall period reduces the comparability of different studies examining the impact of HIV risk behaviors (Catania et al. 1990b) and hinders research in this area. To address this lack of research, the present meta-analysis reviews and extends previous studies by comparing the test-retest reliability of three commonly used recall periods (1, 3, and 6 months). In doing so, our aim was to inform future researchers about differences in reliability and to draw attention to the importance of comparing the reliability of different recall periods, so that researchers may develop optimal self-report instruments for assessing HIV risk behaviors.

Method

Selection of Studies

Papers published in English that examined the test-retest reliability of measures of sex and drug use behaviors were selected. Studies were identified using electronic databases (PsycInfo, PsycArticles, PubMed), and review articles (Noar et al. 2006; Weinhardt et al. 1998). Multiple search terms were used in combination including recall period, reference period, self-report, test-retest, sex, drug use, HIV

risk, and reliability. Authors were contacted to request any relevant published or unpublished data. The references sections of potential articles were checked for additional citations.

All studies considered for inclusion had to meet the following criteria:

1. Studies had to include test-retest reliability of recall of HIV risk-related behaviors. Studies that examined consistency between two partners' recall of behaviors, comparisons of two different approaches to measurement (for example, comparison of diary methods and single-item recall methods), or compared recall of behaviors for the same length of recall period, but for two time periods which did not overlap, were excluded from the analysis.
2. Only studies that reported assessing behaviors over the prior 1, 3, and/or 6 months were considered for inclusion. Studies that did not report a specific recall period in any form were excluded.
3. Measures assessing HIV risk-related behaviors, including sex behaviors and drug use behaviors, were included. Studies that examined the reliability of self reports of attitude, opinion, craving, or substance dependence were excluded.
4. Only studies that reported the reliability of continuous measures of risk behaviors were considered for inclusion. Because of differences in the ways people are likely to recall frequency data (e.g., How many times do you use crack?) and categorical data (e.g., Did you ever use crack?), studies that only assessed categorical data were excluded.
5. Studies that reported Pearson's correlation coefficients or interclass correlations were included in the analysis. Studies that examined ordinal level data (e.g., response options of: once a month, once a week, once a day) were excluded from the analysis.
6. Only studies for which the sample size was available were included.

In total, 28 studies yielded over 300 test-retest effect sizes. Based on the studies that reported the demographics of their samples, ages of those included in the studies ranged from 12 to 74 years old, with the majority being male. The sample included in-treatment and out-of-treatment drug users, sex workers, psychiatric patients, and adolescents. A description of the studies included in the meta-analysis can be found in Tables 1 and 2.

Aggregation of Within-Sample Effect Sizes

The majority of studies included the assessment of test-retest reliability for multiple items. To avoid including multiple statistics from the same study in the meta-analysis,

leading to non-independence of the effect sizes, correlations from the same study were aggregated to provide a mean correlation (Lipsey and Wilson 2001, p. 125). Aggregated effect sizes were calculated separately for drug and sex behaviors, and were used to compute combined effect sizes examining self-report of all drug use behaviors and all sex behaviors. In addition, separate analyses were performed looking at more specific drug and sex behaviors, for example, use of different types of drugs and self-reports of different types of sex behaviors. For studies that reported more than one statistic for one of these more specific behaviors, these statistics were aggregated before being included in the analysis of the different types of behaviors. For example, if a study reported ten items assessing drug use behavior, these items were aggregated for the combined drug use analysis. If the same study reported three items assessing marijuana use, these three items were aggregated for the marijuana analysis. If the sample sizes varied for individual analyses within studies, the mean correlation was calculated by converting the correlations to Fisher's Z, weighting the values by $n-3$, calculating the mean, and then transforming the mean back into a correlation coefficient.

Correlational Analysis

Effect sizes were computed using the procedures outlined in Hedges and Olkin (1985). An effect size was calculated for each behavior by converting relevant correlations to Fisher's Z, weighting the values by the sample size, calculating the mean, and then transforming the mean back into a correlation coefficient. Using the formulas supplied by Hedges and Olkin (1985, p. 227), 95% confidence intervals (95% CI) were calculated. Z tests were used to compare effect sizes for the three recall periods, both for the combined drug and sex variables and for more specific drug-use and sex behaviors.

Results

Results of the meta-analysis are presented in Tables 3 (drug variables) and 4 (sex variables). For each analysis, the population reliability coefficient, number of studies included, total sample size, and 95% CI are reported. Reliability coefficients for the combined-drug variables are provided in Table 3 and labeled "All drug variables." As indicated, these reliabilities are good when a 30-days, 3-, or 6-months recall period are used. Across all drug variables, the test-retest reliability for a recall period of 30 days ($r = .90$) was found to be greater than that of 3 months ($r = .84$; $Z = 4.30$, $P < .001$) and 6 months ($r = .83$; $Z = 4.93$, $P < .001$). The reliability of the data using a

Table 1 Description of studies reporting test-retest reliability of drug use variables

Study	Sample	Measure (recall period)	Reliabilities
Blake et al. (1992)	127 US military personnel	Marijuana use (30 days)	1.00
Carey et al. (2004)	132 psychiatric outpatients. 64% male; mean age 44.1 years (US sample)	TLFB: Marijuana use and total drug use (30 and 90 days)	.94 (30 days) .91–.94 (90 days)
Day et al. (2004)	27 heroin injectors; 70% male; mean age 32 years (Australian sample)	TLFB: Number of days used heroin, cocaine, amphetamines, benzodiazepines, cannabis (6 months)	.67–.88
Dowling-Guyer et al. (1994)[a]	218 out-of-treatment drug users; 74% male; mean age 39.9 years (US samples)	RBA: Crack, cocaine, heroin, marijuana, speedball, amphetamine and methadone use, and injection drug use (30 days)	.69–.87
Ehrman and Robbins (1994)	59 heroin users in an outpatient methadone treatment program; 98% male; mean age 41 years (US sample)	TLFB: Heroin and cocaine use (30 and 180 days)	.77–.82 (30 days) .91–.95 (180 days)
Fals-Stewart et al. (2000)	113 substance abuse outpatients; 71% male; mean age 27.4 years (US sample)	TLFB: Use of amphetamines, cannabis, cocaine, hallucinogens, opiates, sedatives (30 and 90 days)	.72–.93 (30 days) .71–.92 (90 days)
Johnson et al. (2000)	259 out-of-treatment drug users; 67.6% male; mean age 38.2 years (US sample)	RBFA: Use of marijuana, crack, cocaine, heroin, speedball, opiates, amphetamines and injection behavior (30 days)	.48–.96
Krenz et al. (2004)[a]	36 outpatient and inpatient drug users; 68% male; mean age 29.7 years (Swiss sample)	ASI: Heroin, methadone, other opiates, barbiturates, sedatives, cocaine, amphetamines, cannabis & hallucinogens use (30 days)	.80–1.0
Levy et al. (2004)	93 adolescents attending primary-care medical clinics; 28% male; age range 12–19 years (US sample)	TLFB: Marijuana use (30, 90 days)	.70–.89 (30 days) .83–.93 (90 days)
Martin et al. (1998)	103 young adults accessing outpatient and inpatient drug treatment; 81% male; mean age 20.4 years (Canadian sample)	DUHF: Use of cannabis, cocaine and hallucinogens (30, 90, and 180 days)	.30–.88 (30 days) .34–.90 (90 days) .49–.91 (180 days)
Matt et al. (2003)	88 cigarette smokers; 51% male; mean age 29.5 years (US sample)	Use of marijuana, amphetamines, cigarettes and alcohol (30 days)	.90–.99
Miele et al. (2000)	175 inpatient and outpatient drug treatment patients; 62% male; mean age 35.6 years (US sample)	SDSS: Number of days used cocaine, heroin, cannabis and sedatives (30 days)	.50–.82
Myers et al. (1990)	196 IDUs and sex partners of IDUs; 74% male (US sample)	AIA: Frequency of marijuana, crack, cocaine, amphetamine, heroin, speedball, methadone, other opiates use, sharing works (6 months)	−.11–.93
Needle et al. (1995)	214 drug users; 70.3% male; mean age 38 years (US sample)	RBA: Crack, cocaine, heroin use (30 days)	.69–.78
Ross et al. (1995)	23 injecting drug users; 87% male; mean age 28 years (Australian sample)	Number people accepted used needle/syringe from (6 months)	.56
Sacks et al. (2003)	158 homeless adults (US sample)	TLFB: Number days used cocaine, cannabis, any illicit drugs (30 days, 6 months)	.72–.81 (30 days) .89–.93 (6 months)
Scheurich et al. (2005)	30 alcohol-dependent inpatients; 73% male; mean age 44.7 years (German sample)	Form 90: Tranquilizer and sedative use (90 days)	.64–1.00
Slesnick and Tonigan (2004)	37 homeless youth; 49% male; age range 12–17 years (US sample)	Form 90: Cocaine and marijuana use (90 days)	.40–.97
Westerberg et al. (1998)	34 treatment-seeking clients; 53% male; mean age 36.3 years (US sample)	Form 90: Cocaine, opiates, marijuana, stimulants, tranquilizer use (90 days)	.02–.80

Table 1 continued

Study	Sample	Measure (recall period)	Reliabilities
Williams et al. (2000)	392 drug users; 69% male; mean age 36.2 years (US sample)	Computer assisted and face-to-face interviews: Crack, cocaine, heroin, speedball (30 days)	.22–.91

ASI Addiction Severity Index. *DUHF* Drug Use History Form. *AIA* AIDS Initial Assessment Questionnaire. *TLFB* Timeline Followback. *RBA* Risk Behavior Assessment. *RBFA* Risk Behavior Follow-up Assessment. *SDSS* Substance Dependence Severity Scale

[a] Authors from these studies provided additional unpublished data

Table 2 Description of studies reporting test-retest reliability of sex behavior variables

Study	Sample	Measure (recall period)	Reliabilities
Carey et al. (2001)	66 psychiatric outpatients; 50% male; age range 18–60 years (US sample)	TLFB: Number of partners, vaginal and oral sexual events (1, 3 months)	.71–.97 (1 month) .80–.95 (3 months)
Dowling-Guyer et al. (1994)[a]	218 out-of-treatment drug users; 74% male; mean age 39.9 years (US sample)	RBA: Number of partners, frequency of sex (vaginal, oral, anal), condom use (30 days)	.07–1.00
Johnson et al. (2000)	259 out-of-treatment drug users; 67.6% male; mean age 38.2 years (US sample)	RBFA: Number of partners, times had vaginal sex, condom use, times traded sex for money or drugs (30 days)	.57–.87
McKinnon et al. (1993)	16 sexually active psychiatric patients; 66.7% male; age range 18–59 years (US sample)	Sexual Risk Behavior Assessment Schedule: Number of partners, sexual episodes and proportion vaginal intercourse (6 months)	.60–.88
McLaws et al. (1990)	30 men; majority were male prostitutes (Australian sample)	Number of partners, frequency anal sex, oral sex and condom use (30 days)	.08–.98
Myers et al. (1990)	196 IDUs and sex partners of IDUs; 74% male (US sample)	Number of partners, condom use, frequency of vaginal, oral, anal sex for male respondents only (6 months)	.12–1.00
Needle et al. (1995)	214 drug users; 70.3% male; mean age 38 years (US sample)	RBA: Number of partners, IDU partners, times had vaginal sex, days had sex (30 days)	.66–.83
Ross et al. (1995)	23 injecting drug users; 87% male; mean age 28 years (Australian sample)	% time used condoms for vaginal and oral sex (6 months)	−.40–.80
Schrimshaw et al. (2006)	64 gay/lesbian/bisexual youth; 55% male; mean age 18.2 years (US sample)	SERBES: Number of same-sex partners, oral, anal, and vaginal-digital encounters (3 months)	−.01–1.00
Sieving et al. (2005)	152 sexually active 13 to18 year old females seeking reproductive health services (US sample)	Number of partners and frequency of vaginal sex (3 months, 6 months)	.53–.86 (3 months) .48–.82 (6 months)
Sneed et al. (2001)	83 Thai and Korean participants; 51% male; mean age 29 years (US sample)	Adapted version of the NIMH Multisite HIV Prevention Trial survey: Vaginal, anal and oral sex, condom use (90 days)	.97–1.00
Sohler et al. (2000)	39 homeless men with severe mental illness; age range 24–57 years (US sample)	SERBAS: Number of partners, vaginal sex, anal sex, condom use (6 months)	.49–.90
Weinhardt et al. (1998)	110 college students; 53.6% male; mean age 19.7 years	TLFB and single item measures: Number partners and vaginal and oral sex practices (1 month and 3 months)	.85–.98 (30 days) .81–.97 (3 months)
Williams et al. (2000)	392 drug users; 69% male; mean age 36.2 years (US sample)	Computer assisted and face-to-face interviews: Number of partners, drug-using partners, vaginal sex (30 days)	−.13–.99

RBA Risk Behavior Assessment. *RBFA* Risk Behavior Follow-up Assessment. *TLFB* Timeline Followback. *SERBAS* Sexual Risk Behavior Assessment Schedule

[a] Authors from this study provided additional unpublished data

3-months recall period did not differ significantly from the 6-months recall period ($Z = .58$, *ns*).

Marijuana use was found to be most reliably reported when a recall period of 30 days was used ($r = .92$), in comparison to both 3 months ($r = .85$; $Z = 6.49$, $P < .001$) and 6 months ($r = .85$; $Z = 6.65$, $P < .001$). The 3- and 6-months recall periods were not found to differ significantly ($Z = .11$, *ns*). Self-reports of cocaine use were

Table 3 Test-retest reliability of drug use variables

Risk behavior variable	30 days recall	3 months recall	6 months recall
All drug variables[a]	.898	.844	.833
	18 (1718)	7 (458)	6 (440)
	.89–.91	.81–.87	.80–.86
Marijuana use[a]	.923	.848	.846
	12 (1202)	6 (450)	4 (456)
	.91–.93	.82–.87	.82–.87
Cocaine/crack use[a]	.800	.875	.865
	14 (1367)	4(198)	5 (320)
	.78–.82	.84–.90	.83–.89
Amphetamines[a]	.930		.796
	6 (175)		2 (78)
	.91–.95		.69–.87
Heroin[a]	.804		.831
	11 (1002)		3 (181)
	.78–.83		.78–.87
Shared works[a]	.691		.731
	4 (662)		3 (461)
	.65–.73		.69–.77

Cells were left empty where data were not available

[a] Line 1 contains estimated population correlation. Line 2 contains the number of correlations, and the total sample size in parenthesis. Line 3 contains the 95% confidence intervals of the estimated population correlation

Table 4 Test-retest reliability of sex behavior variables

Risk behavior variable	30 days recall	3 months recall	6 months recall
All sex variables[a]	.817	.949	.822
	15 (1040)	6 (361)	5 (289)
	.86–.89	.94–.96	.78–.86
Vaginal sex only[a]	.836	.968	.616
	13 (759)	4 (172)	4 (132)
	.81–.86	.96–.98	.49–.72
Anal sex[a]	.897		.577
	2 (102)		3 (122)
	.85–.93		.44–.69
Oral sex[a]	.768	.897	.610
	4 (215)	5(192)	2 (73)
	.71–.82	.86–.92	.44–.74
Number of partners[a]	.786	.848	.930
	14 (973)	3 (282)	4 (314)
	.76–.81	.81–.88	.91–.94

Cells were left empty where data were not available

[a] Line 1 contains estimated population correlation. Line 2 contains the number of correlations, and the total sample size in parenthesis. Line 3 contains the 95% confidence intervals of the estimated population correlation

found to be more reliable for longer recall periods. Compared to the 30-days recall period ($r = .80$), both the 3-months ($r = .88$; $Z = 3.34$, $P < .001$) and 6-months recall periods ($r = .87$; $Z = 3.44$, $P < .001$) were more reliable, and did not differ significantly from one another ($Z = .45$, ns).

Test-retest reliability of amphetamine use was found to be higher for a recall period of 30 days ($r = .93$) compared to 6 months ($r = .80$; $Z = 4.12$, $P < .001$). The reliability of self reports of heroin use did not differ significantly between the 30 days ($r = .80$) and 6-months recall period ($r = .83$; $Z = -1.00$, ns). Self-reports of sharing works (needles/syringes/cookers/cottons) had the lowest test-retest reliability, but did not differ significantly between 30 days ($r = .69$) and 6 months ($r = .73$; $Z = -1.33$, ns). The literature review revealed too few studies reporting a recall period of 3 months for amphetamines, heroin, and sharing works; thus, for these three variables, analyses were limited to comparing 30-days and 6-months recall periods.

The combined-sex behaviors measures are labeled "All sex variables" in Table 4. Self report of sex behaviors across all items was found to be most reliable when a recall period of 3 months was used ($r = .95$), compared to both 30 days ($r = .82$; $Z = 10.99$, $P < .001$) and 6 months ($r = .82$; $Z = 8.31$, $P < .001$). There was no significant difference in the reliability between the 30-days and 6-months recall periods ($Z = .23$, ns).

A similar pattern of results was found for vaginal sex and oral sex. The recall period of 3 months was most reliable for recall of vaginal sex ($r = .97$), when compared to 30 days ($r = .84$; $Z = 10.01$, $P < .001$) and 6 months ($r = .62$; $Z = 11.47$, $P < .001$). The 30-days recall period was more reliable than the 6-months recall period ($Z = 5.14$, $P < .001$). The recall period of 3 months was most reliable for oral sex ($r = .90$), when compared to 30 days ($r = .77$; $Z = 4.41$, $P < .001$) and 6 months ($r = .61$; $Z = 5.34$, $P < .001$). The 30-days recall period was more reliable than the 6-months recall period for recall of oral sex ($Z = 2.22$, $P < .001$). The 30-days recall period ($r = .90$) was also more reliable than the 6-months recall period for recall of anal sex ($r = .58$; $Z = 5.87$, $P < .001$).

For recall of number of sexual partners, the recall period of 6 months was more reliable ($r = .93$) than both 30 days ($r = .79$; $Z = 9.17$, $P < .001$) and 3 months ($r = .85$; $Z = 4.94$, $P < .001$). The 3-months recall period was more reliable than the 30-days recall period ($Z = 2.77$, $P < .01$).

Discussion

Using meta-analysis, the present study sought to examine the test-retest reliability of commonly used recall periods.

Understanding what influence, if any, the length of recall period has on the reliability of self-report data is important for designing measures. The current analysis demonstrates that the reliability of self-reports of sex and drug behaviors, for different lengths of recall periods, depends upon the particular behavior assessed.

For most drug-use behaviors, all three recall periods (30 days, 3, and 6 months) demonstrated acceptable reliability. Overall, the 30-days recall period produced the most reliable recall period when examining all drug-use behavior items combined. When more specific behaviors were examined, self-report of marijuana was found to be most reliable for shorter recall periods (30 days). This finding is consistent with the suggestion that for more frequent behaviors, shorter recall periods may be more accurate (McFarlane and Lawrence 1999), with marijuana being the most frequently reported illicit drug used in the United States (Office of Applied Studies 2007). Amphetamine use was also found to be more reliable for shorter recall periods. Very few studies were identified that examined the reliability of self-report of amphetamine use. Future studies are needed to examine whether shorter recall periods provide more reliable alternatives for self-reports of amphetamine use.

Although past researchers have suggested that self-reports of drug use may be more reliable with shorter recall periods (Kauth et al. 1991; Martin et al. 1998), the current analysis suggests that this is not always the case for all drugs. For example, whereas length of recall period did not affect the reliability of self-reports of heroin use and sharing of drug-use equipment, cocaine/crack use was more reliably reported when longer recall periods (3 and 6 months) were used. Several reasons may explain why shorter recall periods lead to less reliable self-reports for these drugs. Attenuation due to restricted range may reduce the reliability estimates for shorter periods during which there is less variability in reports of frequency of drug use. Drug use patterns have been found to be highly variable (Samuels et al. 1992), and it may be that longer recall periods are needed to capture some of these behaviors reliably.

Changes in reliability of recall of sex behaviors and partners across the recall periods may also reflect attenuation due to restricted range. The reliability of recall of number of sexual partners was found to increase as the length of recall period increased. This may reflect an increase in variability of number of sexual partners reported as the length of recall increases. For recall of sexual activity, a recall period of 30 days may be too short for some individuals to report having engaged in this behavior. That is, short recall periods may produce little variations in self-reports of frequency of sexual behaviors compared to, for example, a 3-months recall period. On the other hand,

individuals may not be able to accurately recall their sexual behaviors over a longer period of 6-months, thus causing reliability to decrease.

Jaccard et al. (2002) predicted that, for self-reports of sexual behaviors, moderate length recall periods (3 or 6 months) compared to shorter recall periods (1 month) would be more reliable. These researchers argued that moderate-length recall periods may lead to those who engage in sex infrequently providing fairly reliable estimates of sexual behavior using enumeration strategies. In contrast, those who engage in frequent sex may be discouraged from using episodic strategies and instead use rule-based strategies which, for frequent behavior, may be more reliable. Therefore, moderate-length recall periods may maximize accurate recall of sexual behaviors for both those who engage in sex frequently and infrequently. This type of pattern of results is seen for the self-report of vaginal and oral sex, with the 30-days recall period producing lower reliability estimates than the 3-months recall period.

Past research has demonstrated that several factors are likely to influence the reliability of self-report data, including the frequency of behaviors. Recall of less frequent behavior appears to lead to the use of more reliable recall strategies, such as enumeration (Bogart et al. 2007), fewer errors in recall (McLaws et al. 1990) and more reliable recall (Downey et al. 1995). Differences in the patterns of test-retest reliabilities across the recall periods for sex and drug-use behaviors may, in part, reflect differences in frequency of behavior. One limitation of the current study is that data were not available to directly test the hypothesis regarding the interaction between length of recall period and frequency of behavior on the accuracy of recall. Nor were enough data available to allow direct tests of whether some approaches to measurement were more reliable than others, for example, whether the use of the Timeline Followback (TLFB) differed in reliability compared to other methods. Further research is needed to address how different factors interact to influence reliability of recall. Research of this nature would allow researchers to better select a reliable recall period based on, for example, characteristics of the behavior (i.e., frequency, desirability), question format, or cognitive strategy that individuals are likely to employ. Although Jaccard and Wan (1995) have begun to explore one type of research paradigm that would address some of these issues, there continues to be a lack of research in this area.

Test-retest reliability provides one approach to examining the accuracy of self reports of HIV risk behaviors that is not without its limitations. Although consistent self-reports across two time points can result from accurate recall of behavior, it may also be due to recall of responses

provided at the first administration of questions, or some combination of the two. Other approaches, such as diary methods, biological markers, or comparisons of drug-using or sex partners' self-reports, have been used to address the reliability and validity of sex and drug-use self-report data (Darke 1998; Jaccard et al. 2002; Jaccard and Wan 1995; Stopka et al. 2004). To augment the examination of the reliability of self-reported HIV risk behaviors, these methods could be employed to examine the influence of length of recall period on the accuracy of self-report data (Graham et al. 2003).

The current study highlights the need for more data to be collected addressing the reliability of different recall periods. Many studies failed to use or report a specific recall period. This limitation makes it difficult to accumulate data on the influence of length of recall period. Other studies reported only the reliability of combined items, making it difficult to tease apart findings and examine the reliability of items measuring different types of drug use or sexual behaviors. The findings of the present study demonstrate that reliability may differ depending upon the particular drug or sex behavior being assessed, thus making it important to be able to examine the reliability of self-reports of these behaviors separately. The current paper draws attention to the lack of research addressing the optimal length of recall period for assessing self-reports of different HIV risk-related behaviors. For example, few studies investigated the test-retest reliability of anal sex or needle sharing behaviors.

The results of the current meta-analysis support the use of 3-months recall periods for self reports of sexual behaviors, including vaginal and oral sex. Further data are needed to examine whether a 3-months recall period may also provide a reliable approach for self-reports of anal sex. Self-reports of number of sexual partners were more reliable when longer recall periods were used, supporting previous research examining recall of sexual partners (Klinkenberg et al. 2002). Marijuana use was most reliably reported over a 30-days recall period, whereas crack/cocaine self-reports were more reliable over 3- and 6-months recall periods. The most appropriate recall period may depend on a combination of factors including the research question, or the manner of assessment of behaviors (McFarlane and Lawrence 1999). However, understanding what influence the length of recall period may have on the reliability of recall is important for making informed decisions for designing self-report measures.

References

References marked with an asterisk indicate studies included in the meta-analysis

Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *The Journal of Consumer Research, 14*, 280–288. doi:10.1086/209112.

*Blake, S. M., Sharp, E. S., Temoshok, L., & Rundell, J. R. (1992). Methodological considerations in developing measures of HIV risk-relevant behaviors and attitudes: An empirical illustration. *Psychology and Health, 6*, 265–280.

Bogart, L. M., Walt, L. C., Pavlovic, J. D., Ober, A. J., Brown, N., & Kalichman, S. C. (2007). Cognitive strategies affecting recall of sexual behavior among high-risk men and women. *Health Psychology, 26*, 787–793.

Brener, N. D., Billy, J. O. G., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health, 33*, 436–457.

*Carey, M. P., Carey, K. B., Maisto, S. A., Gordon, C. M., & Weinhardt, L. S. (2001). Assessing sexual risk behavior with the Timeline Followback (TLFB) approach: Continued development and psychometric evaluation with psychiatric outpatients. *International Journal of STDs and AIDS, 12*, 365–375.

*Carey, K. B., Carey, M. P., Maisto, S. A., & Henson, J. M. (2004). Temporal stability of the timeline followback interview for alcohol and drug use with psychiatric outpatients. *Journal of Studies on Alcohol, 65*, 774–781.

Catania, J. A., Gibson, D. R., Chitwood, D. D., & Coates, T. J. (1990a). Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. *Psychological Bulletin, 108*, 339–362.

Catania, J. A., Gibson, D. R., Marin, B. V., Coates, T. J., & Greenblatt, R. M. (1990b). Response bias in assessing sexual behaviors relevant to HIV transmission. *Evaluation and Program Planning, 13*, 19–29.

Catania, J. A., Turner, H., Pierce, R. C., Golden, E., Stocking, C., & Binson, D. (1993). Response bias in surveys of AIDS-related sexual behavior. In D. G. Ostrow, R. C. Kessler, et al. (Eds.), *Methodological issues in AIDS behavioral research* (pp. 133–162). New York: Plenum.

Conrad, F. G., Brown, N. R., & Cashman, E. R. (1998). Strategies for estimating behavioural frequency in survey interviews. *Memory, 6*, 339–366.

Darke, S. (1998). Self-report among injection drug users: A review. *Drug and Alcohol Dependence, 51*, 253–263.

*Day, C., Collins, L., Degenhardt, L., Thetford, C., & Maher, L. (2004). Reliability of heroin users' reports of drug use behaviour using a 24 months timeline follow-back technique to assess the impact of the Australian heroin shortage. *Addiction Research and Theory, 12*, 433–443.

*Dowling-Guyer, S., Johnson, M. E., Fisher, D. G., Needle, R., Watters, J., Andersen, M., et al. (1994). Reliability of drug users' self-reported HIV risk behaviors and validity of self-reported recent drug use. *Assessment, 1*, 383–392.

Downey, L., Ryan, R., Roffman, R., & Kulich, M. (1995). How could I forget? Inaccurate memories of sexually intimate moments. *Journal of Sex Research, 32*, 177–191.

*Ehrman, R. N., & Robbins, S. J. (1994). Reliability and validity of 6-months timeline reports of cocaine and heroin use in a methadone population. *Journal of Consulting and Clinical Psychology, 62*, 843–850.

Falck, R., Siegal, H. A., Forney, M. A., & Wang, J. (1992). The validity of injection drug users self-reported use of opiates and cocaine. *Journal of Drug Issues, 22*, 823.

*Fals-Stewart, W., O'Farrell, T. J., Freitas, T. T., McFarlin, S. K., & Rutigliano, P. (2000). The timeline followback reports of psychoactive substance use by drug-abusing patients: Psychometric properties. *Journal of Consulting and Clinical Psychology, 68*, 134–144.

Goldstein, M. F., Friedman, S. R., Neaigus, A., Jose, B., Ildefonso, G., & Curtis, R. (1995). Self-reports of HIV risk behavior by injecting drug users: Are they reliable? *Addiction, 90*, 1097–1104.

Graham, C. A., Catania, J. A., Brand, R., Duong, T., & Canchola, J. A. (2003). Recalling sexual behavior: A methodological analysis of memory recall bias via interview using the diary as the gold standard. *Journal of Sex Research, 40*, 325–332.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.

Hser, Y., Anglin, M. D., & Chou, C. (1992). Reliability of retrospective self-report by narcotics addicts. *Psychological Assessment, 4*, 207–213.

Jaccard, J., McDonald, R., Wan, C. K., Dittus, P. J., & Quinlan, S. (2002). The accuracy of self-reports of condom use and sexual behavior. *Journal of Applied Social Psychology, 32*, 1863–1905.

Jaccard, J., McDonald, R., Wan, C. K., Guilamo-Ramos, V., Dittus, P., & Quinlan, S. (2004). Recalling sexual partners: The accuracy of self-reports. *Journal of Health Psychology, 9*, 699–712.

Jaccard, J., & Wan, C. K. (1995). A paradigm for studying the accuracy of self-reports of risk behavior relevant to AIDS: Empirical perspectives on stability, recall bias, and transitory influences. *Journal of Applied Social Psychology, 25*, 1831–1858.

*Johnson, M. E., Fisher, D. G., Montoya, I., Booth, R., Rhodes, F., Andersen, M., et al. (2000). Reliability and validity of not-in-treatment drug users' follow-up self-report. *AIDS and Behavior, 4*, 373–380.

Kalichman, S. C., Kelly, J. A., & Stevenson, L. Y. (1997). Priming effects of HIV risk assessments on related perceptions and behavior: An experimental field study. *AIDS and Behavior, 1*, 3–8.

Kauth, M., St. Lawrence, J. S., & Kelly, J. A. (1991). Reliability of retrospective assessments of sexual HIV risk behavior: A comparison of biweekly three-months, and 12-months self-reports. *AIDS Education and Prevention, 3*, 207–214.

Klinkenberg, W. D., Calsyn, R. J., Morse, G. A., McCudden, S., Richmond, T. L., & Burger, G. K. (2002). Consistency of recall of sexual and drug-using behaviors for homeless persons with dual diagnosis. *AIDS and Behavior, 6*, 295–307.

*Krenz, S., Dieckmann, S., Favrat, B., Spagnoli, J., Leutwyler, J., Schnyder, C., et al. (2004). French version of the Addiction Severity Index (5th edition): Validity and reliability among Swiss opiate-dependent patients. *European Addiction Research, 10*, 173–179.

Latkin, C. A., Vlahov, D., & Anthony, J. C. (1993). Socially desirable responding and self-reported HIV infection risk behaviors among intravenous drug users. *Addiction, 88*, 517–525.

*Levy, S., Sherritt, L., Harris, S. K., Gates, E. C., Holder, D. W., Kulig, J. W., et al. (2004). Test-retest reliability of adolescents' self-report of substance use. *Alcoholism, Clinical and Experimental Research, 28*, 1236–1241.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*Martin, G. W., Pearlman, S., & Li, S. (1998). The test-retest reliability of the frequency of multiple drug use in young drug users entering treatment. *Journal of Substance Abuse, 10*, 275–290.

*Matt, G. E., Turingan, M. R., Dinh, Q. T., Felsch, J. A., Hovell, M. F., & Gehrman, C. (2003). Improving self-reports of drug-use: Numeric estimates as fuzzy sets. *Addiction, 98*, 1239–1247.

McFarlane, M., & Lawrence, S. J. S. (1999). Adolescents' recall of sexual behavior: Consistency of self-report and effect of variations in recall duration. *Journal of Adolescent Health, 25*, 199–206.

*McKinnon, K., Cournos, F., Meyer-Bahlburg, H. F., Guido, J. R., Caraballo, L. R., Margoshes, E. S., et al. (1993). Reliability of sexual risk behavior interviews with psychiatric patients. *American Journal of Psychiatry, 150*, 972–974.

*McLaws, M., Oldenburg, B., Ross, M. W., & Cooper, D. A. (1990). Sexual behavior in AIDS-related research: Reliability and validity of recall and diary measures. *Journal of Sex Research, 27*, 265–281.

*Miele, G. M., Carpenter, K. M., Smith Cockerham, M., Dietz Trautman, K., Blaine, J., & Hasin, D. S. (2000). Substance dependence severity scale (SDSS): Reliability and validity of a clinician-administered interview for DSM-IV substance use disorders. *Drug and Alcohol Dependence, 59*, 63–75.

*Myers, M. H., Snyder, F. R., Bryant, E. E., & Young, P. A. (1990). *Report on the reliability of the AIDS initial assessment questionnaire*. Bethesda, MD: NOVA Research Company.

*Needle, R., Fisher, D. G., Weatherby, N., Chitwood, D. D., Brown, B., Cesari, H., et al. (1995). The reliability of self-reported HIV risk behaviors of drug users. *Psychology of Addictive Behaviors, 9*, 242–250.

Noar, S. M., Cole, C., & Carlyle, K. (2006). Condom use measurement in 56 studies of sexual risk behavior: Review and recommendations. *Archives of Sexual Behavior, 35*, 327–345.

Office of Applied Studies. (2007). *Results from the 2006 national survey on drug use and health: National findings*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

*Ross, M. W., Stowe, A., Wodak, A., & Gold, J. (1995). Reliability of interview responses of injecting drug users. *Journal of Addictive Diseases, 14*(2), 1–12.

*Sacks, J. A., Drake, R. E., Williams, V. F., Banks, S. M., & Herrell, J. M. (2003). Utility of the time-line follow-back to assess substance use among homeless adults. *Journal of Nervous and Mental Disease, 191*, 145–153.

Samuels, J. F., Vlahov, D., Anthony, J. C., & Chaisson, R. E. (1992). Measurement of HIV risk behaviors among intravenous drug users. *British Journal of Addiction, 87*, 417–428.

*Scheurich, A., Muller, M. J., Anghelescu, I., Lorch, B., Dreher, M., Hautzinger, M., et al. (2005). Reliability and validity of the form 90 interview. *European Addiction Research, 11*, 50–56.

*Schrimshaw, E. W., Rosario, M., Meyer-Bahlburg, H. F. L., & Scharf-Matlick, A. A. (2006). Test-retest reliability of self-reported sexual behavior, sexual orientation, and psychosexual milestones among gay, lesbian, and bisexual youths. *Archives of Sexual Behavior, 35*, 225–234.

Schroder, K. E. E., Carey, M. P., & Vanable, P. A. (2003). Methodological challenges in research on sexual risk behavior: II. *Accuracy of self-reports. Annals of Behavioral Medicine, 26*, 104–123.

*Sieving, R., Hellerstedt, W., McNeely, C., Fee, R., Snyder, J., & Resnick, M. (2005). Reliability of self-reported contraceptive use and sexual behaviors among adolescent girls. *Journal of Sex Research, 42*, 159–166.

*Slesnick, N., & Tonigan, J. S. (2004). Assessment of alcohol and other drug use by runaway youths: A test-retest study of the Form 90. *Alcoholism Treatment Quarterly, 22*(2), 21–34.

*Sneed, C. D., Chin, D., Rotheram-Borus, M. J., Milburn, N. G., Murphy, D. A., Corby, N., et al. (2001). Test-retest reliability for self-reports of sexual behavior among Thai and Korean respondents. *AIDS Education and Prevention, 13*, 302–310.

*Sohler, N., Colson, P. W., Meyer-Bahlburg, H. F. L., & Susser, E. (2000). Reliability of self-reports about sexual risk behavior for HIV among homeless men with severe mental illness. *Psychiatric Services, 57*, 814–816.

Stopka, T. J., Springer, K. W., Khoshnood, K., Shaw, S., & Singer, M. (2004). Writing about risk: Use of daily diaries in understanding drug-user risk behaviors. *AIDS and Behavior, 8*, 73–85.

*Weinhardt, L. S., Carey, M. P., Maisto, S. A., Carey, K. B., Cohen, M. M., & Wickramasinghe, S. M. (1998). Reliability of the timeline follow-back sexual behavior interview. *Annals of Behavioral Medicine, 20*, 25–30.

Weinhardt, L. S., Forsyth, A. D., Carey, M. P., Jaworski, B. C., & Durant, L. E. (1998). Reliability and validity of self-report measures of HIV-related sexual behavior: Progress since 1990 and recommendations for research and practice. *Archives of Sexual Behavior, 27*, 155–180.

*Westerberg, V. S., Tonigan, J. S., & Miller, W. R. (1998). Reliability of form 90D: An instrument for quantifying drug use. *Substance Abuse, 19*, 179–189.

*Williams, M. L., Freeman, R. C., Bowen, A. M., Zhao, Z., Elwood, W. N., Gordon, C., et al. (2000). A comparison of the reliability of self-reported drug use and sexual behaviors using computer-assisted versus face-to-face interviewing. *AIDS Education and Prevention, 12*, 199–213.