Addictive Behaviors 33 (2008) 1131–1139

Contents lists available at ScienceDirect

Addictive Behaviors

Comparison between self-report and hair analysis of illicit drug use in a community sample of middle-aged men

David M. Ledgerwood [a], Bruce A. Goldberger [b,c], Nathan K. Risk [d], Collins E. Lewis [d], Rumi Kato Price [d,*]

[a] Wayne State University School of Medicine, Department of Psychiatry and Behavioral Neurosciences, Detroit, MI 48207, USA
[b] University of Florida College of Medicine, Department of Pathology, Immunology and Laboratory Medicine, Gainesville, FL 32610, USA
[c] University of Florida College of Medicine, Department of Psychiatry, Gainesville, FL 32610, USA
[d] Washington University School of Medicine, Department of Psychiatry, St. Louis, MO 63110-1093, USA

ARTICLE INFO

Keywords:
Hair drug testing
Self-report
Cocaine
Marijuana
Opiates
Methamphetamine

ABSTRACT

Discrepancies between biological assays and self-report of illicit drug use could undermine epidemiological research findings. Two objectives of the present study are to examine the degree of agreement between self-reported illicit drug use and hair analysis in a community sample of middle-aged men, and to identify factors that may predict discrepancies between self-report and hair testing. Male participants followed since 1972 were interviewed about substance use, and hair samples were analyzed for marijuana, cocaine, opiates, phencyclidine (PCP) and methamphetamine using radioimmunoassay and gas chromatography–mass spectrometry (GC–MS) techniques. Self-report and hair testing generally met good, but not excellent, agreement. Apparent underreporting of recent cocaine use was associated with inpatient hospitalization for the participant's most recent quit attempt, younger age, identifying as African American or other, and not having a diagnosis of antisocial personality disorder. The overestimate of marijuana use relative to hair test was associated with frequent use since 1972 and providing an inadequate hair sample. Additional research is needed to identify factors that differentially affect the validity of both hair drug testing and self-report.

1. Introduction

Most large-scale studies have typically assessed illicit drug use prevalence through self-report (e.g., SAMHSA, 2003). Inaccurate reporting may seriously underestimate the prevalence of substance use in community samples (Magura & Kang, 1996), and could lead to erroneous assumptions about the severity of drug problems in society, which in turn may mislead decisions regarding the need for treatment and prevention services. Drug testing is widely accepted in treatment and research settings because it is generally considered more accurate than self-report. Urine testing is the most commonly employed technique for most substances, and it has demonstrated greater sensitivity and specificity than self-report in assessing recent substance use (Kim & Hill, 2003; Lu, Taylor, & Riley, 2001). Urine testing, however, generally can detect drug use only for a short time after consumption, usually 1 to 2 days. An exception is cannabinoids, which are detectable for weeks in chronic users (Hawks & Chiang, 1986). Whereas brief intervals of drug detection using urine may be optimal for drug abuse treatment research, it is not adequate for epidemiologic studies. For these studies, hair drug tests would prove to be more beneficial because of the longer duration of detection and because it is considered less invasive and more readily accepted in community settings.

* Corresponding author. Washington University School of Medicine, Department of Psychiatry, Medical Box 8134, 660 S. Euclid Ave., St. Louis, MO 63110-1093, USA. Tel.: +1 314 286 2282; fax: +1 314 286 2285.
E-mail addresses: dledgerw@med.wayne.edu (D.M. Ledgerwood), bruce-goldberger@ufl.edu (B.A. Goldberger), nathan@rkp.wustl.edu (N.K. Risk), price@rkp.wustl.edu (R. Kato Price).

0306-4603/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.addbeh.2008.04.009

Hair drug testing studies using the gas chromatography–mass spectrometry (GC–MS) technique have demonstrated good sensitivity and specificity (Baumgartner, Hill, & Blahd, 1989; Jurado & Sachs, 2003), particularly when standardized testing procedures are used (Montagna, Polettini, Stramesi, Groppi, & Vignalli, 2002). Hair drug testing is especially sensitive to qualitative (i.e., presence or absence) detection of substances; however, quantitative detection of some drugs and their metabolites (e.g., THC) is still considered difficult (Jurado & Sachs, 2003).

Several studies have compared rates of drug use detected by hair with those detected by self-report. Swartz, Swanson, and Hannon (2003) compared hair, urine, and self-report in a naturalistic study of schizophrenia. They found that 31% had a positive hair test for illicit drugs, while only 16.3% had a positive self-report and 12.4% had a positive urine sample. Another study of heroin users found that 34% did not report concomitant cocaine use that was revealed by hair drug testing (Tassiopoulos et al., 2004). Hair drug testing using the radioimmunoassay procedure was a more sensitive indicator of cocaine use than either self-report or urine testing in pregnant women (Kline, Ng, Schittini, Levin, & Susser, 1997). Using a GC–MS procedure, Farabee and Fredlund (1996) found substantial underreporting of cocaine use among a sample of jail inmates. Prior drug treatment, interviewer ratings of participant truthfulness, and participant understanding of the interview were all significant factors in determining the concordance between self-report and hair drug test result.

Only a handful of studies have employed hair drug testing to validate the prevalence of illicit drug use in general community samples. Fendrich, Johnson, Sudman, Wislar, and Spiehler (1999) found that hair drug testing resulted in a greater prevalence of cocaine use than was found using a self-report survey. The authors noted a trend toward finding more heroin use via hair drug testing than with self-report. A more recent study found that residents of more racially segregated neighborhoods were more likely to underreport lifetime use of cocaine or crack (Richardson, Fendrich, & Johnson, 2003). Another study of a community sample of 627 randomly selected adults ages 18–40 used a computer assisted audio interview, and collected self-report, urine, oral fluid and hair samples to assess substance use (Fendrich, Johnson, Wislar, Hubbell, & Spiehler, 2004). Hair drug testing resulted in a higher estimate of cocaine use than self-report. However, self-report resulted in greater detection of marijuana use than did hair testing.

The present study further explores the relationship between self-report and illicit drug use as indicated by hair testing. Testing was carried out at the third wave (1996–7) of a longitudinal study of a community sample. The participants were all middle-aged men including men who were deployed to Vietnam and a matched non-veteran comparison group who have been followed since 1972 (Price, Risk, Murray, Virgo, & Spitznagel, 2001; Robins & Helzer, 1975). We examined the detection rate and degree of agreement between self-report and results of hair testing for use of marijuana, opiates, cocaine, and methamphetamine, and further examined demographic, substance use, and psychiatric factors associated with the disagreement between self-report and hair testing. We hypothesized that demographic, psychiatric, and drug use factors that predicted reporting disagreements differ for each class of drug. Such findings will delineate the potentially complex relationship between self-report and biological measures of illicit drug use. These data should guide researchers in selecting the appropriate biological measures to be used in conjunction with self-report.

2. Methods

2.1. Study sample

Two previous surveys of the Vietnam Era Study (VES) cohort were conducted in 1972 and 1974, initiated by the White House Special Action Office for Drug Abuse Prevention to assess the level of opiate addiction among enlisted servicemen returning from Vietnam (Robins, Helzer, & Davis, 1975). The original cohort consisted of three samples: "drug-positive" (D+) returning servicemen, "drug-negative" (D−) returning servicemen and a civilian comparison group. About half of the veterans (target $n=500$) were drawn randomly from the list of Army enlisted (E1–E9) returnees who provided a positive urine test for opiates, amphetamines, or barbiturates at the time of departure from Vietnam. The D+ veteran population comprised an estimated 10.5% of the total 13,760 Army enlisted returnees in September 1971. Second, a D− Vietnam veteran sample (target $n=500$) was randomly drawn from the same population of returning servicemen. An overlap sample ($n=39$) was placed in the D+ sample in the current study. The comparison non-veteran sample was ascertained from Selective Service registrations and individually matched to the veteran sample members interviewed in 1974 on draft eligibility, draft board location, age, and education completed by the time of the veteran's entry into service (Robins & Helzer, 1975). The total sample size of individuals selected for interview was 1,227, including 943 (76.9%) veterans and 284 (23.1%) non-veterans. Of the selected veterans, 898 (95.2% of those selected) were interviewed in 1972; non-veteran participants were interviewed in 1974 and totaled 284.

Veterans and non-veterans from the 1972 and 1974 surveys were re-contacted in 1996–7 to study licit and illicit drug use in adulthood. Overall, 10.9% ($n=134$) were found to be deceased between 1971 and 1996 (Price et al., 2001). After a two-decade hiatus, the contact rate was 93.7% ($n=1,024$) for surviving participants; measures from in-person and telephone interviews conducted in 1996–7 were available for 82.1% of the selected cases ($n=841$). The final sample consisted of 323 D+ veterans (38.5%), 319 D− veterans (38.0%), and 197 non-veterans (23.5%). All participants contacted in 1996–7 for interview provided written informed consent, approved by the Washington University School of Medicine Institutional Review Board.

2.2. Self-report measures

The VES interview in 1996–7 included numerous psychiatric, social, vocational, substance use, and service utilization measures as well as demographics. The self-report measures to compare with drug testing results were derived from the questions asking

whether the participant used each of several specific illicit drugs in the preceding 90 days, including "stimulant", "marijuana/hashish", "cocaine/crack", "opiates", and "PCP"; and if so positive, which drugs were used during this time period. The 90-day duration corresponds to the average growth time for the length of hair samples collected for the current analysis (approximately 3.9 cm). Demographic variables included race (African American, Caucasian, or other), education (high school completion or not), marital status (married or cohabitating vs. not), and employment (employed or not).

The Diagnostic Interview Schedule for DSM-IV (DIS-IV) (Robins, Cottler, Bucholz, & Compton, 1995) was modified to obtain annual prevalence rates of substance use disorders and psychiatric disorders since 1972. Participants were asked if they had used each class of drug illicitly – marijuana, cocaine, opiates or methamphetamine – five times or more in a given year since their interview in 1972. If the participant reported illicit use of any drug five times or more since 1972, additional questions were asked about the specific years in which he used the class of drugs "most frequently". Frequent drug use was coded if a participant answered affirmative for "frequent use" for any year between 1972 and 1996–7. Methods of administration were asked for those frequently used years. To assess help-seeking behavior, participants were asked about use of inpatient treatment during their last quit attempts. For each class of drug, inpatient treatment is coded positive if the last quit attempt involved hospitalization. History of psychiatric disorders was also assessed based on DSM-IV criteria (APA, 1994) including posttraumatic stress disorder (PTSD), depression, and antisocial personality disorder (ASPD). For each disorder, meeting the diagnostic criteria for any year between 1972 and 1996–7 constituted a positive diagnosis since 1972 in the current study.

2.3. Hair sampling

The research participants were informed of hair sampling prior to their consent for study participation. A separate written consent was obtained for hair sampling. Of the 839 interviews that yielded analyzable data, 600 were conducted in-person. A total of 565 subjects consented to hair sampling, which represented more than a 90% cooperation rate of in-person interviews. Of the 565 hair samples, two were judged unusable after the standard wash process prior to hair drug testing (see Fig. 1). The remaining 239 interviews were conducted by telephone. Although not required, interviewers collected hair samples from 50 of the long-distance telephone interview cases. A total of 613 hair samples were tested for illicit drug use.

For those interviewed in-person, a hair sample was collected immediately after completion of the interview. The interviewer cut approximately 60–80 strands of hair from each participant. The crown of the head was the primary site; although other parts of the head or body were used if the quantity of hair was insufficient or a participant preferred a sample to be taken elsewhere. Precautions to avoid contamination of hair samples were taken including the interviewer wearing a head cap and the use of surgical gloves and sterilized scissors. The hair sample was wrapped in aluminum foil with root ends marked, sealed, and initialed and dated by both the interviewer and participant. In the case of a telephone interview, the interviewer instructed the participant to follow the same procedure and monitored the process over the telephone. All samples were forwarded to a private laboratory specialized in hair drug testing (Psychemedics Corporation, Culver City, CA). At the time of hair collection, information was collected on hair conditions that were known to affect hair testing, including participant use of hair treatments (i.e., permanent, dye, bleach), use of body rather than head hair, and length of hair sample.

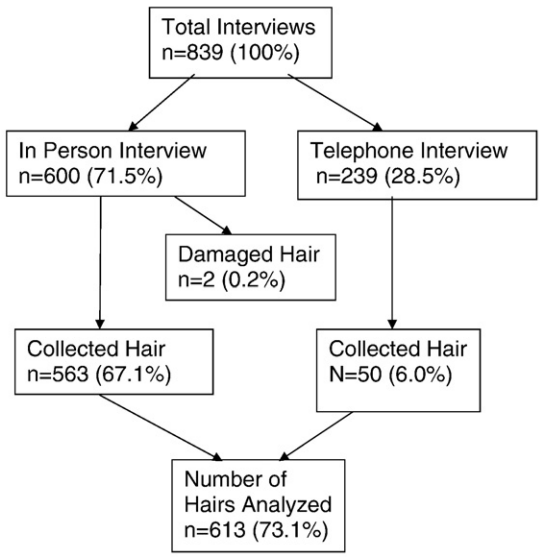

Fig. 1. Flowchart of in-person and telephone interviews and hair samples collected.

2.4. Hair drug testing procedure

The first 3.9 cm of hair was used in hair drug testing corresponding to an average three-month hair growth. For initial screening, radioimmunoassay tests (RIAH®) were applied to unwashed hair for assays of cocaine, opiates, methamphetamine, and phencyclidine (PCP). RIAH was applied separately for detection of marijuana and methadone. A sufficient amount of hair was available for all hair samples to perform screening tests for cocaine, opiates, methamphetamine, and PCP. Confirmation testing was conducted using the sensitive GC–MS assays separately to detect the presence of the following analytes: cocaine, benzoylecgonine (BE), or cocaethylene (CE) for cocaine; codeine, morphine glucuronide, or 6-acetylmorphine (AM) for opiates; amphetamine and methamphetamine for stimulants; and Δ9-carboxy-THC for marijuana. Sub-samples for cases that screened positive for use of all but methadone were used for confirmation testing for each class of drugs separately. In some cases, the samples used for RIAH were resubmitted for confirmation if the quantity of hair was not sufficient (QNS). The cut-off value from RIAH for GC–MS confirmation testing was 5 ng drug/10 mg of hair for cocaine, opiates, methamphetamine, and PCP. Although confirmation testing was carried out for samples with positive screening values less than 5 ng drug/10 mg hair, results were not used for further analysis if this minimum detection quantity was not met, following a standard procedure for hair drug testing analyses. Marijuana RIAH results were reported qualitatively (positive or negative), and all positive samples received quantitative confirmation testing in ng/10 mg hair.

2.5. Hair drug testing measures

For ease of interpretation, only confirmation results are used for the subsequent analyses. For cocaine, opiates, and methamphetamine, confirmation is coded positive if any of the possible confirmation scores were above 5 ng/10 mg hair. For marijuana, any positive confirmation value is coded positive. Any sample with a positive screening but missing conformation score for a drug was considered missing in subsequent analyses. Imputing screening results for cases where confirmation results were missing did not substantially alter results (data available from the last author). Fifty-seven participants had missing values for marijuana screening (9.3%).

Confirmation scores were missing for 18 marijuana (17%), 4 cocaine (5.3%), 0 opiate (0%), and 2 methamphetamine (14.3%) screens. There was no positive screen for PCP, thus it was excluded from the subsequent analyses. For methadone, only RIAH was conducted and it was also excluded from further analyses. Variables constructed from information on the hair conditions with sufficient number of observations were included in analyses to assess their effects on the disagreement between self-report and hair testing.

2.6. Statistical analyses

Using confirmation results, the hair drug testing detection rate for each drug (marijuana, cocaine, methamphetamine and opiates) was compared to the 90-day self-report use. We used McNemar's test for change of state (Conover, 1999) to test if the ratio between the two assessments was significantly different from one. Based on the preliminary analyses, sensitivity and specificity statistics were also computed with both confirmation drug hair test and self-report being the reference. We used the kappa statistic as a measure of agreement above and beyond the level expected by chance. These same statistics were also calculated for the four classes of drugs combined to assess the overall sensitivity and specificity of self-report and hair analysis. We employed the terms *apparent* under- and overreporting because we consider neither self-report nor hair testing to be a "gold standard" for the purposes of this study. Further analyses of opiates and stimulants are not provided due to too few positive self-report and hair test cases.

Initially, bivariate analyses using Chi-square significance tests for prevalence and the t-tests for the means differences were applied to demographics, other self-report measures, and hair condition measures to examine factors related to the disagreement between self-report and hair testing for marijuana and cocaine. Analyses were extended to multivariate logistic regressions to examine relative effects of these disagreement predictors. The magnitudes of their effects were assessed by the adjusted odds ratios (AORs).

3. Results

3.1. Demographic characteristics

Seventy six and a half percent of the participants were veterans and 23.5% were non-veterans. Of the veterans, 50.3% were drug-positive at discharge from the service in 1971 (Table 1). Participants were in their mid- to late-40s at the time of VES follow-up interviews in 1996–7. African Americans constituted 16.3% of the follow-up participants; Caucasians, 78.2%, and other, 5.5%. Most of the study participants (82.3%) reported completing high school or completing some college. Most of the participants were married, and 81.5% were employed. Nearly three quarters (71.5%) of the sample were interviewed in person with the remainder interviewed by telephone. Nearly four fifths (78.9%) of the sample reported that they had used no substances in the 90 days preceding the interview. Self-reported substance use in the preceding 90 days ranged from 1.8% for methamphetamine to 17.0% for marijuana. As expected, interview mode (in-person vs. telephone) was associated with hair sampling since hair sampling was not required with long-distance telephone interviews. Otherwise, the participants with hair samples were not significantly different from those without hair samples.

3.2. Agreement between self-reports and confirmation drug hair testing

Preliminary results showed that drug sampling status (D+ and D−) had no significant effect on drug hair testing agreement. Accordingly, results are presented on the pooled data. The detection ratio (the ratio of hair testing prevalence over self-report 90-

Table 1
Characteristics of the 25-year follow-up sample (N=839)[a]

Characteristics	Total (n=839)		Hair sampled (n=613)		Hair not sampled (n=226)	
	n	%	n	%	n	%
Sample status[b]						
D+ veteran	323	38.5	232	37.9	91	40.3
D− veteran	319	38.0	233	38.0	86	38.1
Non-veteran	197	23.5	148	24.1	49	21.7
Mean age	46.6		46.6		46.6	
Race						
African American	137	16.3	102	16.6	35	15.5
Caucasian	656	78.2	479	78.1	177	78.3
Other	46	5.5	32	5.2	14	6.2
Years of education						
Less than 12	149	17.8	106	17.3	43	19.0
12	312	37.2	228	37.2	84	37.2
Greater than 12	378	45.1	279	45.5	99	43.8
Marital status						
Married	608	72.9	441	72.1	167	75.2
Not married	226	27.1	171	27.9	55	24.8
Employment status						
Employed	682	81.5	497	81.1	185	82.6
Not employed	155	18.5	116	18.9	39	17.4
Mode of interview						
In-person	600	71.5	563	91.8	37	16.4
Telephone	239	28.5	50	8.2	189	83.6
Self-reported drug use past 90 days						
No use	662	78.9	489	79.8	173	76.6
Marijuana	143	17.0	103	16.8	40	17.7
Cocaine	47	5.6	32	5.2	15	6.6
Opiates	22	2.6	16	2.6	6	2.7
Methamphetamine	15	1.8	12	2.0	3	1.3

[a] All measures were based on the 1996/97 interview assessments except the sampling status. Marital status is based on whether or not being married or cohabitated. Employment includes both full-time and part-time. Significance levels for differences between those with a hair sample and those without a hair sample were assessed slightly differently. For the mean age, t-test with unequal variance was used; for methamphetamine, Fisher's exact test was used; for others, Pearson's Chi-square was used. p-values were all insignificant except the mode of interview (p=<.0001).
[b] D+ veterans = drug-positive in 1971; D− veterans = drug-negative in 1971.

day use prevalence) varied from 2.15 for cocaine to .58 for opiates using self-report as the reference (Table 2). The cocaine (p<.0001) and opiate (p<.05) detection ratios were significantly different from one, but the detection ratios for marijuana (p=.68) and methamphetamine (p=.41) were not. Combining the four classes of drugs, the ratio was 1.07, showing that overall hair drug testing was not significantly more likely to detect illicit drug use over the past 90 days. Taking the confirmation hair testing as the reference, specificity was high for self-reports ranging from 94.6 for marijuana, to 100 for cocaine. The overall specificity across the four drugs was 93.4. Sensitivity, however, varied depending on the drug; opiate self-report exhibited the highest sensitivity (77.8), and methamphetamine self-report exhibited the lowest (44.4). In all cases, specificity is much higher than sensitivity, in part

Table 2
Comparisons between self-reports and confirmation drug hair testing (n=613)[a]

Drug class	Sample size	Positive hair test[b]	Positive self-report[c]	Detection ratio[d]	Hair specificity[e]	Hair sensitivity[f]	Self-report specificity[g]	Self-report sensitivity[h]	Kappa
Marijuana	538	14.3	15.4	.93	94.6	75.3	95.8	69.9	.677
Cocaine	609	10.3	4.8	2.15**	100.0	46.0	94.1	100.0	.605
Opiates	613	1.5	2.6	.58*	98.5	77.8	99.7	43.8	.552
Methamphetamine	611	1.5	2.0	.75	98.7	44.4	99.2	33.3	.370
All tested drugs combined	613	21.2	19.9	1.07	93.4	69.2	91.9	73.8	.641

[a] Two observations were dropped from an initial hair sample pool of 615 because two samples were damaged during the washing process. Any observations with missing screening or confirmation hair testing, including quantity-not sufficient (QNS) cases, were also dropped. PCP was not included because there were no positive hair testing cases (n=2 for self-report).
[b] Percent of observations with a positive hair test.
[c] Percent of observations with a positive self-report for drug use in past 90 days prior to interview for the given drug.
[d] Ratio of hair testing prevalence over self-report 90-day use prevalence. *p<.05, **p<.0001.
[e] Percent of total negative self-reports out of negative results from hair testing.
[f] Percent of total positive self-reports out of positive results from hair testing.
[g] Percent of total negative hair reports out of negative results from self-reports.
[h] Percent of total positive hair reports out of positive results from self-reports.

because the proportions of positive test results were small. Using self-report as the reference, the specificities were similar to those using the hair testing as the reference, but the sensitivities were lower for marijuana, opiates, and methamphetamines, while it was perfect (100) for cocaine, and slightly higher (73.8) with four drugs combined.

The kappa values for marijuana (κ=.677) and cocaine (κ=.605) suggest self-report and confirmation hair testing were in substantial agreement (usually .6 to .8 range is considered substantial; above .8 is usually considered near perfect). The kappa for opiates (κ=.552) is considered moderate agreement and the kappa for methamphetamines (κ=.370) is considered fair agreement (Landis & Koch, 1977).

3.3. Predictors of disagreement between self-report and hair testing

Marijuana and cocaine were the two most commonly-used drugs in this sample. The cocaine disagreement was in the direction of apparent underreporting only (i.e., failing to endorse drug use during the last 90 days, with a positive hair drug test for cocaine). Marijuana disagreements were observed for both apparent under- and overreporting relative to hair results.

Bivariate analysis in Table 3 shows that participants with apparent underreporting of past 90-day cocaine use were more likely to be African American (χ^2 (1, n=609)=79.20, p<.001) and were less likely to be Caucasian (χ^2 (1, n=609)=72.31, p<.001), married (χ^2 (1, n=608)=4.90, p<.05), and employed (χ^2 (1, n=609)=11.94, p<.001) (Table 3, left column). They were more likely to report frequent cocaine use since 1972 (χ^2 (1, n=609)=7.69, p<.01), to smoke the substance (χ^2 (1, n=608)=9.94, p<.01), and to have had their last quit attempts involving inpatient care (χ^2 (1, n=609)=12.72, p<.001). These participants were also more likely to have chemically treated hair (χ^2 (1, n=227)=4.15, p<.05) and were more likely to provide a short (i.e., <3 cm) hair sample (χ^2(1, n=606)= 23.21, p<.001). Apparent underreporting of marijuana (i.e., self-report was negative and hair test was positive; Table 3, middle column) was significantly related only to frequent marijuana use since 1972 (χ^2 (1, n=513)=4.83, p<.05). Apparent overestimate of marijuana by self-report (i.e., self-report was positive and hair test was negative; Table 3, right column) was associated with being African American (χ^2 (1, n=519)=17.15, p<.001), with not being Caucasian (χ^2 (1, n=519)=14.85, p<.001), and being unmarried (χ^2 (1, n=518)=11.38, p<.001). It is also associated with frequent marijuana use since 1972 (χ^2 (1, n=518)=27.04, p<.001), having diagnoses of antisocial personality disorder (χ^2 (1, n=519)=7.10, p<.01) or PTSD (χ^2 (1, n=519)=8.37, p<.01) since 1972, and providing a short hair sample (<3 cm) (χ^2 (1, n=519)=17.34, p<.001). All other comparisons were non-significant.

Multivariate analyses for disagreement between self-reported cocaine use and positive cocaine hair result (self-report negative and hair drug test positive; Table 4 left column) yielded an excellent overall fit (Wald Chi-square=62.7 with p<.0001; c=.928).

Table 3
Variables associated with disagreement between self-report and hair testing: cocaine and marijuana[a]

	Cocaine		Marijuana		Marijuana	
	Self = no		Self = no		Self = yes	
	Hair = yes		Hair = yes		Hair = no	
	Agree	Not agree	Agree	Not agree	Agree	Not agree
	n=575	n=34	n=494	n=19	n=494	n=25
Demographics						
Mean age	46.4	44.9	46.4	44.3	46.4	45.2
African American	12.9	70.6***	11.5	21.1	11.5	40.0***
Caucasian	82.1	20.6***	82.8	79.0	82.8	52.0***
Other Race	5.0	8.9	5.7	.0	5.7	8.0
High school education	83.3	73.5	83.8	68.4	83.8	76.0
Married	73.3	55.9*	74.7	57.9	74.7	44.0***
Employed	82.6	58.8***	82.8	79.0	82.8	68.0
Frequent drug use since 1972[b]	23.1	44.1**	42.9	68.4*	42.9	96.0***
Smoking as method of administration	4.9	17.7**				
IV as method of administration	5.6	5.9				
Last quit attempt involving inpatient treatment[c]	3.0	14.7***	2.6	0	2.6	8.0
Psychiatric conditions since 1972[d]						
ASP	31.8	26.5	30.6	26.3	30.6	56.0*
PTSD	21.6	29.4	19.8	31.6	19.8	44.0*
Depression	9.0	2.9	9.1	5.3	9.1	8.0
Hair Condition						
Treatment[e]	3.6	11.8*	4.9	.0	4.9	6.7
Body Hair	7.5	11.8	5.9	5.3	5.9	12.0
Short (less than 3 cm)	12.0	41.2***	9.5	10.5	9.5	36.0***

[a] Any observation with missing screening or confirmation for hair testing were dropped. The disagreement for self-report vs. hair testing was assessed in two ways: apparent underreporting (self-report = no and hair testing = yes); or apparent overreporting (self-report = yes and hair testing = no). There was no case of apparent cocaine overreporting. Thus the total analysis sample size for cocaine=609, and for marijuana=538 (494+19+25). The significance of association was assessed by t-test for age and by Chi-square statistics for other measures. *p<.05, **p<.01, ***p<.001.

[b] The measure is for the specific drug listed.

[c] Quit attempts are for the specific drug involved. Participants reported hospitalization for his last quit attempt for the specific drug.

[d] Based on DSM-IV.

[e] Permanent, dye, bleach, relaxer or rogaine.

Table 4
Adjusted odds ratios (95% confidence intervals) showing major predictors of the disagreement between self-report and hair testing: cocaine and marijuana [a]

	Cocaine	Marijuana	Marijuana
	Self = no	Self = no	Self = yes
	Hair = yes	Hair = yes	Hair = no
	n=609	*n*=513	*n*=519
Demographics			
Age	.76 (.61–.95)*	.89 (.70–1.14)	.99 (.80–1.23)
African American	27.74 (8.79–87.50)***	2.29 (.51–10.19)	2.13 (.60–7.52)
Other	5.16 (1.03–25.80)*	QS	1.47 (.26–8.47)
High school education	.79 (.28–2.26)	.48 (.16–1.41)	1.06 (.34–3.24)
Married	.80 (.32–1.98)	.58 (.21–1.58)	.42 (.17–1.06)
Employed	.39 (.15–1.00)	1.26 (.36–4.46)	.85 (.28–2.56)
Drug use — since 1972 [b]	2.14 (.70–6.51)	2.83 (.98–8.18)	27.5 (3.44–220.46)**
Smoking as method of administration	.47 (.11–2.08)	NA	NA
IV as method of administration	.25 (.04–1.70)	NA	NA
Last quit attempt involving inpatient treatment [c]	5.84 (1.03–32.93)*	QS	1.09 (.19–6.45)
Psychiatric conditions since 1972 [d]			
ASP	.27 (.09–.78)*	.40 (.13–1.26)	1.22 (.47–3.15)
PTSD	.79 (.28–2.19)	1.84 (.59–5.73)	1.46 (.53–4.01)
Depression	.33 (.03–3.20)	.38 (.05–4.57)	.44 (.09–2.14)
Hair Condition			
Treatment [e]	5.89 (.98–35.54)	QS	.97 (.06–17.21)
Body Hair	.38 (.10–1.42)	45 (.04–4.57)	.66 (.14–3.24)
Short (less than 3 cm)	1.18(.42–3.29)	.81 (.14–4.85)	3.86 (1.11–13.44)*

[a] The disagreement for self-report vs. hair testing was assessed in two ways: apparent underreporting (self-report = no and hair testing = yes); or apparent overreporting (self-report = yes and hair testing = no). There was no case of apparent cocaine overreporting. Significance tests were performed on the beta parameters of predictors of the logistic regressions. *$p<.05$, **$p<.01$, ***$p<.001$.
[b] The measure is for specific drug listed. See the text for definitions.
[c] Participant reported hospitalization for last quit attempt for the specific drug involved. QS = quasi-complete separation (all cases of treated inpatient users were in the agreement cell).
[d] Based on DSM-IV.
[e] Permanent, dye, bleach, relaxer or rogaine.

Adjusted odds ratios (AORs) show that this disagreement was significantly associated with younger age, being African American, racial identification as "other", inpatient care for the participant's most recent quit attempt, and not having ASPD (AOR=.76, 27.74, 5.16, 5.84, and .27, respectively). Although measures associated with disagreement at the bivariate level were correlated, diagnostic analyses did not suggest a high level of multicollinearity among significant predictors.

For the apparent underreporting of marijuana use (i.e., self-report negative and hair drug test positive; Table 4 middle column), no individual variable was associated with the disagreement between marijuana self-report and hair test (Wald Chi-square=13.1 with $p=.28$; $c=.739$). The analysis of apparent overestimate of marijuana use (Table 4, right column) yielded an excellent overall fit (Wald Chi-square=29.9 with $p=.0079$; $c=.865$). Participants with positive self-report but negative hair testing were also more likely to report frequent marijuana use at some time since 1972 and more likely to provide a short (i.e., <3 cm) hair sample (AOR=27.5 and 3.86, respectively).

4. Discussion

4.1. Self-report vs. hair drug testing cross-validation findings

Our findings extend previous research on the agreement between hair drug testing and self-report of drug use. Comparisons of self-report and hair results suggest hair drug testing may not be taken as a gold standard for some classes of drugs. Specifically, hair testing appears to offer a more accurate assessment of cocaine use over a 90-day period, but in our study the detection rate was not better than self-report on marijuana, opiates, and methamphetamine. However, the low kappa for methamphetamine may be related to the fact that the hair tests were done on methamphetamine but the question was phrased as "stimulant".

4.2. Factors associated with cocaine discrepancies

The present study also identified predictors of discrepancies between self-report and hair testing. Race was the most salient predictor of cocaine disagreement. Even when other factors were controlled for, the self-report and hair test results for African Americans were more discrepant than for non-African Americans, a finding consistent with past studies (Fendrich et al., 1999; Feucht, Stephens, & Walker, 1994). In a large study of youth (9–20), underreporting of cocaine was documented with urine testing validation as well (Fendrich & Yanchun, 1994) where African Americans in comparison to Caucasians who were urine positive were about 6 times less likely to report cocaine use when other factors are controlled for. Our study extends this finding to a middle-aged

sample of male African American Vietnam veterans and non-veteran community controls. However, African American underreporting may well be a function of the particular sample and not ethnicity. One study of arrestees (Lu et al., 2001) reported that African Americans are more likely to admit using crack cocaine than arrestees of other races. A prior study of veterans recruited through the VA system found no race differences in reporting discrepancies (Calhoun et al., 2000). Further studies are needed to elucidate sources of discrepancy to improve reporting pattern of cocaine.

One possible reason for this ethnic difference is that cocaine metabolites have been detected at a higher rate in African American hair than in other racial groups, which appears independent to specific laboratory procedures (Cone & Joseph, 1996; Welp et al., 2003). The so-called "racial bias" hypothesis (Kidwell, Lee, & DeLauder, 2000) is consistent with the observation that cocaine (but not benzoylecgonine) demonstrates differential affinity to different hair types based on melanin subtype, with dark-color hair demonstrating the greatest affinity (Borges, Roberts, Wilkins, & Rollins, 2003). However, several studies reported conflicting or mixed results on the "racial bias" hypothesis (e.g., Hoffman, 1999; Kelly, Mieczkowski, Sweeney, & Bourland, 2000; Mieczkowski & Newel, 2000).

We attempted to control for several hair conditions. Hair treatment, the use of body hair rather than head hair, and short hair sample were not significantly related to apparent underreporting of cocaine when other measures are simultaneously controlled for. The adjusted odds ratio of 5.89 for hair treatment did not reach a statistical significance ($p=.053$). One critical review concluded that cosmetic hair treatments tend to result in more false negatives (Kintz, Cirimele, & Ludes, 2000). Hair color or melanin information was not collected in our study, however. A consensus is needed on toxicological aspects of hair color and treatment in hair drug testing, given that hair drug testing has become widely-used for employment screening.

Younger age, use of inpatient substance abuse treatment and lack of ASPD were also significant multivariate correlates of inconsistency, while other variables (e.g., depression and PTSD) were not. One recent study of veterans recruited in the VA system found that underreporters of drug use were lower SES and not more likely to have a diagnosis of PTSD. However, they did not differ on the basis of age. An unexpected finding was that participants with antisocial personality disorder were more likely to be consistent in their self-reporting than those without. To our knowledge there is no direct research that has drawn the same conclusion.

4.3. Factors associated with marijuana discrepancies

No variables predicted apparent marijuana underreporting in the multivariate model, and only two variables, marijuana use since 1972 and providing a short hair sample, predicted apparent overreporting. Those who became infrequent marijuana users over time and now smoke only occasionally (i.e., less than monthly) may have estimated that they used marijuana within the past 90 days, when, in fact, it has been longer. Depending on hair growth rate, a shorter hair sample (<3 cm) may not be long enough to detect the 90-day use, especially for infrequent marijuana users. Accurate hair testing of marijuana also continues to be challenging. Recent research has highlighted the difficulties evaluating quantitative levels of marijuana in hair (Jurado & Sachs, 2003). Some techniques for hair testing for marijuana may result in false negative results (Mieczkowski & Newel, 1997; Uhl & Sachs, 2004). Epidemiologic research of marijuana use validation using hair drug testing may consider obtaining longer and more hairs than would be adequate for other drugs.

4.4. Limitations, summary and conclusions

The present study has limitations. First, because only middle-aged men were included, these findings may not generalize to women or different age groups. Second, validity of self-report measures could be questioned. However, the 90-day questions were asked for all classes of drugs tested against hair samples. Therefore, it is unlikely that differential agreement findings across the four classes of drugs reported here reflect self-report bias. Third, the hair testing was conducted in 1996 and 1997, and some changes have occurred in testing techniques and technology since that time (e.g., marijuana; Uhl & Sachs, 2004). Fourth, because confirmation tests were carried out only for positive screening tests, any false negative from screening would have remained a false negative. Thus any specificity based on screening would be as good or better than the specificity based on confirmation and any sensitivity based on confirmation would be only as good as the screening sensitivity. Our preliminary analyses using screening results, however, indicated that the loss of sensitivity was not offset by the gain in specificity. For all classes of drugs except opiates (where all hair reports were identical whether based on screening or confirmation), kappa values based on screening were lower than kappa values based on confirmation. Finally, hairs with insufficient quantity may have occurred disproportionately in polydrug user samples because more hair was required to complete confirmation testing for other drugs. However, we found that those with insufficient quantity of hair for marijuana testing did not differ significantly from those who had positive screening tests for cocaine ($p=.64$) or opiates ($p=.44$), suggesting that insufficient quantities of hair for marijuana may be more likely due to random factors or hair availability.

Despite these limitations, the current study provides useful information regarding self-reported illicit drug use and hair drug testing. The two estimates did not meet standards for excellent agreement, a finding consistent with previous comparisons of these two assessment methods. Hair results suggested more widespread use than did self-report for cocaine only. Several factors were significantly associated with apparent underreporting of cocaine, and fewer were associated with marijuana underreporting. Some factors (e.g., drug use history) were primarily participant-related, while others (e.g., short hair sample) were related to the testing procedures. Race was particularly salient with respect to cocaine underreporting, a finding that underscores the racial bias debate in hair testing that has continued for some time. Our findings further highlight issues related to self-report/hair testing

discrepancies, and address shortcomings of both methods of analysis. Further epidemiological studies may use both methods in combination to determine rates of substance use and interpret discrepancies appropriately.

Acknowledgments

This work was supported in part by the Independent Scientist Award (K02DA00221) and research grants (R01DA09281, MH060961) to the last author. Authors acknowledge contributions by Chris Berka (Psychemedics), Washington University School of Medicine research staff and staff of the St. Louis VA Medical Center, St. Louis Vet Center. Contributions by the study participants were by far the most important.

References

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders*, (4th ed.) Washington, DC: Author.
Baumgartner, W. A., Hill, V. A., & Blahd, W. H. (1989). Hair analysis for drugs of abuse. *Journal of Forensic Science*, *34*, 1433–1453.
Borges, C. R., Roberts, J. C., Wilkins, D. G., & Rollins, D. E. (2003). Cocaine, benzoylecgonine, amphetamine, and *N*-acetylamphetamine binding to melanin subtypes. *Journal of Analytic Toxicology*, *27*, 125–134.
Calhoun, P. S., Sampson, W. S., Bosworth, H. B., Feldman, M. E., Kirby, A. C., Hertzberg, M. A., et al. (2000). Drug use and validity of substance use self-reports in veterans seeking help for posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, *68*, 923–927.
Cone, E. J., & Joseph, R. E. (1996). The potential for bias in hair testing for drugs of abuse. In P. Kintz (Ed.), *Drug testing in hair* (pp. 69–93). Boca Raton, FL: CRC Press.
Conover, W. J. (1999). *Practical nonparametric statistics*, (3rd ed.) New York: Wiley & Sons.
Farabee, D., & Fredlund, E. (1996). Self-reported drug use among recently admitted jail inmates: Estimating prevalence and treatment needs. *Substance Use and Misuse*, *31*, 423–435.
Fendrich, M., Johnson, T. P., Sudman, S., Wislar, J. S., & Spiehler, V. (1999). Validity of drug use reporting in a high-risk community sample: A comparison of cocaine and heroin survey reports with hair samples. *American Journal of Epidemiology*, *149*, 955–962.
Fendrich, M., Johnson, T. P., Wislar, J. S., Hubbell, A., & Spiehler, V. (2004). The utility of drug testing in epidemiological research: Results from a general population survey. *Addiction*, *99*, 197–208.
Fendrich, M., & Yanchun, X. (1994). The validity of drug use reports from juvenile arrestees. *International Journal of the Addictions*, *29*, 971–985.
Feucht, T. E., Stephens, R. C., & Walker, M. L. (1994). Drug use among juvenile arrestees: A comparison of self-report, urinalysis and hair assay. *Journal of Drug Issues*, *24*, 99–116.
Hawks, R. L., & Chiang, N. (1986). Examples of specific drug assays. In R. L. Hawks, & N. Chiang (Eds.), *Urine testing for drugs of abuseNIDA Research Monograph*, *Vol. 73*. (pp. 84–112) Rockville, MD: National Institute of Drug Abuse.
Hoffman, B. H. (1999). Analysis of race effects on drug-test results. *Journal of Occupational and Environmental Medicine*, *41*, 612–616.
Jurado, C., & Sachs, H. (2003). Proficiency test for the analysis of hair for drugs of abuse, organized by the Society of Hair Testing. *Forensic Science International*, *133*, 175–178.
Kelly, R. C., Mieczkowski, T., Sweeney, S. A., & Bourland, J. A. (2000). Hair analysis for drugs of abuse. Hair color and race differentials or systematic differences in drug preferences? *Forensic Science International*, *107*, 63–86.
Kidwell, D. A., Lee, E. H., & DeLauder, S. F. (2000). Evidence for bias in hair testing and procedures to correct bias. *Forensic Science International*, *107*, 39–61.
Kim, M. T., & Hill, M. N. (2003). Validity of self-report of illicit drug use in young hypertensive urban African American males. *Addictive Behaviors*, *28*, 795–802.
Kintz, P., Cirimele, V., & Ludes, B. (2000). Pharmacological criteria that can affect the detection of doping agents in hair. *Forensic Science International*, *107*, 325–334.
Kline, J., Ng, S. K. C., Schittini, M., Levin, B., & Susser, M. (1997). Cocaine use during pregnancy: Sensitive detection by hair assay. *American Journal of Public Health*, *87*, 352–358.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.
Lu, N. T., Taylor, B. G., & Riley, K. J. (2001). The validity of adult arrestee self-reports of crack cocaine use. *American Journal of Drug and Alcohol Abuse*, *27*, 399–419.
Magura, S., & Kang, S. Y. (1996). Validity of self-reported drug use in high risk populations: A meta-analytical review. *Substance Use and Misuse*, *31*, 1131–1153.
Mieczkowski, T., & Newel, R. (1997). Patterns of concordance between hair assays and urinalysis for cocaine: Longitudinal analysis of probationers in Pinellas County, Florida. *NIDA Monograph*, *167*, 161–199.
Mieczkowski, T., & Newel, R. (2000). Statistical examination of hair color as a potential biasing factor in hair analysis. *Forensic Science International*, *107*, 13–38.
Montagna, M., Polettini, A., Stramesi, C., Groppi, A., & Vignali, C. (2002). Hair analysis for opiates, cocaine and metabolites evaluation of a method by interlaboratory comparison. *Forensic Science International*, *128*, 79–83.
Price, R. K., Risk, N. K., Murray, K. S., Virgo, K. S., & Spitznagel, E. L. (2001). Twenty-five year mortality of US servicemen deployed to Vietnam: Predictive utility of early drug use. *Drug and Alcohol Dependence*, *64*, 309–318.
Richardson, J., Fendrich, M., & Johnson, T. P. (2003). Neighborhood effects on drug reporting. *Addiction*, *98*, 1705–1711.
Robins, L. N., & Helzer, J. E. (1975). Drug use among Vietnam veterans — three years later. *Medical World News-Psych*, *16*, 44–49.
Robins, L. N., Helzer, J. E., & Davis, D. H. (1975). Narcotic use in Southeast Asia and afterward: An interview study of 898 Vietnam returnees. *Archives of General Psychiatry*, *32*, 955–961.
Robins, L. N., Cottler, L., Bucholz, K. K., & Compton, W. (1995). *Diagnostic Interview Schedule for DSM-IV (DIS-IV).* St. Louis, MO: Washington University School of Medicine.
Substance Abuse and Mental Health Services Administration (2003). *Overview of findings from the 2002 National Survey on Drug Use and Health.* Rockville, MD: Department of Health and Human Services.
Swartz, M. S., Swanson, J. W., & Hannon, M. J. (2003). Detection of illicit substance use among persons with schizophrenia by radioimmunoassay of hair. *Psychiatric Services*, *54*, 891–895.
Tassiopoulos, K., Bernstein, J., Heeren, T., Levenson, S., Hingson, R., & Bernstein, E. (2004). Hair testing and self-report of cocaine use by heroin users. *Addiction*, *99*, 590–597.
Uhl, M., & Sachs, H. (2004). Cannabinoids in hair: Strategy to prove marijuana/hashish consumption. *Forensic Science International*, *145*, 143–147.
Welp, E. A. E., Bosman, I., Landendam, M. W., Totte, M., Maes, R. A. A., & van Ameijden, E. J. C. (2003). Amount of self-reported illicit drug use compared to quantitative hair test results in community-recruited young drug users in Amsterdam. *Addiction*, *98*, 987–994.