



Published in final edited form as:

J Addict Med. 2009 December ; 3(4): 211–217. doi:10.1097/ADM.0b013e3181a0f5dc.

Temporal and Gender Trends in Concordance of Urine Drug Screens and Self-Reported Use in Cocaine Treatment Studies

Megan S. Schuler, B.S.¹, William V. Lechner, B.S.², Rickey E. Carter, Ph.D.¹, and Robert Malcolm, M.D.²

¹Department of Biostatistics, Bioinformatics and Epidemiology Medical University of South Carolina
135 Cannon Street Suite 303 Charleston, SC 29425

²Center for Drug and Alcohol Programs Medical University of South Carolina 67 President Street,
P.O. Box 250861 Charleston, SC 29425

Abstract

Objectives—To describe temporal trends in concordance, sensitivity, and specificity and to explore demographic trends in concordance in two outpatient treatment studies for cocaine dependence.

Methods—We obtained 2229 urine drug screens from 129 individuals, along with accompanying self-use reports. Paired self-use reports and urine drug screens were considered concordant if the two measures of cocaine use were in agreement. The sensitivity and specificity of the self-use reports in predicting the urine drug screen was also estimated. To model concordance, sensitivity, and specificity as a function of time, generalized estimating equations were used. Demographic effects on concordance among subjects who achieved 100% concordance and subjects who achieved a recently proposed 70% concordance threshold were tested.

Results—Over the course of our studies, both sensitivity and concordance statistically decreased, yet specificity remained relatively constant. Median concordance for all subjects was 88%. Among all subjects, concordance varied significantly by gender, with females achieving significantly higher concordance than males (96% vs. 86%). Similarly, females were almost twice as likely to achieve 100% concordance as males (42% vs. 22%). Finally, 80% of participants achieved the 70% concordance threshold, and no differences among demographic groups with regards to the 70% concordance threshold were observed.

Conclusions—Temporal effects of concordance and sensitivity may have profound repercussions when using self-use reports to gauge efficacy of an experimental intervention. Furthermore, gender may differentially affect concordance. Finally, a substance abuse outcome measure that reliably combines objective and self-report data is promising, but further research is needed.

Keywords

concordance; sensitivity; cocaine; temporal effects; gender

Background

Drug treatment studies have long been plagued by the challenge of accurately monitoring subjects' drug use. Typically, both self-use reports (SUR) and urine drug screens (UDS) are collected in cocaine treatment studies, and the concordance between these two measures has been extensively researched [1,2]. In a clinical drug treatment context, concordance is defined

as the probability an individual's SUR results matching his UDS results (for a given time), meaning both are positive or negative. Following convention, UDS results are taken to be the “gold standard,” and SUR results are compared to these “accepted” UDS results. There are two components of concordance - specificity and sensitivity. Sensitivity is the conditional probability that a subject self-reports cocaine use, given that his urine screen is positive; specificity is the conditional probability that a subject self-reports no cocaine use, given that his urine screen is negative.

It is clear that self-use reports are subjective, and thus cannot be assumed to be an entirely accurate account of drug use. The validity of self-reported data varies from individual to individual and is entirely based on an individual's willingness and ability to accurately recount use patterns, which may be confounded by denial, paranoia, desire to please, cognitive impairment, or unknown motivations. On the other hand, although UDS are generally regarded as the standard measure and as highly objective, UDS does have some pitfalls. Urine drug screening assumes an average metabolic rate, and thus individual variations may affect the validity of UDS. For instance, slow metabolizers may test positive even though they used prior to the standard 72 hour window. Conversely, rapid metabolizers may test negative, even though they used within the standard 72 hour window. Furthermore, the practical utility of UDS is diminished for drugs with a rapid half-life, such as cocaine, since subjects must report to the clinic every 2 to 3 days to accurately monitor use. Clearly, this required frequency results in a substantial amount of missing data, which may be confounded by subjects systematically missing urine screens when they are using. Thus, both SUR and UDS have notable drawbacks.

Clearly, in the ideal cocaine treatment study, UDS would be collected thrice weekly from all subjects (with no missing data), providing a complete picture of use as measured by the gold standard. However, in the typical study, there is extensive missing data among the UDS, and thus the quality of the SUR data becomes crucial. Concordance can be used to quantify the reliability of self-reported data in relation to the accepted standard, urine drug screens. Basically, an unconfirmed SUR from a subject who has “high” individual concordance should be more likely to be judged “reliable” than an unconfirmed SUR from a subject with “low” individual concordance. Clearly, in order to utilize SUR in any fashion more sophisticated than merely taking all SUR at face value, terms such as “high” and “reliable” must be explicitly defined.

Some addiction researchers have proposed the use of a composite outcome, which would benefit from both the completeness of SUR and the objectivity of UDS [3]. However, such an imputation scheme would only be desirable if the SUR data used for imputation was not biased. As proposed by Somoza et al., individual concordance scores are used as an indication of the validity (thus bias) of SUR data. If the subject has a high concordance, defined by Somoza et al. as at least 0.70, then SUR data is accepted in place of missing UDS data. However, if the subject has a concordance less than 0.70, then SUR data is not accepted and the missingness persists.

Again, use of SUR as a primary endpoint in addiction studies would only be desirable if the SUR were not biased. Concordance is one attempt at quantifying the bias in SUR. Many investigators have previously examined concordance in cocaine abusing and dependent populations, and have reported sensitivity (of SUR as confirmed by UDS) ranging from 28% to 75% [4]. Furthermore, there is a well-documented temporal bias in concordance, since subjects systematically under-report use more often at the end of treatment than at admission [5,6,7,8,9]. For example, Hindin et al. report a baseline sensitivity of 89% and a post-treatment sensitivity of only 51% among a treatment-seeking, cocaine dependent population [8].

Demographic trends in concordance have also been explored, yet the findings in the literature have been mixed. Myrick et al. report that age, gender, and race were not associated with concordance among a cocaine-dependent population [10]. However, they found that later onset of cocaine use, heavier use, and presence of an affective illness were associated with concordance, and that completion status was highly associated with concordance. Nyamathi et al. found that the Hispanic ethnicity was associated with lower concordance [11]. Additionally, Tassiopoulos et al. reported that the racial/ethnic pairing of a subject and interviewer can influence concordance, especially for black subjects [5]. Ideally, there would be no temporal or demographic trends in concordance, since such trends raise the possibility of systematic bias among the SUR.

Thus, the goal of this study was to describe both temporal and demographic trends in concordance, sensitivity, and specificity for two outpatient treatment studies for cocaine dependence. Furthermore, a concordance of 0.70 has recently been proposed as threshold for subject reliability by Somoza et al. [3], and we assessed the feasibility and validity of this proposal in relation to our study population. Finally, we characterized subjects who demonstrated a concordance of 1.0, a fundamentally intriguing group of subjects.

Methods

Setting

Observations were drawn from two outpatient clinical trials of pharmacological treatments for cocaine dependence in the Center for Drug and Alcohol Programs at the Medical University of South Carolina - one evaluating the efficacy of Modafinil and the second evaluating the efficacy of N-acetylcysteine (R01 DA019903, R01 DA016368).

Participants

Participants for these two clinical trials were recruited within a 50-mile radius of the Medical University of South Carolina through clinical referrals, flyers, word of mouth, and television, radio, and newspaper advertisements. Qualifying participants were males or females of any race between the ages of 18–65, who met DSM-IV criteria for cocaine dependence, as assessed by the Structured Clinical Interview for DSM-IV [12]. Participants were excluded if they met current dependence criteria for any psychoactive substance other than cocaine, alcohol, nicotine, or marijuana, or demonstrated physiological dependence on alcohol requiring medical detoxification. Participants were also excluded if they met DSM-IV criteria for any current psychiatric disorders. Prior to randomization, each participant received a thorough medical history and physical examination, electrocardiogram, and blood work including a hematology and chemistry panel to ensure physical stability. All participants were actively seeking treatment for cocaine dependence at the time of admission to the programs.

Measures

Self-use reports were collected through Timeline Follow Back interviews administered by research staff at admission, treatment, and follow-up visits [13]. A UDS and a 30-day retrospective SUR was collected at admission. During treatment, a UDS and a calendar-based SUR were collected at each clinical visit (scheduled thrice weekly). Additionally, a UDS and a 30-day retrospective SUR were collected at both the four-week and eight-week follow-up visits; however, these two follow-up visits were not included in our analysis. Self-reported use was queried using statements “Did you use any cocaine on Monday the 19th?”, and each date since last visit was queried.

Urine samples were assessed for BE with an immunoassay and BE levels of 300 ng/ml or higher constituted a cocaine positive specimen. To determine concordance of SUR with a given

UDS results, a subject's SUR for the 72 hours prior to the UDS were analyzed. Note that 72 hours is the standard, conservative estimate of BE persistence. [10] If the given UDS was negative, all SUR within 72 hours must be negative to be concordant. If the UDS was positive, then at least one SUR within 72 hours must be positive to be concordant.

Statistical Methods

A generalized linear model (GLM) framework was used to estimate the concordance of the self-report and urine screen data [14]. The identity link and binomial distribution were specified for the GLMs, and fastidious specification of the dependent variable and predictors was used to estimate the sensitivity and specificity of the self-report data. Both SUR and UDS were coded as binary variables, where 1 defined use and 0 defined no use. When the GLM was constructed to model the probability of SUR = 1, the intercept alone represents the false positive rate (i.e., probability of SUR = 1 given UDS = 0), and the sum of the intercept and the urine screen parameter estimates represents the sensitivity (i.e., probability of SUR = 1 given UDS = 1). Conversely, when the model is constructed to model the probability of SUR = 0, the intercept alone represents specificity (i.e., probability of SUR = 0 given UDS = 0), and the sum of the intercept and the urine screen parameter estimates represents the false negative rate (i.e., probability of SUR = 0 given UDS = 1). The analysis was performed for three time periods: for baseline visit, for final visit, and for overall study duration. When analyzing the overall study data, generalized estimating equations methods (as opposed to frequency table analysis) were necessary to account for repeated measures through the course of the study, as well as to generate statistically-appropriate confidence intervals for these estimates. The correlation of repeated measures within subject was accounted for in the generalized linear model by invoking the robust variance used in generalized estimating equations (GEE) [15].

Concordance was calculated as the percentage of observations for which self-use report and UDS results agreed (either in the positive or the negative). Standard confidence intervals were calculated for baseline and last visit, since repeated measures by subject were not a factor. However, a GEE regression model with subject ID as the repeated measure variable was used to determine the confidence interval for the overall study concordance. Additionally, to assess statistical significance of trends in sensitivity, specificity, and concordance over time, we used separate GEE regression models to regress sensitivity, specificity, and concordance on number of days in study from randomization. For all GEE models, we specified a compound symmetry covariance matrix structure, a conservative approach since the data include a large number of non-uniformly spaced visits per subject.

Furthermore, we sought to describe the demographics of two particularly interesting subgroups, subjects who were 100% concordant (i.e., self-reported use aligned perfectly with UDS results) and subjects who were at least 70% concordant. Seventy percent concordance has recently been proposed as a threshold for sufficient reliability in the literature, yet has not been widely validated [3]. The primary demographic characteristics of interest were race, gender, completion status, and level of baseline cocaine use. For both the 100% and at least 70% concordant classifications, we calculated the percentage of subjects who fell in each demographic category. To test for significance between the concordance classifications with respect to demographic variables, the Wilcoxon-Rank Sum test, a nonparametric alternative to the independent group student's t-test, and chi-square tests were used for continuous and categorical variables, respectively. Additionally, since our data arise from two separate studies, we also conducted a stratified analysis under a Mantel-Haenszel framework, in order to adjust for the potential confounding effects attributable to the two clinical trials. All analyses were conducted with SAS version 9.1. The type I error rate was specified at 0.05 and the reported p-values have not been adjusted for multiple comparisons.

Results

Participant characteristics

Table 1 presents the demographic data for the two clinical trials used in this analysis. The median age for study participants was 42, although the subjects in the Modafinil study were statistically significantly younger than the subjects in the N-acetylcysteine study. Subjects were evenly split between white and black individuals; the two Hispanic subjects were included in the “white” category for purposes of analysis. Eighty percent of subjects were male, and the median number of days of cocaine use in the month preceding admission was 12. At the time of analysis, 66 subjects were enrolled in the N-acetylcysteine study and 63 in the Modafinil study, studies with nearly identical protocol. Fifty-six percent of all subjects completed the eight-week treatment phase, meaning they attended at least one study visit during the eighth week. Overall, the study populations from these two studies are highly comparable.

Cross-sectional measures of concordance

We examined sensitivity and specificity of SUR for all 129 subjects at baseline and at final study visit. Baseline represents first study visit, during the screening phase, and final study visit represents the final treatment visit a subject completed, excluding post-treatment follow-up visits.

In our study population, sensitivity decreased from 83.5% at baseline to 63.2% at final study visit, suggesting the need to more formally examine temporal changes using a repeated measures analysis. Equivalently, the false negative rate increased from 16.5% at baseline to 36.8% at final study visit. Specificity was 90.6% at baseline and 95.2% at final study visit.

In order to determine whether the observed changes in concordance rates from baseline to final visit were statistically significant, we constructed a series of generalized estimating equation (GEE) regression models, which regressed concordance, sensitivity, and specificity, respectively, on the number of days the subject had been enrolled in the study. The day a subject was randomized was taken to be day 0, and thus days in screening (which occurs prior to randomization) were assigned a negative value, while active treatment days were assigned a positive value. Overall, specificity (false positive rate) remained constant throughout the study, yet concordance and sensitivity significantly declined (Figure 1).

Overall Concordance Measures

We introduce a description of concordance throughout the study, to complement the baseline and final visit “snapshots.” Overall concordance measures that utilize all self-reports and urine drug data can be calculated by using generalized estimating equation (GEE) methods to account for the variability unique to repeated measures data [16].

We observe the highest concordance at the beginning of the study (85.3%), lower concordance for the entire study duration, and the lowest concordance (73.6%) at the final visit. Note that all subject visits were included when calculating overall concordance measures, which naturally weighted these estimates to favor subjects who remained in the study the longest. Of our 129 subjects, 56% of subjects completed the study, meaning that they attended at least one visit during treatment week 8.

Concordance Trends among Demographic Groups

Currently, there is no general consensus in the literature defining a reliable level of concordance. Recently, Somoza et al. proposed that subjects who demonstrate an overall concordance of at least 0.70 should be considered to be “reliable.” [3] To assess feasibility and

validity of the proposed 0.70 concordance threshold, we analyzed demographic trends among this group of subjects.

First, median concordance for all subjects was 88%, meaning that 50% of subjects achieved at least 88% concordance between their SUR data their UDS data. The mean concordance for all subjects was 83%, indicating that concordance data for this study population is somewhat left-skewed by subjects with very low concordance. Indeed, the minimum concordance achieved by a given subject was 6%, whereas the maximum concordance was 100%. For overall concordance, we observe a significant gender effect: median male concordance is 86%, whereas median female concordance is 96% (p -value = 0.047). However, completion status, race, baseline cocaine use, and study assignment do not appear to influence concordance.

Second, 80% of subjects exhibited an overall concordance of at least 0.70 throughout their participation in the study. In comparison to 0.70 concordance threshold, 65% of subjects achieved an overall concordance of at least 0.80, and 43% of subjects achieved an overall concordance of at least 0.90. Study assignment did significantly affect whether or not a subject met the 0.70 reliability criterion, with a higher percentage of subjects in the N-acetylcysteine study achieving at least 0.70 concordance. However, race, gender, completion status, and level of baseline cocaine use were not associated with meeting the 0.70 reliability criterion.

Finally, we were interested in classifying the ideal study participants whose SUR results were consistently in agreement with their UDS results throughout the entire study, and thus achieved a concordance level of 1.0. Twenty-six percent of all subjects achieved 100% concordance, as confirmed by UDS. Furthermore, we again observed a significant gender effect, with 42% of female subjects achieving 100% concordance in comparison to 22% of male subjects.

Discussion

Overall, sensitivity of self-use reports (SUR) decreased (and the false negative rate rose) in a statistically significantly fashion over the course of an eight-week pharmacological treatment study for cocaine dependence. A significant gender trend with regards to concordance of SUR with urine drug screens (UDS) was observed among all subjects, and among subjects who achieved 100% concordance. Among all subjects, median concordance among females was 96%, yet only 86% among males. Similarly, 42% of female subjects achieved 100% concordance, while only 22% of male subjects did.

The decreasing temporal trend in sensitivity of SUR and in concordance between SUR and UDS is consistent with other investigators' findings. The results of this study support the hypothesis that subjects are most motivated to report drug use accurately at the beginning of a study, and have declining motivation as the study progresses. [5] Subjects may be more willing to admit drug use at the beginning since current dependence is a required eligibility criterion; furthermore, subjects may believe that an accurate description of their use patterns is essential to successful treatment. As treatment progresses, subjects may be motivated to deny use out of desire to show progress or in an attempt to preserve self-esteem by denying relapse. [5] Furthermore, at study completion, subjects may be reluctant to admit use since the protective atmosphere of the treatment program is terminating [5]. Note that an alternative explanation for a decreasing trend in concordance is that subjects with high concordance drop out early in the study, and subjects with lower concordance complete. However, we analyzed individual concordance rates and individual duration of treatment, and these variables were not strongly associated. That is, subjects with shorter treatment durations did not have statistically higher concordance rates. Thus, we dismiss this hypothesis as an explanation for the observed temporal trend.

A significant gender trend among all subjects and among subjects who achieve 100% concordance was observed; in both cases, females averaged a higher concordance than males. However, the proportion of males and females achieving at least 70% concordance did not statistically differ. Thus, care must be taken when determining a reliability threshold, since demographic differences may exist at certain concordance levels. Although 70% concordance appears robust to demographic factors in our study, it is unclear whether 70% concordance is truly clinically meaningful. Clearly, it is of clinical relevance if a subject is 100% concordant, but to what degree is partial concordance meaningful? Indeed, any adopted concordance threshold may have very little clinical relevance, and only serve as a technical method for addressing missing data. Thus, great care must be taken to establish a set threshold *a priori*, rather than in a post-hoc, data-driven fashion. Finally, this gender difference may be a result of variations in subject-interviewer gender pairing, a phenomenon we plan on investigating further.

In an outpatient clinical trial setting, no perfect measure for assessing illicit drug use currently exists. Since self-report and urine drug screens are the most commonly used measures to gauge illicit drug use, it is imperative to understand the limitations associated with each. The major limitation of UDS is the potential for missing data; furthermore, missing data may be more likely when the subject is using. Thus, from an analysis perspective, when UDS data is missing, it is reasonable to assume that this data are not always missing completely at random (MCAR). When missing data are not MCAR, it is possible that parameter estimates, such as treatment effect, may be biased [17]. However, if subject characteristics (including concordance and SUR) can predict missingness in the UDS, then missing data mechanism may be considered missing at random (MAR), in which case standard likelihood methods would yield unbiased estimates, provided that the predictive characteristics are included in the model. Thus, this research primarily aims to describe a meaningful relationship between SUR results and UDS results, as measured by concordance, with the ultimate goal of developing effective and unbiased methods to handle missing data in addiction clinical trial settings.

Several important limitations of this study deserve mention. First, due to the ongoing nature of these two, blinded clinical trials, tests for a treatment group effect on concordance could not be conducted. Additionally, all research involving benzoylecgonine (BE) drug screens are confounded by the fact that the half-life of BE varies by individual, yet UDS results are interpreted in the context of a standard 72 hour half-life. As such, one must be careful to not directly associate concordance with 'truthfulness.' However, discordance is a clinically relevant phenomenon that should be investigated and described in detail. As previously mentioned, the use of standard metabolic window for BE may result in either false positives or false negatives for subjects with nonstandard metabolic rates. Additionally, in our study self-reported use is assessed in full-day units, and does not cover the fractional day elapsed during the day of treatment. Thus, if a subject has used the morning of treatment, but not the previous three days, his true self-report will be negative. Therefore, some of the false negatives may due to study design, rather than subject inaccuracy. Finally, even if a subject achieves concordance between his urine screens and self-report, he may still be underreporting. If a subject consistently has positive urine screens and reports use once every three days, yet is using every day, he will achieve concordance yet be underreporting his use by 66% [18].

Conclusions

Missing UDS data compromises the results of clinical drug trials. SUR data could potentially be used as a surrogate outcome for missing UDS results; however, the accuracy of SUR may be affected by potential biases. If a standardized method for extracting reliable information regarding use patterns from SUR data could be established, this composite approach would be of great value to clinical drug study research. Missing data methods involving concordance

may prove effective, yet more research is needed to describe and validate concordance-based methods.

Acknowledgments

This research was supported by grants # R01 DA019903 and # R01 DA016368 from the National Institute on Drug Abuse.

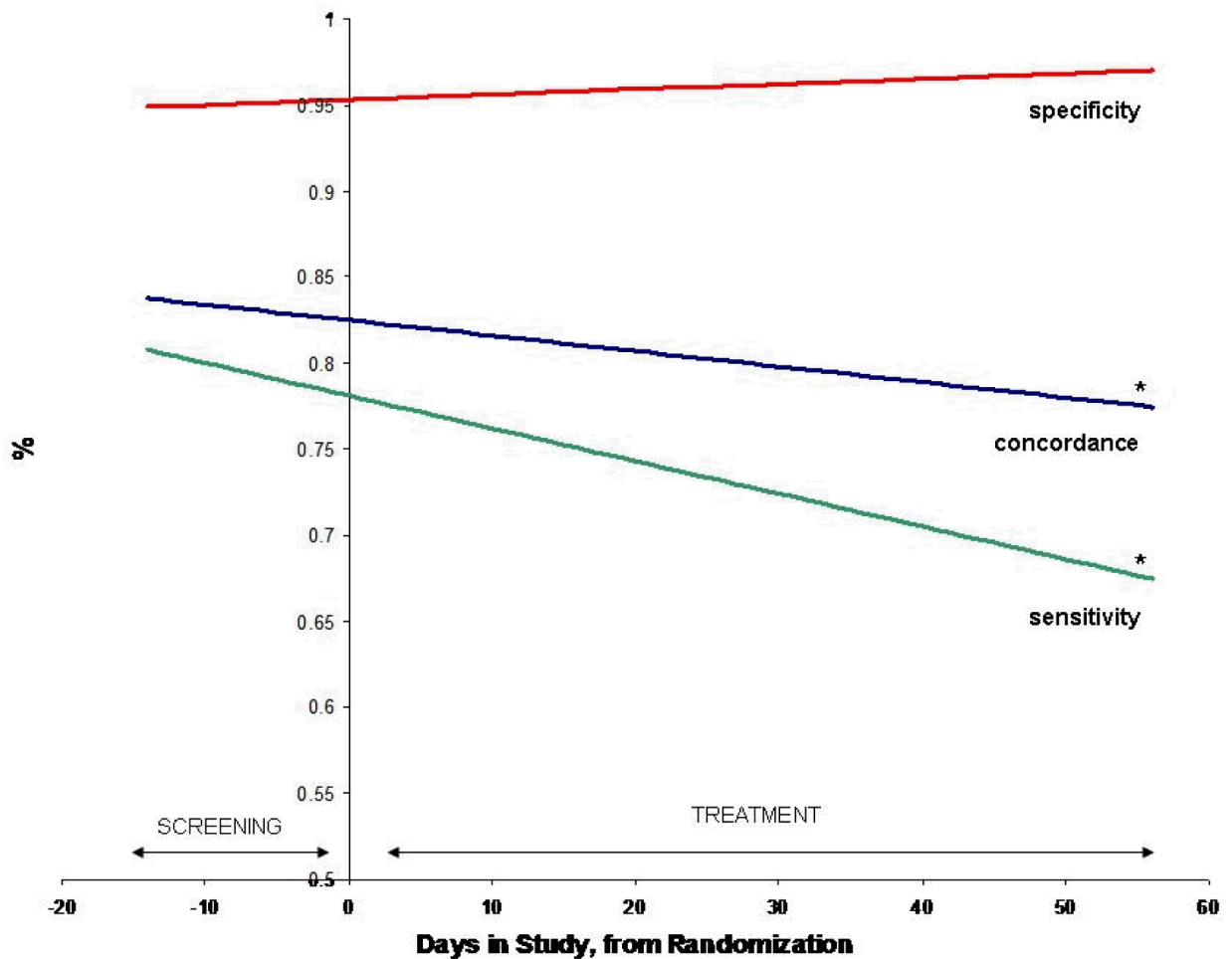
Funding: This research was supported by grants # R01 DA019903 and # R01 DA016368 from the National Institute on Drug Abuse.

References

1. Preston, KL.; Silverman, K.; Schuster, CR., et al. Comparison of Self-Reported Drug Use with Quantitative and Qualitative Urinalysis for Assessment of Drug Use in Treatment Studies. In: Harrison, L.; Hughes, A., editors. *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*. US Department of Health and Human Services; Rockville, MD: 1997. p. 130-412. NIDA Research Monograph 167
2. Magura S, Kang SY. Validity of self-reported drug use in high risk populations: a meta-analytical review. *Substance Use & Misuse* 1996;31:1131–1153. [PubMed: 8853234]
3. Somoza E, Somoza P, Lewis D, et al. The SRPHK1 outcome measure for cocaine-dependence trials combines self-report, urine benzoylcegonine levels, and the concordance between the two to determine a cocaine-use status for each study day. *Drug and Alcohol Dependence* 2008;93:132–140. [PubMed: 18029115]
4. Hser, YI. Self-Reported Drug Use: Results of Selected Empirical Investigations of Validity. In: Harrison, L.; Hughes, A., editors. *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*. US Department of Health and Human Services; Rockville, MD: 1997. p. 320-343. NIDA Research Monograph 167
5. Tassiopoulos K, Bernstein J, Heeren T, et al. Predictors of disclosure of continued cocaine use. *Addictive Behaviors* 2006;31:80–89. [PubMed: 15908136]
6. Greenfield L, Bigelow GE, Brooner RK. Validity of intravenous drug abusers' self-reported changes in hi-HIV-risk drug use behaviors. *Drug and Alcohol Dependence* 1995;39:91–98. [PubMed: 8529537]
7. Hoffman, JA.; Wish, ED.; Koman, JJ., et al. Self-reported drug use compared with hair analysis and urinalysis. Paper presented at the College on Problems of Drug Dependence 56th Annual Scientific Meeting; Palm Beach, FL. 1994, June 18-23;
8. Hindin R, McCusker J, Vickers-Lahti M, et al. Radioimmunoassay of hair for determination of cocaine, heroin, and marijuana exposure: Comparison with self-report. *International Journal of the Addictions* 1994;29:771–789. [PubMed: 8034385]
9. Baumgartner AM, Jones PF, Baumgartner WA, et al. Radioimmunoassay of hair for determining opiate abuse histories. *Journal of Nuclear Medicine* 1979;20:748–752. [PubMed: 541713]
10. Myrick H, Henderson S, Danky B, et al. Clinical characteristics of under-reporters on urine drug screens in a cocaine treatment study. *Am J Addict* 2002;11:255–261. [PubMed: 12584868]
11. Nyamathi A, Leake B, Longshore D, et al. Reliability of homeless women's reports: Concordance between hair assay and self-report of cocaine use. *Nursing Research* 2001;50:165–171. [PubMed: 11393638]
12. First, MB.; Spitzer, RL.; Gibbon, M., et al. *Structured Clinical Interview of Axis I DSM-IV Disorders – Patient Edition (With Psychotic Screen) (SCID-I/P w/ Psychotic Screen, Version 2.0)*. Biometrics Research Department, New York Stat Psychiatric Institute; New York: 1996.
13. Sobell, LC.; Sobell, MB. Timeline follow back: a technique for assessing self-reported ethanol consumption. In: Allen, J.; Litten, R., editors. *Techniques to Assess Alcohol Consumption*. Humana Press Inc; Totowa, NJ: 1992. p. 19-28.
14. McCullagh, P.; Nelder, JA. *Generalized Linear Models*. Chapman & Hall; London: 1989.
15. SAS Institute Inc.. *SAS/STAT 9.1 User's Guide*. SAS Institute Inc; Cary, NC: 1989.

16. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
17. Hedden SL, Woolson RF, Malcolm J. A comparison of missing data methods for hypothesis tests of the treatment effect in substance abuse clinical trials: a Monte-Carlo simulation study. *Substance Abuse Treatment, Prevention, and Policy* 2008;3:13.
18. Morral AR, McCaffrey D, Iguchi MY. Hardcore drug users claim to be occasional drug users: drug use frequency underreporting. *Drug Alcohol Dependence* 2000;57:193–202.

Temporal Trends in Sensitivity, Specificity, and Concordance



Summary Index	Change in summary index for each day on study (i.e., $\hat{\beta}$) Parameter Estimate	Standard Error of Estimated Rate of Change (i.e., $se(\hat{\beta})$)	p-value
Concordance	-0.0009	0.0004	0.002*
Sensitivity	-0.002	0.0005	0.0003*
Specificity	0.0003	0.0003	0.277

Figure 1.

GEE regression model depicting trends in concordance, specificity, and sensitivity during pretreatment screening and eight-weeks of treatment. Specificity remains constant around 0.95, yet both sensitivity and concordance drops statistically significantly during the study. Regression lines marked with an asterisk are significant at the 0.05 level.

Table 1

Demographics of the study population. In general, the N-acetylcysteine and Modafinil subjects are highly comparable, although Modafinil subjects are statistically significantly younger than N-acetylcysteine subjects. P-values marked with an asterisk are significant at the 0.05 level.

	Overall	N-acetylcysteine Study	Modafinil Study	
Total Study Participants	129	66	63	
	<i>median [IQR]</i>	<i>median</i>	<i>median</i>	<i>Wilcoxon Rank Sum p-value</i>
Age	42 [35–47]	44 [35–44]	41 [35–44]	0.037*
Baseline Monthly Cocaine Days	12 [8–19]	12 [8–19]	12 [8–21]	0.888
Days in Study	50 [21–56]	50 [20–55]	51 [23–56]	0.563
	<i>% (n)</i>	<i>% (n)</i>	<i>% (n)</i>	<i>χ^2 p-value</i>
Gender				
Male	80% (103)	80% (53)	79% (50)	0.894
Female	20% (26)	20% (13)	21% (13)	
Race				
Black	49% (63)	47% (31)	51% (32)	0.664
White	51% (66)	53% (35)	49% (31)	
Baseline Cocaine Use				
Light	35% (45)	29% (19)	41% (26)	0.137
Heavy	65% (84)	71% (47)	59% (37)	
Completion Status				
Completers	56% (72)	53% (35)	59% (37)	0.515
Non-Completers	44% (57)	47% (31)	41% (37)	

Table 2

Measures of concordance at baseline (first screening visit) and at final study visit for each of the 129 subjects. Sensitivity and specificity values are reported, as well as the false negative and false positive rates.

Baseline	Urine Drug Screen		Total	Final Study Visit	Urine Drug Screen		Total
	-	+			-	+	
Self-Report							
-	29 Specificity = 90.6%	16 False Neg = 16.5%	45	-	40 Specificity = 95.2%	32 False Neg = 36.8%	72
+	3 False Pos = 9.4%	81 Sensitivity = 83.5%	84	+	2 False Pos = 4.8%	55 Sensitivity = 63.2%	57
Total	32	97	129	Total	42	87	129

Table 3

Concordance, sensitivity, and specificity at baseline, final visit, and overall. The overall concordance measures yield the narrowest confidence intervals, since they utilize the most subject observations.

	Concordance		Sensitivity		Specificity	
	Point Estimate	95% CI	Point Estimate	95% CI	Point Estimate	95% CI
Baseline	85.3%	[79.1%, 91.5%]	83.5%	[76.1%, 90.1%]	90.6%	[80.5%, 100%]
Overall	82.0%	[78.8%, 85.2%]	74.8%	[70.1%, 79.6%]	96.1%	[94.3%, 97.9%]
Final Visit	73.6%	[65.9%, 81.3%]	63.2%	[53.1%, 73.4%]	95.2%	[88.8%, 100%]

Table 4

Description of concordance by various demographic groups for three groups of interest: total study population, subjects that achieve 70% concordance, and subjects that achieve 100% concordance. We observe a significant gender effect for the study population overall, and for subjects that achieve 100% concordance. P-values marked with an asterisk are significant at the 0.05 level.

	Overall		70% Concordance		100% Concordance	
	Median Concordance % (n)	Wilcoxon-Rank Sum p-value	% of subjects achieving 70% concordance (n)	χ^2 p-value	% of subjects achieving 100% concordance (n)	χ^2 p-value
Completion Status						
Completers	87% (72)	0.900	79% (57)	0.829	25% (18)	0.694
Non-Completers	88% (57)		81% (46)		28% (16)	
Gender						
Male	86% (103)	0.047*	80% (82)	0.896	22% (23)	0.039*
Female	96% (26)		81% (21)		42% (11)	
Race						
Black	88% (63)	0.754	78% (49)	0.567	25% (16)	0.809
White	88% (66)		82% (54)		27% (18)	
Baseline Cocaine Use						
Light	83% (45)	0.401	78% (35)	0.668	24% (11)	0.718
Heavy	89% (84)		81% (68)		27% (23)	
Study						
N-acetylcysteine	87% (66)	0.711	73% (48)	0.039*	30% (20)	0.298
Modafinil	88% (63)		87% (55)		22% (14)	