THE USE OF SELF-REPORT AND DRUGS TESTS IN THE MEASUREMENT OF ILLICIT DRUG CONSUMPTION[$]

Ziggy MacDonald and Stephen Pudney[*]

Department of Economics
University of Leicester

March 2003

Summary

We use data from the New England and Wales Arrestee Drug Abuse Monitoring (NEW-ADAM) programme to assess the validity of self-report measures of illicit drug use and to evaluate the use of alternative drug testing strategies within survey enquiries. Our analysis of the NEW-ADAM data reveals that bio-assay measurements of drug use tend not to be very sensitive to the cut-off levels selected for screening tests, a result that holds for cannabis, cocaine and opiates. We also show that a self-reported history of previous drug use can be used as a way of identifying individuals who are potential under-reporters of current drug use. This suggests a selective drug testing strategy which can reduce dramatically the cost of drug testing without comprising the accuracy of measurements of illicit drug use.

Keywords: Drugs testing, Self-reports, Concordance, Testing Strategies

[$] This work was funded by the Home Office as part of a feasibility study of redesign of NEW-ADAM, conducted jointly with the National Centre for Social Research. We are grateful to colleagues at the Home Office and the National Centre for their help and advice. The opinions expressed in this paper are in no way an official view of the Home Office; neither should they be considered an indication of Home Office policy. Remaining errors and omissions are the sole responsibility of the authors.

[*] Address for correspondence: Stephen Pudney, Department of Economics, University of Leicester, University Road, Leicester LE1 7RH. E-mail sep2@le.ac.uk

1. Introduction

Illicit drug use and associated crime touch many people's lives in modern society, with those affected ranging from the users and sellers of drugs, to parents, social groups and ultimately government. Most research into the causes and consequences of illicit drug use makes use of self-reported consumption data, usually collected via social surveys (for example the British Crime Survey in the UK and the National Survey on Drug Abuse in the US). The difficulties associated with self-reported data are widely recognised, particularly when the subject matter is sensitive. Typically we expect some form of under-reporting from questions about drug use, and this is likely to vary between socio-economic groups. There is a growing literature that considers the validity of self-reported drug use by comparison with drugs tests on biological samples. Magura and Kang (1996) present a meta-analysis of 24 studies published since 1985. This review suggests that there tends to be only moderate agreement between self-reports and drug test results. For the original Drug Use Forecasting programme (the forerunner to the US Arrestee Drug Abuse Monitoring [US-ADAM] programme), Feucht *et al*. (1994) report that juvenile male arrestees in Cleveland had very low self-report rates, with only 6% of those with cocaine-positive hair samples actually admitting cocaine use in the reference period.

More recent studies include that by Lu *et al*. (2001), who compared the self-report information of arrestees from the US-ADAM programme with the associated urinalysis data, focussing specifically on crack cocaine, but also cannabinoids, methamphetamines and opiates. The authors present a number of measures of concordance that reveal false-negative reporting to be the biggest problem, although the discrepancies are not as large as those reported by Feucht *et al*. (1994). The authors found that apart from cannabis, less than 50% of arrestees who tested positive for drugs actually admitted it. The rate for cannabis was 63.6%.[1] Lu *et al*. also found that true-negative reporting rates were very high, with the majority of detainees who tested negative for drugs being accurate in their reporting of non-use. They also carried out a logit analysis of false reporting that revealed little difference between ethnic groups in reporting behaviour, except that Black arrestees were found to be less likely to under-report crack consumption than whites or Hispanics. A final point to note from this

[1] Although it is interesting to note that in a study into the drug use habits of a small sample of 182 HIV infected or high-risk non-infected adolescents, Murphy *et al*. (2000) found that self-reported cannabis rates tended to be slightly *higher* than urinalysis rates.

study is that other factors found to influence truthful reporting included past experience of drug treatment programmes and previous experience of arrest.

In another recent study, Mieczkowski *et al*. (1998) revisited the site originally studied by Feucht *et al*. (Cleveland), and a new site in Florida. Hair and urine samples were collected from 426 juvenile detainees in the 14-18 age range. They found that more drug use was revealed in samples than in self-reports, although this was less pronounced for cannabis. Like the Feucht *et al*. study, they also found that hair analysis revealed a greater rate of cocaine use than was revealed by urine testing. Substantial under-reporting of cocaine was also found by Appel *et al*. (2001), who compared the results of hair analysis to the self-reported drug use of a sample of 179 homeless/transient adults from New York, surveyed in 1994. The rate of self-reported cocaine use was one third of that detected by hair analysis (64% of the sample had cocaine-positive hair samples whilst only 18% of the sample reported use in the past 30 days).

In this paper we investigate the validity of self-reported drug use in Britain, using drug-testing information collected through the New England and Wales Arrestee Drug Abuse Monitoring (NEW-ADAM) survey conducted in 1999-2001. We consider the sensitivity of false reporting rates to the calibration of assay-type drug screening tests. We then analyse the extent to which false reporting of illicit drug use can be predicted given information about the individual's personal characteristics and history of drug use. We go beyond the existing literature by considering the implication of this for survey design, by developing two alternative selective drug-testing strategies, which can reduce survey costs by implementing tests only in cases where there is predicted to be a significant danger of under-reporting.

In the next section the practicalities of drugs testing are discussed, including the types of tests that can be done, the biological samples that can be used, the differences between screening and confirmatory testing, and the calibration of drugs tests. We then analyse the validity of self-reported drug use using data from the NEW-ADAM programme. Section 4 proposes some selective drug testing strategies and assesses their accuracy, Section 5 concludes.

2. Bio-assay as a measurement device

2.1. Screening and Confirmatory Testing

The presence of drugs in a biological sample is typically detected via immunoassay or chromatography, which are used to detect the metabolites of the drugs produced by the body

following consumption. The former is a screening test that is usually used to test for the presence of metabolised drugs in a sample of blood, urine, saliva, hair or sweat. A biological sample is taken and subjected to the screening test. The test is calibrated against pre-determined cut-offs that indicate the presence of drugs in the biological sample (see below for more details of the cut-offs). Whereas other ADAM programmes such as those in the U.S. (US-ADAM) and Australia (DUMA) use the Enzymes Multiplied Immune Testing (EMIT) screening test (Makkai, 2000), NEW-ADAM used the 'On-Line' Kinetic Interaction of Micro-Particles (KIMS) test. The choice of the KIMS test for NEW-ADAM was based on 'a balance of its cheapness and acceptable levels of accuracy' (Bennett, 1998, p.13).

Screening tests usually involve the use of a competitive binding immunoassay to detect the presence of drugs in a biological matrix. The choice of biological sample and the particular testing kit used will determine what drugs of abuse can be tested for. The urine samples in NEW-ADAM were tested by the Forensic Science Service (FSS), for seven drug types including opiates, cocaine, canabinoids and amphetamines. Although screening tests are relatively inexpensive (currently around £10-£15 per test for saliva and urine samples), they are subject to two limitations: specificity and cross-reactivity (Makkai, 2000; Bennett, 1998). Typically, screening tests do not identify specific chemical compounds present in biological samples, and hence it is not possible to separately identify cocaine from crack, and amphetamine from ecstasy. This problem of specificity is not unique to KIMS (see discussion below), although it is suggested that the EMIT test 'is considered to have high sensitivity and moderate specificity' (Makkai, 2000, p. 13). Cross-reactivity is perhaps a bigger concern when using screening tests in the context of the Criminal Justice System. This problem relates to false-positive results, whereby consumption of legal substances (e.g. codeine-based pain killers or poppy seeds) can give rise to detectable traces of morphine or codeine in urine and hence a positive screening test (this is discussed in more detail below).

The typical approach to overcoming the uncertainties of screening tests is to subject samples that test positive to a confirmatory test (the alternative is to conduct a confirmatory test on randomly selected samples). Confirmatory tests are far more accurate than screening tests, but considerably more expensive. They were not used in NEW-ADAM (Bennett, 1998, p.91), but have been used in the Australian DUMA for opiates, amphetamines and benzodiazepines. The typical confirmatory test is undertaken by use of Gas Chromatography-Mass Spectrometry (GCMS), although there are recent developments such as capillary electrophoresis-mass spectrometry that offer an acceptable alternative (Ramseier *et al*., 2000).

In this section we give a brief description of the metabolites from the common drugs of abuse that are detected by the standard GC-MS confirmatory tests, concentrating on cannabinoids, cocaine and the opiates (see Goldberger and Cone, 1994 for further detail).

2.1.1. Cannabinoids

The most prevalent metabolite of tetrahydrocannabonal (THC) present in cannabis and the cannabinoids is 11-nor-Δ^9-tetrahydrocannabinol-9-carboxylic acid (THC-COOH). This will be detected in most biological samples regardless of the method of ingestion. It is unlikely that THC-COOH will be detected in biological samples taken from individuals who have passively consumed cannabis smoke, unless this has happened for a long duration in a totally unventilated room.

2.1.2. Cocaine

The main metabolite of cocaine is benzoylecgonine and the body rapidly creates this through hydrolisation following ingestion. The extraction of benzoylecgonine from biological samples is very straightforward and it can be present in some samples for several days after ingestion. It is possible that benzoylecgonine will be detected in biological samples following consumption of certain 'health teas' that are prepared with coca leaf, although these are quite rare. There is a possibility that benzoylecgonine can also be detected in the urine of individuals who have been passively exposed to the vapours of cocaine base (crack) as it is smoked. Generally it is difficult to determine whether cocaine metabolites found in specimens are from the powder form that has been insufflated ('snorted') or injected from the freebase form (crack). Lu *et al*. (2001) suggest that testing for anhydroecgonine methyl ester can give direct evidence of smoking crack as this is produced as crack cocaine is heated and is rarely present when the powder form is consumed.

2.1.3. Opiates

Following consumption, heroin (diacetylemorphine) is metabolised first into 6-monoacetylmorphine (MAM) and then into morphine. Morphine is further metabolised into morphine-glucuronide and normorphine. The primary metabolites found in biological samples are morphine and morphine-glucuronide, although MAM may also be present (a definite

indication of heroin consumption). Codeine is often found as it is frequently present in heroin street samples. When consumed, codeine is metabolised to codeine-glucuronide, morphine and norcodeine, although codeine and morphine are the primary metabolites found in biological samples. The main problem with opiates is that although the GC-MS analysis will separately identify, say, codeine and morphine, it usually fails to differentiate between morphine and codeine deriving from the ingestion of some over-the-counter medicines and even poppy seeds, which are widely used in baking. Alternative indicators of heroin such as acetylcodeine, which is an impurity of heroin synthesis, have been considered in the literature (see for example Staub *et al*., 2001), but these tend to be more useful in detecting illicit use of street heroin in individuals engaged in treatment programmes.

2.2. Biological samples

There are five alternative samples that can be taken for drug testing: blood, urine, saliva, sweat, and hair. Each of these has its relative merits in terms of cost, ease of collection, openness to adulteration, and the window of exposure. There is an extensive literature on the use of different biological samples for drugs testing (e.g. Baer and Booher, 1994; Kerrigan and Phillips, 2001; Kidwell *et al*., 1998; Kintz, 1996; Makkai, 2000; Navarro *et al*., 2001; Osselton *et al*., 2000; Pichini *et al*., 2002; Sachs and Kintz, 1998; Skopp and Potsch, 1999), a synthesis of which is presented below.

2.2.1. Blood

Most drug types and their metabolites are passed into the blood system, although their concentration can be quite low and give rise to a window of detection of only 24-36 hours. The biggest drawback with blood is that screening can only be done in the laboratory and given its invasive nature can only be taken as a sample by medically trained staff. In the context of drugs testing of arrestees, blood sampling is probably not feasible given its cost, the potential for infection, and the environment in which it would have to be taken.

2.2.2. Urine

Urine is the most widely used biological sample for drugs testing in the Criminal Justice System and in the workplace. Drugs tend to be relatively more concentrated in urine (e.g. 10-

100 times more concentrated than in blood) due to the way in which the kidney filtrates the blood and concentrates the by-products. Urine testing has a cost advantage as much of the preparation of samples can be automated. One drawback is that sample collection and handling is relatively intrusive and unsavoury, which can be problematic for survey interviewers. A further problem with urine is that it is relatively easy to adulterate the sample (for example, excessive dilution through water consumption or through the use of commercial test evasion products). These are not usually serious concerns in the survey context.

2.2.3. Saliva

Saliva can be used to screen for the presence of heroin, dihydrocodeine, codeine, morphine, cocaine (and its principle breakdown product), and benzoylecgonine. Is not as useful for detecting cannabis, as the molecules tend to be too big to pass into the saliva. The window of detection for most drugs in saliva is similar to that for blood as oral fluid is formed from the circulating blood (drug metabolites pass into saliva through tissue membranes from the blood plasma). The amount of drug metabolite present in saliva will depend on the method of consumption, the acidity of the oral fluid, and the elapsed time since the dose was taken. It is difficult to adulterate saliva as it is continuously produced in the mouth then swallowed and collection can be easily observed without invading privacy. The main difficulty is with 'dry-mouth', which can be caused by cannabis, amphetamine and opiate use. Consuming some over-the-counter travel sickness preparations can also produce a dry mouth. The implication of 'dry mouth' is that it takes longer to achieve an appropriate volume of sample and may therefore threaten the response rate in voluntary surveys.

2.2.4. Sweat

Most drugs of abuse are excreted by the body in sweat, and in some case heroin molecules, in addition to the metabolites, have been detected in this matrix. The window of detection is not as long as urine, but certainly better than blood. The biggest problem with sweat is its collection, since one cannot expect an arrestee to secrete a sufficient sample of sweat within the usual time frame of an interview. One practical method that has been developed is the 'sweat patch', which acts as a container for the non-volatile components of sweat. Presuming that the patch cannot be tampered with, it is fairly secure against adulteration as the material used does not allow non-volatile substances from the environment to penetrate it.

Unfortunately the sweat patch is not practical in the context of the Criminal Justice System, as it has to be worn for several days before a sufficient volume of sweat is collected for analysis.

2.2.5. Hair

Hair can be used to detect most of the main drugs of concern including the main cocaine and heroin metabolites, and has the benefit of detecting THC-COOH rather than just THC, which can come from environmental exposure. One advantage of hair is that is can give a fairly accurate picture of the individual's recent drug history, plus it can give a more accurate measure of concentration than, say, urine. It would also appear that hair samples give a much more accurate indicator of cocaine, and that individuals who use cocaine infrequently or at low dosage levels are less likely to be screened positive with urine samples (Mieczkowski *et al*., 1998). A further benefit of using hair is that it is fairly difficult to adulterate the sample (although some intensive hair preparations can affect the analysis). Against this, hair testing is expensive relative to urine and cannot be tested on-site. There are also obvious problems if the detainee does not have any visible hair. The biggest advantage in using hair is the wide window of detection (a two-inch strand can capture approximately four months' worth of drug exposure), although it should be noted that there is a latency period of between three to seven days before drugs are deposited in the hair (Appel *et al*., 2000).

2.3. Choosing the cut-off

Screening tests usually require the analyst to visually detect the presence or absence of a coloured band or line on a chemically impregnated membrane following exposure to the sample. This is then calibrated against a pre-determined scale according to the colour of the band that gives an indication of the concentration of the drug in the sample. A positive or negative result is given by reference to cut-off levels of concentration (this may be achieved automatically with an electronic test reader). The cut-off levels used in the NEW-ADAM and DUMA studies are given in Table 1 (also shown are the cut-offs used in Leino *et al*.'s (2001) comparison of eight on-site screening devices). The NEW-ADAM cut-offs were agreed in consultation with the FSS, whilst the DUMA cut-offs were set by the official body Standards Australia. There is some variation in these cut-off levels across the three studies, particularly for amphetamines, but also for cocaine where the NEW-ADAM cut-off is half the level used in DUMA and by Leino *et al*. Bennett (1998, 2000) does not define the criteria used to

determine these cut-offs, although it appears to follow common practice of setting cut-offs quite high to minimise false-positive results. Although this is appropriate in legal contexts, it is a potential source of bias for a prevalence survey, where one would like the cut-off to give a reasonable balance between false negatives and false positives.

Table 1. Cut-offs used in NEW-ADAM and DUMA*

Drug Type	NEW-ADAM	DUMA	Leino *et al*.
Amphetamines	500	300	1000
Benzodiazepines	100	100	-
Cannabis	50	50	50
Cocaine	150	300	300
Methadone	300	-	-
Opiates	300	300	300

* Source: Bennett (2000) and Makkai (2000).

In the following section we use data from the NEW-ADAM study to consider further the sensitivity of test results and their concordance with self-reporting to the specified cut-off levels for cannabis, cocaine and opiates.

3. An analysis of NEW-ADAM data

3.1. The NEW-ADAM data

The NEW-ADAM survey was conducted annually (from 1999 to 2001) over a thirty-day period, and yields information on approximately 1500 arrestees collected at eight police custody suites, rotated on a two-year cycle. Interviewing of arrestees was by consent, and a typical interview lasted for twenty minutes after which a urine sample was collected from the individual (again, by consent, given at the start of the interview process). The interview covered areas such as self-reported drug and alcohol use, criminality and contact with drug treatment services. For more details of the NEW-ADAM methodology see Holloway and Bennett (2003).

With respect to the sampling process, all arrestees entering the participating custody suites were assessed by the interviewers for eligibility to be interviewed, although in some cases the custody officer will have made this assessment on the basis of interviewer safety. Typical

reasons for non-eligibility were that the arrestee was not fit due to alcohol or that there is insufficient time in the custody process for an interview to take place. In addition, juveniles, prison transfers and those held only for breath tests were ineligible, as were arrestees with mental health problems or insufficient English language skills. The result of this selection process is that the pool of eligible arrestees was typically less than 50% of the throughput of the participating custody suites. Moreover, on average only 50% of eligible arrestees were interviewed for NEW-ADAM, although this varied considerably across participating sites. In addition to response rate problems, NEW-ADAM is also considered to be unrepresentative of arrestees. This is mostly because NEW-ADAM only involved a small number of not necessarily representative custody suites, which tended to be relatively large ones situated in areas with relatively high levels of social deprivation and drug abuse. The full extent of the limitations of the NEW-ADAM data are discussed in Pudney *et al*. (2003), whilst a detailed feasibility study to improve the design of the survey is reported in Boreham *et al*. (2003). Some of the problems arising from the NEW-ADAM design are also discussed in Bird *et al*. (2002).

3.2. Robustness of test results and conflict with self-reported drug use

Given the nature of the NEW-ADAM process, it is possible that our results will be affected by the arrestees' prior knowledge that a drugs test will be done at the end of the interview (on average, 95% of arrestees consenting to be interviewed also gave consent for a urine sample to be taken). Conceivably, if an arrestee knew that he or she would be tested for drugs after answering questions about drug use, then that individual has a greater incentive not to false report. However, as we shall see later, despite this, there is still a large amount of apparent misreporting of drug use in the NEW-ADAM sample. Before we compare the drug test results with self-reported drug use, we begin by considering the dispersion of screening test results in the NEW-ADAM sample. The distributions of NEW-ADAM tests scores (calibrated in Ng/ml) for cannabis, cocaine, and opiates are shown in Figures 1-3 respectively.[2] What these show is that the distribution of test results tends to be bimodal, with a large number of arrestees providing samples with a reading around zero and another group providing samples at a much higher positive value at or beyond the NEW-ADAM cut-off level. This is particularly pronounced for cannabis, although still apparent for cocaine and opiates.

[2] Note that the test scores are calibrated in such a way that negative values are possible. Figures 1-3 show non-parametric kernel density estimates.

Figure 1. The distribution of test scores for cannabis (cut-off 50 Ng/ml)

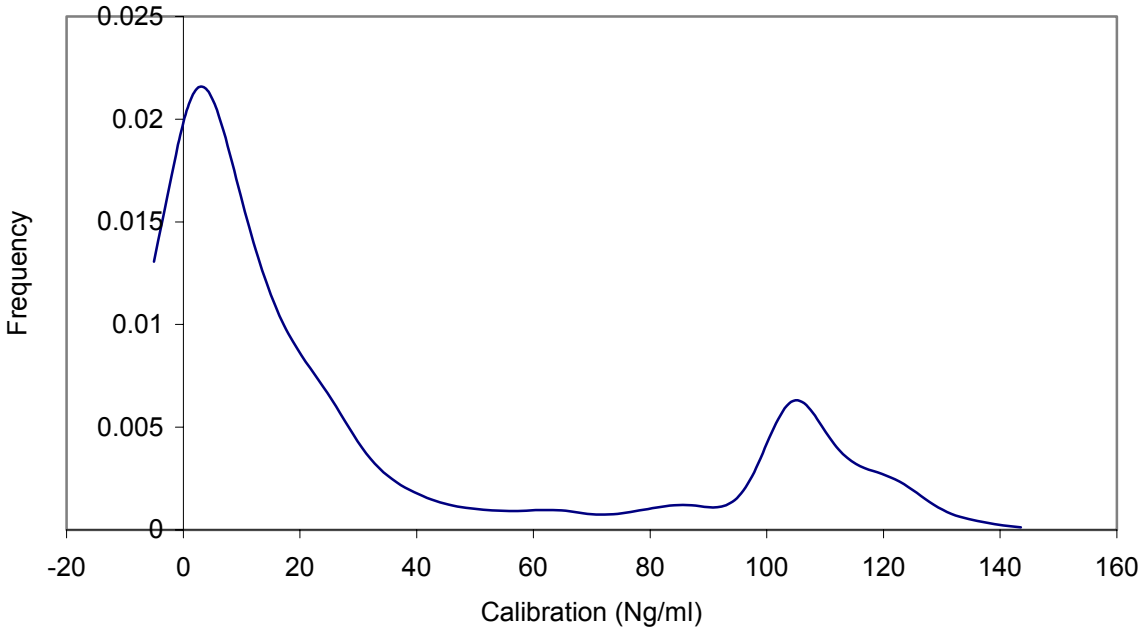


Figure 2. The distribution of test scores for cocaine (cut-off 150 Ng/ml)

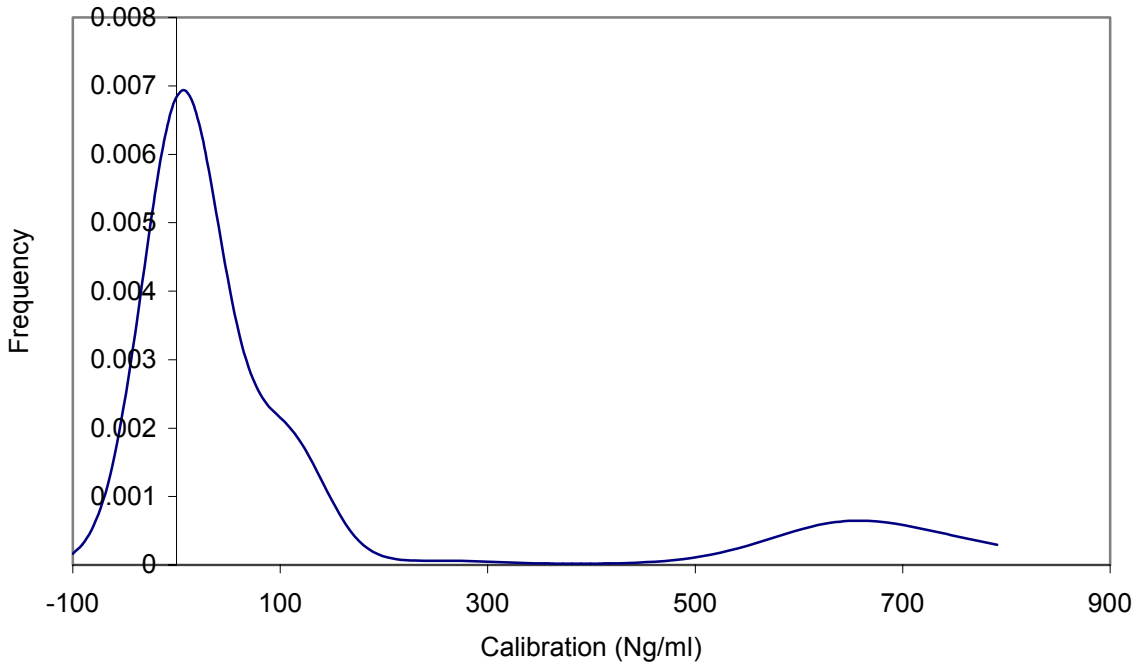
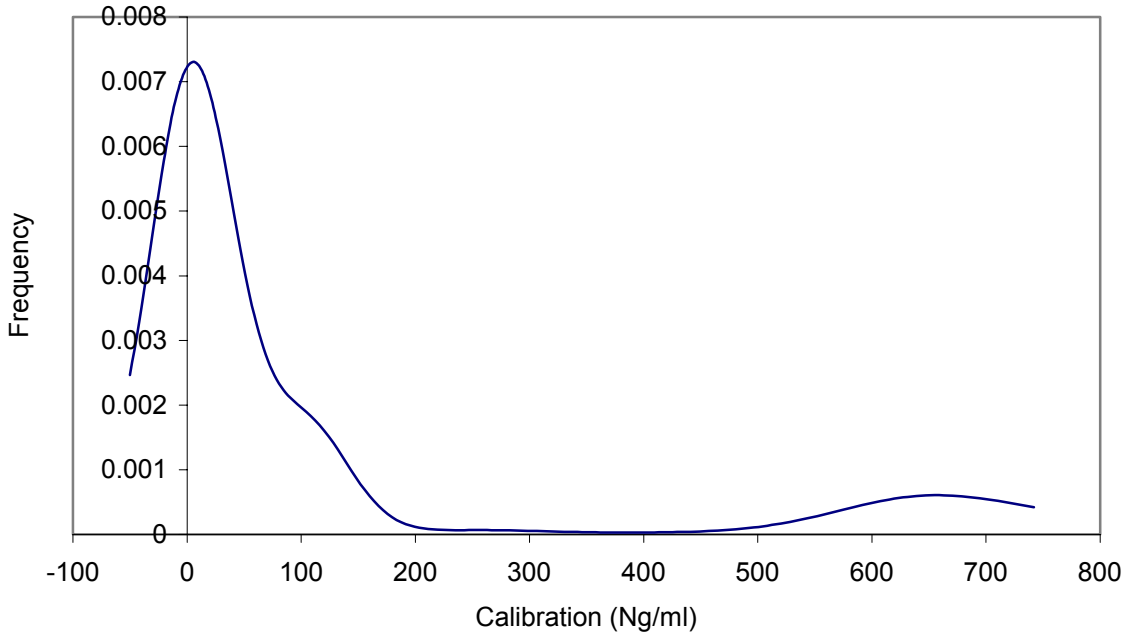


Figure 3. The distribution of test scores for opiates (cut-off 300 Ng/ml)



Our next concern is the concordance between the self-reported drug use of NEW-ADAM interviewees[3] and the drugs test results. Given the distribution of test scores revealed in Figures 1-3, we are also interested in how sensitive concordance is to the cut-off levels used to determine a positive screening test. Table 2 shows the relationship between reporting rates and test results for cannabis, cocaine and opiates, covering a number of recall periods. It should be noted, however, that concordance is most relevant when screening test results are compared to the past three days recall period, rather than longer periods such as the past 12 months, although this longer recall period is of relevance in determining an optimal testing strategy (see below). The results in Table 2 are given for five different cut-off levels centred on the NEW-ADAM level. In general we find that as the cut-off level is increased the percentage of samples testing positive for any of the drugs falls, although there is not a great deal of variation in the rates for opiates and cocaine. However, there is more noticeable variation for cannabis. False-negative reporting is particularly high for cocaine at all the cut-offs used, with a 35% average rate of positive tests for individuals reporting no use in the past three days. This may reflect the window of detection for cocaine, but it should be noticed that there is still a 20% rate of false-negative reporting for individuals who report no use of the drug in the past month, and a 9% rate for those who claim no use ever. This pattern is also seen for opiate use, although it is slightly less pronounced. Cannabis users appear to be the most truthful (as reported earlier) with only 2-3% of those claiming no use providing a positive test result, although the false-negative reporting is quite high (around 20%) for individuals who claim no use in the past three days. This could be partly a window of detection problem, but it should be noted that false-positive reporting is most pronounced for cannabis, with around 16% of those claiming some use in the past three days testing negative (a range of 13.3% to 20% depending on the cut-off level used). Overall these results do not compare well with the reported concordance rates for the Scottish pilot of ADAM (McKeganey *et al*., 2000). Although the Scottish ADAM involved a much lower number of participants (there were 427 interviews yielding 280 useable urine samples), comparison of self-reported drug use and urinalysis results show very high rates of concordance for heroin, cocaine and heroin in the range of 86% to 97%.

[3] The NEW-ADAM survey used traditional paper-based interviewing. For a discussion of the relative merits of computer-aided or paper-based interviewing see Flood-Page *et al*. (2000).

Table 2. Sensitivity of test results to cut-off levels

Cut-off	% testing positive	% of those testing positive claiming no use in last 3 days	% of those testing negative claiming some use in last 3 days	% of those testing positive claiming no use in last month	% of those testing positive claiming no use ever
			Cannabis		
30	53.7	24.3	13.3	10.4	3.2
40	50.6	22.2	15.2	9.5	2.8
50	47.9	20.4	16.7	8.5	2.5
60	45.9	19.9	18.7	8.2	2.6
70	44.0	18.9	20.0	7.7	2.2
			*Cocaine & Crack**		
100	23.3	37.2	2.5	21.3	10.2
125	22.6	35.8	2.7	20.6	9.9
150	21.9	35.2	3.0	20.4	9.4
175	21.6	34.8	3.2	20.2	9.5
200	21.4	34.6	3.3	20.1	9.5
			Opiates		
200	31.7	23.2	1.0	18.3	10.6
250	31.5	22.6	1.0	17.8	10.0
300	30.9	21.4	1.1	16.6	8.9
350	30.8	21.2	1.2	16.5	8.8
400	30.5	20.8	1.3	16.4	8.7

* These include individuals who report recent use of prescription medicines; exclusion of these cases makes little difference to the sample proportions.

To explore the relationship between false reporting and the screening test cut-off levels further, we now examine the proportions of test results that are either very close to or very far from the cut-off levels. Table 3 relates to the results for those individuals who claim no drug use in the past three days but who test positive for drugs, whilst Table 4 covers those who test negative but report drug use in the past three days. What these figures reveal is that most apparently false reporting is not in cases where test results are very close to the cut-off level. In table 4 only a tiny proportion of positive tests are within 10% of the cut-off levels, whilst most are at least 150% above the cut-off. Similarly, for negative test results for those claiming drug use in the past three days, only a small proportion of the tests are within 10% of the cut-off level, whilst the majority, particularly for cocaine and opiates, are below 50% of the cut-off level for all the ranges of cut-offs we have used. This is strong evidence that apparent misreporting is indeed that - a definite attempt to conceal or exaggerate recent drug use, rather than an artefact of the testing process.

Table 3. Test scores of individuals who test positive but claim no drug use within last 3 days

Cut-off	Number of cases	Mean test score	% within 10% of cut-off	% above 150% of cut-off
		Cannabis		
30	373	83	4	81
40	321	91	3	81
50	280	97	3	84
60	261	101	4	75
70	238	104	2	42
		Cocaine & Crack		
100	248	562	2	89
125	231	595	2	92
150	221	616	2	94
175	215	628	1	91
200	212	635	2	87
		Opiates		
200	210	587	1	90
250	203	599	3	89
300	189	623	1	95
350	186	628	3	95
400	181	635	0	95

Table 4. Test scores of individuals who test negative but claim drug use within last 3 days

Cut-off	Number of cases	Mean test score	% within 10% of cut-off	% below 50% of cut-off
		Cannabis		
30	176	14	8	54
40	214	18	7	59
50	249	22	7	59
60	289	26	9	58
70	320	30	7	60
		Cocaine & Crack		
100	55	7	2	95
125	59	15	3	88
150	68	31	7	78
175	71	37	4	76
200	73	41	3	75
		Opiates		
200	19	30	0	95
250	20	41	5	90
300	22	64	9	82
350	23	75	4	78
400	25	98	8	76

3.2. Who are the under- and over-reporters?

One issue that can be explored with the NEW-ADAM data is whether or not individual false reporting can be predicted. To do this we estimate two logit models of false reporting. In the first we consider the probability of generating a positive test result given no reported drug use in the past three days, whilst in the second we consider the occurrence of a negative test result for those who claim to have used drugs in the past three days. Note that we are modelling the probability of misreporting conditional on self-reported drug use, not conditional on the test result. The reason for this is that we want to explore the use of the model in survey design, as the basis for a selective testing strategy. During the interview, self-reported drug use information is immediately available and can be used automatically to determine whether or not a subsequent drug test should be administered.

The dependent variable is a binary indicator of a having positive/negative test result and we include explanatory variables to control for self-reported drug use, offending behaviour and the demographic characteristics of the respondent. This specification is similar to Lu *et al*. (2001), although other covariates that were not statistically significant in any of the models have been dropped (e.g. gender and offence class for the first model and some of the self-reported drug use in the second model). These were also found to be insignificant in the Lu *et al*. study. The model was estimated for cannabis, cocaine, opiates and a composite category of 'any hard drug'. The results for the false-negatives model are reported in Table 5 and those for false-positives in Table 6.

Table 5. Logit coefficients for the occurrence of a positive drug test among those who report no drug use within last three days

Covariate	Cannabis	Cocaine & Crack	Opiates	Any hard drug
Intercept	-25.130 ***	-42.422 ***	-16.367 *	-28.308 ***
Self-report use in last 30 days	0.117 ***	1.203 ***	0.056 **	0.804 ***
Self-report use in last year	0.910 ***	0.153	1.552 ***	0.657 ***
Self-report use ever	0.209	0.589 **	0.985 ***	0.469 **
Suspected of drug offence	-0.025	0.487 **	-0.133	0.214
ln(age)	14.486 ***	23.886 ***	7.457	15.178 ***
$\ln(age)^2$	-2.287 ***	-3.594 ***	-1.023	-2.209 ***
N	1519	2384	2142	1983
Goodness of fit $\chi^2(8)$	13.4	8.5	9.0	3.7
Mean predicted probability	0.184	0.093	0.088	0.133
Range of predicted probabilities	0.006 – 0.920	0.002 – 0.426	0.027 – 0.800	0.020 – 0.476

Note: *** = $p \leq 0.01$; ** = $p \leq 0.05$; * = $p \leq 0.10$

Table 6. Logit coefficients for the occurrence of a negative drug test among those who report drug use within last three days

Covariate	Cannabis	Cocaine & Crack	Opiates	Any hard drug
Intercept	-1.160	3.860 **	0.913	4.231
At least 24 offences in last year	0.611 ***	-0.183	0.519	-0.394
Self-report use in last 30 days	-0.051 ***	-	-0.098 ***	-
ln(age)	0.125	-1.653 ***	-0.777	-2.175 ***
Female	0.168	-1.249 ***	0.097	-0.862
N	1339	473	716	874
Goodness of fit $\chi^2(8)$	11.7	4.2	3.7	10.3
Mean predicted probability	0.186	0.144	0.031	0.043
Range of predicted probabilities	0.088 – 0.511	0.020 – 0.305	0.006 – 0.278	0.005 – 0.127

Considering first the results in Table 5, it is clear that self-reported past drug use does provide useful predictors of a positive drug test for those claiming no drug use in the past three days. In all cases the estimated effects of declared drug use in the last month, year or ever are positive and mostly statistically significant. In other words, if an arrestee claims not to have used a drug in the past three days, but reports some past drug use, then that individual has a much higher probability of generating a false negative observation. Age has a significant nonlinear impact, with the probability of a false negative rising up to a critical age and then falling thereafter. The peak age is at 24 years for cannabis but 31 years for the combined hard drug category (cocaine plus opiates).

For individuals testing negative but claiming the use of the drug in the past three days, the results reported in Table 6 do not suggest a clear pattern across drug types. For example, whereas an individual who has carried out at least 24 offences in the past year is more likely to falsely report cannabis in the past three days, this is not the case for cocaine and opiates. For cocaine, females are less likely than males to falsely report use in the past three days, whereas individuals who report drug use in the past 30 days are less likely to falsely report opiate use in the past three days, a result that also holds for cannabis. There is some evidence of a declining probability of a false positive with age and also a smaller probability for women. Although this corresponds with the common perception of young males as the group most prone to bragging about their illicit exploits, the evidence is far from consistent.

Using the estimated coefficients from the model, we can calculate the predicted probability of misreporting for each individual, conditional on the explanatory covariates. Tables 5 and 6 report the range of these predicted probabilities for all the individuals in the data. Looking first at the probability of a false negative (a positive test result with no self-reported drug use in the past three days), we see from Table 5 that the range of predicted probabilities is wide, particularly for cannabis, where the probability of conflict between test and self-report varies from 0.006 to near-certainty (0.920), depending on the individual's characteristics. The range is narrower for cocaine, but still substantial. The spread of predicted probabilities is generally less in Table 6 for false positives. Again cannabis shows the widest range. These large variations in individual probabilities of misreporting suggest that there is considerable scope for using selective drug-testing in surveys like NEW-ADAM, where the cost of drug testing is a major constraint on survey design.

4. Drug testing strategies

In this section we consider ways in which selective drug testing can be implemented to give acceptably accurate measurement of drug use by arrestees at lower cost than a universal testing policy. The question here is whether it is possible to devise a 'smart' version of Computer Assisted Personal Interviewing, in which the programme determines whether or not a sample should be taken for drug testing based on the arrestee's previous responses to questions on drug use history and other relevant characteristics.

4.1. Full and partial testing strategies

We now consider the accuracy, relative to the NEW-ADAM urine test results, that would have been yielded by alternative partial testing strategies. In doing this, we concentrate on the problem of false negative responses, since false positives are relatively infrequent and less clearly related to observable attributes of the individual.

A partial testing strategy results in the following composite indicator of recent drug use:

$$D_i = \begin{cases} 1 & \text{if not selected for drug test, but self - reported drug use within last 3 days} \\ 1 & \text{if selected for drug test and the test result is positive} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

A major focus of surveys like NEW-ADAM is the measurement of the volume of crime committed by problem drug users. To examine this, we consider the measurement of the prevalence amongst arrestees of drug-using repeat offenders, identified by the following qualitative variable:

$$R_i = \begin{cases} D_i & \text{if the respondent claims to have committed at least 24 offences} \\ & \text{in the previous year} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The prevalence of drug use and drug-related crime is then measured by the sample means of the binary variables D_i and R_i.

We evaluate five different strategies based on the following alternative criteria for selection for drug testing:

- No testing
- Selected for test if claims no drug use in last 3 days
- Selected for test if claims no drug use in last 3 days but some in the last month
- Selected for test if claims no drug use in last 3 days but some in the last year
- Selected for test if claims no drug use in last 3 days but some use ever.

The analysis is done for a single "hard" drug category consisting of cocaine (including crack) and opiates combined and the results are summarised in Table 7. This shows clearly the

trade-off between the number of cases that need to be tested and the number of false negatives generated by the process. If all arrestees who report no drug use in the last three days are tested then this has the biggest impact on the number that need to be tested (69% of the sample), but reduces the number of false-negatives in the composite process to zero. As we tighten the selection criterion for testing by including only a subset of those reporting no hard drug use in the last three days, the number of required tests increases slowly, from 8% of the sample for the 'last month' criterion to 27% for the 'ever' criterion. At the same time, the total number of false negatives is reduced by more than half (from 178 to 88).

The discrepancy in measured prevalence (relative to universal testing) is also shown in Table 7. If no-one is tested then we underestimate prevalence by 20.5% due to false negative self-reporting of drug use in the last three days. This discrepancy is almost halved by testing the 8% of the sample who report no hard drug use in the last three days but some use in the last month. By testing all those who admit to some past hard drug use, the discrepancy can be reduced to only 4.7%. For the corresponding measure of drug-related crime (the mean of R_i) the under-estimation is reduced from the low figure of 4.4% without drug testing to zero by testing all self-declared 'last month' users. More extensive testing results in apparent over-adjustment since false positives are not corrected by this strategy. Note that, unlike drug use, there is no external check available for self-reported offending rates, so the apparently accurate estimate of the prevalence of drug-related crime may still be heavily contaminated by reporting error.

Table 7. The effect of selective drug testing on the measured proportion of recent hard drug users and of hard drug-using repeat offenders (n = 2858)

Testing strategy	Proportion of cases where a drug test is needed	Hard drug users (Sample frequency = 38.5%)			Hard drug-using repeat offenders (Sample frequency = 17.4%)		
		Discrepancy relative to full testing	Number of false negatives	Number of false positives	Discrepancy relative to full testing	Number of false negatives	Number of false positives
No testing	0%	-20.5%	263	38	-4.4%	40	18
Test all who self-report no use in last 3 days	69%	+3.1%	0	38	+3.6%	0	18
Test all who self-report no use in last 3 days but some use in last month	8%	-12.8%	178	38	0.0%	18	18
Test all who self-report no use in last 3 days but some use in last year	18%	-7.6%	120	38	+2.0%	8	18
Test all who self-report no use in last 3 days but some use ever	27%	-4.7%	88	38	+2.6%	5	18

4.2. A model-based testing strategy

The previous analysis does not make use of any information we have about the individual arrestees, apart from their self-reported drug use. An alternative strategy is to use our predictions of the probability of a false negative self-report as a basis for testing. For each individual who claims no drug use within the last 3 days, we use the predicted logit probability (see Table 5) of a positive test result conditional on personal characteristics and self-reported drug history as an indicator of the need to carry out a screening test. To implement this, we need to choose a cut-off level for the predicted probability, below which no test will be done. Thus the composite indicator of drug use for individual *i* is now:

$$D_i = \begin{cases} 1 & \text{if self - reported use in last 3 days} \\ 1 & \text{if no self - report for last 3 days and } \hat{P}_i > C \text{ and test result is positive)} \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

where $\hat{P}_i$ is the predicted probability of a positive drug test and C is the selected cut-off.

A higher cut-off will mean fewer tests, lower survey costs but a greater risk of under-estimation of prevalence. Table 8 summarises the effect of using this selective strategy for alternative choices of the cut-off C. The figures quoted are (i) the proportion of cases where a test is required; (ii) the proportionate discrepancy between the estimate of prevalence produced by this approach and the result for a universal testing strategy; (iii) the sample number of false negatives ($D_i = 0$ and positive test result); and (iv) the sample number of false positives ($D_i = 1$ and negative test result). This is carried out as a measure of drug prevalence alone (with D_i defined by (3)) and as a measure of drug-using crime, with R_i defined by (2).

Table 8. The effect of selective drug testing on the measured proportion of recent hard drug users and of hard drug-using repeat offenders (n = 2858)

Cut-off for predicted probability	Proportion of cases where a drug test is needed	Hard drug users (Sample frequency = 38.5%)			Hard drug-using repeat offenders (Sample frequency = 17.4%)		
		Discrep-ancy relative to full testing	Number of false negatives	Number of false positives	Discrep-ancy relative to full testing	Number of false negatives	Number of false positives
0.1	27%	-5.0%	87	38	+2.4%	6	18
0.2	14%	-9.4%	141	38	+0.4%	16	18
0.3	7%	-13.3%	184	38	-0.6%	21	18
0.4	3%	-17%	225	38	-3.0%	33	18
1.0 (no testing)	0%	-20.5%	263	38	-4.4%	40	18

The choice of cut-off is clearly important. Without drug testing, prevalence of drug use is under-estimated by 20.5% and the prevalence of drug use and repeat offending by 4.4%. Predicted misreporting probabilities are generally low, so it requires a tight cut-off of C = 0.1 before the degree of under-estimation falls below 5%. Even so, this reduces the required number of bio-assay tests by nearly three-quarters, relative to universal testing, giving very large potential cost savings. For the prevalence of drugs and repeat offending, a cut-off of 0.3 (implying the need to test only 7% of respondents) would suffice to eliminate under-estimation almost completely.

It should be emphasised that this sort of selective testing rests heavily on the reliability and robustness of the statistical model of misreporting. This is not something that can be taken for granted, so if selective testing is to be used within a survey like NEW-ADAM, then there should always be a randomly-selected subset of survey subjects who are subject to universal testing and who can serve as a check on the validity of the selectively-administered tests.

5. Conclusions

In this paper we have used drug-testing information collected through the New England and Wales Arrestee Drug Abuse Monitoring (NEW-ADAM) programme to investigate the validity of self-reported drug use. A review of the technical issues for drug testing suggested that urine and saliva were the most appropriate biological samples to use in the context of surveys of drug use. We examined the rates of dissonance between self-reported drug use within the previous 3 days and the NEW-ADAM urine test results. Although we acknowledge

that the NEW-ADAM data are potentially problematic due to their unrepresentative nature, there were substantial differences for the major drug types, with false negatives (claims of no recent drug use coinciding with a positive test result) predominating. Misreporting was much more serious for cocaine and opiates than for cannabis and respondents appeared more ready to report accurately drug use in the distant than the recent past. This result lends support for the use of a relatively long recall period (e.g. 12 months) when using self-report drug information to monitor trends in drug prevalence. An analysis of the sensitivity of false reporting rates to the calibration of drugs screening tests suggested that that most false reporting is not due to test results being very close to the cut-off levels used in the tests. For example, only a tiny proportion of positive tests were within 10% of the NEW-ADAM cut-off levels, whilst the majority were at least 150% above the cut-off.

We examined the extent to which apparently false reporting of illicit drug use can be predicted, given information about the individual's personal characteristics and history of previous drug use. We found that the risk of a false negative self-report is relatively high when an individual admits to some use in the past (despite claiming no use in the last three days). We considered a number of selective drug-testing strategies in which drug tests are administered to those identified as being at risk of supplying false negative self-reports, using personal characteristics and self-reported information on past drug use as predictors. On this evidence, such schemes appear capable of achieving acceptable accuracy in the measurement of drug use whilst reducing survey costs by eliminating a large proportion of the drug tests that would otherwise be required.

References

Appel, P.W., Hoffman, J.H., Blane, H.T., Frank, B., Oldak, R. and Burke, M. (2001), 'Comparison of Self-Report and Hair Analysis in Detecting Cocaine Use in a Homeless/Transient Sample', *Journal of Psychoactive Drugs*, vol. 33, pp. 47-55.

Baer, J.D. and Booher, J. (1994), 'The Patch: A New Alternative for Drug Testing in the Criminal Justice System', *Federal Probation*, vol. 58., pp. 29-33.

Bennett, T.H. (1998), *Drugs and Crime: the Results of Research on Drug Testing and Interviewing Arrestees*, Home Office Research Study 183, London: Home Office.

Bennett, T.H. (2000), *Drugs and Crime: The Results of the Second Developmental Stage of the NEW-ADAM programme*, Home Office Research Study 205, London: Home Office.

Bird. S.M., Pearson, G. and Strang, J. (2002), 'Rationale and Cost-efficiency Compared for Urine or Saliva Testing and Behavioural Inquiry among UK Offender Populations: Injectors, Arrestees and Prisoners', *Journal of Cancer Epidemiology and Prevention*, vol. 7, pp. 37-47.

Boreham, R., MacDonald, Z., Nicolaas, G., Pudney, S. and Tipping, S. (2003), *NEW-ADAM Feasibility Studies*, Home Office Online Paper, London: Home Office, forthcoming.

Feucht, T.E., Stephens, R.C. and Walker, M.L. (1994), 'Drug Use Among Juvenile Arrestees: A Comparison of Self-Report, Urinalysis and Hair Assay', *Journal of Drug Issues*, vol. 24, pp. 99-116.

Flood-Page, C., Campbell, S., Harrington, V. and Miller, J. (2000), *Youth Crime, Findings from the 1998/99 Youth Lifestyles Survey*, Home Office Research Study 209, London: Home Office.

Goldberger, B.A. and Cone, E.J. (1994), 'Confirmatory Tests for Drugs in the Workplace by Gas Chromatography-Mass Spectrometry', *Journal of Chromatography A*, vol. 674, pp. 73-86.

Holloway, K and Bennett, T. (2003), *The results of the first two years of the NEW-ADAM programme*, Home Office Research Study, London: Home Office, forthcoming.

Kerrigan, S. and Phillips, W.H. (2001), 'Comparison of ELISAs for Opiates, Methamphetamine, Cocaine Metabolite, Benzodiazepines, Phencyclidine, and Cannabinoids in Whole Blood and Urine', *Clinical Chemistry*, vol. 47, pp. 540-547.

Kidwell, D.A., Holland, J.C. and Athanaselis, S. (1998), 'Testing for Drugs of Abuse in Saliva and Sweat', *Journal of Chromatography B*, vol. 713, pp. 111-135.

Kintz, P. (1996), 'Drug Testing in Addicts: A Comparison Between Urine, Sweat, and Hair, *Therapeutic Drug Monitoring*, vol. 18, pp. 450-45.

Leino, A., Saarimies, J., Gronholm, M. and Lillsunde, P. (2001), 'Comparison of Eight Commercial On-Site Screening Devices for Drugs-of-Abuse Testing', *Scandinavian Journal of Clinical & Laboratory Investigation*, vol. 61, pp. 325-332.

Lu, N.T., Taylor, B.G. and Riley, K.J. (2001), 'The Validity of Adult Arrestee Self-Reports of Crack Cocaine Use', *American Journal of Alcohol Abuse*, vol. 27, pp. 399-419.

Magura, S. and Kang, S.Y. (1996), 'Validity of Self-Reported Drug Use in High Risk Populations: A Meta-Analytical review', *Substance Use and Misuse*, vol. 31, pp. 1131-1153.

Makkai, T. (2000), *Drug Use Monitoring in Australia (DUMA): Drug Detection Testing*, Research and Public Policy Series No. 25, Canberra: Australian Institute of Criminology.

McKeganey, N., Connelly, C., Knepil, J., Norrie, J. and Reid, L. (2000), *Interviewing and Drug Testing of Arrestees in Scotland: A Pilot of the Arrestee Drug Abuse Monitoring (ADAM) Methodology*, The Scottish Executive Central Research Unit.

Mieczkowski, T., Newel, R. and Wraight, B. (1998), 'Using Hair Analysis, Urinalysis, and Self-Reports to Estimate Drug Use in a Sample of Detained Juveniles', *Substance Use and Misuse*, vol. 33, pp. 1547-1567.

Murphy, D.A., Durako, S., Muenz, L.R. and Wilson, C.M (2000), 'Marijuana Use Among HIV-Positive and High-Risk Adolescents: A Comparison of Self-Report through Audio Computer-Assisted Self-Administered Interviewing and Urinalysis', *American Journal of Epidemiology*, vol. 152, pp. 805-813.

Navarro, M., Pichini, S., Farre, M., Ortuno, J., Roset, P.N., Segura, J. and de la Torre, R. (2001), 'Usefulness of Saliva for Measurement of 3,4-Methylenedioxymethamphetamine and its Metabolites: Correlation with Plasma Drug Concentrations and Effect of Salivary pH', *Clinical Chemistry*, vol. 47, pp. 1788-1795.

Osselton, M.D., Cox, A.R., Reddick, S.R. and Robinson, S.M. (2000), *Drugs in oral Fluid: An Evaluation of Point of Contact Drug Testing*, London: Forensic Science Service.

Pichin, S., Navarro, M., Farre, M., Ortuno, J., Roset, P.N., Pacifici, R., Zuccaro, P., Segura, J. and de la Torre, R. (2002), 'On-Site Testing of 3,4-Methylenedioxymethamphetamine (Ecstasy) in Saliva with Drugwipe and Drugread: A Controlled Study in Recreational Users', *Clinical Chemistry*, vol. 48, pp. 174-176.

Pudney, S., Goulden, C., Sharp, C. and Williams, T. (2003), *A Review of NEW-ADAM and Proposals for Future Developments*, Home Office Online Paper, London: Home Office, forthcoming.

Ramseier, A., Siethoff, C., Caslavska, J. and Thornmann, W. (2000), 'Confirmation Testing of Amphetamines and Designer Drugs in Human Urine by Capillary Electrophoresis-ion Trap Mass Spectrometry', *Electrophoresis*, vol. 21, pp. 380-387.

Sachs, H. and Kintz, P. (1998), 'Testing for Drugs in Hair: Critical Review of Chromatographic Procedures Since 1992', *Journal of Chromatography B*, vol. 713, pp. 147-161.

Skopp, G. and Potsch, L. (1999), 'Perspiration versus Saliva: Basic Aspects Concerning their use in Roadside Drug Testing', *International Journal of Legal Medicine*, vol. 112, pp. 213-321.