Technical Review: Hair Testing: just how accurate is it?

Jason Ditton

Abstract

Extensive forensic examination of the hair of 209 “ecstasy” (MDMA) users demonstrated virtually no correlation between self-reported tablet use, and traces of MDMA in the hair of users. Why should this be so? Three answers are possible, and all true. First, self-report is fallible; second, tablet strength varies enormously; and third, forensic analysis is of unknown accuracy. The first two are well known. Forensic analysis, however, typically presents itself as impeccably precise. The review demonstrates that not only is this claim spectacularly untrue, but also that validation of forensic analysis (and, thus, indirectly, self-report) lies in the very blind intra- and inter-laboratory comparisons that are never undertaken.

Introduction

Although the results of testing hair for the presence of poisons have been accepted - at least in American courts - for over a century (Mieczkowski, 1996: 61), hair testing for evidence of the use of drugs has not yet reached that degree of full judicial acceptability (Huestis, 1996a), although there are various reports of quasi-judicial use (Tagliaro, et al., 1997; Moeller, 1996; Moeller, et al., 1993; and Lewis, et al., 1997).

Hair testing began slowly about 20 years ago, initiated, perhaps, by Baumgartner's pioneering article (Baumgartner, et al., 1978). Sachs (1997: 8-9) suggests an erratic growth thereafter, with a "gold rush" period between 1986 and 1992 typified by relatively uncritical use of hair testing, followed by a "hang over" period between 1992 and 1996 characterised by more critical reflection. It seems that a renewed "gold rush" - at least of published papers if not of conclusive results - began thereafter.

Hair allegedly offers one crucial potential advantage when compared to, for example, urine as a medium for divination - the long time window. Whereas urine can only indicate use for 2-3 days for opiates, cocaine and the amphetamines, hair offers the possibility of retrospective use examination for at least 3 months, and possibly as long as 12 months. Indeed, Baumgartner, et al. (1989:1447-8) claim to have tested 5 hairs, 7.6cm
in length, from Keats, and, 167 years later, discovered the presence of opiates therein. Specifically, hair offers the opportunity to examine evidence of drug use for each of the months separately, rather than merely for all of the months together. A number of other advantages of hair are sometimes cited (difficult to falsify, easy to store, lengthy shelf life, low body invasion), but for others, these apparent benefits merely offer more opportunity to trammel with people's civil liberties for longer periods (Huestis, 1996; Evans, et al., 1994; Kidwell & Blank, 1996).


Substances have been detected in beard hair (Nakahara, et al., 1993; Mangin, 1996), animal hair (Arnold & Puschel, 1981; Harrison, et al., 1974; Kikura & Nakahara, 1995; Kikura, et al., 1997; Nakahara, 1995; Nakahara & Kikura, 1997), fingernails (Cirimele, et al., 1995), toenails (Suzuki, et al., 1984a), sweat (Hossain & Johnson, 1994) and, on more than one occasion, in single human head hair strands (Matsuno, et al., 1990; Smith & Pomposine, 1981; Suzuki, et al., 1984; Takayama, et al., 1997). The latter appears particularly impressive and perhaps mistakenly gives the impression of microscopic endeavour coupled to astounding precision.

**The Case of “Ecstasy” (MDMA)**

Previous work on MDMA (3,4-methylenedioxymethamphetamine) in hair is of some relevance here for reasons that will become apparent shortly (although Mieczkowski, 1996: 60, considers that MDMA is only incorporated in hair to a "moderate degree"), and thus deserves a slightly more detailed review. The case reported in Moeller, et al. (1992) is unhappily almost entirely in Japanese, and interpretation is beyond the current author. Cirimele et al. (1995) discovered AP (amphetamine), MDA (3,4-methylenedioxyamphetamine) and MDMA in both the fingernails and head hair of a 22 year old male "known to be a drug user" (68). Kikura and colleagues analysed hair...
samples from 3 rats that had been given MDMA (amongst other drugs) and also analysed head hair from 7 human "MDMA abusers", in whose hair, MDMA was indeed discovered (Kikura, et al., 1997).

Kintz & Cirimele (1997) report a methodological trial involving interlaboratory comparisons of the analysis of a hair sample containing MDMA, MDA and AP, and the laboratory comparison will be discussed in a later section of this paper. At this point, it is merely worth noting that the sample came from "a male subject, aged 24 years, known to be a stimulant abuser and deceased from fatal methylenedioxyamphetamine overdose" (152).

Röhrich & Kauert (1997) spiked drug-free hair samples with MDA, MDMA and MDEA (3,4-methylenedioxyethylamphetamine) for calibration purposes, and subsequently tested "authentic hair samples from [303] individuals with suspected drug abuse" (180). They discovered that "28 of these samples (9.3%) contained amphetamine and/or methylenedioxy-amphetamine-derivatives" (187). AP was detected in 23 samples, MDMA in 9, MDA in 6 and MDEA in 5. They conclude (ibid.) "Considering the increasing use of amphetamine and particularly of 'Ecstasy', this number of 28 amphetamine-positive samples seems remarkably low."

Finally, Rothe, et al., (1997) examined (proximate 3 cm length) hair samples from "20 regular 'speed' or 'ecstasy' users of the techno scene...who anonymously reported their drug history in a drugs advice centre." Inevitably, while these "abusers" could report the number of tablets consumed, they could not comment on their relative purity (another issue to be taken up later). The 20 were classified as follows: 7 "insignificant abusers" (up to 1 tablet per month); 5 "moderate abusers" (1-3 tablets per month); 3 "strong abusers" (3-8 tablets per month); and 5 "very strong abusers" (over 8 tablets per month). The mean discovered ng/mg (nanograms per microgram) levels matched their rank order by tablets consumed, although there was considerable variation within each group. For example, in the very strong abuser group, one subject delivered no ng/mg positivity, whereas another delivered over 16 ng/mg. Their ability to discover measurable concentrations of MDMA in the hair of subjects who had consumed less than 1 tablet per month persuades the authors of the sensitivity of hair analysis; indeed, they suggest "in cases of ecstasy abuse the intake of 1 tablet may lead to a hair concentration in the order of 1 ng/mg" (126). However, they recognise considerable interindividual variation; and, separately, that participation in the "techno or rave" scene, and associated sweating, can lead to false positives through deposition, and chronological misinterpretation from sweat-assisted migration along the hair shaft.

I have summarised these studies at some length as they provide a background to some hair testing of "ecstasy" users that I conducted with colleagues in the mid 1990s, and which is reported at length in Cooper et al. (2000) and Ditton et al. (2000). The aim of that study was to compare self reported "ecstasy" use with the results of the analysis of hair harvested from the same users. Subjects were recruited by multi-site chain-referral sampling within the 1994-95 “dance scene” in Glasgow. One hundred subjects donated hair after completing a lengthy interviewer-administered questionnaire. Overall gross
concordance between self reported "ecstasy" use, and discovery of MDMA (or related compounds) in analysed hair did not surpass 59 per cent, and no relationship had a Cohen's Kappa of more than 0.08. Within the positive concordant data set (n=52), scatter was considerable, with no correlation being significant, and none more strongly positive than -0.0518. These results indicate that, as far as MDMA is concerned, if judged by self-report, hair does not reach a level of apparent accuracy that would permit its use as a general population estimator.

**Interpretation**

Why should this be so? There are three main possible problems, and some or all of them may have led to the low concordance between self-report and forensic examination. They are: fallibility of self-report, variation in tablet strength, and inaccuracy of forensic analysis.

**Self-report**

These “ecstasy” users were volunteers, who were given credible guarantees of anonymity and confidentiality, and who had no obvious motive for exaggerating or minimizing their prior drug use. Nevertheless, experimentation has shown both that over reporting is more common than underreporting (Fendrich, et al., 1997) and vice versa (Wish, et al., 1997). This is sometimes explained by reference to bravado and/or reticence, although it may be more appropriate instead to consider what experimental and other evidence exists related to the capability of memory even for those with no wish to report levels of drug use higher or lower than those actually experienced.

There is an extensive literature on memory problems, and this has been exquisitely reviewed – recently and with reference to drugs research – by Hammersley (1994). There is no space to summarise his (for believers in self report) gloomy review here, but it is difficult to maintain unalloyed faith in the procedure after reading it.

Nevertheless, complete disregard of self report (the stance adopted, inter alia, by Marques, et al., 1993: 172, who comment: "the widespread use of self report data to validate technical measures of drug exposure is roughly analogous to having a patient self report temperature in order to validate a thermometer reading") probably goes too far.

**Tablet strength**

Independently of these problems of truthful recall and memory difficulty, are problems of substance purity (although the latter are only rarely referred to in the literature on hair analysis). Street purchased MDMA is believed to vary significantly in terms of the proportions of MDMA and other inert components or psychoactive adulterants. Indeed, as part of the general research design of which hair analysis was a component, the research team obtained 15 substances bought as "ecstasy", and these were analysed (as were the hair samples) at Glasgow University's Department of Forensic Medicine and Science. MDMA content ranged from a "trace", to a high of 70 per cent (with a mean of 34 per cent).
Forensic analysis

Turning now to problems associated with forensic analysis, perhaps the main one is an unswerving belief in the validity of clinically generated data, which, when produced by GC/MS (Gas Chromatography / Mass Spectrometry), is often "assumed to be virtually 100 percent accurate" (Harrison & Hughes, 1997: 20). However, forensic analysis of hair for the presence of drugs is not without its own problems. What follows is a digest of recognised problems. Issues relating to laboratory error and/or laboratory contamination of assayed samples are never addressed.

First, there are problems of comparison. Hair preparation procedures adopted prior to testing itself vary widely, and new ones are developed with a regularity unjustified by citation of conclusive evidence of superiority. Subsequent forensic tests selected are from a variety of different options which are known to vary in their ability to detect the presence of various substances. Further, cut-off points, which crucially determine the ratio of false positives to false negatives, are not always reported, and when they are, not always similar.

Next, there are occasionally problems of credibility. It may not have escaped attention that, of the five English language reports of forensic testing for MDMA use, one analysed hair from somebody "known to be a drug user", a second analysed hair posted from overseas where it had been harvested from "MDMA abusers", and a third analysed a sample from somebody "known to be a stimulant abuser". Only the samples used by Röhrich & Kauert (1997) and Rothe, et al., (1997) come anywhere near sizes or sampling techniques that would permit any acceptable generalisations.

Finally, there exist a group of problems which are known to exist, but for which precise parameters have yet to be calibrated. First, rates of hair growth. A norm currently is to assume a growth rate of 1 cm/month (the assumption on which the technical analysis in this paper was based). Not so long ago, however, the American assumption was half an inch/month. Twelve cm is the equivalent of four and three-quarter inches. On the current assumption, this would indicate 12 months of hair growth: on the earlier one, nine and a half months. In addition, rates of hair growth are believed to vary individually from, by one estimate, from 0.6-3.36 cm/month (Harkey, 1993). For an individual at the low end of this scale, 12 cm of hair would reflect 20 months of hair growth; for one at the high end, 12 cm of hair would imply just over three and a half months of hair growth. This is a major, yet typically unaddressed problem.

These latter are only averages: the situation for individuals is more complex as hair growth varies by the length of the individual's hair (with "in general, the longer the hair is, the longer the growing phase is", Harkey & Henderson, 1989: 306); by race (Caucasian hair grows faster than Asian hair); by sex (scalp hair in women grows faster than in men); age (hair growth generally decreases with age); by position on the scalp (hair grows faster in the vertex region than elsewhere); by general location (compared to scalp hair, pubic hair grows more slowly, and beard hair more slowly still); by hair colour (darker hair is more likely to reveal substances or their metabolites); and possibly by hair
texture (whether "coarse" or "fine"). In addition, a certain proportion of head hair - usually estimated at 15 per cent of it - is dormant at any one time, and could not, thus, be any sort of calendrical record of consumption.

Second, perhaps surprisingly, "hair is a very complex part of the anatomy whose biology is only partially understood" (Harkey, 1993: 9). This leads to two more problems. First, rates of incorporation of substances into hair are believed to vary by many of the dimensions that affect rates of hair growth as noted above. Kidwell & Blank (1996: 39) suggest that cocaine incorporation ratios for Africans:Whites are 2.9:1, and for Koreans:Whites, 6.8:1. Second, here is the more immediately relevant difficulty: the problem of the possibility of external contamination of hair (Pötsch & Moeller, 1996). Harkey (ibid.: 16) comments: "the hair shaft is exposed directly to sebaceous secretions before it emerges from the skin. Scalp hair is also exposed to sweat secretions as well as contaminants in air, water, or dust."

Post-harvesting wash procedures are designed to eliminate traces on the external surface of the hair. The extent to which they do so is clearly related to the thoroughness and frequency of washing, with Baumgartner, et al., (1989: 791) indicating that washing has a variable effect with, in one case, removal of 88 per cent of cocaine from a hair sample, and in another, 41 per cent. It is sometimes believed that tracing drug metabolites (rather than the drugs themselves) in hair is indicative of consumption rather than of contamination, but Kidwell & Blank (1993: 140) offer evidence that metabolites have been traced in street samples.

The washing problem is akin to the cut-off problem noted above. Too much washing might lead to too many false negatives: too little, too many false positives. It has been noted, by Blank & Kidwell (1993: 149) that, even after "substantial washing", two cocaine spiked samples still retained 211 ng of cocaine or its metabolites/10mg hair. This is equivalent to amounts found in the hair of those classified as heavy cocaine users. Finally, the degree of external contamination will relate to the mechanism and context of original drug delivery. Drugs which are smoked or chased will, presumably, stand a greater chance of externally contaminating hair than those which are consumed orally or by injection.

However, most drugs can be delivered by a variety of different methods, and might be consumed differently by one individual on different occasions. Tablets, such as those of MDMA (which is also available in powdered - as well as capsule and tablet - form, and has been known to be injected) are generally handled by those not wearing surgical gloves, who may thereafter run their hands through their own, or somebody else's, hair. If MDMA is consumed in hot dance venues, which is typical in the UK, then the problem of sweat contamination is presumably greater than it may be in cooler drug consumption environments. On the other hand, handling (and subsequent hair-fondling) has been challenged by Mieczkowski (1996: 62) who refers to his work with undercover narcotics officers "who are chronically exposed to cocaine", but who, even after "several thousand" exposures, have "near-zero levels of contamination [and] are readily differentiated from cocaine users".
The effects of the differential regularity and frequency of individual non-experimental washing, dyeing, perming and otherwise cosmeticising hair are unknown, but cannot be dismissed. Jurado, et al., (1997) examined the hair of those with similar drug consumption histories, and concluded that "in all cases the drug content in hair that had undergone cosmetic treatment decreased in comparison to untreated hair." (159) They also conclude that shampooing may not have much effect on drug retention, but that bleaching, dyeing and streaking all do: albeit with different individual variations. Skopp, et al., (1997: 47) appear to disagree. Their analysis of bleaching indicates that: "in general, the drug concentrations in virgin hair were slightly lower compared to cosmetically treated hair."

In the introduction to this chapter, reference was made to Sachs' (1997) distinction between a "gold rush" phase for hair analysis, followed by a "hang over" one. One implication is that, in brief, during the "gold rush" era, hair was believed to possess the probable potential to be an absolute "dosimeter" for drug consumption, whereby discovered nanograms of any substance would reliably indicate precise quantities of prior consumption over periods of up to 12 months. The period of retreat typifying the subsequent "hang over" era has forced a stocktaking, and a revision of expectations down to hoping that hair may offer what Baumgartner & Hill (1996: 252) latterly have referred to as a "chronometrically operating relativistic dosimeter": that is, that levels of reported use (for example, light, intermediate, heavy) might correlate broadly but not exactly, with equivalent levels of discovered nanograms.

That even this has not proved possible in the analysis presented above does not necessary mean a further retreat is called for. Instead, improved research protocols might well be more appropriate, and it is to this that the Conclusion attends.

**Conclusion**

Whilst the analysis of the materials presented here offers no obvious source of optimism for the future of hair analysis in the confirmation of MDMA (or related compound) use, the failure to generate impressively positive results may well be due to the failure of self report and/or forensic analysis. How could faith in each be enhanced?

Self report would seem to be more reliable if samples are large, voluntary, and not derived from criminal justice populations. Self report could be more credible if it were prospective rather than retrospective, and limited, perhaps, to a three month period. The issue of the consumption of variously adulterated substances is a difficult one to attend to, but until sampled users can be permitted to return samples of substances along with their self reported consumption details, little impact on this key issue will be possible. An alternative is for the sort of controlled dosage studies that Mieczkowski, et al., (1991) has called for, although these may be harder to organise.
Forensic analysis would, in principle, be surprisingly easy to check for reliability, and if a high degree can be demonstrated to exist, then the need for research into self report itself declines. There would appear to be room for improvement in forensic analysis, and this would be facilitated if considerably heavier hair samples were harvested in future. This would enable double testing of samples: something which should become the norm rather than merely an ideal. Another part of the problem is the proliferation of different washing procedures, assays, etc., on which, Mieczkowski (1991: 68) comments:

There is not yet a universal standard for washing the specimens, conducting the assay or establishing quantitative outcomes which are readily extrapolated to consumption levels. Systematic comparison of hair analysis values will require a standardised assay method, and be greatly facilitated by a standard system of reporting outcomes.

The establishment of a single method for each could be the basis of a sound interlaboratory and intralaboratory research programme. This has been tried, in part, but only with small samples of rather dubious authenticity, or with possibly flawed sample spiking. Welsh, et al., (1993) conducted two exercises with "hair from suspected drug users" and "drug-free hair". In the first exercise, 8 laboratories were sent 8 powdered samples, and 7 laboratories returned quantitative results. Ninety four per cent correctly identified positives, and all correctly identified negatives. However, "scatter in quantitative results was high" (295). In a second exercise, ten laboratories were sent eight hair segments rather than powder. The qualitative results were not as good (78 per cent correctly identified positives, 94 per cent correctly identified negatives), and they comment "quantitatively, the results were more scattered for the second exercise when the hair was in form of short segments." (303)

Mieczkowski, et al., (1997) report the despatch of cocaine positive hair samples from 141 individuals to two laboratories. The two laboratories used different assay methods (one RIAH, the other GC/MS) and different cut-offs. None of those deemed negative by GC/MS were deemed positive by RIAH, but 16 of those deemed positive by GC/MS were deemed negative by RIAH. Finally, Kintz & Cirimele (1997) sent a hair sample spiked with AP, MDA and MDMA to 16 different laboratories. The laboratories all used GC/MS (but used a variety of washing procedures) and were asked to test for AP, MA, MDA and MDMA. Two laboratories detected MA (with which the samples were not spiked), three failed to detect AP, two failed to detect MDA and three failed to detect MDMA. For the remainder, the scatter in the concordant set was: for AP (13 positive results) 3.3-17.5; for MA (2 positive results) 0.8-1.8; for MDA (14 positive results) 2.5-19.5; and for MDMA (14 positive results) 3.3-100.00.

Such interlaboratory testing for MDMA (and related compound) use as has been conducted so far does not inspire confidence, but divergent results may be because of insufficient sample size (Welsh, et al., 1993), different assay methods (Mieczkowski, et al., 1997), different washing procedures (Kintz & Cirimele, 1997), and/or because the amphetamines are notoriously volatile in the testing process itself (Sachs, 1997a). Clearly, if the first can be boosted, the middle two standardised, and the latter overcome,
then an appropriate level of faith in hair analysis might be calibrated. Until then, to reply to Strang, et al., (1993), hair analysis seems to be more “ethical quagmire” than “technological breakthrough”.

References


Kikura, R., Y. Nakahara, T. Mieczkowski and F. Tagliaro (1997) Hair Analysis for Drug Abuse. XV. Disposition of 3,4-methylenedioxymethamphetamine (MDMA) and its related compounds into rat hair and application to hair analysis for MDMA abuse. *Forensic Science International*, 84: 165-177.


