Presumptive drug testing – the importance of considering prior probabilities

Peter D. Maskell
Graham Jackson

Presumptive drug testing – the importance of considering prior probabilities

Peter D Maskell[1*], Graham Jackson[1,2]

[1]School of Applied Science, Abertay University, Dundee. Scotland. UK

[2]Advance Forensic Science, St. Andrews, Scotland, UK

*Author to whom correspondence should be addressed

Abstract

Presumptive drug testing is commonly used in both the clinical and forensic fields to allow rapid identification of the presence and/or usage of drugs. Because the tests generally have a high sensitivity and specificity (often >90%), then a positive test result may be taken to mean there is a high probability that a targeted drug is present. This assumption is, however, incorrect. This paper demonstrates how, in order to assess the positive predictive value (PPV) of a test, it is necessary to take into account, along with the sensitivity and specificity of the test, the prevalence of the drug in the population being investigated. We demonstrate how an alternative, Bayesian approach to assessing the posterior probability of a drug being present mimics the conventional calculation of PPVs but, because a Bayesian approach requires case-specific prior probabilities, the posterior probabilities are more meaningful than PPV in any one specific case. The effectiveness of presumptive test results in cases such as drink-spiking, drug-driving, testing of drugs during seizures and the confirmation of initial presumptive test results is explored.

In order to exploit the potential of presumptive drug testing, it is important that the prevalence of the targeted drugs in relevant populations is understood but, more importantly, it is important to consider using a Bayesian approach in order to tailor results to the specific individual or drug batch being tested.

Introduction

Presumptive/screening tests are used to provide information to facilitate decisions about any further analyses that may be required. These preliminary tests are usually carried out as they are quicker and cheaper than more advanced laboratory tests or because the testing needs to be carried out in locations (such as a crime scene or at the bedside) which are usually not as easily accessible to more advanced (and usually less portable) analytical equipment. Presumptive screening tests are usually either a) a simple chemical reaction, where the targeted substance reacts with a chemical to give a certain colour (Harper, Powell, & Pijl, 2017) or b) an immunoassay, where an antibody specific to the substance or chemical group/ moiety binds to give a colour result on binding (Harper et al., 2017). More recently, Raman spectroscopy and electrochemical sensors have been developed for presumptive drug testing that allow the determination of the presence of a wider range of drugs but are more expensive than colour and immunoassay testing (Florea, de Jong, & De Wael, 2018; West & Went, 2011). The ideal aim in drug screening would be to identify correctly all of the cases in which the substance being targeted is present and not incorrectly give a result when the targeted substance is not present. There are four possible outcomes to a screening test: 1) a true positive (TP) where an individual has taken the targeted drug and the test indicates a positive; 2) a true negative (TN) where an individual has not taken the targeted drug and the test give a negative result; 3) a false positive (FP) where an individual has not taken the targeted drug but the test returns a positive result; 4) a false negative (FN) where an individual has taken the targeted drug but the test gives a negative result.

The conventional approach in toxicology to assess the performance of a test has been to calculate two metrics: -

a) Sensitivity - the proportion of individuals who were correctly identified by the screening test as having taken the drug (true positives only) among all individuals who have taken the drug (true positives and false negatives)

$$\text{Sensitivity} = TP/(TP+FN) \quad (1)$$

b) Specificity - the proportion of individuals who were correctly identified by the screening test as not having taken the drug (true negatives only) among all those who have not taken the drug (true negatives and false positives)

$$\text{Specificity} = TN/(TN+FP) \quad (2)$$

Modern drug screening methods generally have sensitivities and specificities above 90 – 95%. Intuitively, given these figures, a user of these tests might well assume that, for any one given individual with a positive result, it was ~90 – 95% certain the individual had consumed the drug and only ~5 – 10% probability that they had not. However, in practice this is not true – a frequently overlooked but important factor affecting the prior probability that an individual has (or has not) taken a drug is the prevalence of use of that drug in the population of which the tested individual is a member.

How prevalence is important in determining the effectiveness of a screening test

Let us assume that we have a screening test (Test A) that has a 98.7 % true-positive rate (sensitivity) and a 2.5 % false-positive rate (97.5 % specificity). We wish to assess the effectiveness of this test in assessing if an individual has consumed a drug. For this example, we focus on cocaine use among drivers. The European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) compiled epidemiological data on the

prevalence of drug abuse in drivers in Europe (EMCDDA, 2014). From this study, is it possible to estimate the mean prevalence of cocaine in the general driving population as 0.42 %. This statistic, along with the metrics of sensitivity and specificity, allow us to determine the false discovery rate (FDR) (Colquhoun, 2014), i.e. the rate at which, if people test positive, they have not actually consumed the drug, and its complement, the positive predictive value (PPV), i.e. the rate at which, if people test positive, they have actually consumed the drug. A simple way to illustrate these two characteristics of the test is through the use of a tree diagram, as shown in Figure 1. It shows that, if a driver is taken at random from the European population and tests positive using this screening test, they have only a 14% probability (positive predictive value (PPV)) of actually having consumed cocaine, very different from the intuitive assumption that, given a person had tested positive, then there would be a high probability (~90 – 95%) that that person had consumed cocaine.

Bayes Theorem as an alternative model to assess the effectiveness of a presumptive test

Bayes Theorem has been proposed extensively in forensic science and other domains as a logical means of dealing with uncertainty (such as (Institutes, 2015; Linnet, Bossuyt, Moons, & Reitsma, 2012; Robertson, Vignaux, & Berger, 2016)). Bayes Theorem shows how the posterior probability (the probability after the test result is known) of the drug being present in the individual's sample depends not just on the metrics of sensitivity and specificity for the test but also on the prior probability of the drug being present in that individual.

A common depiction of Bayes Theorem in forensic settings is in terms of odds (mathematically related to probability, see equations 3 and 4).

$$Odds = \frac{Probability}{1-Probability} \quad (3)$$

$$Probability = \frac{Odds}{1+Odds} \quad (4)$$

In the case of cocaine consumption, the Theorem may be written as:

$$Odds[C|T_{pos}] = Odds[C] \times LR[T_{pos}] \quad (5)$$

where $Odds[C|T_{pos}]$ is the posterior odds of cocaine being present, given a positive test result; Odds[C] is the prior odds of cocaine being present; and $LR[T_{pos}]$ is the likelihood ratio (LR) provided by the positive test result. The prior odds of cocaine being present in the individual's sample (Odds[C]) should be assigned on the basis of what is known about that person but, in the absence of any specific information on that person that would influence these odds, a pragmatic estimate could be obtained from the prevalence of cocaine use in the general population of the country in question. However, it should be stressed that a probability based on general prevalence may misleadingly inflate or deflate the prior probability against or in favour of the individual. From the published data (EMCDDA, 2014), a prior probability that cocaine would be present in a sample from someone chosen at random in Europe could be assigned a value of 0.42%. The prior odds of C (Odds[C]) would then simply be the ratio of this probability to its complement, i.e. 0.0042/0.9958, or 0.00422.

The likelihood ratio for a positive test result ($LR[T_{pos}]$) is the ratio of the probability of observing that test result given that cocaine was present to the probability of the test

result given that cocaine was not present. These two probabilities can be assigned essentially from the sensitivity and specificity metrics, i.e. 0.987 and 0.025. Tables 1 -3 show how the metrics for sensitivity and specificity map across to the calculation of an LR. The LR for a positive test result is given by dividing the two values of probability in the horizontal line for a positive test result (these two values represent, respectively, specificity and 1 - sensitivity). The LR is then 0.987/0.025, or 39.48.

The posterior odds of a person having consumed cocaine given a positive test result ($Odds[C|T_{pos}]$) is then the product of the prior odds and the LR, i.e 0.00422 x 39.48. This gives a value of 0.166 for the posterior odds.

The posterior odds can be converted to posterior probability using equation 3. Converting odds of 0.166 back to probability gives a posterior probability of cocaine being present as $\frac{0.166}{1+0.166}$ or 0.142 (14.2 %). This means that, even given a positive test result, it is unlikely that cocaine is actually present. Note that this probability has the same value the positive predictive value as given by the conventional calculation using prevalence, sensitivity and specificity detailed earlier.

The LR values shown in Table 3 include not only the LR that is provided by a positive test result (39.48) but also that provided by a negative test result (0.13). While obtaining a negative test result will very likely mean no further action will be taken, it is interesting to see the impact of such a result on the odds of cocaine being present. If we start with the same prior odds of cocaine being present as before (0.00422), then an LR of 0.13 provided by the negative test will give posterior odds of 0.00055 for cocaine being present. This equates to a posterior probability of 0.00055 (rounded) or 0.055%.

The interplay of prevalence and predictive value

Returning to the conventional calculation for the moment, it is possible to demonstrate how prevalence influences the positive predictive value of the cocaine test. Figure 2 illustrates how the positive and negative predictive values for the cocaine test (sensitivity 0.987; specificity 0.975) are altered by the prevalence of the number of people consuming cocaine. These data show that to have PPV of 90% (individuals being correctly identified as having consumed cocaine given a positive test result) will require a prevalence in the population being sampled of ~20% and that the prevalence in the population will need to be 50% for the PPV to equal the sensitivity and specificity of the test.

These two figures for prevalence can again be confirmed using Bayes Theorem. A prevalence of 20% is equivalent to a prior probability of 0.20 (i.e. prior odds of ¼ from equation 4: $\frac{0.20}{1-0.20} = 0.25$). If these prior odds of 0.25 are multiplied by the LR (for the positive test result) of 39.48, the resulting posterior odds are 9.87, giving a posterior probability of 0.91. A prevalence of 50% is equivalent to a prior probability of 0.5 (i.e. prior odds of 1). When multiplied by an LR of 39.48, the posterior odds are 39.48 and the posterior probability 0.975. These two values for the posterior probability are the same as the PPV values calculated earlier using prevalence, sensitivity and specificity.

Using an appropriate prevalence in the population is important to assess the positive predictive value of a test. In the above example, we assumed the individuals being tested were being selected at random. In most legal jurisdictions however, law enforcement will not randomly stop drivers. They will stop individuals on the basis that they believe them to be driving under the influence of drugs (DUID) and this is likely to increase the positive predictive value of the screening tests because the prevalence

in such a group may well be greater than in the general population. This theory is backed up from field studies in which drivers suspected of DUID (cocaine) by law enforcement exhibited a prevalence of cocaine use of between 2 and 14% (Gjerde, Clausen, Andreassen, & Furuhaugen, 2018; Lema-Atán, de Castro, Lendoiro, López-Rivadulla, & Cruz, 2019; Musshoff, Hokamp, Bott, & Madea, 2014; Strano-Rossi et al., 2012; Tang et al., 2018). These suspected DUID drivers showed a prevalence of cocaine use around 4 to 28 times higher than the general population. These data demonstrate that studies should be carried out to assess the prevalence of drug usage in the targeted population in order to give a better-informed view of the positive predictive value of any test.

In Bayesian terms, the probability that the individual being tested had, *a priori*, consumed cocaine should be assigned according to what is known about that individual. That information may increase or decrease the value indicated by prevalence studies. A benefit of Bayes Theorem is in this individual approach to assigning priors that, in turn, result in case-specific, posterior probabilities.

Prior probability and confirmatory testing

It is accepted in the forensic and clinical fields that all presumptive screening tests should be confirmed with a secondary test based on a different chemical principal to increase the confidence that the initial result was not a false positive (Elliott, Stephen, & Paterson, 2018). This is particularly important in the forensic field where an incorrect positive testing result could lead to an individual losing their job, losing custody of their child or being incarcerated. Although it is generally assumed that a so-called confirmatory test would give absolute certainty that a positive presumptive test result

actually means the drug is present, this is not always the case - the degree of certainty (or positive predictive value) of the two tests would again depend on the prevalence and the performance characteristics, in terms of sensitivity and specificity, of the two tests. However, because there will always be some level of correlation between the two tests, likely to be low due to the use of tests with different chemical principal, the PPV of the combination of the two tests will most likely be an overestimation. Ideally, the combined PPV of the two testing methods would be determined by running a study with the tests in series and combining the overall results.

The overall PPV of combined tests could be illustrated using the conventional terminology and calculations that we presented earlier. However, we believe that the alternative approach of conditional probabilities within Bayes Theorem provides a more tractable and transparent means of illustration.

Let us assume that the presumptive test to be used is the one described earlier – we will call this Test A. The probabilities of the observations (the positive and negative test results), and LRs provided by these observations, for Test A are shown in Table 3. Remember that the probabilities reflect the rates of sensitivity and specificity of 98.7% and 97.5% respectively. However, unlike our earlier example, a prior probability of 0.5 (50%) for the drug being present in the tested individual will be used instead of a probability of 0.0042 (0.42%) that was based on prevalence studies.

Using formula (3) for Bayes Theorem, and substituting the values described, we have

$$\frac{0.5}{0.5} \; x \; \frac{0.987}{0.025} = 39.48$$

Posterior odds of 39.48 equates to a posterior probability of 0.975 (97.5%) for the drug being present. So, the impact of the positive test result has been to change the probability of drug being present from 0.5 (50%) to 0.975 (97.5%).

Assume now that we wish to use a so-called confirmatory test, Test B. Also assume that, in trials of Test B, not associated with Test A, Test B has been assessed as having a sensitivity of 99.0 % and a specificity of 99.0 %. The probabilities for the test observations, and the LRs derived from them, based on these metrics are shown in Table 4. It might be tempting, if a positive result is obtained with Test B, to follow the same steps as for Test A and multiply prior odds of 0.5/0.5 (a prior probability of 0.5) by the LR value of 99 (Table 4) for a positive observation to give posterior odds of 99 (equivalent to posterior probability of approx. 0.99 (99%)) for the drug being present. However, this would be wrong on two counts. Firstly, the prior odds for Test B are not 0.5/0.5 because the outcome of Test A has changed those odds to 39.48 (probability 97.5%). Therefore, the prior odds for Test B should be 39.48. Secondly, the metrics of sensitivity and specificity for Test B were assessed independently from any results obtained using Test A, i.e. they were not conditioned on the outcome of Test A. Therefore, the probabilities assigned from these metrics will not take into account any dependency between the tests. To assess logically and safely the combined effect of Test A and Test B results, we need to ensure that any dependency between them has been accommodated. Ideally, we need data from studies of the performance of Test B on samples that were positive in Test A. Table 5 illustrates probabilities based on such data which, in this instance, are hypothetical. Note the apparent sensitivity and specificity are now 99.9% and 90% instead of 99% and 99% as in Table 4. Note also the impact on the magnitude of the LRs – for a positive test result, the LR has been reduced from 99 to approximately 10, an order of magnitude lower. In actual practice,

tests A and B will be chosen on the basis that they identify the analyte concerned using different chemical principals and, therefore, the assumption is that the results are not correlated. Our example shows there is a degree of correlation between the two tests that may overestimate that which occurs in practice but we stress the only way to assess the degree of correlation, even when using orthogonal tests, is to conduct controlled experiments to acquire the relevant data to justify the metrics used when evaluating the results of a combination of tests.

We can now calculate revised posterior odds for the combination of positive results in both tests. In this new calculation, the posterior odds resulting from a positive test A become the new prior odds for the result of Test B. These odds are multiplied by the conditional LR for a positive Test B result (approx.10) to give new posterior odds:

$$39.48\ x\ \frac{0.999}{0.1} = 394$$

Posterior odds of 394 equate to a posterior probability of 0.997 (99.7%). So, the outcome of Test B has changed the posterior odds by about an order of magnitude from 39.48 to 394.

Had we used the unconditional LR of 99 for the positive Test B, the posterior odds would have resulted in a misleading hundred-fold increase in the posterior odds from 39.48 to 3,908.

Table 6 illustrates the impact of different values for the prior probability of the drug being present in an individual for Test A in isolation and for the combination of Tests A and B. At a low value of prior probability of 1 %, a secondary confirmatory test will only give 29% certainty of the drug being present, given positive results in both tests.

Increasing values of prior probability will, of course, increase the posterior probability but it is only when the prior probability is greater than 2% does the posterior probability exceed 90% for the combination of both tests. A prior probability of 50% will result in a posterior probability of 99.8% - presumably a value that would satisfy the decision-maker that the drug was indeed present. The overall posterior probability would of course vary with the sensitivity and specificity of the testing methods being used at each stage, but it is important for forensic and clinical practitioners to understand that a confirmatory test may not be as good as it appears, particularly if the prior probability of the drug being present in the individual being tested is low. It is also vital that any conditionality between the outcomes of separate tests run in series is accommodated in the evaluation of the results.

Further comments on low priors - screening of drinks for potential "spiking"

In forensic toxicology, there is a lack of knowledge of the prevalence of drugs in certain groups or populations. As we have shown, this knowledge is essential to inform the assigning of realistic, justifiable, prior probabilities which, in turn, facilitate a better understanding of the relevance of test results. One such area of limited knowledge of prevalence is that of alleged spiking of drinks. Because of the clandestine nature of drug-facilitated sexual assault (DFSA), it is very difficult to determine the incidence of not only drug-facilitated DFSA in individual countries but also of the drugs that have been utilised. Some information on the frequency of use of specific drugs in alleged cases of DFSA is provided by a US study (March 2015- June 2016) of 1000 cases of suspected DFSA (Fiorentin & Logan, 2019). The authors provide data on the frequency of detection of a very wide range of drugs in blood and/or urine samples

from DFSA complainants. We will select just three drugs from this study - γ-hydroxybutyrate (GHB), ketamine and diazepam – to illustrate the principles of the assessment of posterior probabilities of a drink having been spiked. The frequencies of detection of these three drugs are quoted as 5.9%, 0.5%, and 3.7% respectively and are summarised, along with the frequencies of some other DFSA drugs, in Table 7. An earlier study in the UK of 1014 cases of alleged DFSA (Scott-Ham & Burton, 2005), again based on analysis of complainants' blood/urine samples, found that a sedative or disinhibiting drug was detected in 18% of cases but, of these, only 2% were considered, based on other evidence, as involuntary ingestion. There is no breakdown for individual drugs in this category.

For any Bayesian analysis, it is vital that the issue being investigated is clearly identified and specified from the outset. Clarity about the issue leads to clarity about the specific hypotheses (and their negations) that are being evaluated. In drink-spiking cases, one issue is that of whether the complainant had consumed a DFSA drug. However, because we are dealing here with presumptive tests for "drink-spiking", the issue has to be one of whether a drink had been "spiked". The issue of whether a complainant had consumed a DFSA drug would be best addressed through analysis of body fluid or tissue samples from the complainant. Given that the issue to be addressed here is that of whether the drink had been "spiked", then the specific hypothesis would be "The drink had been spiked" and its negation "The drink had not been spiked". The question now is how to assign a prior probability for that hypothesis. Information from studies on the incidence of DFSA drugs in complainants' blood/urine samples could be helpful but that information is at least one step removed from the central issue of whether the drink was spiked. There is very little, if any, information on the occurrence of DFSA drugs actually in the drinks of complainants to help inform

prior probabilities. If absolutely no information were available, then a prior of 50% would be appropriate to reflect maximum uncertainty about whether the drink had been spiked. However, given that there is some information to help, we could assign priors based on data from the US and/or UK studies, according to jurisdiction. However, only the US study breaks down prevalence according to drug type and, accordingly, we will base our prior probabilities on these data. For comparative purposes, we will also use a 50% prior for each drug to illustrate the effect of assuming maximum uncertainty. Values of prior probability are, of course, open to adjustments to reflect the specific conditions of the case.

Having assigned prior probabilities, we now need to assign conditional probabilities of observing positive and negative outcomes of the tests given a drink had or had not been spiked. Information to inform these probabilities can be obtained from independent studies, such as using the sensitivity and specificity of three presumptive "drink spiking" tests - Drinksafe 1.2; Drink Guard and Drink Detective (Beynon, Sumnall, McVeigh, Cole, & Bellis, 2006; Quest & Horsley, 2007). Table 8 shows the sensitivity and specificity data from these sources based on independent testing. Table 9 shows the probabilities assigned from these data and the LRs that result from the probabilities. Overall, the values of LRs for positive observations are not particularly high. The impact of such LRs on the prior probabilities will therefore not be great. The largest LR, of value 8.26, is that for Drink Detective test for ketamine, while the lowest LR is of value approximately 1 for Drink Guard's test for Ketamine. An LR value of 1 means the test result does not help at all in updating the prior probability – basically, the positive test result is uninformative. With regard to negative test results, with one exception, the values of LR range from approximately 1 to 0.21 (1/5). An LR value of one, as just explained, provides no help in updating uncertainty

while an LR of 1/5 (0.21) means that the prior odds (the ratio of the prior probability to its complement) of the drug being present will be multiplied by a factor of 1/5 given a negative test result. The exception for negative results is that for Drink Detective's test for Ketamine, where the LR is of value zero. This value has resulted from the zero value that has been assigned to the probability of obtaining a negative test result if the drug is indeed present. In probabilistic reasoning, a value of zero for an event means that that event is impossible to observe given the condition. However, it is unjustifiable to assert a probability of zero from observations taken from a limited sample, especially with a sample size of only 20. It might be that, if a sufficiently large sample were to be studied, that event will be observed. It is more justified, therefore, to assign a small probability for that event. For the purposes of this paper, however, we will proceed on the basis that the value of zero can be justified.

Equation 5 can now be used to give the posterior probability of a drink having been spiked with each of the three drugs, given positive test observations and given two values of prior odds. Table 10 shows the posterior probabilities given prior probabilities that reflect prevalence of the drugs as shown in Table 7 while Table 11 uses a prior probability that reflects maximum uncertainty.

The posterior probabilities shown in Table 10 range from 0.40 (40%) down to 0.005 (0.5%). So, even with the best performing test in terms of the magnitude of the LR, i.e. Drink Detective's test for ketamine (LR=8.26), the posterior probability of the drug being present is only approximately 4%. Drink Safe 1.2's test for GHB (LR=5.95) gives a higher posterior probability of 27% but that is because the prior probability is that much higher for GHB (5.9%) than for diazepam (3.7%). A positive result for Ketamine using the Drink Guard test does not change the prior probability at all because the LR

of value 1 is uninformative – the probability of the drug being present (0.5%) is unchanged by the positive test result.

In general, even if the user of a "drink spiking" presumptive test gets a positive result then, for all of the drug/test kit combinations shown in Table 10, it is unlikely (in some cases, very unlikely) that the drug is truly present in the drink. It must be stressed that these values for posterior probabilities for the drug being present in the drink are predicated not just on the reliability of the performance metrics of sensitivity and specificity but also on the appropriateness of the assigned prior probabilities. If we increase the priors to reflect maximum uncertainty, then the effect on posterior probability can be seen in Table 11. A positive result using the best performing test (in terms of LR magnitude) the Drink Detective diazepam test now provides a posterior probability of 89% for the drug being present in the drink. A weaker-performing test, such as Drink Detective's test for GHB, now provides a posterior probability of 59% for the drug being present. As explained earlier, test results that provide LRs of value 1 are uninformative and therefore do not change the prior probabilities at all.

Aside from apparently ineffective tests, i.e. those providing LRs of 1, the presumptive tests examined in this paper do not provide what may be described as very high posterior probabilities of a drink being spiked unless the prior probability is also relatively high.

However, while the posterior probabilities may not be high, the best advice in the interest of safety, given the prevalence figures from the literature, would still seem to be to discard drinks that give positive test results.

Turning now to the meaning of negative test result, we show in Tables 12 and 13 the posterior probability of a drug being present given that a negative result has been

observed and given two different prior probabilities (as in Tables 10 and 11). With the exception of results that give LRs of value 1 or zero, all posterior probabilities are lower than the prior probabilities when a negative result is obtained. Table 12 shows that, if prior probabilities based on prevalence are used, the posterior probabilities of drug being present are very low, ranging from 5.9% down to 0.26%. For those results that have LR values of 1, the prior probability remains unaltered. For the test result that has an LR of value zero, Drink Detective's Ketamine test, a negative result gives a posterior probability of value zero, meaning the drug is certainly not present in the drink. But, as we have mentioned, this inference is predicated on a probability, and hence an LR, of zero being justified.

Table 13 is equivalent to Table 12 but with a prior probability that reflects maximum uncertainty. As would be expected, the posterior probabilities of a drug being present, given a negative result, are lower than their prior probabilities, with the exception of those tests that have an uninformative LR of 1. Posterior probabilities range from 17% to 39% for those tests that have an informative LR. Those values may not be reassuring to someone using those tests to check whether a drink is clear of a DFSA drug.

Overall, a decision on whether a drink has been spiked depends critically on two factors – 1) the reliability of the performance metrics of the test and 2) the prior probability of the drink having been spiked. The reliability of the performance metrics will depend on how closely the conditions of the performance study reflects real-life situations and on the size of the sample. The reliability of prior probabilities will depend on how closely those probabilities reflect the individual's circumstances. For example, if the drink being tested was that of a young male in a nightclub in Dundee, Scotland, it would be preferable to base the prior probability on data about the prevalence and

identity of drugs used for surreptitious drugging in the drinks of males between the ages of, say, 18 – 25 in nightclubs in Dundee. Often though, data of this kind, tailored to the specific case in hand, is just not available. A pragmatic approach would be to take whatever data there may be on prevalence and adjust those data for what is known about the circumstances of the case in hand. The sensitivity of the choice of priors can be assessed by exploring alternative values for those priors but, in the absolute absence of any expert insight into what may be an appropriate value for the prior, a value of 0.5 for the prior probability, reflecting maximum uncertainty, could be adopted.

Comment on large priors - drug seizures and cost-benefit analysis.

Presumptive testing is often used on a questioned powder or tablet when an initial determination of its identity is required before additional, more expensive testing is carried out. For a substance such as cocaine, the most common presumptive test being used is the Scott colour test. As can been seen in table 14, the sensitivity and specificity of this test are 0.68, 0.75 respectively (de Jong et al., 2018), giving an LR for a positive result of 2.72 (table 15). As shown in the earlier examples, if we wish to assess the posterior probabilities provided by a specific test, we need to assign relevant, reliable prior probabilities (which could be informed by studies on prevalence) in the circumstances being examined. In cases where there is a seizure of a substance, especially in powdered form, the priors could well be higher than 50% (50% reflecting maximum uncertainty). Based on the circumstances of the seizure, on the packaging, its appearance and the place of storage of the powder or substance (for example, a sealed, clear package containing white powder in a toilet cistern), the

probability that the substance being seized was a controlled substance of a specific type would seem to be relatively high. Reliable data to help assign priors in such circumstances do not appear to be available and therefore the expert needs to rely on their own knowledge and experience of drug seizures. For this particular example, we have assigned an illustrative prior probability of 0.8 (80%). The posterior probabilities resulting from positive test results can now be obtained by combining the prior probabilities with the LRs that were assigned from sensitivity and specificity data. Table 16 shows that, with a prior probability of 50% reflecting maximum uncertainty, the posterior probability is 73% when Test 1 is used. This posterior probability, as expected, increases to 92% when the prior probability is raised to 80% (because the prior information within the case circumstances caused the expert to assign a high probability of the seized substance being cocaine). The posterior probabilities of cocaine being present given a negative result, as shown in table 17, are not particularly low. A posterior probability of 30% arises when the prior probability is 50% (maximum uncertainty); this posterior probability increases to 63% with a prior probability of 80%. A positive colour test provides a good probability that the substance is cocaine but, even with a negative result and a prior probability of 80%, there is a still a good probability that the substance is cocaine. Because of this risk of misleading inferences from negative results, the development of presumptive tests has concentrated on improving the sensitivity and specificity (and hence the magnitude of the LRs) of such tests. For example, an electrochemical test for the detection of cocaine has improved the sensitivity and specificity of cocaine detection to 0.93 and 0.86 respectively (de Jong et al., 2018) (Table 14). This leads to an increased LR for positive detection (6.64 for the electrochemical test compared to 2.72 for the colour test) and a decrease in the LR for a negative detection (0.08 for the electrochemical test compared to 0.43 for

the colour test). Tables 16 and 17 show that the electrochemical test has improved performance for both positive and negative results at prior probabilities of 50% and 80%.

If a decision was required on which of the two presumptive tests to use, the choice would seem to be a simple one, based on these data alone. However, an additional consideration is the cost of the presumptive test and the cost of any confirmatory testing (see the earlier discussion of confirmatory testing and the requirement for proof "beyond a reasonable doubt"). If we assume that both of the presumptive tests cost £1 and a confirmatory test costs £100, then we can determine the total costs of the testing regimes for 10,000 cases, assuming a prior probability of the substance being cocaine as 80%. In all cases, it was assumed confirmatory tests were only carried out if the preliminary test had given a positive result. Figure 3 and Table 18 shows that total costs of the colour test in this scenario are £604,000 and the electrochemical test £782,000. This difference in total costs is due to the electrochemical test detecting more positives correctly than the colour test. In this scenario, the use of a specific presumptive test (either the colour test or the electrochemical test) becomes more of a philosophical question that would need to be answered. Is the increase in the predictive power of one presumptive test over another worth the increased long-term cost of using, in our example, the electrochemical test rather than the colour test? If a suspect is equally likely to plead guilty given either presumptive test, then irrespective of that test's effectiveness, the electrochemical test would, of course, be a better test to use. However, if all of the positive presumptive tests went to confirmatory tests, a decision would need to be made of cost versus benefit. Is the better test worth the increased costs that would be associated with it? Of course, the cost of a test should not come into question as justice should not come with a price,

but often it does. For this reason, in many jurisdictions now, even with a guilty plea, the confirmatory test will always be completed.

Conclusions

When considering whether to carry out a particular presumptive test, it is important to take into account the prior probability of the substance being present. That probability should be assigned according to what is known about the circumstances of the case and the person in question. In the absence of anything specific about the case or person, one could default to a general prior for the group of which they are a member. In the complete absence of any information, then an uninformative prior probability of 0.5, reflecting maximum uncertainty, could be adopted. This is not to say that a practitioner should just adopt a 0.5 prior because they are uncertain about an appropriate prior for the case in question. It is very rare that there would be absolutely no knowledge or data to inform the prior. Every attempt should be made to gather data and information that would help assign a case-specific prior to avoid either a misleadingly high, or misleadingly low, resultant posterior probability. It is important for all forensic practitioners to understand the role that prior probability plays in drug screening and confirmation in order to interpret reliably and logically the results of toxicological testing.

Acknowledgments

The authors received no funding for this work and have no conflicts of interest.

References

Beynon, C. M., Sumnall, H. R., McVeigh, J., Cole, J. C., & Bellis, M. A. (2006). The ability of two commercially available quick test kits to detect drug-facilitated sexual assault drugs in beverages. *Addiction (Abingdon, England)*, *101*(10), 1413–1420. https://doi.org/10.1111/j.1360-0443.2006.01420.x

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science*, *1*(3), 140216–140216. https://doi.org/10.1098/rsos.140216

de Jong, M., Florea, A., Eliaerts, J., Van Durme, F., Samyn, N., & De Wael, K. (2018). Tackling Poor Specificity of Cocaine Color Tests by Electrochemical Strategies. *Analytical Chemistry*, *90*(11), 6811–6819. https://doi.org/10.1021/acs.analchem.8b00876

Elliott, S. P., Stephen, D. W. S., & Paterson, S. (2018). The United Kingdom and Ireland association of forensic toxicologists forensic toxicology laboratory guidelines (2018). https://doi.org/10.1016/j.scijus.2018.05.004

EMCDDA. (2014). *Driving Under the Influence of Drugs, Alcohol and Medicines in Europe — findings from the DRUID project*. Retrieved from http://www.emcdda.europa.eu/attachements.cfm/att_192773_EN_TDXA12006ENN.pdf

Fiorentin, T. R., & Logan, B. K. (2019). Toxicological findings in 1000 cases of suspected drug facilitated sexual assault in the United States. *Journal of Forensic and Legal Medicine*, *61*, 56–64. https://doi.org/10.1016/J.JFLM.2018.11.006

Florea, A., de Jong, M., & De Wael, K. (2018). Electrochemical strategies for the detection of forensic drugs. *Current Opinion in Electrochemistry*, *11*, 34–40. https://doi.org/10.1016/J.COELEC.2018.06.014

Gjerde, H., Clausen, G. B., Andreassen, E., & Furuhaugen, H. (2018). Evaluation of Dräger DrugTest 5000 in a Naturalistic Setting. *Journal of Analytical Toxicology*, *42*(4), 248–254. https://doi.org/10.1093/jat/bky003

Harper, L., Powell, J., & Pijl, E. M. (2017). An overview of forensic drug testing

methods and their suitability for harm reduction point-of-care services. *Harm Reduction Journal*, *14*(1), 52. https://doi.org/10.1186/s12954-017-0179-5

Institutes, E. N. of F. S. (2015). *ENFSI guIdElINE For EvaluatIvE Strengthening the Evaluation of Forensic Results across Europe (STEOFRAE)*. Retrieved from http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf

Lema-Atán, J. Á., de Castro, A., Lendoiro, E., López-Rivadulla, M., & Cruz, A. (2019). Toxicological oral fluid results among Spanish drivers testing positive on on-site drug controls from 2013 to 2015. *Drug and Alcohol Dependence*, *195*, 106–113. https://doi.org/10.1016/j.drugalcdep.2018.12.003

Linnet, K., Bossuyt, P. M. M., Moons, K. G. M., & Reitsma, J. B. R. (2012). Quantifying the accuracy of a diagnostic test or marker. *Clinical Chemistry*, *58*(9), 1292–1301. https://doi.org/10.1373/clinchem.2012.182543

Musshoff, F., Hokamp, E. G., Bott, U., & Madea, B. (2014). Performance evaluation of on-site oral fluid drug screening devices in normal police procedure in Germany. *Forensic Science International*, *238*, 120–124. https://doi.org/10.1016/j.forsciint.2014.02.005

Quest, D. W., & Horsley, J. (2007). Field-test of a date-rape drug detection device. *Journal of Analytical Toxicology*, *31*(6), 354–357. https://doi.org/10.1093/jat/31.6.354

Robertson, B., Vignaux, G. A., & Berger, C. E. H. (2016). *Interpreting Evidence: Evaluating Forensic Science in the Courtroom: Second Edition*. *Interpreting Evidence: Evaluating Forensic Science in the Courtroom: Second Edition*. Wiley-Blackwell. https://doi.org/10.1002/9781118492475

Scott-Ham, M., & Burton, F. C. (2005). Toxicological findings in cases of alleged drug-facilitated sexual assault in the United Kingdom over a 3-year period. *Journal of Clinical Forensic Medicine*, *12*(4), 175–186. https://doi.org/10.1016/j.jcfm.2005.03.009

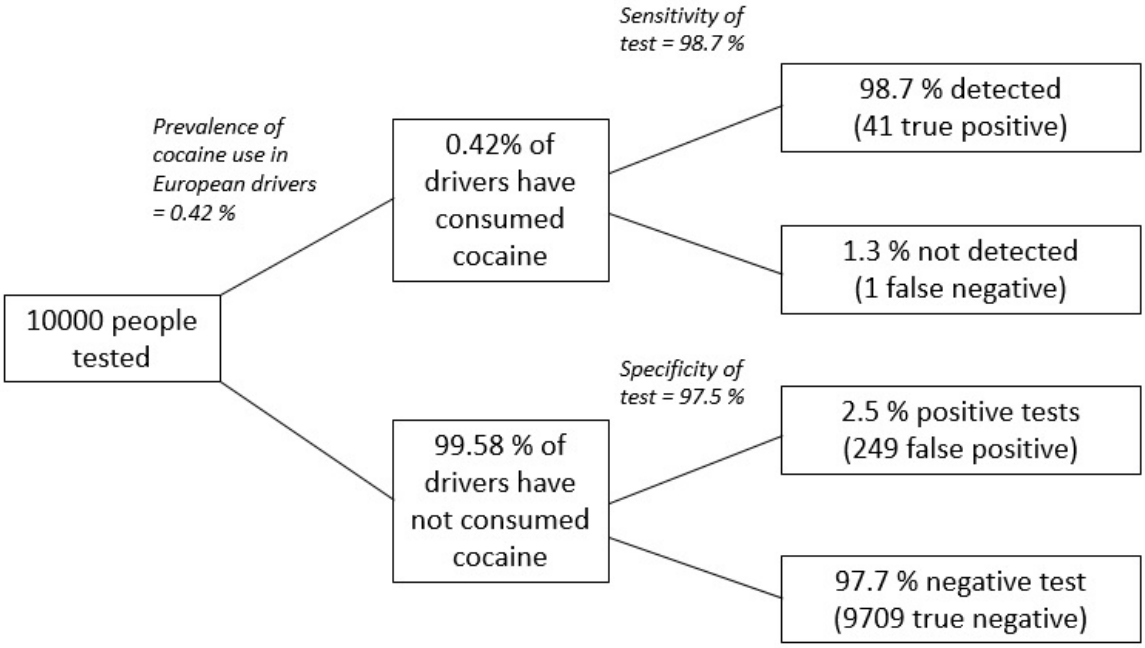
Strano-Rossi, S., Castrignanò, E., Anzillotti, L., Serpelloni, G., Mollica, R., Tagliaro, F., … Chiarotti, M. (2012). Evaluation of four oral fluid devices (DDS®, Drugtest 5000®, Drugwipe 5+® and RapidSTAT®) for on-site monitoring drugged driving

in comparison with UHPLC-MS/MS analysis. *Forensic Science International*, *221*(1–3), 70–76. https://doi.org/10.1016/j.forsciint.2012.04.003

Tang, M. H. Y., Ching, C. K., Poon, S., Chan, S. S. S., Ng, W. Y., Lam, M., … Mak, T. W. L. (2018). Evaluation of three rapid oral fluid test devices on the screening of multiple drugs of abuse including ketamine. *Forensic Science International*, *286*, 113–120. https://doi.org/10.1016/j.forsciint.2018.03.004

West, M. J., & Went, M. J. (2011). Detection of drugs of abuse by Raman spectroscopy. *Drug Testing and Analysis*, *3*(9), 532–538. https://doi.org/10.1002/dta.217

Figure 1: Tree diagram of the false discovery rate of a cocaine screening test using fingermarks. This example assumes that 10,000 people living in Europe are randomly screened using the cocaine screening test. The prevalence of cocaine abuse in drivers is 0.42 %, the specificity of the test is 97.5 % and the sensitivity is 98.7 %. Out of the 10,000 people screened, 41+249 = 290 positive tests. 249 are false positives so the false discovery rate is 14%.



Total Number of positive tests = 41 + 249 = 290

False discovery rate $\frac{249}{41+249}$ = 0.86 or 86 %

True positive rate is 1- false discovery rate, i.e. 14 %

Figure 2: Graphical representation of the changes in the positive predictive value (PPV; the proportion of individuals with positive test results who are correctly identified as having consumed cocaine) and the negative predictive value (NPV; the proportion of individuals with positive test results who are incorrectly identified as having consumed cocaine). The sensitivity of the test is set at 0.987, the specificity as 0.975.

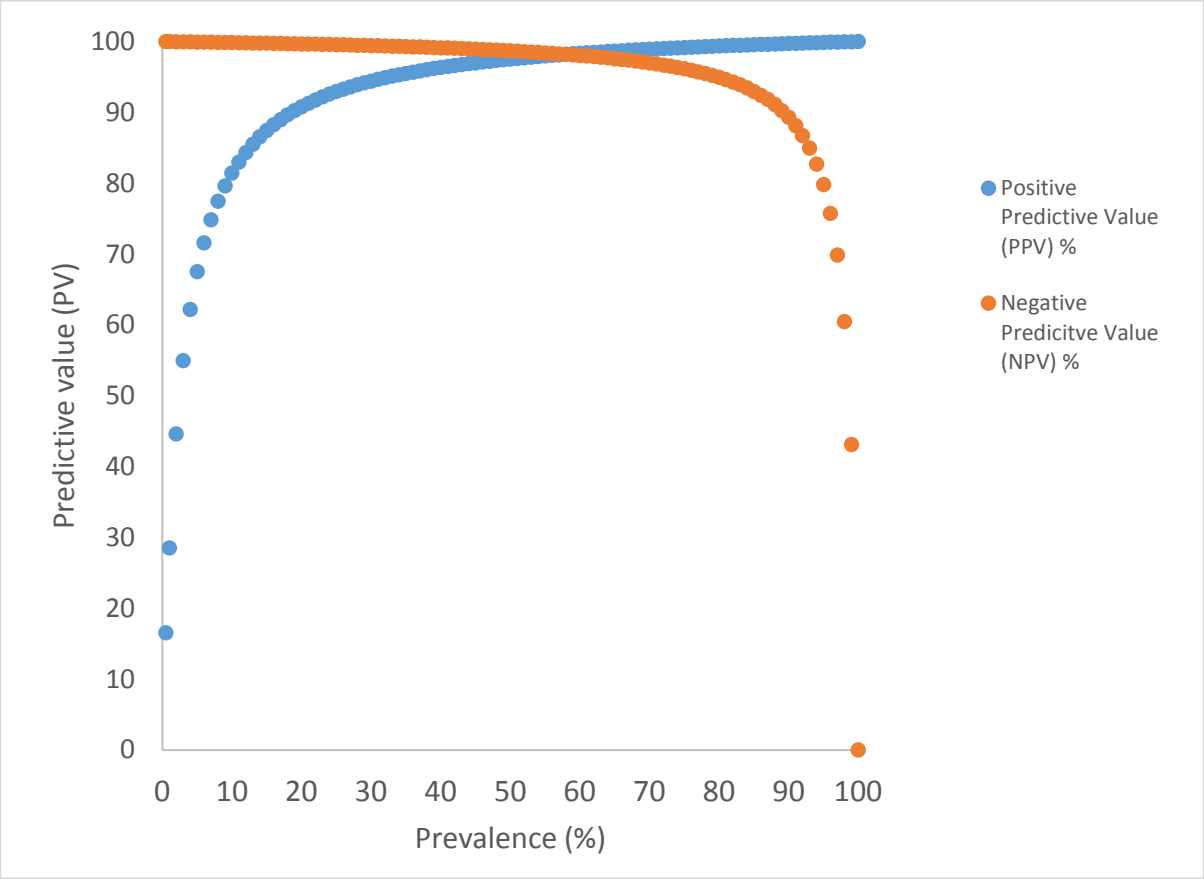
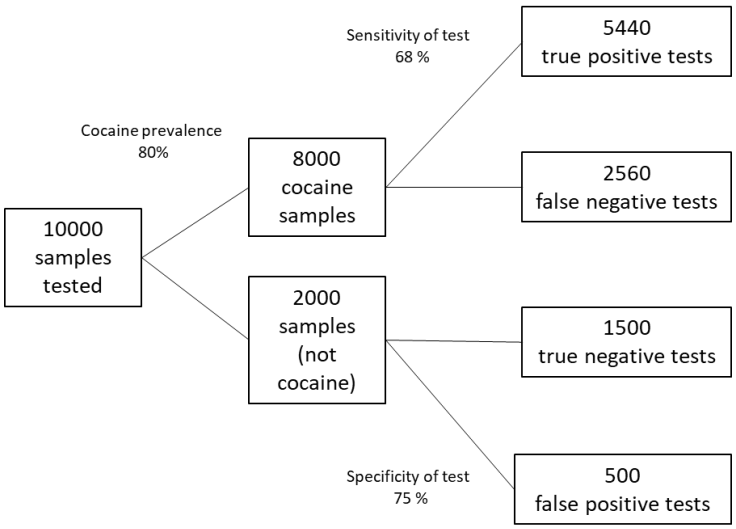


Figure 3: Tree diagram for a) the colour test (sensitivity 0.68 and specificity 0.75) and b) the electrochemical test (sensitivity 0.93 and 0.86) for cocaine. The example assume that 10,000 tests were carried out with 80% of the samples truly being cocaine. Cost of screening test £1, Cost of confirmatory test £100. Each positive test is confirmed.

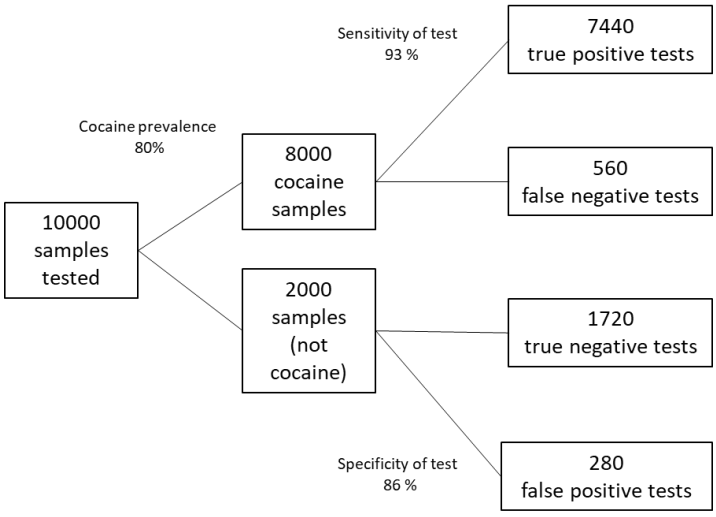
a)



Total number of positive tests = 500+5440 = 5940

Cost of Screening tests = 5940 * £100 = £594,000

b)



Total number of positive tests = 280+7440 = 7720

Cost of Screening tests = 7720 * £100 = £772,000

Table 1. Rates of sensitivity and specificity

Observation	Drug present	Drug not present	
Positive test result	Sensitivity	1 - specificity	
Negative test result	1 - sensitivity	Specificity	

Table 2. Probability of the potential observations using test A, given the conditions of "Drug present' and "Drug not present" and based on sensitivity and specificity data

Observation	Probability of observation, given drug present	Probability of observation, given drug not present	
Positive test result	0.987	0.025	
Negative test result	0.013	0.975	
	1	1	

Table 3. To show likelihood ratios (LRs) provided by the potential observations using test A

Observation	Probability of observation, given drug present	Probability of observation, given drug not present	Likelihood ratio
Positive test result	0.987	0.025	39.48
Negative test result	0.013	0.975	0.013
	1	1	

Table 4. Probability of the potential observations, and their respective likelihood ratios, for test B, based on metrics for sensitivity 99% and specificity of 99% respectively,

Observation	Probability of observation, given drug present	Probability of observation, given drug not present	Likelihood ratio
Positive test result	0.99	0.01	99
Negative test result	0.01	0.99	Approx. 0.01
	1	1	

Table 5. Probability of the potential observations, and their respective LRs, for test B, conditioned on a positive test observation with test A,

Observation	Probability of observation, given drug present and given +ve test A	Probability of observation, given drug not present and given +ve test A	Likelihood ratio
Positive test result	0.999	0.1	Approx. 10
Negative test result	0.001	0.9	Approx. 0.001
	1	1	

Table 6: Effect of various values of prior probability on the posterior probability of the drug being present when using a presumptive test (test A: sensitivity 98.7%, specificity 97.5% LR for a +ve result = 39.48) only and when combined with a confirmatory test (test B: conditional sensitivity 99.9%, conditional specificity 90%, conditional LR for a +ve result = 10).

Prior probability (%)	Posterior probability of drug being present given +ve Test A (%)	Posterior probability of drug being present given +ve Test A and positive test B (%)
0.1	4	29
0.5	17	67
1.0	28	80
2.0	45	89
5.0	68	96
50.0	98	99.8

Table 7: The frequency of detection of a selection of drugs in alleged DFSA cases. Data taken from (Fiorentin & Logan, 2019)

Drug Detected	Percentage of Cases
GHB	5.9
Ketamine	0.5
Diazepam	3.7
Flurazepam	0.2
Temazepam	0.9
Benzodiazepines	20.9
Ethanol	30.9

Table 8: Sensitivity and Specificity of various "drink spiking" presumptive tests for three DFSA drugs

Source	Drug	Sensitivity	Specificity
(Quest & Horsley, 2007)	**Drink Safe 1.2**		
	GHB	**0.5**	**0.916**
	Ketamine	**0.5**	**0.916**
(Beynon et al., 2006)	**Drink Guard**		
	GHB	**0.5**	**0.766**
	Ketamine	**0.25**	**0.766**
(Beynon et al., 2006)	**Drink Detective**		
	Diazepam	**0.813**	**0.879**
	GHB	**0.175**	**0.879**
	Ketamine	**1**	**0.879**

Table 9. Probability of the potential observations, and their respective LRs, based on data in Table 8, for three drugs and three different tests

Test/Drug	Test observation	Drug present	Drug absent	LR
Drink Safe 1.2				
GHB				
	Positive	0.5	0.084	5.95
	Negative	0.5	0.916	0.54
		1	1	
Ketamine				
	Positive	0.5	0.084	5.95
	Negative	0.5	0.916	0.54
		1	1	
Drink Guard				
GHB				
	Positive	0.5	0.234	2.14
	Negative	0.5	0.766	0.65
		1	1	
Ketamine				
	Positive	0.25	0.234	approx. 1
	Negative	0.75	0.766	approx. 1
		1	1	
Drink Detective				
Diazepam				
	Positive	0.813	0.121	6.72
	Negative	0.187	0.879	0.21
		1	1	
GHB				
	Positive	0.175	0.121	1.44
	Negative	0.825	0.879	approx. 1
		1	1	
Ketamine				
	Positive	1	0.121	8.26
	Negative	0	0.879	0
		1	1	

Table 10. Posterior odds and probabilities of the presence of the targeted drug, given a positive test observation and given prior probabilities based on prevalence shown in Table 7

Test/Drug	Prior probability	Prior odds	LR	Posterior odds	Posterior probability
Drink Safe 1.2					
GHB	**0.059**	**0.063**	**5.95**	**0.37**	**0.27**
Ketamine	**0.005**	**0.005**	**5.95**	**0.03**	**0.03**
Drink Guard					
GHB	**0.059**	**0.063**	**2.14**	**0.13**	**0.12**
Ketamine	**0.005**	**0.005**	**≈1**	**0.005**	**0.005**
Drink Detective					
Diazepam	**0.037**	**0.038**	**6.72**	**0.26**	**0.205**
GHB	**0.059**	**0.063**	**1.44**	**0.09**	**0.083**
Ketamine	**0.005**	**0.005**	**8.26**	**0.042**	**0.040**

Table 11. Posterior odds and probabilities of the presence of the targeted drug, given a positive test observation and given a prior probability that reflects maximum uncertainty

Test/Drug	Prior probability	Prior odds	LR	Posterior odds	Posterior probability
Drink Safe 1.2					
GHB	0.5	1	5.95	5.95	0.86
Ketamine	0.5	1	5.95	5.95	0.86
Drink Guard					
GHB	0.5	1	2.14	2.14	0.68
Ketamine	0.5	1	≈1	1	0.50
Drink Detective					
Diazepam	0.5	1	6.72	6.72	0.87
GHB	0.5	1	1.44	1.44	0.59
Ketamine	0.5	1	8.26	8.26	0.89

Table 12. Posterior odds and probabilities of the presence of the targeted drug, given a negative test observation and given prior probabilities based on prevalence shown in Table 7

Test/Drug	Prior probability	Prior odds	LR	Posterior odds	Posterior probability
Drink Safe 1.2					
GHB	**0.059**	**0.063**	**0.54**	**0.034**	**0.0327**
Ketamine	**0.005**	**0.005**	**0.54**	**0.0027**	**0.0026**
Drink Guard					
GHB	**0.059**	**0.063**	**0.65**	**0.0408**	**0.039**
Ketamine	**0.005**	**0.005**	**≈1**	**0.005**	**0.005**
Drink Detective					
Diazepam	**0.037**	**0.038**	**0.21**	**0.0081**	**0.0075**
GHB	**0.059**	**0.063**	**≈1**	**0.063**	**0.059**
Ketamine	**0.005**	**0.005**	**0**	**0**	**0**

Table 13. Posterior odds and probabilities of the presence of the targeted drug, given a negative test observation and given a prior probability reflecting maximum uncertainty

Test/Drug	Prior probability	Prior odds	LR	Posterior odds	Posterior probability
Drink Safe 1.2					
GHB	0.5	1	0.54	0.54	0.35
Ketamine	0.5	1	0.54	0.54	0.35
Drink Guard					
GHB	0.5	1	0.65	0.65	0.39
Ketamine	0.5	1	≈1	1	0.5
Drink Detective					
Diazepam	0.5	1	0.21	0.21	0.17
GHB	0.5	1	≈1	1	0.5
Ketamine	0.5	1	0	0	0

Table 14: Sensitivity and specificity of 2 presumptive tests for cocaine, a colour test and electrochemical test. Data taken from (de Jong et al., 2018)**.**

	Colour Test (test 1)	Electrochemical (Test 2)
Sensitivity	0.68	0.93
Specificity	0.75	0.86

Table 15. Probability of the potential observations, and their respective LRs, based on data in Table 14, for the two presumptive tests for cocaine.

Test/Drug	Test observation	Drug present	Drug absent	LR
Colour Test	**Positive**	**0.68**	**0.25**	**2.72**
	Negative	**0.32**	**0.75**	**0.43**
		1	**1**	
Electrochemical Test	**Positive**	**0.93**	**0.14**	**6.64**
	Negative	**0.07**	**0.86**	**0.08**
		1	**1**	

Table 16: Posterior odds and probabilities of the presence of the targeted drug (cocaine), given a positive test observation and given prior probabilities that reflect 1) maximum uncertainty, i.e. 50%, and 2) an expert judgement, given the circumstances, of 80%.

	Prior probability	Prior odds	LR	Posterior odds	Posterior probability
colour	0.50	1	2.72	2.72	0.73
	0.80	4	2.72	10.88	0.92
electro	0.50	1	6.64	6.64	0.87
	0.80	4	6.64	26.56	0.96

Table 17: Posterior odds and probabilities of the presence of the targeted drug (cocaine), given a negative test observation and given prior probabilities that reflect 1) maximum uncertainty, i.e. 50%, and 2) an expert judgement, given the circumstances, of 80%.

	Prior probability	Prior odds	LR	Posterior odds	Posterior probability
colour	0.50	1	0.43	0.43	0.30
	0.80	4	0.43	1.72	0.63
electro	0.50	1	0.08	0.08	0.07
	0.80	4	0.08	0.32	0.24

Table 18: Predicted total cost of each test (test 1 - colour test; test 2 – electrochemical test) based on 10,000 tests. Sensitivity and specificity are taken from table 14. The prior probability was assigned as 80%.

Test	Cost Per screening test (£)	Cost Per confirm test (£)	Total Screening Cost (£)	Total Confirmation Cost (£)	Total Overall Cost (£)
Colour Test	1	£100	£10,000	£594,000	£604,000
Electrochemical Test	1	£100	£10,000	£772,000	£782,000