Sally A. Coulson, ${ }^{1}$ Ph.D.; Anne Coxon, ${ }^{1}$ Ph.D.; and John S. Buckleton, ${ }^{1}$ Ph.D.

## How Many Samples from a Drug Seizure Need to Be Analyzed?

REFERENCE: Coulson SA, Coxon A, Buckleton JS. How many samples from a drug seizure need to be analyzed? J Forensic Sci 2001;46(6):1456-1461.


#### Abstract

Recently Aitken (1) introduced an outstanding advance in the approach to decision making regarding drugs sampling. Unfortunately this approach has not, as yet, been widely implemented despite being based on a solid mathematical foundation. In this paper we advocate a Bayesian approach along the lines of that outlined by Aitken but designed to be both easily understood with less mathematical sophistication and implementable using standard EXCEL $^{\circledR}$ software. The emphasis is placed on encouraging the application of this methodology to routine case work by explaining the statistics involved. Minor differences exist between this approach and that of Aitken in both the modeling of the prior probability and in dealing with the discrete nature of the samples. These differences in no way detract from the sound mathematical foundation of the approach.


KEYWORDS: forensic science, drug sampling, Bayesian inference, probability

Drug analysts are regularly faced with the problem of how many samples to analyze when a large number of apparently similar items are submitted. These samples can include bags of powder, tablets, or pieces of impregnated paper (tabs).

Historically, a range of approaches has been taken to determine how many samples should be analyzed and are admirably discussed by Aitken (1). These approaches do not include any knowledge based on visual examination and experience of the homogeneity of the seizure, and as such do not make full use of the available information. It might be thought that such knowledge is not "scientific," but this is a misconception. Although not necessarily numerical, this knowledge is as scientific as any other.

Here we describe a Bayesian approach to this problem, following Aitken. It is hoped that the statistical explanations are understandable to most drug analysts and to the court.

The Bayesian approach differs fundamentally from more classical approaches and is preferred in this context, both because it is more practical (in that it tends to suggest more realistic sample sizes) and because it makes better use of the information available. We will therefore take some time to introduce it. Al-

[^0]though most of the modern forensic literature on interpretation $(2,3)$, especially in the fields of DNA statistics (4) and glass evidence $(5,6)$, utilizes this approach, it may be less familiar to drug analysts. It was first suggested to us, however, by practicing drug analysts who perceived they were analyzing relatively large numbers of seemingly similar samples for little gain (personal communication).

It is worthwhile asking "what is the question we are trying to answer?" The first answer that may come to mind is "How many samples do I need to analyze?" This question is linked to the question "analyze to show what?" and "with what level of certainty?" If we set our requirement at, say, "How many samples out of a seizure of 1000 do I need to analyze to show that they are all LSD with $100 \%$ certainty?" then the answer is "all of them." There is no statistical approach that offers any alternative. However if we rephrase the question to a more realistic one, "How many samples out of a seizure of 1000 do I need to analyze to show that there was a dealing limit present (of say 25 tablets) of LSD with $99 \%$ certainty?" This question is now amenable to statistical analysis.

If some thought is given to this then it is obvious that we seek to make a probability statement about the seizure (the 1000) from the sample. It is unfortunate that most statistical methods offered are better suited to making a statement about the sample from the seizure.

The Bayesian approach offers a coherent and mathematically sound way of making exactly the type of statement we desire.

Implementation of this approach will require an understanding of Bayes' theorem and the hypergeometric distribution. These are treated in standard texts but will be introduced briefly here with a particular emphasis on their application to the question at hand.

## The Hypergeometric Distribution

This distribution is particularly suited to sampling without replacement. This is likely to be exactly the situation in drug sampling.

We assume here that a seizure of size $N$ has been received. The analyst has examined $n$ of these of which $m$ have proved to be positive. Obviously, all the sampled items may have been positive in which case $n=m$. However, in general we consider that there may have been a mixture of both positive and negative samples.

The general equation where both positive and negative samples are found is ${ }^{2}$

$$
\begin{equation*}
\operatorname{Pr}(X=m)=\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} \tag{1}
\end{equation*}
$$

This should be read as: the probability that the number of positives $(M)$ in the sample $(N)$ is $m$ equals:

$$
\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}
$$

Where $\binom{M}{m}$ is the number of ways that $m$ objects may be chosen from $M$ and is written as:

$$
\frac{M!}{m!(M-m)!}
$$

$M$ is the total number of positive samples in the seizure (of size $N$ ).

## Bayes' Theorem

The hypergeometric distribution discussed above describes the probability of $m$ positive objects given that we know $N, M$ and $n$, that is the size of the seizure, the number of positive items in the seizure and the size of the sample. Whilst we will typically know the size of the seizure $(N)$ and we can choose the size of the sample ( $n$ ), we will typically not know the number of positives in the seizure $(M)$. In fact if we knew that then why would we be sampling at all. The hypergeometric distribution therefore answers the question the wrong way around for our problem. We seek to make a statement about $M$ from $n, m$, and $N$. This can be achieved using Bayes' theorem.

In general Bayes' theorem tells us that, for a set of events Hi and some evidence, $E$ :

$$
\operatorname{Pr}(H i \mid E, I)=\frac{\operatorname{Pr}(E \mid H i, I) \operatorname{Pr}(H i \mid I)}{\sum_{i} \operatorname{Pr}(E \mid H i, I) \operatorname{Pr}(H i \mid I)}
$$

where $I$ is background information.
If we have $N$ items in the seizure then there are $N+1$ possible things that may happen. For instance all may be positive. In which case we have an $N: 0$ partition of positive to negative. Then we could have $N-1: 1 \ldots N-M: M \ldots 0: N$. These are all the possible events (values $M$ could take) and we could call them $H_{M}$. Read this as the hypothesis that the number of positives is $M$.

If we write the evidence as $m$, the background information as $N$ and $n$ then:

$$
\operatorname{Pr}\left(H_{M} \mid N, n, m\right)=\frac{\operatorname{Pr}\left(m \mid H_{M}, n, N\right) \operatorname{Pr}\left(H_{M} \mid n, N\right)}{\sum_{i} \operatorname{Pr}\left(m \mid H_{M}, n, N\right) \operatorname{Pr}\left(H_{M} \mid n, N\right)}
$$

It seems reasonable to assume that $\operatorname{Pr}\left(H_{M} \mid n, N\right)=\operatorname{Pr}\left(H_{M} \mid N\right)$. This assumption can be read as: The number of samples we choose will not change the true number of positives in the seizure. Hence,

$$
\begin{equation*}
\operatorname{Pr}\left(H_{M} \mid N, n, m\right)=\frac{\operatorname{Pr}\left(m \mid H_{M}, n, N\right) \operatorname{Pr}\left(H_{M} \mid N\right)}{\sum_{i} \operatorname{Pr}\left(m \mid H_{m}, n, N\right) \operatorname{Pr}\left(H_{M} \mid N\right)} \tag{2}
\end{equation*}
$$

[^1]TABLE 1—The possible hypotheses.
\(\left.\begin{array}{cc}\hline Hypothesis H_{M} <br>

(positive:negative)\end{array}\right]\)| $12: 0$ |
| :---: |
| $11: 1$ |
| $10: 2$ |
| $9: 3$ |
| $8: 4$ |
| $7: 5$ |
| $6: 6$ |
| $5: 7$ |
| $4: 8$ |
| $3: 9$ |
| $2: 10$ |
| $1: 1$ |
| $0: 12$ |

This has the potential to answer any question regarding the unknown number of positives in the seizure ( $M$ ) given knowledge of $N, m$, and $n$. We can use it to set $n$ to a value that gives us the level of probability regarding any statement about $M$ that we require.

## Practical Implementation

In this section we will set out a step by step implementation of Eq 2 that is particularly amenable to programming in EXCEL ${ }^{\circledR}$.

Imagine a simple case where 12 suspected LSD tabs are submitted for examination.

1. The 12 tabs are visually analyzed to assure they all appear to be homogeneous.
2. Prior to any chemical analysis, the analyst estimates the homogeneity of the 12 tabs: i.e., the probability of whether they are all positive or all negative (e.g., $\operatorname{Pr}(\mathrm{hom})=0.99$ ). Many analysts will feel uncomfortable with the subjective nature of this assessment.
3. The analyst also estimates how likely it is that the tabs will be positive if the sample is homogeneous (e.g., $\operatorname{Pr}(\operatorname{pos} \mid$ hom $)=$ 0.8).

These estimates come from the previous experience and knowledge of the analyst.
4. Given the 12 tabs, we have 13 possible scenarios. All 12 tabs are positive, all 12 tabs are negative and combinations of positive and negative tabs (Table 1). These alternative scenarios are the different hypotheses.
5. Assessment of the prior probability.

The prior probability is a term for the probability of the hypothesis before any chemical analysis has taken place. ${ }^{3}$ This is based, in this paper, on the visual examination of the tabs and the experience of the analyst.
The probability that all 12 tabs will be positive $=\operatorname{Pr}($ hom $) \times$ $\operatorname{Pr}(\operatorname{pos} \mid$ hom $)=0.99 \times 0.8=0.792$.
The probability that all 12 tabs will be negative $=\operatorname{Pr}($ hom $) \times$ $[1-\operatorname{Pr}(\operatorname{pos} \mid h o m)]=0.99 \times 0.2=0.19$.
The probability for the two homogeneous options adds to 0.99 . The remaining prior probability $(1-0.99=0.01)$ is assigned equally over the other hypotheses of which there are 11. ( $0.01 / 11=0.000909091$ ) (Table 2). Other distributions may be preferred for other cases, although the exact distribution has little effect.

[^2]TABLE 2-Assigning prior probabilities.

| Hypothesis $H_{M}$ | Prior Probability <br> $\operatorname{Pr}\left(H_{m} \mid N\right)$ |
| :---: | :---: |
| $12: 0$ | 0.792 |
| $11: 1$ | 0.000909091 |
| $10: 2$ | 0.000909091 |
| $9: 3$ | 0.000909091 |
| $8: 4$ | 0.000909091 |
| $7: 5$ | 0.000909091 |
| $6: 6$ | 0.000909091 |
| $5: 7$ | 0.000909091 |
| $4: 8$ | 0.000909091 |
| $3: 9$ | 0.000909091 |
| $2: 10$ | 0.000909091 |
| $1: 1$ | 0.000909091 |
| $0: 12$ | 0.198 |

TABLE 3-Assigning p (samplelhypothesis).

| Hypothesis $H_{M}$ <br> $M: N-M$ | Prior Probability <br> $\operatorname{Pr}\left(H_{M} \mid N\right)$ | $\operatorname{Pr}\left(m \mid H_{M}, n, N\right)$ |
| :---: | :--- | :--- |
| $12: 0$ | 0.792 | 1 |
| $11: 1$ | 0.000909091 | 0.25 |
| $10: 2$ | 0.000909091 | 0.045454545 |
| $9: 3$ | 0.000909091 | 0.004545455 |
| $8: 4$ | 0.000909091 | 0 |
| $7: 5$ | 0.000909091 | 0 |
| $6: 6$ | 0.000909091 | 0 |
| $5: 7$ | 0.000099091 | 0 |
| $4: 8$ | 0.0090909091 | 0 |
| $3: 9$ | 0.000909091 | 0 |
| $2: 10$ | 0.000909091 | 0 |
| $1: 1$ | 0.000909091 | 0 |
| $0: 12$ | 0.198 | 0 |

6. If we used the UN guidelines (as used by this laboratory) a sample size of nine would be analyzed ( 7,8 ). UN guidelines state that for sample sizes between 11 and 27, three-quarters of the samples should be analyzed. The probability of all nine of the samples being positive given each hypothesis needs to be calculated.
For example, if the hypothesis that all 12 tabs are positive is true, then the probability of the nine samples being positive is 1 . Conversely, given the hypothesis that all 12 tabs are negative, there is no way all nine tabs can be positive and the probability is 0 . Similarly, the probabilities for all hypotheses with eight or less tabs being positive must also be 0 (Table 3).
7. The remaining probabilities can be calculated by using the hypergeometric distribution, which has a function programmed into EXCEL ${ }^{\circledR}$. For example, if 11 of the tabs are positive and one is negative, the probability of the first tab sampled being positive is $11 / 12$.
The probability of the second tab sampled being positive $=$ 10/11.
The probability of the third tab sampled being positive $=9 / 10$, etc.
Therefore, the probability of all nine tabs sampled being positive $=11 / 12 \times 10 / 11 \times 9 / 10 \times 8 / 9 \times 7 / 8 \times 6 / 7 \times 5 / 6 \times 4 / 5$ $\times 3 / 4=3 / 12=0.25$

This formula can be expressed as:

$$
\frac{11!(12-9)!}{(11-9)!12!}
$$

This is a simple version of the hypergeometric distribution. In this case, where all the samples tested are positive the general formula can be expressed as:

$$
\operatorname{Pr}(m \mid n, M, N)=\frac{\binom{M}{m}}{\binom{N}{n}}=\frac{\frac{11!}{(11-9)!9!}}{\frac{12!}{9!(12-9)!}}
$$

where

```
\(N=\) the total number of submitted items \((=12)\)
\(M=\) the total number of positive items under the hypothesis
        ( \(=11\) )
    \(n=\) the number of items sampled ( \(=9\) )
    \(m=\) the number of positive samples \((=9)\)
```

This is a simplified version of equation 1 with $m=n$ and hence the term

$$
\binom{N-M}{n-m}=1
$$

The probability of the nine samples being positive under the hypothesis that 10 of the 12 tabs are positive can be calculated using the same approach.
$=10 / 12 \times 9 / 11 \times 8 / 10 \times 7 / 9 \times 6 / 8 \times 5 / 7 \times 4 / 6 \times 3 / 5 \times 2 / 4$
$=0.0454545$
Or using the equation: $=\frac{\frac{10!}{9!(10-9)!}}{\frac{12!}{9!(12-9)!}}$
8. Now we calculate the product of the second and third column and sum this column (Table 4).
9. The figures in column 4 are not true probabilities as they do not add up to 1 . The posterior probability, as defined by Bayes' rule, is gained by dividing each entry in the product column by the sum of the product. This normalizes the product (i.e., sum $=1)$ and takes in to account all the hypotheses (Table 5).
10. The posterior probability shows that the chance of all 12 tabs being positive if nine sampled tabs are positive is 0.999656 .
11. Assume the dealing limit for LSD is ten tabs. The probability that ten or more of the tabs are positive, given the nine sampled tabs were positive, can be determined by summing the probabilities of 12 tabs being positive, 11 tabs being positive, and 10 tabs being positive.

$$
=0.999655766+0.000286862+5.121567 \mathrm{E}-05=0.99995
$$

It is interesting to see how the results change for this example if we use smaller sample sizes (e.g., five tabs and two tabs). Using the same approach as above, the probability of all 12 tabs being positive if five tabs taken as a sample are positive is 0.998663 and the probability of all 12 tabs being positive if two tabs taken as a sample are positive is 0.996188 . It is up to the analyst's discretion to decide the level of probability that they would be satisfied with.

We next present some case examples to illustrate this approach.

TABLE 4-Calculating the product.

| $\begin{aligned} & \text { Hypothesis } H_{M} \\ & M: N-M \end{aligned}$ | Prior Probability $\operatorname{Pr}\left(H_{M} \mid N\right)$ | $\operatorname{Pr}\left(m \mid H_{M}, n, N\right)$ | $\begin{gathered} \text { Product } \\ \operatorname{Pr}\left(m \mid H_{M, n, N)}\right. \\ \operatorname{Pr}\left(H_{M} \mid N\right) \end{gathered}$ | $\begin{gathered} \text { Posterior } \\ \text { Probability } \\ \operatorname{Pr}\left(H_{M} \mid N, n, m\right) \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: |
| 12:0 | 0.792 | 1 | 0.792 |  |
| 11:1 | 0.000909091 | 0.25 | 0.00023 |  |
| 10:2 | 0.000909091 | 0.045454545 | $4.1 \mathrm{E}-05$ |  |
| 9:3 | 0.000909091 | 0.004545455 | 4.1E-06 |  |
| 8:4 | 0.000909091 | 0 | 0 |  |
| 7:5 | 0.000909091 | 0 | 0 |  |
| 6:6 | 0.000909091 | 0 | 0 |  |
| 5:7 | 0.000909091 | 0 | 0 |  |
| 4:8 | 0.000909091 | 0 | 0 |  |
| 3:9 | 0.000909091 | 0 | 0 |  |
| 2:10 | 0.000909091 | 0 | 0 |  |
| 1:11 | 0.000909091 | 0 | 0 |  |
| 0:12 | 0.198 | 0 | 0 |  |
|  |  | Sum | 0.79227 |  |

TABLE 5-Calculation of posterior probabilities.

| $\begin{aligned} & \text { Hypothesis } H_{M} \\ & M: N-M \end{aligned}$ | Prior Probability $\operatorname{Pr}\left(H_{M} \mid N\right)$ | $\operatorname{Pr}\left(m \mid H_{M}, n, N\right)$ | $\begin{gathered} \text { Product } \\ \operatorname{Pr}\left(m \mid H_{M, n}, N\right) \\ \operatorname{Pr}\left(H_{M} \mid N\right) \end{gathered}$ | $\begin{gathered} \text { Posterior } \\ \text { Probability } \\ \operatorname{Pr}\left(H_{M} \mid N, n, m\right) \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: |
| 12:0 | 0.792* | $1 \dagger$ | 0.792 $\ddagger$ | $0.999655766 \S$ |
| 11:1 | $0.000909091{ }^{\text {II }}$ | 0.25 | 0.00023 | 0.000286862 |
| 10:2 | 0.000909091 | 0.045454545 | 4.1E-05 | $5.21567 \mathrm{E}-05$ |
| 9:3 | 0.000909091 | 0.004545455 | 4.1E-06 | $5.21567 \mathrm{E}-06$ |
| 8:4 | 0.000909091 | 0 | 0 | 0 |
| 7:5 | 0.000909091 | 0 | 0 | 0 |
| 6:6 | 0.000909091 | 0 | 0 | 0 |
| 5:7 | 0.000909091 | 0 | 0 | 0 |
| 4:8 | 0.000909091 | 0 | 0 | 0 |
| 3:9 | 0.000909091 | 0 | 0 | 0 |
| 2:10 | 0.000909091 | 0 | 0 | 0 |
| 1:11 | 0.000909091 | 0 | 0 | 0 |
| 0:12 | $0.198^{\text {III }}$ | 0 | 0 | 0 |
|  |  |  | $\sum_{\operatorname{Pr}\left(H_{M} \mid N\right)} \operatorname{Pr}\left(m \mid H_{M}, n, N\right)$ |  |

$*=\operatorname{Pr}(\mathrm{hom}) * \operatorname{Pr}(\mathrm{pos} \mid \mathrm{hom})$
$\dagger=$ HYPGEOMDIST(m.n.M.N)
$\ddagger=$ Column 2* Column 3
$\S=$ Column 4/Sum(Column 4)
$\|=(1-\operatorname{Pr}($ hom $)) /(N-2)$
$\mathbb{I}=\operatorname{Pr}(\mathrm{hom}) * \mid 1-\operatorname{Pr}($ pos $\mid \mathrm{hom}) \mid$

## Case 1

Imagine a case where 100 tabs are received from a seizure. These tabs are visually similar and an estimate is made as to their homogeneity $(\operatorname{Pr}(\mathrm{hom})=0.99)$. The analyst also predicts from their experience the probability that the tabs are positive given that they are homogeneous ( $\operatorname{Pr}($ pos $\mid$ hom $)=0.6)$.

Theoretically, there are 101 outcomes for this case. Using the same methodology as presented above, the probability of all 100 tabs being positive and all 100 tabs being negative can be calculated. The remaining probability is spread over the other 99 outcomes. Table 6 lists some of these prior probabilities.

If a sample of 15 tabs was selected, then the probability of the 15 tabs being positive given each hypothesis is calculated. The posterior probability is then calculated using the method described above (Table 6).

This shows that the chance that all 100 tabs are positive given the 15 tabs sampled are positive is 0.999097 .

It is interesting to see how the final probability changes as the number of tabs sampled varies. Table 7 shows the probability that all 100 tabs are positive given a range of sample sizes. This table also shows that a large decrease in the sample number corresponds to only a small decrease in the posterior probability.

## Case 2

Next, we consider the same scenario as case one, but where only 14 of the 15 tabs sampled give positive results. Table 8 details the relevant statistics for the first 13 hypotheses.

For this type of case, where only 14 of the 15 tabs sampled gave positive results Eq 1 should be used in full.

It is impossible for all 100 tabs to be positive given that one of the sampled tabs was negative. As can be seen from Table 8 the result of one negative tab in the sample drastically changes the posterior probabilities, as one would intuitively expect. For this case there is no single hypothesis that stands out as being the most

TABLE 6-Calculation of posterior probabilities for case one.

| $\begin{aligned} & \text { Hypothesis } H_{M} \\ & M: N-M \end{aligned}$ | Prior Probability $\operatorname{Pr}\left(H_{M} \mid N\right)$ | $\operatorname{Pr}\left(m \mid H_{M}, n, N\right)$ | $\begin{gathered} \text { Product } \\ \operatorname{Pr}\left(m \mid H_{M, n}, n, N\right) \\ \operatorname{Pr}\left(H_{M} \mid N\right) \end{gathered}$ | $\begin{gathered} \text { Posterior } \\ \text { Probability } \\ \operatorname{Pr}\left(H_{M} \mid N, n, m\right) \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: |
| 100:0 | $0.6 \times 0.99=0.594$ | 1 | 0.594 | 0.999097 |
| 99:1 | 0.0001 | 0.8500 | $5.59 \times 10-5$ | 0.000144 |
| 98:2 | 0.0001 | 0.7212 | $7.28 \times 10-5$ | 0.000123 |
| 97:3 | 0.0001 | 0.6108 | $6.17 \times 10-5$ | 0.000104 |
| 96:4 | 0.0001 | 0.5164 | $5.22 \times 10-5$ | 0.000088 |
| 95:5 | 0.0001 | 0.4357 | $4.40 \times 10-5$ | 0.000074 |
|  | 0.0001 |  |  |  |
| $\ldots$ | 0.0001 |  |  |  |
| 3:97 | 0.0001 | 0 | 0 | 0 |
| 2:98 | 0.0001 | 0 | 0 | 0 |
| 1:99 | 0.0001 | 0 | 0 | 0 |
| 0:100 | $0.4 \times 0.99=0.396$ | 0 | 0 | 0 |
|  |  | Sum | 0.594537 |  |

TABLE 7-Variation in posterior probability with changing sample size.

| Number of <br> Tabs Sampled | Posterior Probability <br> for 100:0 |
| :---: | :---: |
| 5 | 0.9973 |
| 10 | 0.9986 |
| 15 | 0.9991 |
| 20 | 0.9994 |
| 25 | 0.9995 |
| 50 | 0.9998 |

likely. However, if a minimum number of positive tabs was to be calculated, the probabilities are increased (Table 9).

## Case 3

The examples discussed so far have involved relatively small drug seizures. Imagine that 1000 tabs, which are visually the same, have been received in one shipment. The analyst predicts the following:

$$
\begin{gathered}
\operatorname{Pr}(\text { hom })=0.95 \\
\operatorname{Pr}(\operatorname{pos} \mid \text { hom })=0.90
\end{gathered}
$$

Table 10 shows the posterior probabilities for all 1000 tabs being positive given a varying number of tabs sampled (where all sampled tabs give positive results). These results illustrate the small number of tabs that need to be analyzed to give high probabilities that all 1000 tabs are positive.

## Computer Programs

The methodology presented above can be easily programmed into an EXCEL ${ }^{\circledR}$ spreadsheet. A visual basic program run through EXCEL $^{\circledR 8}$ has been programmed by Professor B. Hibbert and is available from him. ${ }^{4}$

[^3]TABLE 8-Case 2 posterior probability calculations.

|  | Prior <br> Hypothesis | Prob <br> (sample/ <br> Probability | Pypothesis) |
| :---: | :---: | :---: | :---: | :---: | Product | Posterior |
| :---: |
| Probability |

TABLE 9—Case 2: the probability that a minimum number of tabs are positive given the result that 14 of the 15 sampled tabs are positive.

| Minimum Number of Positive Samples | Posterior Probability |
| :---: | :---: |
| 90 | 0.5508 |
| 80 | 0.8960 |
| 70 | 0.9844 |
| 60 | 0.9985 |
| 50 | 0.9999 |

TABLE 10—Case 3: Variation in posterior probability.

| Number of Tabs Sampled <br> (all give positive result) | Posterior Probability for 1000 <br> Tabs Being Positive Given the <br> Number of Sampled Tabs |
| :---: | :---: |
| 10 | 0.994759 |
| 25 | 0.997810 |
| 50 | 0.998911 |
| 100 | 0.999479 |
| 250 | 0.999825 |
| 500 | 0.999942 |

## Discussion

When drug analysts are initially presented with this methodology they may feel hesitant about assigning the prior probabilities. Each case will be different, however, the analyst should have the confidence to estimate a value. The fact that different analysts may propose different values from their experience is to be expected, as they will be relying on different knowledge and points of view.

Drug analysts in this laboratory were presented with two suspected LSD cases and asked to assign prior probabilities.

The first case comprised of a seizure of 500 suspected LSD tabs bearing the "Man on Bicycle 2000" design, a design frequently encountered by all analysts. Before any analyses, the predictions for $\operatorname{Pr}$ (hom) ranged from 0.95 to 0.99 and for $\operatorname{Pr}($ pos $\mid$ hom $)$ ranged from 0.75 to 0.95 . These values are the prior probabilities.

Even using the most conservative prediction $(\operatorname{Pr}(\mathrm{hom})=0.95$ and $\operatorname{Pr}(\operatorname{pos} \mid$ hom $)=0.75)$, the number of samples needed to be analyzed to indicate all tabs were positive with a probability of 0.995 was 12 . This is a much smaller sample size than the one currently taken by this laboratory ( 50 tabs as determined by UN guidelines (8)). Obviously, the higher values for $\operatorname{Pr}$ (hom) and $\operatorname{Pr}$ (pos | hom) predicted by other analysts would result in smaller sample sizes. The dealing limit for LSD in New Zealand is 25 tabs. Using the same prior probabilities and sampling just one tab, the certainty that at least 25 tabs in the seizure of 500 were positive for LSD is 0.997 .

The second case involved a design not previously seen by all analysts, a geometric line design in blue on one side and yellow on the reverse. The predictions for $\operatorname{Pr}$ (hom) ranged from 0.75 to 0.95 . As would be expected for a less familiar design, the $\operatorname{Pr}$ (pos | hom) was lower and over a greater range than for the previous design. Values for $\operatorname{Pr}$ (pos | hom) ranged from 0.5 to 0.8 .

The most conservative prediction given by an analyst was $\operatorname{Pr}($ hom $)=0.75$ and $\operatorname{Pr}($ pos $\mid$ hom $)=0.5$. To be able to say all 100 tabs in a seizure were positive with a probability of 0.995 using these prior probabilities would require 57 tabs to be sampled. This is a greater amount than recommended by UN guidelines (8).

However, it is suggested that the value for $\operatorname{Pr}$ (hom) for the suspected LSD tabs was a little cautious and all other analysts predicted a value of 0.90 or higher. Table 11 shows the number of tabs that would have to be analyzed to be able to say that all 100 tabs were positive with 0.995 probability and a value of $\operatorname{Pr}(\operatorname{pos} \mid$ hom $)$ $=0.5$ (i.e., it is not known whether the tabs will be positive or negative).

As a new design becomes more familiar to the analysts within a laboratory, the number of tabs required to be sampled will reduce as values assigned to $\operatorname{Pr}(\mathrm{hom})$ and $\operatorname{Pr}$ (pos $\mid$ hom) increase. It will only be for new designs or inexperienced analysts that relatively large sample sizes will be required. Sample sizes could be expected

TABLE 11—Sample size required for differing $\operatorname{Pr}$ (hom) if $\operatorname{Pr}$ (pos|hom) $=0.5$ and $\mathrm{N}=100$.

| $\operatorname{Pr}($ hom $)$ | Sample Size $(n)$ |
| :---: | :---: |
| 0.90 | 31 |
| 0.95 | 17 |
| 0.99 | 3 |

to reduce with increasing frequency of a design and developing experience of the analyst.

Although LSD cases were used in this survey, this method can, of course, be applied to any type of drug case where the items in a seizure are apparently similar.

## Conclusion

The methodology described applies a logical framework to select sample size when presented with a drug seizure consisting of multiple items that are apparently similar. It relies on the experience of the drug analyst and takes into account case-specific information. It is hoped the statistical explanations of this method are understandable to drug analysts.

## Acknowledgment

We are indebted to Richard Pinchin, Graeme Sutherland, and Ian Evett who participated in work in this area in the early 1990s.

## References

1. Aitken CGG. Sampling-How big a sample? J Forensic Sci 1999;44(4): 750-60.
2. Aitken, CGG. Statistics and the evaluation of evidence for forensic scientists. Chichester, UK: John Wiley \& Sons Ltd, 1995.
3. Robertson BW, Vignaux GA. Interpreting evidence: evaluating forensic science in the courtroom. Auckland, NZ: John Wiley \& Sons Ltd, 1995.
4. Evett, IW, Weir, BS. Interpreting DNA evidence. Sunderland, MA: Sinauer, 1998.
5. Curran JM, Hicks TN, Buckleton JS. Forensic interpretation of glass evidence. London: CRC Press, 2000.
6. Curran JM, Triggs CM, Buckleton JS. Sampling in forensic comparison problems. Sci and Jus 1998;38:101-7.
7. United Nations. Recommended methods for testing. New York, 1989.
8. United Nations. Recommended methods for testing Lysergide (LSD). New York, 1989.

Additional information and reprint requests:
Sally Coulson, Ph.D.
Forensic Scientist
ESR
Private Bag 92021
Auckland, New Zealand


[^0]:    ${ }^{1}$ ESR, Private Bag 92021, Auckland, New Zealand.
    Received 11 Nov. 2000; and in revised form 15 Feb. 2001; accepted 15 Feb. 2001.

[^1]:    ${ }^{2}$ HYPGEOMDIST from EXCEL ${ }^{\circledR}$ may be used and appears under the Help function.
    Syntax
    HYPGEOMDIST (sample_s.number_sample.population_s.number_population).
    Sample_s is the number of successes in the sample ( $m$ ).
    Number_sample is the size of the sample ( $n$ ).
    Population_s is the number of successes in the population $(M)$.
    Number_population is the population size ( $N$ ).

[^2]:    ${ }^{3}$ This is present in Eq 2 as the $\operatorname{Pr}\left(H_{M} \mid N\right)$ terms.

[^3]:    ${ }^{4}$ D. Brynn Hibbert, Professor of Analytical Chemistry, School of Chemistry, University of New South Wales, Sydney, NSW 2052 Australia
    Tel: + 61293854713
    Fax: +61293856141
    http://www.chem.unsw.edu.au
    http://www.chem.unsw.edu.au/research/groups/esc/ESCgroup.htm b.hibbert@unsw.edu.au

