TECHNICAL NOTE

Myriam Azoury,[1] *M.Sc.; David Grader-Sageev,*[2] *M.A.; and Shlomit Avraham,*[1] *M.Sc.*

Evaluation of a Sampling Procedure for Heroin Street Doses

REFERENCE: Azoury M, Grader-Sageev D, Avraham S. Evaluation of a sampling procedure for heroin street doses. J Forensic Sci 1998;43(6):1203–1207.

ABSTRACT: New legislation regarding methods of drug sampling was passed in Israel in 1991. According to this law, the qualitative result (i.e., identification of the drug) as well as the estimated weight for the total exhibit, based on random sampling, are applied to the total exhibit and may be accepted as evidence. Since then, it has become standard procedure in the Analytical Chemistry Laboratory to open, weigh and analyze only a part of a larger number of drug street doses while the indictment is based on the estimated weight of the total exhibit. In this study, the routine sampling method for heroin street doses used in the laboratory is described and evaluated. For this purpose, 48 exhibits, including about 1300 street doses of powder which had been sampled and examined in the past, have been collected. The previously unanalyzed street doses of each exhibit were weighed and the true total weight of each exhibit was compared with the original estimated total weight. The relative sampling error of the original estimates is about 5% and these tend to be lower than the true weight by about 0.7%. Additional random sampling was also performed on the 48 exhibits, creating for each exhibit four new samples from the unanalyzed street doses. The additional estimates have been compared with the original estimated and the true total weight. Heroin was detected in all the previously unanalyzed street doses.

KEYWORDS: forensic science, drug sampling, estimated weight

Drug exhibits containing street doses of heroin are commonly analyzed in the Analytical Chemistry Laboratory in the Division of Identification and Forensic Sciences in Israel. Generally, these are wrapped in plastic which is heat sealed, making it time-consuming to open, weigh and analyze the enclosed powder. Therefore, when the exhibit contains a large number of street doses, a practice has been adopted to representatively sample the exhibit and to estimate its weight and its composition, rather than to test every unit (1).

In Israel, new legislation regarding methods of drug sampling was passed in 1991 (2). According to this law, the qualitative results (i.e., the identification of the drug) as well as the estimated weight for the total exhibit based on random sampling are applied to the total exhibit and may be accepted as evidence. The legal meaning of this law is that the burden of proof to contradict the results which are based on sampling is shifted to the defendant. The statistical basis for the new legislation was published by Tsidony and Ravreby (3). Following this, a routine sampling method based on this statistical premise was established (4) and will be described below. The method is based on the use of binomial and hypergeometric probability distributions to determine a lower limit for the proportion of units, in a population which contains a drug, necessary to reach a 95% confidence level.

The new sampling procedure has been applied routinely to the multiple-units exhibits received by the Analytical Chemistry Laboratory since 1991. After several years of using this method, it seemed appropriate to evaluate it. The idea was to compare the results based on the new sampling procedure with the results that would be obtained if all the units were analyzed. For this purpose, 48 exhibits, including about 1300 street doses of previously sampled powder, were collected. The previously unanalyzed street doses were weighed and qualitatively analyzed. The true total weight of each exhibit (i.e., the sum of the true weight of all the doses in this exhibit) was compared to with the original estimated total weight (the original estimate) that was submitted by the expert to the court. In addition, four additional random samples were created from the unanalyzed street doses in each of the 48 exhibits (''the additional samples''). The additional estimates were compared with the original estimates and with the true weight.

The main objectives of this study were:

- To evaluate the precision of the original estimates.
- To compare the achieved precision to the statistical theoretical basis.
- To find out if there is a bias in the original estimates.
- To test the presence of heroin in the previously unexamined units.

Methodology

Routine Sampling Method

The routine sampling procedure, as defined by law, can be performed on a population of units with sufficient similar external characteristics (e.g., size, color). The decision is left to the discretion of the examiner. The sample size (n) from the total population (N) depends on the population size: 5 of a population up to 15 units, 6 between 16 and 50 units, and 7 if the population exceeds 51 units. The sampling procedure is never applied on a population under 5 units, where all the units must be examined.

[1] Deputy head and forensic chemists, Analytical Chemistry Laboratory, Division of Identification and Forensic Science, Israel Police National Headquarters, Jerusalem.

[2] Statistician, Statistical Unit, Planning and Organization Department, Israel Police National Headquarters, Jerusalem.

Received 10 Oct. 1997; and in revised form 11 March 1998; accepted 24 March 1998.

Each unit of the random sample is weighed and analyzed. For easy routine use, the analytical balance has been directly connected to the main computer of the Analytical Chemistry Laboratory. The weight of each unit is entered and the average weight of a drug unit, the standard deviation (S) between the sample units, the estimate and the confidence interval (C.I.) for the total weight at a 95% confidence level are automatically calculated and saved. This C.I. defines the interval which has a 95% probability of containing the true value. A report is also generated.

Acceptance Criterion

As a rule, the sampling results are accepted if half the C.I. is under approximately 15% of the estimate. If the value exceeds 15%, the sampling results are rejected. In this case, the examiner has to choose between two options: weighing the entire population if its size (i.e., total number of units) is relatively small, or increasing the sample size (i.e., the number of units taken for sampling) and once again performing the sampling. As an example, the calculated relative C.I. at 95% confidence level for an estimate of 2.0 g and a corresponding sampling error of 0.1 g for the total weight is $(1.96 \times 0.1)/2.0 = 9.8\%$, which is less than 15%. In this case the estimate is accepted. On the other hand, for an estimate of 3.0 g and a sampling error of 0.3 g, the relative C.I. is $(1.96 \times 0.3)/3.0 = 19.6\%$, which is greater than 15% and, therefore, is rejected.

Exhibits

The exhibits examined in this study consisted of multiple packages of units (street doses containing heroin) which had been sampled and examined in the past in the Analytical Chemistry Laboratory. Expert opinion based on estimated weight and qualitative analysis of a sample were presented to the court. After the cases were closed (i.e., on the completion of all the legal proceedings), the exhibits were recalled for the purpose of this study. Forty-eight exhibits, including about 1300 drug units, were collected between June and November 1996.

In addition, computerized data of 18 exhibits (including a total of 292 drug units) in which the original sample was rejected according to the above routine acceptance criterion were analyzed in this study. The data included the original estimated weight, the total true weight of the exhibit (as finally all the drug units were weighed) and the weight of each unit of the exhibit.

Sampling Methodology in this Study

The unexamined drug units of each exhibit were weighed. The sum was added to the weight of the corresponding original sample so the true total weight was obtained. The standard deviation between the population units was also calculated.

Also, additional random sampling (with no replacement) was performed, creating for each exhibit four new samples from the unanalyzed drug units (a total of 192 additional samples). The sample size was the same as the original sample submitted to the court. In small populations, the last samples were randomly created using weight data from all the samples, including the original sample. The routine sampling method described above was performed on each new sample.

The data, including the original estimate, the four additional estimates and the true weight of the exhibit were merged into one record and analyzed using Focus software.

The same methodology was used on the previously rejected samples, creating in addition to the original sample for random additional samples from the saved data.

Qualitative Analysis

All the previously unexamined units were tested for the presence of heroin using the thin layer chromatography method (dioxane : xylenes : ethanol : amonia-40 : 30 : 5 : 5). The plates were observed under ultraviolet light, sprayed with ninhydrin solution (10% in ethanol) and dried. A blue spot similar to the heroin standard was considered to be a positive result.

Results

Precision of the Estimates

Comparison Between the Estimates and the True Total Weight—The true total weights of the exhibits were compared with their corresponding original and additional estimates. As shown in Table 1, the deviation from the true value of each estimate was calculated relatively to the true value (in %), then presented as an absolute value. For example: for an estimate of 2.15 g and a true value of 2.00 g, the absolute relative deviation would be $|(2.15–2.00)/2.00| = 7.5\%$, which is within the (6 to 10%) deviation range.

Table 1 shows that in general, the precision of the estimates is very good. The precision of the original estimates is quite similar to that of the additional estimates, except for the 16 to 20% deviation range.

Location of the Real Weight within the Calculated C.I. Around the Estimate—As seen in Table 2, about 95% of the true weights are within the 95% C.I. around the additional estimates, while in the original samples, only 88% are within the 95% C.I. This implies that, in general, the routine sampling method matches the theoretical statistical basis. However, the smaller number of original sample estimates that fall within the C.I. may originate from a bias in the sampling method (see below). As seen from Tables 1 and 2, the acceptance criterion adopted in the Analytical Chemistry Laboratory, based on a confidence interval at 95%, is satisfactory.

Differences are observed between ''small exhibits'' (under 29 drug units) and ''large exhibits'' (over 29 units). These definitions are arbitrary and made only for the purpose of this study. In the original sample, the precision of the estimates in ''small exhibits'' is higher than in larger ones. In the additional samples, only small

TABLE 1—*Absolute deviation distribution of estimates of original and additional samples vs. true total weight (in %).*

	Deviation range, %				Total	
Samples	0–5	6–10	11–15	16–20	%	No. of Estimates
Original	70.8	20.8	4.2	4.2	100.0	48
Additional	72.4	20.8	5.7	0.1	100.0	192

TABLE 2—*Distribution of true total weights of original and additional samples according to their location within the calculated confidence intervals (C.I.) of 65 and 95% around estimates (in %).*

Samples	Within 65% C	Within 95% C	>95% C.I.	No. of Estimates
Original	75.0	87.5	12.5	48
Additional	74.5	95.3	4.7	192

differences between small and large exhibits can be detected, with high precision in the large exhibits.

Relative Sampling Errors of the Estimates—The relative sampling error of an estimate is defined as the sampling error divided by the estimate. For example, for an estimate of 2.0 g and a sampling error of 0.1 g, the relative sampling error would be (0.1/2.0) = 5%. The median relative sampling error was calculated after ranking the relative sampling error values calculated as above. In addition, the interquartile range, which is a measure of dispersion, was calculated as follows: the ranked values were grouped into four classes, each of them including 25% of the relative sampling errors; then the interquatile range was defined as the two middles classes, which include 50% of the relative sampling errors (from the lower boundary of the second class to the upper boundary of the third class). Therefore, the 25% lowest values and the 25% highest values were excluded.

Table 3 shows that the relative sampling error of the estimates is around 5% (median), which may be considered a small sampling error. The relative sampling error of 50% of the estimates is between 3 and 7%.

The Bias of the Estimates

Theoretically, several factors involved in the sampling procedure can lead to bias. The random selection of the samples by the examiners from the street doses of the exhibit is done manually. It could well be that this procedure is not totally ''blind'' and is susceptible to bias. The examiner has a ''hidden interest'' to select a sample containing similar-sized units. In this case, the standard deviation between the sample units will tend to be small, decreasing the possibility of sample rejection according to the acceptance criterion. As previously mentioned, the rejection of sampling results leads to serious additional work.

Comparison Between the Estimated Weight and the True Weight—As seen in Table 4, the original estimates tend to be lower by about 0.7% than the true weight. On the other hand, the additional estimates tend to be higher as compared to the true weight by about 0.16%. The interquartile interval length is similar in both sampling procedures and is about 5.6%. The distribution of the deviations around the median is relatively symmetric. On the other hand, the distribution of the original estimates tends to be consistently shifted down. However, the median deviation of the original sample is the lowest of the five estimates, but some relatively high positive deviations were also found within the additional estimates.

The results demonstrate that the original estimates tend to be shifted to lower values compared to their respective true weights. Nevertheless, it cannot be concluded that the only reason is a constant bias and it seems that it can be also partially explained by the weight variation between the samples units. However, even if there is any bias, its median value does not exceeds 1%.

TABLE 3—*Relative sampling errors of estimates in original and additional samples and lower and upper limits of interquartile range (in %).*

Samples	Median Relative Sampling Error	Interquartile Range		
		Lower Limit	Upper Limit	Range
Original	4.5	3.1	6.8	3.7
Additional	4.9	3.2	7.3	4.1

TABLE 4—*Relative deviation of estimates of original and additional samples vs. true total weights and lower and upper limits of interquartile interval (in %).*

Samples	Median Deviation	Interquartile Range			No. of Estimates
		Lower Limit	Upper Limit	Range	
Original	−0.68	−3.82	1.89	5.71	48
Additional	0.16	−2.72	2.80	5.52	192

TABLE 5—*Distribution of original estimates by their rank (1—the lowest, 5—the highest) compared with predicted distribution for the 48 exhibits.*

Rank	% of Original Estimate		Predicted %
1	25.0	72.9	60.0
2	22.9		
3	25.0		
4	14.6	27.1	40.0
5	12.5		
Total 48 exhibit	100.0	100.0	100.0

TABLE 6—*Extreme relative deviations of original estimates from true weight according to their order (in %).*

Extreme Deviation	lst	2nd	3rd	4th
Negatives (lowest decile)	−15.4	−13.2	−10.7	−7.5
Positives (highest decile)	13.4	6.9	6.9	6.3

Distribution of the Ranked Estimates—For each exhibit, the four additional estimates were compared with their original corresponding estimates and all five estimates were then ranked according to their values (from the smaller to the higher weight). The 48 original estimates were distributed by their ranks and the corresponding distribution (in %) is shown in Table 5.

The results in Table 5 show that the % of cases in which the original estimate has the lower value is actually higher than predicted—about 73% instead of 60%. Consequently, one can suspect the possibility of bias in the original estimate.

The use of the Chi-square (4) test does not lend meaning to the finding. The number of exhibits is too small to reveal significant difference (theoretically, the number of exhibits should be increased to about 130, to lead to significant results at a confidence limit of 95%).

Extreme Values Distribution—The extreme values of the negative and the positive deviations around the median of the 48 original samples were compared and are presented in Table 6. The deviations were ranked from the lowest value to the highest. The extreme negative deviations were the four lowest, which are about 10% of the 48 estimates, and are referred to as the lowest decile. The extreme positive deviations were the four highest, referred as the highest decile. The data show clearly that the distribution is not symmetric and tends to negative values. The same finding was

observed in the additional samples, but was less extreme. These results support, once again, the suggestion of bias (lower values) in the original estimate. We assume that this finding may be explained by the use of a non-random sampling procedure (small units) by the examiner.

Standard Deviation (S)

The *S* between the units of an exhibit is an important factor when calculating the sampling error and the confidence interval. Therefore it is important to estimate *S* and the precision of the estimates acquired by the routine sampling procedure. In addition, information about *S* may help to estimate and improve the routine sampling procedure.

Comparison Between Estimated and True S—Comparison between the deviation of the *S* in the original and the additional samples versus the true *S* is presented in Table 7. The *S* between the units of the original sample tends to be lower than the *S* between the units of the additional samples. A lower value of *S* in the original sample causes an overestimated accuracy of the estimates. In this case, the examiner may take a wrong decision and accept the sampling results, even though the accuracy is really lower. These findings support once again the suspicion of bias in the original sample selection.

Relative Standard Deviation—As seen in Table 8, the relative *S* is of the order of 10% for the entire population as well as for the original and additional estimates. This is considered as a small relative *S* and indicates that the units in the exhibits are homogeneous. The limits of the interquartile range and their range are similar. The stability of this parameter is also observed in the four additional samples. It can also be seen that the *S* distribution around the median is not symmetric, and is shifted down in the samples.

TABLE 7—*Comparison between median deviation of* S *in original and additional samples vs. true* S *and lower and upper limits of interquartile range (in %).*

Samples	Median Deviation	Interquartile Range		
		Lower Limit	Upper Limit	Range
Original	−7.0	−29.1	16.3	45.4
Additional	−2.8	−22.3	15.5	37.8

TABLE 8—*Comparison between true relative* S *and relative* S *of original and additional samples and lower and upper limits of interquartile range as their range (in %).*

Relative S	Median Deviation	Interquartile Range		
		Lower Limit	Upper Limit	Range
Entire population	10.5	8.0	16.6	8.6
Original samples	10.0	6.8	15.0	8.2
Additional samples	10.7	7.2	16.2	9.0

Acceptance Criterion

Presently, the acceptance criterion is around 15% of the estimate. This means that an estimate is accepted if half the confidence interval length at 95% is smaller than approximately 15% of the estimate. This criterion results in the acceptance of estimates having a relative sampling error of up to 7.5% and the rejection of those having a relative sampling error above 7.5%.

It seems that the acceptance criterion is satisfactory. About 93% of the original estimates submitted to court using this criterion differ by less than 10% from the true weight.

However, one of the purposes of this study was to test the influence of changing this criterion. For example, if the criterion limit is set higher, will the number of acceptable estimates with high accuracy go up, or will the number of ''bad estimates'' with lower accuracy go up? It should be emphasized that decrease in the number of rejected estimates has an operative meaning in the routine work because fewer drug units are examined.

Based on the additional samples, it is possible to anticipate the influence of a change in acceptance criterion. As shown in Table 9, over 90% of the estimates which would have been rejected according to the acceptance criterion within the range of 16 and 20% (26 of 28) deviate by less than 10% from the true weight. Based on these results, we can anticipate that raising of the acceptance criterion from about 15 to 20% will add a significant number of estimates with a sufficient precision level. On the other hand, the benefit anticipated by raising the acceptance criterion to more than 20% will be small and not efficient.

Qualitative Results

Based on thin-layer chromatography, all the units examined contained heroin as in the original sample presented in court.

Some Limitations of this Study

Representative Exhibits—One drawback of this study is the fact that the exhibits could not be examined before the end of the legal proceedings. As a consequence, the elapsed time between the original analysis and the end of the trial is highly variable. We think

TABLE 9—*Additional estimates distribution according to their precision (relative deviation of estimate from real weight), grouping by relative C.I. (2 sampling errors vs. estimate), accepted, rejected and additional accepted estimates by raising acceptance criterion.*

Deviation of the Estimate from the True Weight, %	Relative C.I., %							
	0–5	6–10	11–15	16–20	21–25	26–30	>31	Total
0–5	24	68	28	13	4	. . .	2	139
6–10	. . .	11	13	13	2	1	. . .	40
11–15	1	1	3	2	1	2	1	11
16–20	. . .	1	. . .	. . .	. . .	. . .	1	2
Total	25	81	44	28	2	3	4	192
%		78.1		14.6		7.3		100
acceptance by criterion		accepted estimates		additional accepted estimates		rejected estimates		

that this drawback does not influence the main results of this study—precision and bias—and we can anticipate that the same conclusions would have been derived if fully representative exhibits were examined.

Weighing Accuracy—Theoretically, several factors could influence the weight results:

- Weighing at different periods of time.
- Weighing by different examiners.
- Weighing on different balances.

Again, it is our opinion that the influence from all these factors is negligible. The street dose is wrapped in heat-sealed plastic, and no loss of weight is anticipated. Although the original sample and the additional ones were weighed by different examiners, the weighing procedure as well as the balance are the same.

Number of Exhibits—Approximately a third of the exhibits to which sampling is applied in our laboratory include 10 to 15 drug units. The present research was carried out on exhibits including more than 15 drug units, in order to create four additional new samples. As a consequence, the results and the conclusions derived from this study do not theoretically represent this group of exhibits (10 to 15 drug units). Despite that, one can anticipate that the precision of the estimates is even better for the small exhibits than for the larger exhibits examined in the present study.

The small number of exhibits examined (48) does not allow the performance of standard statistics tests such as Chi-square or others. For the same reason, comprehensive analysis based on sample size (5, 6 or 7) was not carried out nor was comprehensive comparison between ''large'' and ''small'' exhibits.

The statistical analysis of originally rejected samples in the 18 additional exhibits was helpful in evaluating the acceptance criterion. However, due to the small number of exhibits, a comprehensive statistical analysis could not be carried out.

Additional Samples—The additional samples were created using the remaining unexamined units of each exhibit. Therefore if the original sample was not randomly selected and included more from the relatively small units, then as a direct result the additional samples will include more from the relatively large units. As a consequence, the additional samples are not totally random. In our opinion, this is merely a minor drawback which does not significantly affect the results of this study.

Conclusions

In this study, the routine sampling method used in the Analytical Chemistry Laboratory has been evaluated by comparing the estimated weight with the true weight of exhibits. The relative sampling error of the original estimates is about 5%, which can be considered as a small sampling error. The majority of the original estimates deviate slightly from the true weight: about 71% of them deviate by less than 5% and about 92% of them by less than 10%.

For about 88% of the exhibits, the true weight is within the C.I. at 95% around the original estimate. Higher precision was found in smaller exhibits (up to 29 drug units): the true weight of about 93% of the exhibits was within the 95% C.I. In larger exhibits (over 29 drug units), the true weight of only about 80% of the exhibits was within the 95% C.I. On the other hand, the precision of the additional estimates tends to be higher than the original estimate submitted to the court (95% instead of 88%). There is a possibility that this result points to the fact that there is some bias in the sampling method.

It was also found that the original estimates tend to be lower than the true weight by about 0.7% (median deviation). The asymmetric distribution of the extreme deviation of the estimates and the shifted deviation of *S* supports the possibility of bias in the sampling procedure. It should be remembered, however, that this bias is of the order of less than 1%. Also, there is no significant indication that this bias is the only reason for the above-mentioned finding. It can possibly be explained also by sampling errors.

The sampling method described in this paper should be improved so that the sample will be selected in a completely random way and so the selection will not be influenced by the examiner's choice. This change is expected to minimize the bias effect on the estimate. A simple method to make sample selection more random can be achieved by performing sampling from a ''black box.'' This procedure will be ''blind,'' since the examiner does not have the possibility of selecting a sample containing similar sized units. A more sophisticated alternative is using a random computerized selection. Units have to be numbered before selection; a completely random selection can be processed by the computer based upon population size.

From this study, it can be concluded that the precision of the actual sampling procedure is satisfactory and matches the statistical theoretical basis. Improvement in the sampling procedure should allow an increase in the acceptance criterion from about 15% to about 20%, gaining additional accurate acceptable estimates.

Acknowledgments

The authors gratefully acknowledge the technical assistance of Sara Berman in the Analytical Chemistry Laboratory and as well as their colleagues in the Computer & Electronics Laboratories. Thanks are also given to Dr. Shmuel Zitrin and Dr. Zafrir Goren for their helpful comments and suggestions.

References

1. Frank RS, Hinkley SW, Hoffman CG. Representative sampling of drug seizures in multiple containers. J Forensic Sci 1991;350–6.
2. Dangerous drug regulations (drug sampling), published in: Laws of the State of Israel, No. 5339, 8.3. 1991;704–5.
3. Tsidony D, Ravreby M. A statistical approach to drug sampling: a case study. J Forensic Sci 1992;1541–9.
4. Tsidony D, Yaeger M, Ravreby M. Drug sampling—laboratory manual. Internal report 1991 (in Hebrew).

Additional information and reprint requests:
Myriam Azoury
Analytical Chemistry Laboratory
Division of Identification and Forensic Science
Israel Police National Headquarters
Jerusalem, Israel