Projectnr.: 99998.0010

Report 91.62 December 1991

A statistical model for inspection procedures in forensic and environmental analysis

W.G. de Ruig[1], A.A.M. Jansen[2] and F.A. Huf[1]

[1] State Institute for Quality Control of Agricultural Products, P.O. Box 230, 6700 AE Wageningen, The Netherlands

[2] Agricultural Mathematics Group, Wageningen, The Netherlands

DLO-State Institute for Quality Control of Agricultural Products (RIKILT-DLO)
Bornsesteeg 45, NL-6708 PD Wageningen
Postbus 230, NL-6700 AE Wageningen
Telefoon +31 (0)8370-75400
Telex 75180 RIKIL
Telefax +31 (0)8370-17717

MAILING LIST

INTERNAL:
Director
Heads of Research Departments
Coördinator Chemometrics
Programme Management and Public Relations (2x)
Circulation
Library (3x)

EXTERNAL:
Agricultural Research Department
Department for Science and Technology
Department for the Environment Quality and Nutrition

ABSTRACT

A statistical model for inspection procedures in forensic and environmental analysis

Report 91.62 December 1991

W.G. de Ruig[1], A.A.M. Jansen[2] and F.A. Huf[1]

[1] State Institute for Quality Control of Agricultural Products (RIKILT-DLO), P.O. Box 230, 6700 AE Wageningen, The Netherlands

[2] Agricultural Mathematics Group (GLW-DLO), Wageningen, The Netherlands

Legal regulations on composition and safety of food products require inspection programs to maintain the decisions or to get insight in the real situation. Such programs consist of a sampling procedure followed by analysis of the samples. In monitoring of environmental contamination similar programs are required.

When large quantities of samples for monitoring of contaminants in environment or in food chains have to be analysed, and the majority of the samples fulfills the legal requirements, a two stage control system will be attractive. In the first stage samples are analysed by a simple 'screening method', to sift out the large number of samples fulfilling the requirements ('negative result'). The minor part of samples not fulfilling the requirements ('positive result') are further investigated by a sophisticated 'confirmatory method', specific for the analyte(s) of interest, for an ultimate judgement.

For this kind of inspection a simple model is used in this study, in order to calculate the fraction of false negative results in the inspection. It is assumed, that, when the quality criteria for the confirmatory method are appropriate, the fraction of false positive results is negligibly small. The quality of the inspection is expressed in terms of the fraction f of false results with respect to the real positive samples and is related to labour and costs. In an example given, the costs of analysing 40 000 samples with f = 6.6 % are ecu 7 800 000. Improving f down to 3.2 % causes ecu 16 000 000 additional costs.

The offered model may be not sufficiently realistic, but more realistic models can be implemented.

The approach can be applied to determine the quality of an inspection for a given method of analysis, or, in reverse, to determine the requirements for a method of analysis for a given inspection quality.

Keywords: Inspection, residues, forensic control, screening, confirmation, false results.

CONTENTS page

1 INTRODUCTION

In national and supra-national legislation, requirements as to composition and safety of food products have been stated. For the environment, there are also regulations concerning environmental contaminants. Among these are Directives of the European Community (EC) on health aspects in fresh meat, on hormones, residues and pesticides. Legal regulations require effective control. This is executed by inspection, consisting of a well described sampling procedure, followed by analysis of the samples and resulting in an inspection decision. Therefore methods of analysing and sampling are laid down, in EC e.g. on foodstuffs, milk, surface water, hormones and heavy metals. In many cases, stated regulations are not yet accompanied by inspection procedures. And for existing procedures, there is lack of information concerning the quality of the inspection.
What is the probability of false positive results and of false negative results for a given inspection procedure? And what will be the cost of an inspection procedure of a desired quality?
In this study we present a tool to give a first answer to these questions. Although our attempt is focused on control of application of banned growth promoters in meat production, such as hormones and ß-agonists, it is generally applicable in the field of food and environmental control.

If most of the samples to be inspected fulfill the legal requirements, a two-stage analysis procedure will be attractive. Firstly, all samples are investigated by a cheap method with a high throughput, a so called 'screening method', to filter out the majority of samples that fulfill the requirements ('negative results'). The smaller part of samples found not to fulfil the requirements ('positive results') is further investigated by a 'confirmatory method'. A confirmatory method is focussed on preventing false positive results, which is necessary since the moral and financial consequences of the inspection decision may be tremendous.
Such a model is worked out and quantified in this study.

The concepts 'screening method', 'confirmative methods', 'positive' and 'negative results' are used here in accordance with EC regulations (EEC, 1991), see Scheme 1.

The inspection result, *i.e.* the final result of sampling, screening and confirmation is a category 'positive' and a category 'negative'. In fact, there are four categories: 'true negative', 'false negative', 'true positive' and 'false positive', although the public analyst does not know which results are 'true' and which ones 'false'. The quality of the inspection procedure relates to the size of the false categories.

Scheme 1. EC definitions

Methods used for screening purposes (**screening methods**) are methods which are used to detect the presence of an analyte or class of analytes at the level of interest. These methods have a high sample throughput capacity and are used to sift a large numbers of samples for potential positives. They are aimed at preventing false negative results.

Methods for confirmatory purposes (**confirmatory methods**) are methods which provide unequivocal identification of the analyte at the level of interest. These methods are aimed at preventing false positive results as well as having an acceptable low probability of false negative results.

Positive result

The presence of the analyte in the sample is proved, according to the analytical procedure, when the general criteria, and the criteria specified for the individual detection method, are fulfilled.

(a) For substances with a zero tolerance, the result of the analysis is "positive" if the analyte is present at or above the limit of determination of the method.

(b) For substances with an established maximum residue limit, the result of the analysis is "positive" if the measured content of the analyte in the sample is equal to or greater than that established maximum residue limit plus n times the standard deviation which the method produces for a sample at that level.

Note The value of n should be defined according to the acceptable probability of obtaining false positive or false negative results.

Negative result

The result of the analysis is regarded as "negative" if the criteria specified for the procedure are not fulfilled or:

(a) in the case of substances for which there is a zero tolerance, the analysis does not indicate the presence of the analyte in the sample above the limit of determination; or

(b) in the case of substances with an established maximum residue limit, the measured content of the analyte in the sample is below the level specified above.

Note A negative result does not prove in case (a) that the analyte is absent from the sample, or in case (b) that the true content of the analyte is below the maximum

The aim of this study is, to present a first approximation to quantify the fraction 'false negative'. As a quality parameter, the *sensitivity*, defined as the ratio f of 'false negative' to 'real positive' will be used. In a following study we will refine the estimation of the fraction 'false negative' and pay attention to the fraction 'false positive' as well.
The statistical model will be illustrated with a practical inspection problem, concerning growth promotors.

In the EC the use of hormones and of ß-agonists in animal food production is banned. Within EC the numbers of animals yearly slaughtered are (1989) : 22 000 000 cows,
6 000 000 calves, 160 000 000 pigs and 96 000 000 sheep and goats. These animals have to be controlled on illegal use of drugs.
Most of the farmers are not consorted with illegal practices, i.e. most of the animals are not treated with banned drugs. Therefore, the two stage approach of screening and confirmation will be attractive from the viewpoint of cost and effort.
The banned drugs consist of a series of chemical compounds. These are partly strongly related from chemical point of view, but partly not. The composition and structure of a lot of them is known and reference substances are commercially available. New compounds, unknown to the inspecting departments so far, may be present as well. This complicates the whole exercise, but we simply assume that the the inspection aims to detect one known xeno-biotic analyte. Then, the analyte content in samples from untreated animals will be zero in all cases.

Our approach can be applied in two directions. Starting with a given analytical method, the quality of the inspection procedure, based upon this method, can be calculated, as is done hereafter. In reverse, in a given inspection procedure, the demands to be put upon the characteristics of the analytical method can be determined. If not already available, such a method can be searched for, and it is now known, which characteristics it must fulfil.

2 A SIMPLE MODEL FOR A TWO STAGE INSPECTION PROCEDURE

The aim of the inspection is to identify as many real positive samples as possible. The samples are considered to be obtained by random sampling from the relevant populations. In this study, we use the following concepts and nomenclature, closely following the treatment for the one stage case. The approach is summarized in Scheme 2. In the real situation two categories are present:

Analyte containing, originated from treated animals =	Real positive	RP
Analyte free, originated from untreated animals =	Real negative	RN.

After inspection, the samples are assessed to be positive or negative, but due to inspection errors four categories should be distinguished:

True positive	TP
True negative	TN
False positive	FP
False negative	FN.

The inspection results are based upon the results of the screening and confirmation analyses. In each stage results can be true or false, *i.e.* we should distinguish:

True positive in the screening	TP_S	True positive in the confirmation	TP_C
False positive in the sreening	FP_S	False positive in the confirmation	FP_C
True negative in the screening	TN_S	True negative in the confirmation	TN_C
False negative in the screening	FN_S	False negative in the confirmation	FN_C.

The inspection results are related to the screening and confirmation results in the following way:

$$TP = TP_C$$
$$FP = FP_C$$
$$TN = TN_S + TN_C$$
$$FN = FN_S + FN_C.$$

The fractions of each category will be denoted by ϕ and the appropriate subscript, e.g. ϕ_{TP_S} means the fraction of true positive results in the screening and ϕ_{FN_C} the false negative fraction in the confirmation.

In the model the distribution of the analyte in the population and the distribution of measurement errors have to be specified. By way of example the following conditions are assumed.
The true content x of the analyte of interest in the fraction of analyte containing samples, originating from treated animals, is normally distributed with a mean of μ and standard deviation σ: $x_{RP} \sim N(\mu,\sigma^2)$. In the samples, originating from untreated animals, the analyte content is always zero.
In the measuring procedures there is a random error, but not a systematic error. The measurement errors do not depend on x and are independently distributed for the screening and the confirmatory method, with standard deviation τ_s and τ_c respectively. Thus:

Screening result: $y = x + \epsilon_s$ $\quad \epsilon_s \sim N(0,\tau_s^2)$

Confirmation result: $z = x + \epsilon_c$ $\quad \epsilon_c \sim N(0,\tau_c^2)$

The limit of determination, above which the result is denoted 'positive', is D.

$y > D_S$ Screening result: positive

$y \leq D_S$ Screening result: negative

$z > D_C$ Confirmation result: positive

$z \leq D_C$ Confirmation result: negative

The inspection concerns one analyte only.

The *screening results* are the following.

- A fraction ϕ_{0a} of the *real negative* samples is found positive in the screening due to aspecificity of the screening method; that means, other substances than the analyte can give a response.
- A fraction ϕ_{0r} of the real negative samples is found positive in the screening due to random errors; $\phi_{0r} = P(y > D|RN) = P(u > D/\tau)$, where u has a standard normal distribution.
  Together, a fraction $\phi_{FP_S} = \phi_{RN} * (\phi_{0a} + \phi_{0r})$ is becoming *false positive*.

- Out of the *real positive* samples, a fraction ϕ_1 is found to be negative and a fraction $(1 - \phi_1)$ positive. ϕ_1 is the probability of an observation less than D for the distribution of measuring results with mean μ and standard deviation $\sigma_* = \sqrt{(\sigma^2 + \tau_S^2)}$, thus

  $$\phi_1 = P(y \leq D|RP) = P[u \leq (D - \mu)/\sigma_*], \quad (1)$$

  where u has a standard normal distribution.
  For all samples in the screening a fraction $\phi_{FN_S} = \phi_{RP} * \phi_1$ is *false negative* and $\phi_{TP_S} = \phi_{RP} * (1 - \phi_1)$ *true positive*.

The samples found negative in the screening are not further analysed.

The samples found positive in the screening, true or false, are further analysed by a confirmatory method.

The results of the *confirmation of the positive screening* samples are the following.

- The *real negative* samples, wrongly denoted positive by the screening method, will be found *negative* with a very large probability now, due to the sharp selectivity criteria of the confirmatory method. In the model, we assume $\phi_{FP_C} = 0$.
- Out of the *true positive* fraction of the screening, ϕ_{TP_S}, a fraction ϕ_2 becomes negative by random errors. This is causing an additional contribution of $\phi_{FN_C} = \phi_{TP_S} * \phi_2 = \phi_{RP} * (1 - \phi_1) * \phi_2$ *false negative* results.
- True positive samples can fail to fulfil the stated criteria of the confirmatory method, due to interferences. In fact, this concerns a noise problem. It can raise a substantial number of false negative results. While it can be quantified for a specific case only, where sufficient data are available, this contribution is ignored in the model.

Scheme 2. Positive and negative results in screening, confirmation and inspection decision

		Screening result		
		+		–
		Confirmation result		No confirmation
		+	–	
Real situation	+ **Real positives** RP	True positives in screening TP_S		False negatives in screening FN_S
		True positives in confirmation TP_C	False negatives in confirmation FN_C	
	– **Real negatives** RN	False positives in screening FP_S		True negatives in screening TN_S
		False positives in confirmation FP_C	True negatives in confirmation TN_C	
		+	–	
		Decision = Result inspection		

The *final result* of the inspection is:

true negative	$\phi_{TN} = \phi_{RN}$	(2)
true positive	$\phi_{TP} = \phi_{RP}\,(1 - \phi_1)\,(1 - \phi_2)$	(3)
false negative	$\phi_{FN} = \phi_{FN_S} + \phi_{FN_C} = \phi_{RP}[\phi_1 + (1 - \phi_1)\phi_2]$	(4)
false positive	$\phi_{FP} = 0$	(5)

The size of ϕ_2 can be calculated as follows, see figure 1.
For simplicity of calculation τ^2 and D are kept equal for the screening and confirmation stage. This may be not realistic and should be adapted for practical application.
When y, the screening result, and z, the result of the confirmation, are determined for all real positive samples, then y and z are bivariate normally distributed with mean (μ,μ), variances (σ_*^2,σ_*^2) and correlation coefficient $\rho = \sigma^2/\sigma_*^2$. Now we have to calculate the probability that z $\leq$ D, given y >D, (i.e. the normal distribution, integrated over area III in Figure 1), thus

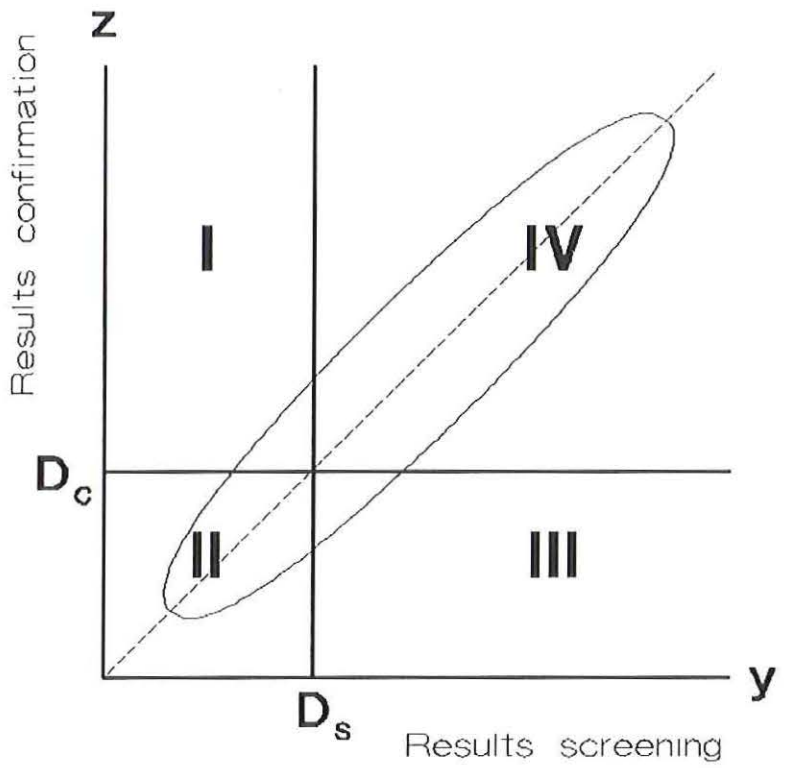


Figure 1. Calculation of ϕ_2.
y = results of screening method.
z = results of confirmatory method.
II + III = P ($z \leq D_C$)
II = P ($z \leq D_C$ and $y \leq D_S$)
III = P ($z \leq D_C$ and $y > D_S$).

$$\begin{aligned}\phi_2 &= \phi_{FN_c} / \phi_{TP_s} \\ &= P(z \le D \mid y > D \text{ and } RP) \\ &= [P(z \le D|RP) - P(z \le D \text{ and } y \le D|RP)] / [1 - P(y \le D|RP)] \\ &= [\phi_1 - P(u_1 \le (D-\mu)/\sigma_* \text{ and } u_2 \le (D-\mu)/\sigma_*)] / [1 - \phi_1] \end{aligned} \quad (6)$$

where u_1 and u_2 are correlated standard normal variables. The bivariate normal probabilities in (6) may be calculated using e.g. Algorithm AS 76 (Young et al.,1974), based on Owens' T function. A program for doing this is e.g. provided by StaTable (Metha et al., 1990).

All N samples of the population have to be investigated by the screening method. The number of samples to be reanalysed by the confirmatory method is the sum of the false and the true positive screening samples:

$$N_C = [\phi_{RN} (\phi_{0a} + \phi_{0r}) + \phi_{RP} (1 - \phi_1)] \; N. \quad (7)$$

This strongly concerns the work load in the laboratory, and thus the costs of the analyses.

A criterion for the quality of the inspection procedure is the number of false negative results with respect to the number of real positive samples $f = \phi_{FN} / \phi_{RP}$. From (4), this quality measure is easily found to be

$$f = \phi_1 + (1 - \phi_1)\,\phi_2 = 2\phi_1 - P(u_1 \le \frac{D-\mu}{\sigma_*} \text{ and } \mu_2 \le \frac{D-\mu}{\sigma_*}) \quad (8)$$

This criterion should not be too large, where the meaning of 'not too large' has to be defined further in respect of the requirements of the inspection.

The assumption of normality is not essential but has been made as a convenient first approximation, which can be used if transformation to normality is appropriate e.g. in case of a log normal distribution. For a running procedure a realistic distribution can be obtained from the data. It may be required to adapt the presented calculations in order to accomodate for this other distribution.

3 APPLICATION OF THE MODEL ON A PRACTICAL INSPECTION PROBLEM

3.1 Control of illegal application of drugs in animal fattening

The approach can be demonstrated by the following example, see Figure 2 and Scheme 3.

Let us assume that a representative sample of 40000 items has been taken for analysis in the laboratory. This is a realistic amount in case of control on illegal application of drugs on the farms. Let 15 % be treated, thus $\phi_{RP} = 0.15$ and $\phi_{RN} = 0.85$. Out of the 40000 samples 34000 are thus originated from untreated and 6000 from treated animals.

Suppose that in the set of samples, originating from treated animals, the analyt content is normally distributed, with a mean value μ = 5 µg/kg and a dispersion σ = 2.5 µg/kg. Then 95 % of the true values is between 0 and 10 µg/kg, which is not unrealistic in case of hormones. (In this simple example we ignore the non-existence of negative contents.)
The measurement standard deviation is assumed to be τ = 0.33 µg/kg, both for the screening and the confirmatory method. The limit of determination is D = 1 µg/kg.
The analyte is assumed to be xeno-biotic, thus in case of untreated animals the analyte is absent in all cases, thus $\mu_0 = 0$ and $\sigma_0 = 0$.
Due to random errors, in the measurement values unequal to 0 will be found. The probability that values > 1 µg/kg are found is 0.0013, that means 44 false positive results in the screening.
Further, by lack of specificity, systematic positive results can be found. Let in
0.5 % of the real negatives a value > 1 µg/kg be found, or 170 samples. For 34000 negative samples 44 + 170 = 214 samples will be found *false positive* and 33786 *true negative* in the screening procedure.
For the analysis of samples of treated animals, $\sigma = \sqrt{(2.5^2 + 0.33^2)} = 2.52$. The expectation remains μ = 5 µg/kg. Thus the fraction of the results expected to be below 1 µg/kg, and thus denoted negative is

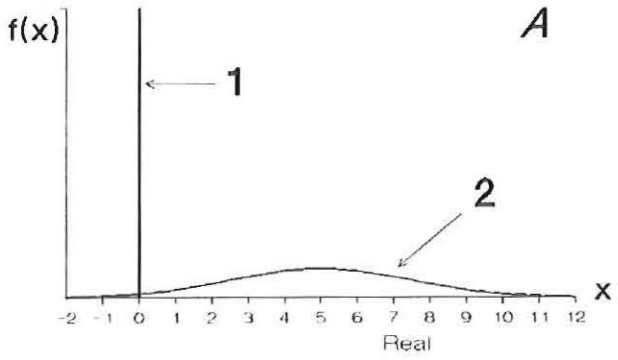
$$\phi_1 = P\ (u < (D-\mu)/\sigma_*) = P\ (u < (1-5)/2.52) = 0.0562,$$

or 5,71 %. The result out of 6000 samples will thus be 343 samples *false negative* and 5657 *true positive*.
The result of the screening procedure is thus 33786 + 343 = 34129 samples denoted negative and 214 + 5657 = 5871 samples positive.

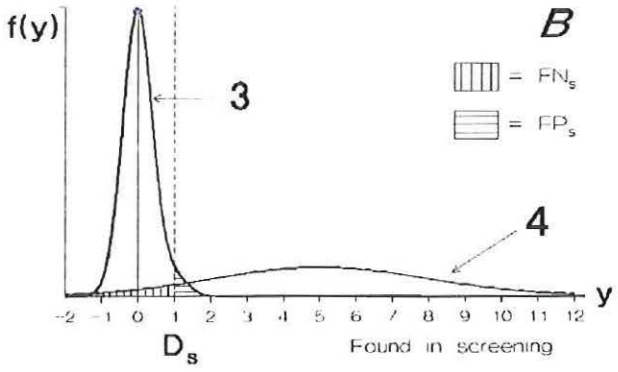
The samples found negative in the screening (false or true) are not further investigated.
The samples, found positive in the screening procedure, are further investigated by the confirmatory method. A fraction ϕ_2 of these will be found to give a result smaller than 1 now and thus become false negative. The numerator of ϕ_2 can be calculated to be 0.0085, the denominator equals $(1 - \phi_1) = 0.9438$ thus
$\phi_2 = 0.0085 / 0.9438 = 0.0090$, or 0.9 %. Out of the 5657 real positive results, due to random errors, 51 samples will be found false negative.

The aim of a confirmatory method is, to prevent false positive results. By confirmation using an MS method, if its criteria are formulated sufficiently sharp, this aim can be reached in practice. Then, the 214 false positive samples from the screening procedure, will become true negative now.
The final result of the inspection is, that 33786 + 214 + 51 + 343 = 34394 samples or 86 % are found to be negative and 5606 samples or 14 % positive. Out of the negative samples about 1 % is false negative. In this model there are no false positive results.

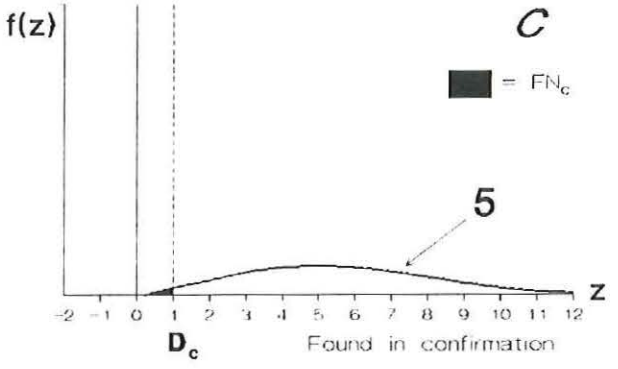
Figure 2. Relation between the real situation (A), the results of screening (B) and the results of confirmation (C).



A. Real situation.
1 = Distribution of real content x of the analyte in 'negative' samples. In case of a xenobiotic analyte x = 0 in all cases.
2 = Distribution of the real content x of the analyte in 'positive' samples. Assumed here is a normal distribution with $\mu = 5$ and $\sigma = 2.5$.



B. Screening results. Limit of determination $D_S = 1$.
3 = Distribution of screening results for the 'real negative' samples.
FP_S = false positive results in the screening.
4 = Distribution of screening results for the 'real positive' samples.
FN_S = false negative results in the screening.



C. Confirmation results. Limit of determination $D_C = 1$.
5 = Distribution of confirmation results for the samples found 'positive' in the screening.
FN_C = false negative results in the confirmation.

Scheme 3. True and false positive and negative results in case of inspection of 40000 samples.
$\phi_{RP} = 0.15 = 1 - \phi_{RN}$; $\mu = 5$; $\sigma = 2.5$; $\tau = 0.33$; D =1

		Screening result		
		+		–
		Confirmation result		No confirmation
		+	–	
Real situation N = 40 000	+ **RP** = 6000 (15 %)	TP_S = 5657		**FN_S** = 343
		TP_C = 5606	**FN_C** = 51	
	– **RN** = 34 000 (85 %)	**FP_S** = 170 + 44 = 214 (syst) (random)		**TN_S** = 33786
		FP_C = 0	**TN_C** = 214	
		+	–	
		TP = 5606 **FP** = 0	**TN** = 33786 + 214 = 34000 **FN** = 343 + 51 = 394	
		Positive 5606 (14.0 %)	**Negative** 34394 (86.0 %)	
		Decision = Result inspection		

Quality measure: f = 394 / 6000 = 6.6 %

The quality measure, 'sensitivity', i.e. the fraction of real positive samples that remains undetected is given by

$$f = \phi_{FN} / \phi_{RP} = 394 / 6000 * 100\% = 6.6\%.$$

According to current rates, the price of one screening analysis, using RIA or EIA techniques, is ecu 20 and of one confirmation analysis by GC-MS ecu 1200. The total cost of the screening is ecu 800 000 and of the confirmation ecu 7 000 000 (1 ecu = US$ 1.15).

3.2 Effect of lowering of the limit of determination of the screening method

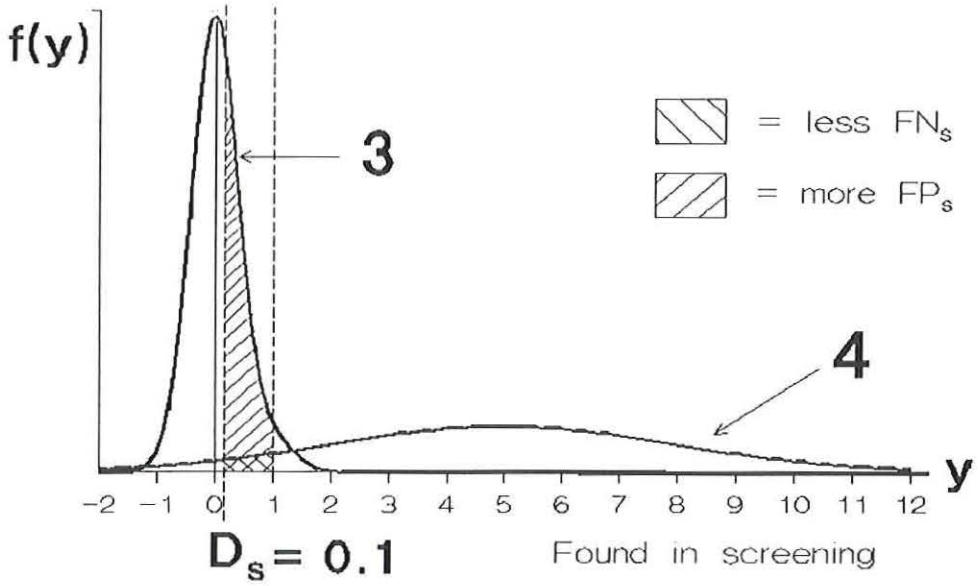


Figure 3. Effect of lowering the limit of determination for the screening method from D_S = 1 to D_S = 0.1. Other parameters as in Figure 3.
\\\\ = less false negative screening results.
//// = more false positive screening results.

In the example, we assumed that the limit of determination is 1 µg/kg for both the screening and the confirmatory method. In reality, for RIA or EIA techniques, lower values can be determined. Let us consider the effect of lowering the limit of detection for the screening method down to 0.1 µg/kg, Figure 3 and Scheme 4.

Now the distance between limit of determination and the population mean of the real positive

samples is enhanced to 5 - 0.1 = 4.9 μg/kg or 4.9/ 2.52 = 1.994 s. This means that the fraction

Scheme 4. True and false positive and negative results in case of inspection of 40000 samples.

$\phi_{RP} = 0.15 = 1 - \phi_{RN}$; $\mu = 5$; $\sigma = 2.5$; $\tau = 0.33$; $D_S = 0.1$; $D_C = 1$

		Screening result		
		+		–
		Confirmation result		No confirmation
		+	–	
Real situation N = 40 000	+ **RP** = 6000 (15 %)	TP_S = 5862		FN_S = 138
		TP_C = 5810	FN_C = 52	
	– **RN** = 34 000 (85 %)	FP_S = 680 + 12920 = 13600 (syst) (random)		TN_S = 20400
		FP_C = 0	TN_C = **136000**	
		+	–	
		TP = 5810	**TN** = 20400 + 13600 = 34000	
		FP = 0	**FN** = 138 + 52 = 190	
		Positive 5810 (14.5 %)	**Negative 34190** (85.5 %)	
		Decision = Result inspection		

Quality measure: f = 190 / 6000 = 3.2 %

ϕ_1 of false negative results lowers from 5.5 to 2.3 %; for 6000 positive samples thus 138 samples. Thus 343 - 138 = 205 less positive samples are lost in the screening procedure. The quality measure is now f = 190 / 6000 = 3.2 %.
This gain has to be paid for by a substantially higher amount of false positive samples in the screening. Now by random errors 38 % or 12920 negative samples will be found false positive in the screening (0.1 µg/kg = 0.1/0.33 = 0.3 s; a fraction 0.38 is above 0.1 µg/kg in the normal distribution). This will be corrected in the confirmation, but the number of confirmations has increased dramatically.
The cost of the screening remains unaltered, but the confirmation requires ecu 23 000 000 now.
The applied model may not be realistic. Yet, as a general rule, it may be expected that the fraction of false positive results will be enhanced substantially, when the limit of determination is lowered; this is caused by both lack of accuracy and lack of specificity.

4 DISCUSSION

4.1 Characteristics of the inspection

The characteristics of an inspection procedure are fully determined by:

- the appointed *inspection rules*, including prescriptions on sampling, analyses and acceptance and rejection limits.
- the *quality of the analysis*. This can be expressed in terms of accuracy (trueness and precision) for the determination of the quantity to be measured. Here trueness refers to systematic measuring errors (level) and precision to random measuring errors (dispersion) (ISO 5725, 1990).

4.2 Inspection rules

These may be laid down from desk, but the quality of a given inspection procedure is only known by research in practice. Therefore, the usefulness of a model, as presented here, has to be checked by data, obtained in practice. Then the parameters of the model can be adapted and optimised by feed back of the information. So, for a given inspection, a more realistic distribution model can be estimated.

Practical information should be obtained with respect to the following aspects.

- The distribution of results as function of the analyte content. This quantitative aspect can be

investigated by control samples with different levels of analyte content.

- The probability of false positive and of false negative results in relation to the operating decision criteria. To this end the criteria have to be applied to the established distributions of results. Then the consequences of alterations in the criteria can be can be examined.

In our model, the false negative results of the screening method will not be further investigated: these are lost. Therefore a quality requirement has to be, that the fraction of false negative results in the screening is not too large in relation to forensic security (see also EC definition for 'screening method').

As all positive results of the screening will be confirmed, true or false, false positive screening results are less harmfull: they will be corrected in the confirmation. However, the need of additional confirmation amounts to additional labour and costs.

The confirmatory technique is a more sophisticated, more expensive method, preferably based upon spectrometric principles. In practice, GC-MS is widely used nowadays. The conclusion 'analyte is present' is made using well defined criteria. When these criteria are properly defined, GC-MS can offer a specific decision. That means, that only the searched analyte can fulfill the criteria with large probability for not too small concentrations and all other substances do not.

As an example, in Scheme 5 the criteria are denoted as specified for low-resolution mass spectrometry in Commission Decision 89/610/EEC (EEC, 1989).

Scheme 5. Criteria specified for gas chromatography - low-resolution mass spectrometry, according to Commission Decision 89/610/EEC

1 The intenties of at least four diagnostic ions must be measured. If the compound does not yield four diagnostic ions with the method used, then identification of the analyte should be based on the results of at least two independent GC-LRMS methods with different derivatives and/or ionization techniques, each producing two or three diagnostic ions.

2 The molecular ion should preferably be one of the diagnostic ions selected.

3 The relative abundances of all diagnostic ions monitored from the analyte should match those of the standard analyte.

4 The relative intensities of the diagnostic ions detected, expressed as a perentage of the intensity of the base peak, must be the same as those for the standard analyte within a margin of ± 10 % (EI mode) or ± 20 % (CI mode).

4.2.1 Relation between false positive results, false negative results and criteria

An interaction exists between false positive and false negative results in relation to the application of criteria. When the criteria are formulated very strictly, a substantial number of real positive samples will fail in fulfilling the criteria, and thus become false negative. If that is unacceptable, the criteria can be made less strict, but that will give rise to a higher amount of false positive results.
A solution, to overcome the dilemma, can be the use of two decision criteria in the confirmation: one that eliminates largely the occurrence of false positive results and one that is focussed on a low probability of false negative results. Then, in between, a third category of samples is found, that can be denoted as 'suspect'.
A test procedure can be optimised in both directions, dependent on the requirements.
In this study, although we assumed that false positive results were fully excluded, we ignored the consequence of this strict assumption on the occurrence of a lot of false negative results.

4.2.2 The case of MS criteria

Criteria for MS (De Ruig et el, 1989, EEC, 1989) as a confirmatory method, include the requirements that 4 mass peaks have to be present, and their peak ratios have to be equal to those of a reference sample within a margin of ± 10% (EI mode). A new EC draft Decision[12] relativates the last requirement to 'preferably 10%'; this moves the responsibility concerning the decision to the laboratory.
What are the fractions false positive and false negative results of given criteria and what are the consequences when these criteria are changed? And is that acceptable? These questions can be answered in an actual case when the data are evaluated and alterations in the applied criteria are performed.

4.2.3 Pooling

Another modification is pooling of samples. If a screening method is used, having a much lower limit of detection than the confirmatory method (as is the case for RIA or EIA with respect to MS), in the screening pooled samples can be analysed, which will lower the cost of the screening considerably. When a pool is found to be 'positive', the individual samples may be rescreened. This may be done, when it is plausible that if positive samples are present in the pool, all or the majority of samples in the pool will be positive. This could be the case when one couple of treated animals from one farm is investigated.

4.3 Quality of the analysis

4.3.1 Quality of inspection, related to quality of analysis

As explained before, the fraction f of false negative results in the real positive samples, can be used as quality measure. The sensitivity of f for τ can be calculated, i.e. the increase of f if the precision of the analysis decreases:

$$f = 1 - [1-\phi_1(\tau)]\,[1-\phi_2(\tau)]$$
$$= \phi_1(\tau) + \phi_2(\tau) - \phi_1(\tau)\,\phi_2(\tau)$$

To this end the function f can be calculated for various values of τ. Some results are presented in Table 1. In this example, f is increasing about 10 % when τ increases 0.17.

Table 1. Fraction f of false negative results in the number of real positive samples as function of the measuring standard deviation τ (μ = 5, σ = 2.5, D = 1)

τ	0	0.17	0.33	0.50	0.67
f	0.055	0.060	0.066	0.073	0.081

A less precise method may be cheaper. The dependence of f from τ can be used to judge how far such a cheaper method is acceptable. On the other hand it can be seen which improvements can be expected from increasing measurement precision.

4.3.2 Quality assurance inspecting laboratory

In a laboratory having the appropriate knowledge and experience on a method of analysis, the quality assurance of the inspection has to be focussed on unbiassedness and (intermediate (ISO,1991)) reproducibility of the measuring results. This can be executed by analysing control samples by two institutes: the inspection institute the quality of which has to be assured and the assuring reference institute. Duplicate samples are send to both laboratoria, distributed in time and received as single samples. The samples are randomized among the normal sample series and should not be identifiable as control samples. By analysis of variance, from the sum of the duplicate results information about the systematic differences can be obtained, and from the differences information about the (intermediate) reproducibilities.

5 CONCLUSION

The approach, presented in this study, can be useful in evaluating the quality of inspection procedures for forensic and environmental purposes. When implementing the model for a given case, insight can be gained concerning the fraction of false negative results and the cost of the procedure.

6 ACKNOWLEDGEMENT

The authors thank L.M.H. Frijns, J. van der Greef, R. Schilt, R.W. Stephany and H. van der Voet for comments and valuable discussions.

6 LITERATURE

EEC 1989
Commission Decision of 14 November 1989 (89/610/EEC).
Official J. Europ. Comm. 1989, L 351, 39.

EEC 1991
Draft Commission Decision.
To be published.

ISO/DIS 5725-1990. Accuracy (trueness and precision) of measurements and results - Part 1, 2, 4.
Int. Org. for Standardization, Geneva.

ISO/DIS 5725-1991. Accuracy (trueness and precision) of measurements and results - Part 3.
Int. Org. for Standardization, Geneva.

Metha, C.R. and N.R. Patel
StaTable: Electronic tables for statisticians and engineers
Cytel software Corporation, Cambridge, Maasachussets, USA, 1990.

De Ruig, W.G., R.W. Stephany and G. Dijkstra
Criteria for the detection of analytes in test samples
J. Assoc. Off. Anal. Chem., 1989, 72, 487 - 490.

Young, J.C. and C.E. Minder
Algorithm 76, An integral useful in calculating non-central t and bivariate normal probabilities
Applied Statistics 1974, 23, 455.