Pergamon
PII: S0003-4878(00)00087-9

Printed in Great Britain.
0003-4878/01/$20.00

A Proposal for Harmonising Laboratory Performance Assessment Criteria in National Asbestos Fibre Counting Schemes

M. C. ARROYO* and J. M. ROJO

Instituto Nacional de Seguridad e Higiene en el Trabajo. Centro Nacional de Verificación de Maquinaria de Vizcaya, La Dinamita s/n, 48903 Baracaldo, Spain

Four European national asbestos fibre counting proficiency testing schemes have been studied in order to compare their criteria for the assessment of laboratory performance. Performance assessment is based on each laboratory's results after counting a certain number of samples. Two methods are currently being applied. To be classified 'satisfactory' laboratories must obtain at least 75% of normalised counts lying within defined performance limits (in three schemes), or the median and coefficient of variation of normalised counts must be within performance limits (in the fourth scheme). Differences in the numbers of test samples mean that the schemes are operating with different selectivity in assessing their laboratories' performances. Differences in the percentage of laboratory results falling within performance limits indicate that the schemes do not operate the same confidence probability in correctly assessing individual counts. It means that some schemes may be more lenient than others. This paper discusses two proposals to move towards harmonisation of the asbestos fibre counting proficiency testing schemes: (i) standardisation of the number of samples used for laboratory assessment and (ii) changes to the criteria to establish the limits of satisfactory performance.

Keywords: asbestos; proficiency testing scheme; harmonisation; fibres; quality control

INTRODUCTION

The need for improvements in the quality of analysis of samples in asbestos exposure assessment led to the development of several national and international proficiency testing schemes for asbestos fibre counting laboratories (LeBel, 1992). Although these schemes were primarily organised for the benefit of the laboratories, at present several European states require a specific minimum performance to be obtained in their own national proficiency testing schemes for the analysis of asbestos fibres, and accreditation of asbestos laboratories is required by law in some countries.

The national asbestos fibre counting schemes currently operate with different types of test samples and different methods of laboratory assessment criteria. Differences among the schemes' laboratory performance assessment criteria can be expected, taking into account that there has been no formal co-ordination, co-operation or communication between the schemes (Arroyo and Rojo, 1998). Harmonisation of asbestos fibre counting proficiency testing schemes is seen as beneficial by the parties involved. Any laboratory interested in obtaining accreditation must participate in their national scheme, although laboratories may also take part in schemes in other countries. Harmonisation would also benefit participants by avoiding artificial discrepancies between schemes in the evaluation of laboratory performance. So, with harmonisation, any laboratory could join in any scheme and obtain a rating which was directly comparable with other laboratories and understood by its customers, accreditation bodies, etc. Ideally, if the proficiency testing schemes were completely harmonised, a client might be able to send samples to any proficient counting laboratory and be assured that the counting results would not differ substantially from those which would have been obtained from any other proficient counting laboratory.

Received 11 July 2000; in final form 9 October 2000.
*Author to whom correspondence should be addressed. Tel.: +34-4-944-990211; fax: +34-4-944-990678; e-mail: evaluacalidad@mtas.es

A network to co-ordinate proficiency testing activity in Europe for occupational and environmental analysis of air samples was financially supported by the Standards, Measurement and Testing Programme (contract number SMT 4-CT96-7504) from the EC Commission. The main objective of the proposal was to co-ordinate activity between several national proficiency testing schemes towards harmonising of proficiency testing in a European context (Tylee, 1998). The partners in this project represent proficiency testing schemes in the United Kingdom, Germany, Denmark, the Netherlands, Belgium, Finland, France, Norway and Spain.

As part of the project a comparison between four national asbestos proficiency testing schemes was planned. A brief summary of this study and conclusions were given to participants in the fourth meeting of the project held in Geilo (Norway) in February 1999. The whole study and the resulting proposal for harmonisation are presented here.

DESCRIPTION OF THE SCHEMES

Four European national asbestos fibre counting proficiency testing schemes (AFCPT schemes) were available: the Belgian scheme (Grosjean, 1998), the French scheme (Kauffer, 1989), the United Kingdom scheme (RICE) (Crawford *et al.*, 1992) and the Spanish scheme (PICC-FA) (Arroyo, 1991).

The study was planned to get a detailed description of the operational characteristics, including laboratory performance assessment criteria, of AFCPT schemes in order to examine them in greater detail and to discover the points where harmonisation might first be focused. A representative body of data was required from each scheme. The data used in this work were those from the last laboratory performance assessments, and were obtained from the respective scheme co-ordinators.

A summary giving the main characteristics of these national schemes is given in Table 1. All four schemes operate by circulating test samples to each participating laboratory at regular intervals. Each test sample is associated with a reference count, which is some form of consensus of the counts obtained by scheme participants. The laboratories' results (counts) are compared with the reference counts, and the performance of the laboratory can then be
the criteria employed by that schem
assessment is repeated after each rou
calculated from the accumulated resul
one round.

The test samples consist of fibre-bea
filters permanently mounted on micro
the French scheme the filter is cut i
get six replicate slides of the same sa
three schemes use individual test sa
come from asbestos removal operatio
industry and airborne dust generated
containing materials in the laboratory

The RICE Scheme with 232 par
tories is the oldest and the biggest o
other cases, participation is notably
from 24 to 37 laboratories.

The period studied in this work
(1996–1997). The total number of sar
and the corresponding number of in
are also indicated in Table 1.

ANALYSIS OF DATA: LABOR PERFORMANCE ASSESSMENT

There are many aspects to be co
AFCPT scheme. To begin with we dec
the criteria used for the laborator
assessment. That is also the first point
ences between schemes can objective

If we examine these criteria, which
in Table 2, we can see that, apart fr
all the schemes transform the counts
values (a normalised value being the
tory count to reference count), ther
characteristics in which we can find
For example:

- The assessment is for the labora
schemes, but one scheme assesses
counters.
- Two schemes use the same numbe
laboratory assessment (N=32). T
schemes use fewer samples (range
- In three out of four schemes the
formance assessment is based on
individual counts lying within cert

Table 1. Characteristics of the European national asbestos fibre counting proficiency testing s

	Scheme			
	UK (RICE)	Belgium	France	Sp
Starting date	1984	1988	1986	
National regulation	Mandatory	Mandatory	Mandatory	
Test sample	Whole filter	Whole filter	Replicate segment	
Number of participant laboratories	232	26–27	33–37	
Period evaluated (in this study)	1997	1996–97	1996–97	
Number of samples circulated	824	32–40	6 lots of 21	
Number of individual counts	9463	1153	3486	

Table 2. Laboratory performance assessment criteria

	UK	Belgium	France	Spain
Subject	Laboratory	Laboratory	Individual counter	Laboratory
Number of samples (N) used for the laboratory assessment	32	20–24	21	32
Number of counts used for the reference value	At least 15	At least 15	Reference group	Overall counts
Statistical parameter for the reference value	Median	Median	Mean	Mean
Counts reduced to normalised values for statistical treatment	Yes	Yes	Yes	Yes
Criteria for satisfactory classification of laboratories				
≥75% of N normalised counts within limits	Yes	Yes		Yes
Mean and CV of normalised counts within limits			Yes	
Type of laboratory classification				
Satisfactory Non-satisfactory				Yes
Satisfactory Class 1 Satisfactory Class 2 Non-satisfactory	Yes	Yes	Yes	

its. The other scheme takes the mean and coefficient of variation of all a laboratory's counts, and requires these summary values to be within defined limits.

- Three schemes classify the laboratory performance into three classes or groups (two classes for satisfactory performance and one for non-satisfactory). The other scheme uses only two classes: satisfactory or non-satisfactory.
- The Belgian and British schemes have differently calculated limits for samples of low fibre densities. The French and Spanish schemes do not have this kind of limit but, as we will see later, they do not use samples of low densities.

It is interesting to analyse the differences among schemes regarding the main components of laboratory performance assessment, which are as follows.

Sample densities and distributions of counts

Figure 1 describes the sample densities in each AFCPT scheme. Sample densities higher than 450 fibres per mm^2 of sample surface are rarely used as test samples. The proportion of low-density samples (defined as less than 64 fibres/mm^2) is markedly different between schemes. This value is the density corresponding to counting 100 fibres in 200 Walton–Beckett graticule areas (Crawford *et al.*, 1991). The Spanish PICC-FA scheme does not include low-density samples while the Belgian scheme includes a large proportion of this kind of sample. Samples with low fibre density are affected by larger variability of counts in proportion to their density. This is the rea-

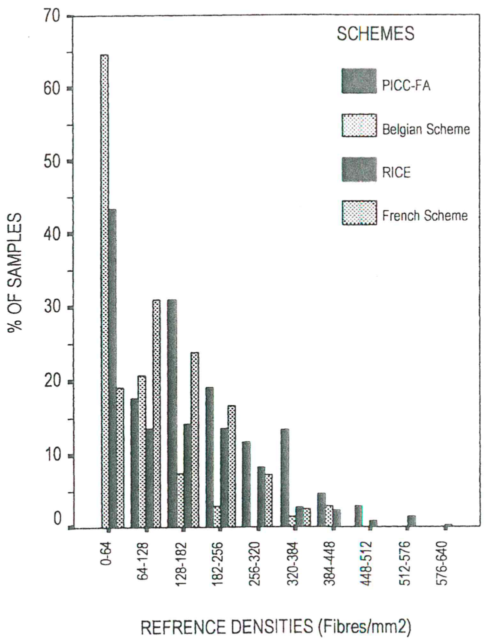

Fig. 1. Range of reference densities.

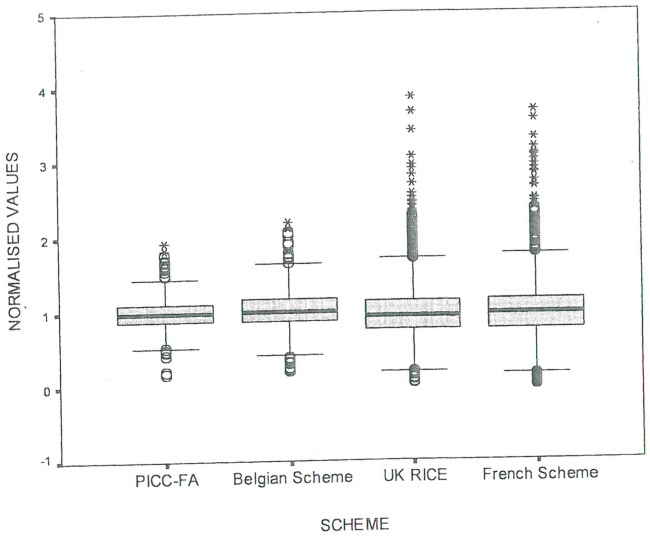

Fig. 2. Description of normalised values.

son why the British and Belgian schemes apply different limits in these cases (Crawford *et al.*, 1991). In order to simplify this study and to deal with a homogeneous set of data, it was considered better to discard counts with reference values less than 64 fibres/mm^2. The reduction affects mainly the UK RICE and Belgian schemes but the numbers of counts remaining were large enough to obtain statistical significance.

All the schemes transform the individual counts to normalised values. The overall normalised values obtained in the AFCPT schemes are represented in the box-plots in Fig. 2. A statistical description of these data before and after discarding counts with reference values under 64 fibres/mm^2 is shown in Table 3.

It is seen that the discarded cou the extreme values. The standard de consequently but the mean and me no or little change.

The Kolmogorov Smirnoff test w count distributions. The hypothesis c bution has to be rejected. The two-ta 0.05. The normal distribution coul for the Spanish scheme ($P = 0.00$ statistical distribution of the asbest is not relevant for this work.

Reference values

The procedure to calculate the ref fers from one scheme to another as i 2. The Belgian and British scheme

Table 3. Statistical summary of normalised counts

	All data Scheme[a]				Data from high density Scheme[a]		
	1	2	3	4	1	2	3
Number of cases	715	741[b]	8947[c]	3486[d]	715	435	50
Mean	0.99	1.03	1.05	1.00	0.99	1.02	0.9
SD	0.20	0.35	2.45	0.56	0.20	0.31	0.4
Median	0.99	1.00	0.93	0.96	0.99	1.00	0.9
Ranges of values	0.17–1.92	0.21–2.91	0.00–195.26	0.00–15.62	0.17–1.92	0.21–2.20	0.03–
Interquartile range	0.24	0.38	0.43	0.42	0.24	0.31	0.
K-S Z	0.038	1.957	33.44	8.165	0.038	0.065	0.0
2-tailed *P*	0.016	0.001	0.000	0.000	0.016	0.000	0.0

[a]1: PICC-FA; 2: Belgian scheme; 3: RICE; 4: French scheme
[b]412 missing normalised values (without reference value)
[c]516 missing normalised values (without reference value)
[d]1 missing result
[e]After discarding normalised values of counts with reference value <64 fibres/mm^2

of at least 15 laboratory counts as the reference count. The French and Spanish schemes use different types of mean values. The Spanish scheme uses a mean of all available counts excluding outliers, whereas the French scheme takes the mean of a 'reference group' of experienced counters.

Applying different procedures of course leads to slightly different reference values. However, the important question is whether or not the differences affect the classification of the resulting normalised values. In an earlier sample exchange between the international scheme 'AFRICA' and PICC-FA, we found discrepancies in the classification of counts due to the different position of performance limits in each scheme, but the laboratory performance assessment was not affected when the assessment criteria of each scheme (calculation procedure for reference counts and performance limits) were imposed on the data of the other scheme (Arroyo and Rojo, 1998). For the purpose of this present paper we will assume this premise also applies between the four schemes being studied.

Number of samples and criteria used for laboratory performance assessment

The laboratory performance assessment is based on the laboratory results after counting a certain number of samples. At the time of this study, the numbers of samples used in the AFCPT schemes are: 32 for UK RICE and Spanish PICC-FA, 21 for the French scheme and between 20 and 24 in the Belgian scheme (see Table 2).

To be classified 'satisfactory' laboratories must obtain a minimum of 75% of normalised counts lying within the performance limits described below. This criterion, initially applied by the UK RICE scheme, was adopted afterwards by the Belgian and Spanish PICC-FA schemes. The French scheme bases the categorisation of each laboratory on using the mean and coefficient of variation of all the normalised counts of that laboratory. These summary values must then lie within performance limits.

Performance limits

The performance limits for each of the AFCPT schemes are indicated in Table 4. Figure 3(a) is a representation of these limits, shown graphically in order to make the differences between them clear. The boxes represent the bands of satisfactory performance. The shaded part corresponds to the inner limits, which are employed by three of the schemes (the Belgian scheme, the UK RICE scheme and the French scheme) to distinguish between different grades of satisfactory performance. For the French scheme we have to take into consideration that the limits apply to the mean of several values instead of to the individual values.

It is interesting to compare the percentages of laboratories meeting the requirement for satisfactory performance in each AFCPT scheme, represented by the bars in Fig. 3(b). This percentage is very similar (81–87%) in PICC-FA, the Belgian and the French schemes. All laboratories in the RICE scheme get a satisfactory classification in this particular dataset. On the other hand, this scheme has the broader performance limits. In an asbestos counting scheme the performance limits should represent a practical compromise to match the minimum variation with the maximum number of laboratories (Arroyo, 1991). Hopefully not all laboratories are considered satisfactory.

For the schemes with two classes of satisfactory performance the ratio

$$\frac{Class\ 1}{Class\ 2} Labs$$

is quite different. There is no *a priori* reason for systematic differences between laboratories in different countries. These discrepancies in laboratory classification therefore suggest that some schemes are more lenient than others when they assess laboratory performance.

DISCUSSION ON SCHEMES' CRITERIA FOR LABORATORY PERFORMANCE ASSESSMENT

All the schemes assess their laboratories' performances on the laboratory results on N consecutive test samples. There are two methods for laboratory classification; one is based on the percentage of within-limits values, the other uses the mean and coefficient of variation of all values. The first method is applied

Table 4. Performance limits of AFCPT schemes expressed as normalised values or (for low density samples) in terms of the reference value V_{ref}

Limits for individual counts				Limits for mean and CV
Spanish PICC-FA		Belgian scheme	UK RICE	French scheme
0.65–1.35	$V_{ref} \geq 64$	Inner: 0.75–1.33 Outer: 0.6–1.67	Inner: 0.65–1.55 Outer: 0.5–2.00	For mean — Inner: 0.75–1.33 Outer: 0.50–2.00
	$V_{ref} < 64$	Inner: $(\sqrt{V_{ref}} - 1.07)^2$ to $(\sqrt{V_{ref}} + 1.24)^2$ Outer: $(\sqrt{V_{ref}} - 1.80)^2$ to $(\sqrt{V_{ref}} + 2.33)^2$	Inner: $(\sqrt{V_{ref}} - 1.57)^2$ to $(\sqrt{V_{ref}} + 1.96)^2$ Outer: $(\sqrt{V_{ref}} - 2.24)^2$ to $(\sqrt{V_{ref}} + 3.30)^2$	For CV: ±40%

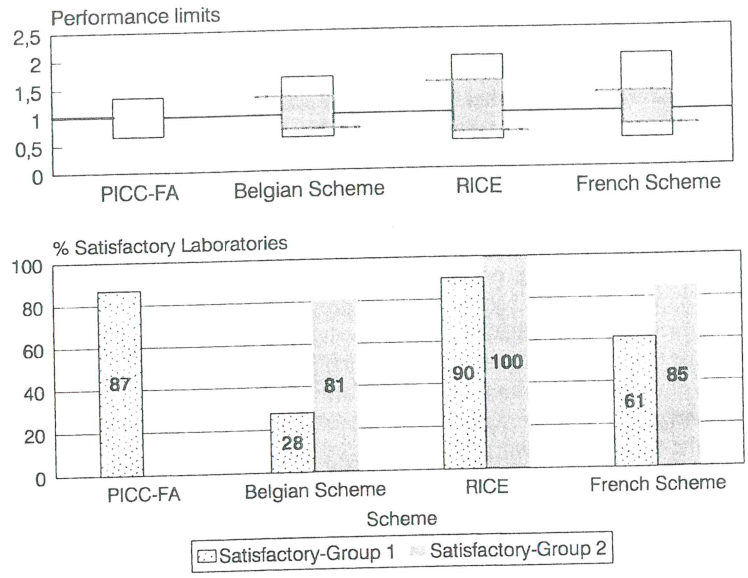

Fig. 3. (a) Band of satisfactory performance, and (b) percentage of laboratories that met the requirement fo
fication (individual counters in the French scheme).

by three schemes, so we will centre the discussion on this method although comparisons will be extended to the fourth scheme. The two main points to discuss are then the number of samples (N) and the position of the performance limits.

The number of samples used in the laboratory performance assessment determines the selectivity of the scheme, i.e. its capability to discriminate between satisfactory and unsatisfactory laboratories. Ogden (1984) described the selectivity of a proficiency testing scheme as the ratio between the percentages of routine samples that a laboratory must have in the band of satisfactory performance required by the scheme, to have respectively more than 95% and less than 5% probability of success in the scheme (in this case, success means having at least 75% of values within the limits for the N consecutive control samples). The more samples used for the laboratory performance assessment, the better selectivity will be obtained by the scheme, assuming that the test samples are representative of routine samples. Comparing the selectivity in an asbestos proficiency testing scheme using 8, 16, 32 and 64 samples, Ogden provided the rationale for using 32 samples in the RICE scheme. Between 8 and 16 and between 16 and 32 the selectivity of the test sees an important increase. However, doubling the number of samples to 64 did not produce a significant improvement in the selectivity of the scheme. To have more than a 95% chance of passing a RICE-type trial of 32 slides picked at random from its routine slides a laboratory must get more than about 85% of its counts within the same performance limits. On the other hand, a laboratory with less than about 60% of its results within those performance limits will have less than 5% probability of passing the trials.

The number of samples used for th
formance assessment cannot be c
because it determines the selectivit
Schemes. As the selectivity is a
characteristic of the schemes, it seem
pose that all the ACFPT schemes r
selectivity. Taking into account that
already using 32 samples and that t
good practical number, the proposal
two schemes should increase thei
samples for laboratory performance
this figure.

Regarding the performance limits
to look at them as tolerance limits th
fraction of the population with a c
Taking into account that the normal
not be generally assumed for the res
fibre counting it is necessary to ap
limiting the acceptable variability. M
be convenient that all the schemes
same probability in correctly ass
counts. This can be achieved if the p
include a similar percentage of the r
of normalised counts in each schen

Assuming that all the schemes us
laboratory performance assessment,
which figure might be suitable for
the population of counts that shoul
the performance limits. From the
85% seems to be a suitable figure.
limits including 85% of the populati
that, by taking 32 values randoml
probability of getting at least 75% c
within limits.

Now, it would be interesting to
happen if this criterion were applie

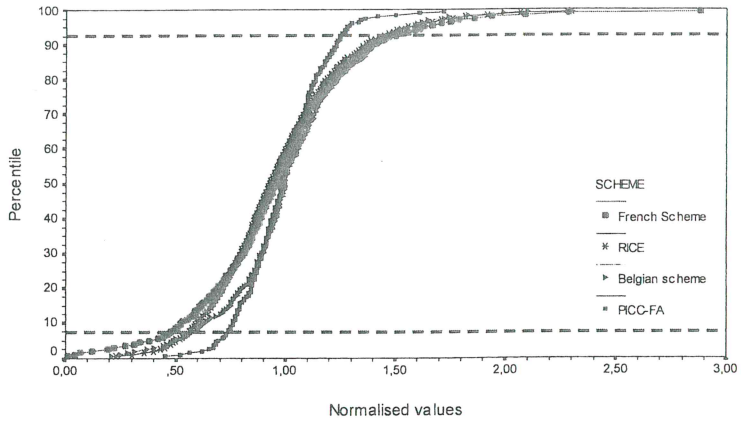

Fig. 4. Percentile distribution of normalised values. Lines represent the proposed band for satisfactory performance.

data from the four schemes. The description of populations of data in each AFCPT scheme is shown in the form of percentiles in Fig. 4. The symmetrical limits for the 85% of data corresponding to percentile 92.5 (upper limit) and percentile 7.5 (lower limit) are also drawn. It allows us to find easily the normalised values that match the resulting performance limits. Note that this procedure allows us to take the central 85% of the data regardless of the shape of the distribution, normal or otherwise.

The values for the new limits based on this criterion are indicated in Table 5 and Fig. 5 together with the actual performance limits. The new limits are presented only as the outer limits (i.e. those corresponding to the boundary between satisfactory and non-satisfactory performance). For inner limits, an appropriate percentage could also be selected if required.

We can note two things. Firstly, there is a tightening of the band of satisfactory performance in all schemes, which should encourage a reduction of uncertainty in the results from the asbestos counting method. Secondly, there is now agreement between schemes on the positions of the performance limits. This agreement is now remarkably good in the case of the Belgian and RICE schemes, which with their present criteria are quite different. It confirms our suspicion that, at present, some AFCPT schemes are more lenient than others in terms of judging whether a laboratory performance is satisfactory or not. These discrepancies between schemes should be avoided.

RECOMMENDATIONS

The aim of any asbestos fibre counting scheme is to provide the asbestos fibre counting laboratories with a standard by which to assess and control the quality of their counts. As well as existing for the purposes of improving the performance of laboratories, these schemes are being used increasingly by third parties including national accreditation bodies and national or international regulatory authorities. It is not necessary to remark on the importance of harmonisation: the only questions now are how and from where to start?

The ISO Guide 43 'Proficiency testing by laboratory intercomparisons' recommends procedures for proficiency testing schemes. This standard can help to bring together the asbestos counting schemes in general terms. However, the particularities of asbestos counting make it necessary to establish other complementary criteria to be incorporated by the AFCPT schemes in order to get comparability between them.

At the present time, the assessment of performance of a laboratory in any AFCPT scheme is made by comparison with the performance of the body of par-

Table 5. Present and proposed bands for satisfactory performance

	Present Scheme[a]				Proposed Scheme[a]			
	1	2	3	4	1	2	3	4
Outer limits	1.35 0.65	1.67 0.60	2.00 0.5	–	1.26 0.74	1.46 0.57	1.45 0.56	1.50 0.48
Percentiles of population counts	97 2.5	97 8	99 5	–	92.5 7.5	92.5 7.5	92.5 7.5	92.5 7.5
% of within limits values for outer limits	94.5	89	94	–	85	85	85	85

[a]1: PICC-FA; 2: Belgian scheme; 3: RICE; 4: French scheme

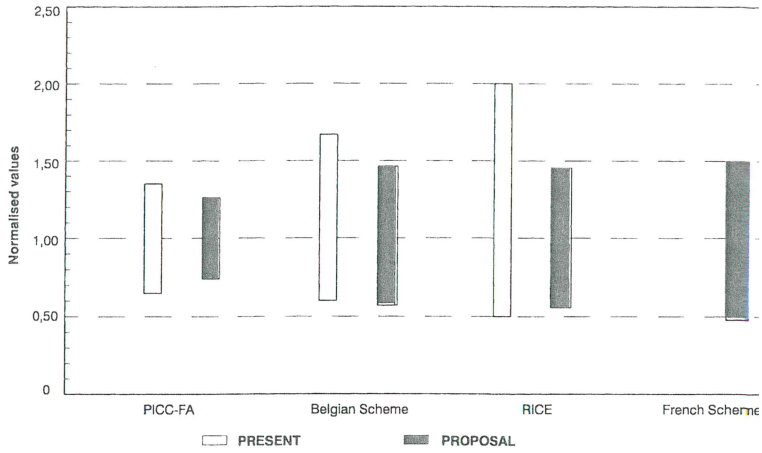

Fig. 5. Graphical representation of present and proposed bands for satisfactory performa

ticipant laboratories. In other words, the criteria to assess the laboratory performance applied by the schemes must be in accordance with the collective performance of the laboratory membership. If the contributing laboratories have poor performance, the limits of the scheme have to be broad. On the contrary, a group of high-performing laboratories should imply narrow limits.

The proposals that arise from this work for harmonising AFCPT schemes are:

- To use the same number of samples (N=32) for the laboratory performance assessment, to get the schemes operating with the same efficiency to discriminate between satisfactory and unsatisfatory laboratories
- To establish the position of performance limits so as to include the same percentage of the population of results (85%), to get the schemes running with the same probability of a performance assessment of 'satisfactory'.

Both aspects should be important steps towards AFCPT schemes' harmonisation, and the advantages have been demonstrated. On the other hand, neither represents a major change in the practical operation of the present schemes. This is a matter to be taken into account to avoid proposals which could be difficult or even unrealisable in their implementation.

In the simulation of harmonised criteria for performance limits we discovered that it was possible to achieve good agreement in the position of the limits. We think that the remaining differences are mainly due to the particular natures of the test samples in each scheme. All aspects (type, densities, quality control, calculation procedure for reference values, etc.) related with test samples in the AFCPT schemes should be carefully studied. That could also be a very important step forward in harmonisation. However, modifications regarding test samples would have a more pronounced effect on the present schemes. This and other future actions should acceptance of this initial proposal.

In any case, it must be drawn at that implementation of this propos fully preparation and gradual intr relationship as provided by the Eur Proficiency Testing Schemes in Che Workplace and Environmental Sa maintained.

Acknowledgements—The authors acknov ator [B. Tylee of the HSE (UK)] and all Project European Network of Proficienc Special thanks are due to Dr N. P. Crawfc R. Grosjean of the LIT (Belgium) and E. (France) for providing the data used in th also thank P. Brown of the IOM (UK) fo bution to the manuscript.

REFERENCES

Arroyo MC. Quality control of airborne a in Spain. In: Clean air at work. Luxe tember 1991 Royal Society of Chem cation no. 108, 1991:440–4.

Arroyo MC, Rojo JM. National versus i fibre counting schemes: comparison b interlaboratory quality control program Asbestos Fibre Regular Informal Co (AFRICA). Ann Occup Hyg 1998;42:9

Crawford NP, Brown PW, Jones AD, M BG. Towards development of RICE to samples from asbestos clearance operat work. Luxembourg, 9–13 September 19 Chemistry, Special Publication no. 108

Crawford NP, Brown P, Cowie AJ. The schemes for asbestos fibre counting. 1992;36:59–69.

Grosjean R. Proficiency testing scher measurements, In: European News. S Steering Committee, Health and Safety

Kauffer E. The French asbestis quality c Sixth International Colloquium on Dust nique And Strategy, Jersey, Channel I ember 1989. Paris: Asbestos Internationa p. 156–63.

LeBel J. A reference scheme for quality assurance of fibre counts by phase contrast optical microscopy. Working document prepared for the WHO–IFGS working group of experts for harmonisation of methods and quality assurance in the evaluation of exposure to airborne fibres in work environment. The Asbestos Institute. Sherbrooke, 1992.

Ogden TL. Statistical uncertainties in asbestos — laboratory classification through the RICE scheme. Ann Occup Hyg 1984;4:449–52.

Tylee B. European network, In: European News. Spring 1998. WASP Steering Committee, Health and Safety Executive, 1998.