Applied Industrial Hygiene

ISSN: 0882-8032 (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uaph20

The Quality of Fiber Count Data

Martin T. Abell , Stanley A. Shulman & Paul A. Baron

To cite this article: Martin T. Abell , Stanley A. Shulman & Paul A. Baron (1989) The Quality of Fiber Count Data, Applied Industrial Hygiene, 4:11, 273-285, DOI: 10.1080/08828032.1989.10390653

To link to this article: https://doi.org/10.1080/08828032.1989.10390653

Published online: 25 Feb 2011.

The Quality of Fiber Count Data

Martin T. Abell, Stanley A. Shulman, and Paul A. Baron
National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226

Optical fiber counts are used to determine asbestos exposure, so it is important to assess, control, and document the quality of those counts. These functions are the responsibility of the quality assurance (QA) coordinator in each laboratory. The QA coordinator must recognize that, compared to the analytical results for other substances, fiber count data are much more variable and have different statistical properties. These data, therefore, warrant special treatment. This article discusses the need to recount some samples, the procedures for determining bias and variability from these recount data, and the use of these statistics to test analytical results or assign confidence limits to them. Three kinds of bias and variability must be considered: intracounter, intralaboratory, and interlaboratory. As data pairs (count and recount) are obtained, the first consideration is whether bias is present. If bias is detected in a set of data, that data should not be used for any purpose until the source of bias is investigated. Bias can be difficult to detect, and when not detected, the differences in the data are assumed to be variability. The procedures recommended in this article for determining variability are based on the relatively simple calculations of NIOSH Method 7400, but alternative calculations are also discussed. Variability is expressed as s_r, which is an estimate of relative standard deviation.

An example calculation of intracounter s_r is given, along with an example of how to use this s_r to test the quality of a fiber count for an individual sample. This test, which does not have great power, is meant to detect differences between a counter's historically established s_r and the s_r for the test sample. Such a difference indicates a problem with the sample or the analytical procedure, and as a guideline, fiber counts that fail the test should not be used to evaluate exposure. In an extension of the test given in NIOSH Method 7400, a guideline is given for deciding whether an entire sample set should be rejected based on the number of individual fiber counts rejected. Intralaboratory s_r is an indicator of differences among counters within a laboratory, but these differences should be investigated for identifiable biases due to differences in training, visual acuity, or equipment. Interlaboratory s_r can be calculated if samples are exchanged between laboratories. Interlaboratory s_r is used to calculate a confidence interval about each analytical result, and that confidence interval should be reported as the analytical result. Analytical results reported for other substances do not include a confidence interval, but the analytical methods for other substances can be calibrated with readily available reference materials, thereby reducing the differences among laboratories. The nature of the fiber counting method and the lack of reference materials means that biases between laboratories are not easily corrected, or even identified. The confidence interval informs the person reading an analytical report about the differences among laboratories.

Abell, M. T.; Shulman, S. A.; Baron, P. A.: The Quality of Fiber Count Data. Appl. Ind. Hyg. 4:273–285; 1989.

Definitions

The following acronyms are defined here and are not defined again in the text.

AAR	Asbestos Analysts Registry program (a quality assurance program for individuals who count fibers)
ACGIH	American Conference of Governmental Industrial Hygienists
AIHA	American Industrial Hygiene Association
HSE/NPL	UK Health and Safety Executive/UK National Physical Laboratory
NIOSH	National Institute for Occupational Safety and Health
OSHA	Occupational Safety and Health Administration
PAT	Proficiency Analytical Testing program (a laboratory quality audit program)
PEL	Permissible Exposure Limit (issued by OSHA)
REL	Recommended Exposure Limit (NIOSH)
TLV	Threshold Limit Value (ACGIH)

The following words and symbols are defined as used in this article. Most are more thoroughly described in the text.

bias: systematic difference between two sets of measurements.

confidence interval: an interval which contains the parameter under study with high probability; a range of fibers/mm^2 values (or fibers/cc) that includes the true mean for the sample with high probability.

component of variance: one of several sources of total variability of a measurement.

count: (verb) to determine the number of fibers/mm^2 on a filter by phase contrast microscopy and by following the counting rules of a specific method such as NIOSH Method 7400.

count (fiber count): (noun) the datum produced as a result of counting the fibers on a single filter sample.

data pair: two numbers; the count and recount data for a sample.

fiber: for the purpose of this paper, any particle counted according to the rules of NIOSH Method 7400.[1]

interlaboratory RSD: the true RSD pertaining to measurements made by a randomly chosen laboratory.

interlaboratory s_r: statistic that describes the variability between/among laboratories (see s_r); an estimate of interlaboratory RSD.

intracounter RSD: the true RSD pertaining to counts by a given counter.

intracounter s_r: statistic that describes the variability of data pairs obtained by a single counter (see s_r); an estimate of intracounter RSD.

intralaboratory RSD: the true RSD pertaining to measurements made by a randomly chosen counter in a given laboratory.

intralaboratory s_r: statistic that describes the variability between/among counters in a laboratory (see s_r); an estimate of intralaboratory RSD.

lognormal: refers to data that, after transformation to the log scale, conforms to the normal distribution.

Poisson component: that part of the variability of count data due to the random distribution of fibers on a filter surface.

pooled: data that are combined into one set because they can be considered to measure the same thing.

power: the ability of a statistical test to detect differences between two groups of data; the probability of rejecting the null hypothesis of a statistical test, as a function of the value assumed by the parameter under study.

QA Coordinator: person in each laboratory responsible for assessing, controlling, and documenting the quality of fiber count data.

recount: (verb) to count the fibers on a sample filter for the second (or third, etc.) time.

recount: (noun) the datum produced as a result of performing a recount.

reference sample: samples that have already been prepared and counted more than once and are recounted for the purpose of calculating or testing a statistic.

reference value: statistic based on the best available data for reference samples.

RSD: the true value of the relative standard deviation, also known as coefficient of variation, or CV; the true standard deviation divided by the true mean.

s_r: estimate of relative standard deviation[2] (see RSD); a statistic that indicates the variability of a set of data.

$s_{r,s}$: subjective component of s_r, that is, s_r with the Poisson component of variability removed. Applied only to interlaboratory variability in this article.

sample categories: groups of samples judged by the counter to be similar for quality control purposes (i.e., the count data are expected to exhibit homogeneous variability). A sample, or the data pair for it, is usually assigned to a category on the basis of fiber loading (fibers/mm^2), but other characteristics should be considered.

sample set: incoming samples grouped together for quality control purposes. Samples are usually assigned to a set based on the fact that they came from a similar source and will be analyzed by one counter.

stopping rule: A rule for counting fibers, stated in NIOSH Method 7400 as follows: "Count enough graticule fields to yield 100 fibers. Count a minimum of 20 fields. Stop at 100 graticule fields regardless of count."

transform: a mathematical operation (e.g., taking a log or square root) performed on all the elements of a set of data.

variability: random differences among numbers in a set, where all the numbers in the set are presumed to measure the same quantity (such as the fibers/mm^2 on a filter.) It is measured by the variance.

Introduction

Many decisions regarding worker exposure to asbestos or other fibrous materials are made on the basis of fiber count data. Therefore, it is important to know the quality of these data. Each laboratory should have a Quality Assurance (QA) Coordinator who is responsible for assessing, controlling, and reporting the quality of the laboratory's fiber count data. This means that the variability and bias of count data should be determined and reduced when possible and that the data should be reported with some indication of its variability.

NIOSH Method 7400[1] outlines procedures for determining variability and reporting it. Methods for other analytes[3] do not discuss such procedures, and the procedures are not necessarily the same as would be used for those other methods. However, optical fiber counting is sufficiently different to warrant special treatment.[4] It is more subjective than other methods, involving complex decisions by the analyst, as well as depending on the visual acuity, training, motivation, and experience of the analyst.

The primary objective of this article is to explain in greater detail the quality control procedures and calculations outlined in NIOSH Method 7400. These recommended procedures are summarized in the "Conclusions and Recommendations." The authors believe that widespread use of the practices described will ultimately improve the quality of fiber count data being reported. However, the simple procedures of NIOSH Method 7400 will not be appropriate for all situations, and some alternative statistical procedures are discussed briefly in the appendix. The investigation of the procedures described in the appendix, and several others, failed to find any for general use that were clearly superior to those presented in the text and in NIOSH Method 7400. Although this article is meant as a practical guide, it concerns research questions still open for discussion.

Variability and Bias

The quality of fiber count data is measured by the variability and bias of that data. There are many elements in the procedure used for determining a fiber count that can contribute to variability and bias. For example, the samples brought into the laboratory for analysis may be of poor quality (poor filter quality, poor sampling conditions, etc.) without anyone being aware of it, causing the variability of the fiber counts to increase. This section discusses some of the sources of variability and bias, particularly those which may be quantified.

One source of variability is the random distribution of fibers over the filter surface, only part of which is analyzed. Random distribution affects even high quality samples. It means that if a filter surface is divided into small, equal areas, there is equal probability of a fiber depositing on any given area, but areas selected at random will not necessarily have the same number of fibers. This distribution of fibers on the filter surface can be described by Poisson statistics.[5] However, it has also been shown that the distribution can be more variable than Poisson, being better described by a negative binomial[6] or lognormal[7] distribution. In NIOSH Method 7400 and in this article, the variability

of a fiber count due to this random distribution is called the Poisson component of variance.

Another source of variability is dependent on the actions of the individual counter. As a counter gains experience, that person will usually adopt routines that decrease variability.[8] On the other hand, counting is still somewhat subjective, and each result is influenced by the "mind-set" of the counter at the time. A statistic that indicates a single counter's ability to reproduce previous results is the intracounter s_r, which is an estimate of intracounter RSD. This is obtained from repeat counts of samples by a counter; the procedure for calculating this statistic is given below.

Repeat counts produced by one counter usually match each other more closely than they match those produced by some other counter for the same sample. This additional source of variation, the difference among counters in a single laboratory, is included in the intralaboratory s_r, an estimate of intralaboratory RSD. Finally, laboratories differ from each other, a source of variation included in interlaboratory RSD. A laboratory is defined in this context as a group of counters who participate in a single QA program. Note that in this discussion, each kind of variability includes those previously mentioned. That is, the Poisson component of variance is included in the intracounter RSD, which is included in intralaboratory RSD, which is included in interlaboratory RSD.

Bias means systematic difference, but the word is used in different ways. Usually, bias is thought of as the difference between a measured result and some "true" value. However, there is no reference method for counting fibers that gives the true value, so this is not the meaning of bias that will be used here. Fiber counting can be "calibrated" only to the extent of using a test slide (HSE/NPL Phase Contrast Test Slide, PTR Optics, Waltham, Massachusetts) to check the microscope or counting reference samples, as discussed below, to compare results with those obtained in the past (usually by other laboratories). While these steps are highly recommended, they are not comparable to the calibration that can be achieved in methods for other analytes.

Bias is used in another way which deserves brief mention here and is discussed at length elsewhere.[9] This use of the word has to do with the linearity of the method over the range for which it is used. NIOSH Method 7400 recommends that loadings be in the range of 100 to 1300 fibers/mm^2 of filter surface area so that reported fiber counts will be proportional to loading. It has been found that loadings below this range often result in concentration estimates that are positively biased, while loadings above this range result in concentration estimates that are negatively biased.[10] It is recognized that measurements are often made at low concentrations in order to document compliance with established exposure limits (e.g., a PEL, REL or TLV). Even if the results of these measurements are positively biased, they are useful indicators of compliance if their upper confidence limits are below these exposure limits.

In the rest of this article, the word bias means consistent differences between the results obtained by a single counter or laboratory and the reference values for those samples. A reference value is usually based on the average result obtained by a group of competent counters. The use of reference samples and reference values for various purposes is discussed throughout the remainder of this article.

In practice, bias may not be easy to distinguish from variability. If one counter always reports much higher counts than the other counters in a laboratory, there is clearly a bias and something should be done about it. But if a consistent difference is smaller, it may not be noticed so easily and becomes a component of the intralaboratory RSD. Each of the three components of variability mentioned above, along with a related kind of bias, is covered in a separate subsection of the "Calculations and Discussion" section below.

The different types of variability, or RSD, are listed in the center column of Figure 1. To the left, some possible sources of variability or bias that contribute to each type are shown. To the right, the uses of each type of RSD are listed. In the "Use" column, an "Indicator" is a statistic that can be included in an analytical report. A better use for these indicators is to identify problems to be corrected, linking the "Use" column back to the "Source" column.

Quality Assurance Program

Determining the different kinds of variability and bias defined above can help document the quality of fiber counts and is the first step to identifying the source of poor quality fiber counts. It is recommended that the QA Coordinator in each laboratory assume the responsibility of determining and tracking these measures of variability and bias. To accomplish this, the QA Coordinator supplies each counter with QA samples in addition to the routine workload. These additional samples are provided blind to the analysts to ensure that they have no idea of the expected answers for the samples. As illustrated in Figure 2, QA samples are of two types: recount samples and reference samples.

Recount samples are the most frequently counted QA samples introduced into the sample stream. These are samples from a set that are counted a second time and constitute at least 10 percent of each set. They provide a check on both sample quality and analyst performance. The count and recount data for the samples are called data pairs and are used to calculate various statistics. Further discussions of recounts are given under the "Intracounter Bias and s_r" and "Sample Quality Test" sections below.

Reference samples should constitute an additional 5 percent of the workload of each counter. These samples come from a previously prepared bank of samples for which the mean and variability have been historically established. These samples are taken from the normal workload and are selected to uniformly cover the range of concentrations normally encountered by the laboratory. Note that samples mounted by the recommended acetone–triacetin technique are not permanent and have been observed to degrade in time periods ranging from six months to several years.[11] Reference samples should therefore be checked for filter integrity and replaced at least every two years.

Reference sample data are used to determine several statistics. First, the data for reference samples already counted by the same counter can be used for determining intracounter bias. This is mentioned in the "Intracounter Bias and s_r" section, although that section is primarily concerned with recount samples and variability. Next, the data for reference samples counted by other counters in the same laboratory can be used to determine the statistics discussed in the "Intralaboratory Bias and s_r" section. Finally, samples that have been counted by analysts from other laboratories in an exchange program can be used to estimate biases between laboratories and interlaboratory s_r. If sufficient data are collected to provide a good estimate of interlaboratory s_r, this value can be used by the laboratories in the exchange group to calculate confidence limits on reported results. The calculation of these statistics is described under the "Interlaboratory Bias and s_r" section.

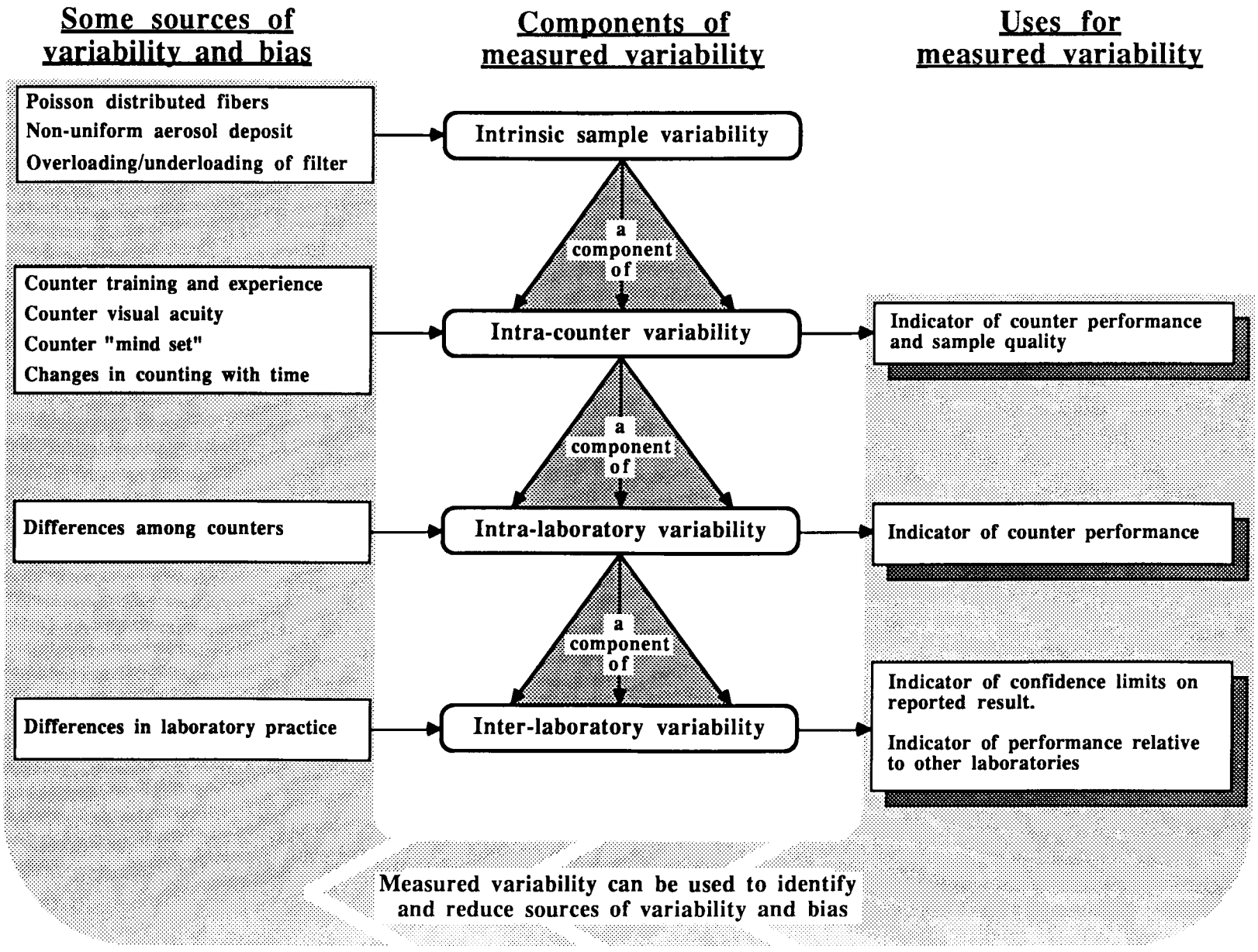

FIGURE 1. Variability components, their sources, and their uses.

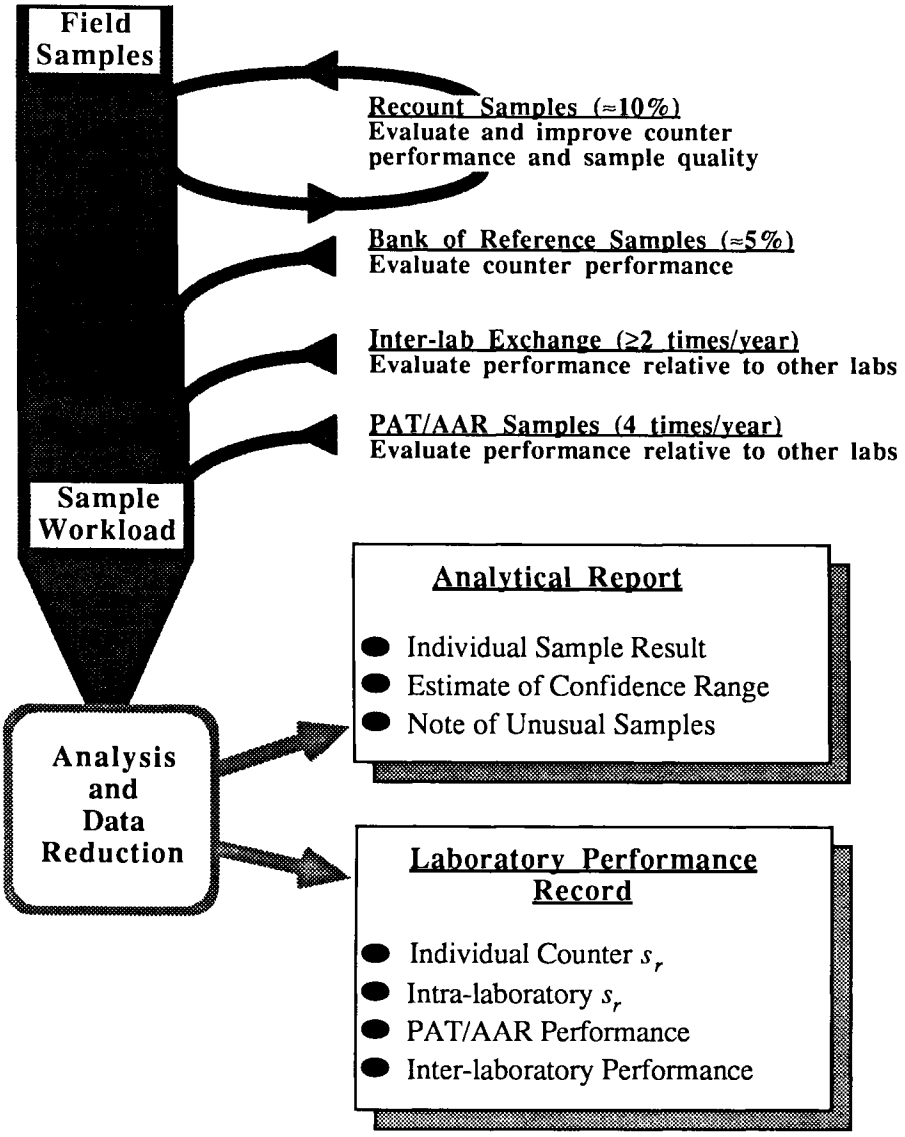

FIGURE 2. The total asbestos sample workload includes quality assurance samples so that data quality can be determined.

Calculations and Discussion

In this section, the terms "sample set" and "sample category" are used frequently. A sample set is any number of samples arriving in a laboratory which are grouped together for quality control purposes. The quality control elements of concern are the assignment of blank and recount samples and some of the calculations discussed in the rest of this article. Ordinarily, the assignment of samples to a set is based on the fact that they came from a similar source and will be analyzed by one counter. For example, samples collected in one building, where there is only one source of fibers, could constitute a set. However, if a laboratory must report results on a daily basis, the samples analyzed in a day may be taken as a set, even if they represent only a subset of the samples from a source.

A sample category refers to samples with similar characteristics as judged by the person counting them. The data pairs obtained from the recounted samples in each category are used to calculate an intracounter s_r for the category; a counter will achieve different s_r's for different categories of samples. The characteristics that can affect variability are primarily fiber loading, as well as fiber type and nonfibrous-particulate loading. The number of categories needed will depend on the variety of samples received by the laboratory, but a good starting point is to differentiate samples into three categories based on loading (total fiber count): 5 to 20 fibers, 20.5 to 50 fibers, and > 50.5 fibers. A laboratory may decide to split these three categories into six, based on whether the particulate loading is considered low or high, or to differentiate categories by fiber type, etc.

The above descriptions of set and category point up several differences. Samples are assigned to the same set only if they are judged to be similar based on sampling conditions; samples are assigned to a category based on what the analyst observes about them through a microscope. Samples are assigned to sets as they arrive, and the set is usually of little interest to the laboratory after the data from the set have been reported; sample data are assigned to categories and may be the basis of statistical calculations. Sample sets may be divided into smaller sets, usually to accommodate other than quality considerations; sample categories are subdivided primarily in the interest of refining quality control calculations.

Intracounter Bias and RSD

This section is primarily concerned with the variability in the

fiber count data of individual counters, as estimated by intracounter s_r. However, if fiber count data include some differences due to bias, those differences should not be incorporated into s_r. Bias for a single counter means bias over time. That is, intracounter bias is the difference in the true fiber counts produced by one counter when counting the same sample at two different times. Of course, bias cannot be distinguished from variability for a single sample, but if enough data are available, bias can be detected.

It is even possible to detect intracounter bias within a single sample set if enough filters are recounted and the bias is great enough. When these conditions apply, bias can be tested using a two-tailed t-test that compares an average difference to a true mean of zero. The procedure is described in the "Bias" section of the Appendix. If bias is detected, its cause should be determined and eliminated before the data are used in any other way.

Intracounter bias may also occur over a longer period of time, analogous to "drift" in an instrumental result. Such bias can be detected from recounts of past reference samples introduced into the sample stream by the QA Coordinator. It is possible to do a t-test, as before, or to simply plot the count differences for reference samples as a function of time (date). If the differences are calculated by subtracting the original count from the recount, positive differences indicate a drift toward higher counts and negative differences toward lower counts. If a plot indicates that drift is occurring, it may be more instructive to simply proceed with an investigation of the possible causes for such a drift instead of testing for significant bias. This is particularly true if the counter's results are diverging from reference values established externally, e.g., in the PAT program.

If bias is not detected in the fiber count data produced by an individual counter, the remaining differences in the data are considered to be variability, specifically, intracounter RSD, or its estimate, intracounter s_r. Theoretically, the minimum intracounter RSD for optimally loaded samples is approximately 0.1, based solely on the distribution of fibers on the filter.[6,12] In practice, other sources of variability increase this minimum. The intracounter s_r statistic should be calculated by the QA Coordinator for each counter in a laboratory so that the quality of sample data can be determined by the "Sample Quality Test" below.

The following is a suggested scheme for initially determining an analyst's intracounter s_r. If available, field samples should be used to determine intracounter s_r since they will match the samples analyzed in the future better than laboratory generated samples would. Otherwise, samples from another source, such as the PAT program, can be used. Based on the initial fiber count, randomly select 15 or more representative field samples in each of three ranges of total fiber count: 5 to 20 fibers, 20.5 to 50 fibers, and >50.5 fibers. (The procedures described here may not be applicable for counts less than 10, but they are recommended unless the applicability of an alternative can be established. The appendix describes some alternative procedures.) Recount each of these (already mounted) samples and record the result as fiber/mm^2 and total fibers. Use the average of the total fibers counted to make the final assignment to a category. Use the fibers/mm^2 value for the calculation of intracounter s_r and in the "Sample Quality Test" below. The calculation of intracounter s_r can only be done in terms of fibers counted if all the counts were done on the same number of fields, a condition that may not be met because of the stopping rule given under "Definitions" above. Calculate the average and standard deviation for each sample. In this case, where there are only two measurements:

$$\bar{x} = \frac{(x_1 + x_2)}{2} \text{ and } s = 0.707 \cdot |x_1 - x_2| \quad (1)$$

where:

x_1, x_2 = independent counts by 1 counter (fiber/mm^2)
$\bar{x}$ = average fiber count (fibers/mm^2)
s = estimate of standard deviation (fibers/mm^2)

Divide the estimate of the standard deviation, s, by the average, $\bar{x}$, to obtain s_r for each sample:

$$s_r = \frac{s}{\bar{x}} \quad (2)$$

Square the s_r values. The square root of the average of these squared s_r values is the pooled intracounter s_r:

$$s_r \text{ (pooled)} = \sqrt{\frac{s_r^2 + s_r^2 + \ldots}{n}} \quad (3)$$

where:

n = the number of s_r values pooled

Values are pooled within each category. The more values that are pooled, the better the pooled s_r estimates the true RSD value for that category. The final step is to convert the pooled s_r value to a pooled s_r value on the square root scale for use in the "Sample Quality Test" (Equation 5) below. This is accomplished by simply dividing it in half:

$$s_r \text{ (pooled, sq root scale)} = \frac{s_r \text{ (pooled, original scale)}}{2} \quad (4)$$

If the true RSD of a measurement is less than 0.3, dividing by 2 gives a good approximation to the true RSD on the square root scale. Thus, Equation 4 is a good approximation for the estimate of RSD, s_r, if enough values of s_r are pooled together.

Table I shows a sample calculation of intracounter s_r using artificial data. The data are in two categories, differentiated by

TABLE I. Determination of Intracounter s_r

Low Range Example: 5 to 20.5 Total Fibers Counted

Orig. Count (f/mm²)	Recount (f/mm²)	Average (f/mm²)	Average* (total)	Std. Dev. (f/mm²)	s_r
18	32	25	20	9.90	0.396
10	5	7.5	6	3.54	0.471
18	9	13.5	11	6.36	0.471
9	21	15.0	12	8.48	0.566

$s_r = \sqrt{\Sigma (s_r^2)/4} = 0.48$

s_r (pooled, square root scale) = 0.48/2 = 0.24

High Range Example: ≥ 50.5 Total Fibers Counted

Orig. Count (f/mm²)	Recount (f/mm²)	Average (f/mm²)	Average* (total)	Std. Dev. (f/mm²)	s_r
318	253	285.5	100	46.0	0.161
90	118	104.0	82	19.8	0.190
68	97	82.5	65	20.5	0.249
108	84	96.0	75	17.0	0.177
83	61	72.0	57	15.6	0.216

s_r (pooled) = $\sqrt{\Sigma (s_r^2)/5} = 0.20$

s_r (pooled, square root scale) = 0.20/2 = 0.10

*Data are assigned to ranges based on the average of total fibers counted

loading range. There are four data pairs given under "Low Range" and five pairs under "High Range." Actual calculations should be based on about 15 or more pairs of counts since values of s_r based on too few counts may be quite biased.[13] The third column of Table I lists the averages of the pairs of results in fibers/mm^2. The fourth column gives the averages in terms of total fibers, which is the basis for assigning the data to the loading categories. The fifth column lists the standard deviations for the count pair. The sixth column gives an s_r for each filter, the result of dividing the average into the standard deviation. When the values in this last column are squared, averaged, and the square root of that result calculated, the result is the pooled relative standard deviation (0.48 for the low-range example). On the square root scale, the pooled s_r is approximately 0.48/2, or 0.24. The calculations for the high range are the same. The first sample in the high range differs from the rest in that only 40 fields were counted instead of 100, which is why the ratio of total fibers counted to fibers/mm^2 is not the same.

The initial estimate of pooled, intracounter s_r may be based on a limited amount of data. However, the "Sample Quality Test" given below requires that 10 percent of the field samples in each set be recounted, so the number of data pairs available for the s_r calculation constantly increases. Using this data to frequently recalculate s_r for each counter in the laboratory is a valuable quality assessment tool in itself. The pooled s_r value obtained from Equation 3 should have immediate meaning to most analysts since it is a relative standard deviation on the original scale.

There are only three reasons for not including all of these new data pairs in a recalculation of the intracounter s_r. The first reason is that there are so many new data pairs (more than one-fifth of all the available data pairs) that the value of s_r will depend primarily on that data. In that case, a randomly selected subset of the data may be chosen. The second reason applies when there are definite reasons for suspecting that the samples actually belong in a new category. In that case, a new category could be created. The third reason is that individual values of s_r differ greatly from the others in a category even though the samples seem to fit in that category. An s_r that differs greatly from the others is sufficient reason to exclude it from the pooled data. There are formal tests for poolability of standard deviations[14] (which are approximately applicable for relative standard deviations), but they will not be discussed here.

As additional data pairs from the "Sample Quality Test" become available, that data can be used either to increase the amount of data in each sample category or to create more sample categories. Several decisions need to be made. A first consideration is that, since data generated by a counter in the past may not represent the current capability of that counter, a balance should be struck between eliminating older data and having enough data to determine s_r accurately. When there are 100 data pairs for a sample category, they should be assigned to smaller ranges (5 to 15, 15.5 to 25 fibers counted, etc.). The top range should be 80 to 100 fibers, even if the 100 fibers are counted in less than 100 fields, since the s_r values throughout this range are approximately equal. And when the number of ranges for a given sample type exceeds 5, it is advisable to establish a curve of s_r as a function of sample loading and to estimate s_r values from this curve.

Sample Quality Test

All samples and the procedures used to collect, transport, store, mount, and analyze them should be examined often to detect any problems. If there is reason to suspect a problem for any sample, the data for that sample should not be reported or used in any further calculations. Otherwise, sample data should only be rejected if they fail the test given here.

The intracounter s_r is needed to test the quality of the results obtained for a sample set. Also needed are data pairs for some of the samples in the set. NIOSH Method 7400 states that 10 percent of the samples in every set should be randomly selected for recounting. The QA Coordinator, or someone designated by the coordinator, should select and relabel the samples. The samples may then be combined with other relabeled samples, e.g., those from other counters for determining intralaboratory s_r, and given to the counter (or counters) for recounting. The data for each of the samples that have been counted twice by the same counter are then tested as follows:

$$\text{If } |y_1 - y_2| > 2.8\,\bar{y} \cdot s_r \text{ (pooled, sq root scale), reject sample} \qquad (5)$$

where:

$$y_1 = \sqrt{\text{original count}}$$

$$y_2 = \sqrt{\text{recount}}$$

$$\bar{y} = \frac{y_1 + y_2}{2}$$

This test differs from that given in earlier versions of NIOSH Method 7400 primarily in that the data are converted to the square root scale first. Revision 3 of NIOSH Method 7400 gives the test in the form given here. The derivation of Equation 5 is explained in "Remarks on Equation 5" in the appendix, and alternative tests are also described in the appendix. Table II gives two examples of how to evaluate data pairs using the values of s_r from Table I. Since the test is done on the square root scale, the intracounter s_r [pooled, square root scale] from Equation 4 is used.

The justification for converting to the square root scale is given in the "Distribution of Fiber Count Data" section of the Appendix. The conclusion given there is that Equation 5, which assumes normality, will be more appropriate on the square root scale for most loadings. If the data are left on the original scale, as in Revisions 1 and 2 of NIOSH Method 7400, the test results will only be different when the samples are lightly loaded and the counts differ widely. Even if the data are better described as lognormally distributed, as long as the intracounter RSD is less than 0.3, there is little difference in the power of the test in Equation 5 and an analogous test based on log-normality as described under "Test on the Log Scale" in the Appendix.

TABLE II. Evaluating Recounts Using Intracounter s_r

Low Range	
Square Root of First Count: (10 fibers = 13 f/mm^2)	3.57
Square Root of Second Count (29 fibers = 37 f/mm^2)	6.08
Average ($\bar{y}$)	4.82
Difference:	2.51
Historical s_r (low range):	0.24
$2.8 \cdot \bar{y} \cdot s_r$	3.24
Difference < 3.24:	accept
High Range	
Square Root of First Count: (65 fibers = 83 f/mm^2)	9.10
Square Root of Second Count (46 fibers = 59 f/mm^2)	7.65
Average ($\bar{y}$):	8.38
Difference:	1.44
Historical s_r (low range):	0.10
$2.8 \cdot \bar{y} \cdot s_r$	2.35
Difference < 2.35:	accept

TABLE III. Determining the Need for 100% Recounts

Number of Samples in 10% Recount	Number of Rejected Samples (Equation 5) that Indicate Need for 100% Recount
2– 7	≥ 2
8–16	≥ 3
17–28	≥ 4
29–40	≥ 5

If 15 samples are recounted, there is an 80% chance of deciding that the RSD for the set is greater than the established (historical) intracounter RSD, when the true set RSD is 75% greater than the established intracounter RSD.

The factor 2.8 in Equation 5 provides an upper limit for $|y_1 - y_2|$ that will be exceeded not more than 5 percent of the time when the actual RSD for the sample set equals the counter's established RSD. When that limit is exceeded, the sample is rejected. About 5 percent of the samples recounted will be rejected even when all the samples are of good quality. However, when too many of the recounted samples are rejected by Equation 5, then the RSD for this sample set may be higher than the established intracounter RSD. That conclusion can be drawn when the number of rejected samples match that listed in Table III. This test, based on all the count-recount data pairs that have been determined for a sample set, is an extension of the test in NIOSH Method 7400, which is performed for each recounted sample.

It should be noted that this test does not have high power to detect differences between the sample set RSD and the established intracounter RSD when the number of recounted samples is small. For example, even if as many as 15 samples are recounted from a sample set that has an actual RSD that is 75 percent greater than the established intracounter RSD, there is only an 80 percent chance that this test will detect a difference. The chance of detecting a difference drops for fewer recounts, becoming 40 percent when five samples in a set are recounted. Although the Appendix gives some alternatives to Equation 5, the alternative tests have similar power. Equation 5 is simple to apply, once the recounts have been performed, and it will confirm that there is a problem with the worst data.

When a sample set has high variability as indicated by Equation 5 and Table III, there are several courses of action, which depend primarily on how the data are to be used. In some cases, it is desirable to identify and eliminate the source of variation. If the microscope is functioning properly and the counter has not changed procedures, then it may be necessary to obtain new samples. The samples may have been collected under unusual circumstances, for example, with a highly charged cassette, or they may have a matrix that is very difficult to count, or they may have been improperly mounted. In other cases, it may be sufficient to report that the variability was high. Often, it is advisable to recount all the samples and perform the "sample quality test" for each. When a data pair is rejected by Equation 5, both numbers should be reported with the note that the data are more variable than normal for that kind of sample. The reason for reporting the data is that the data may still be useful for decision making. For example, when the count and recount differ greatly but both are below the exposure limit, a decision may still be made using these data.

Intralaboratory Bias and RSD

A laboratory is defined here as a number of analysts grouped together for quality assurance purposes. In a laboratory with only one counter, intralaboratory bias and intralaboratory s_r are the same as intracounter bias and intracounter s_r, so the following discussion assumes that the laboratory has two or more counters.

Intralaboratory bias can be detected by having all the counters in the laboratory count the same reference samples. As mentioned, these reference samples can be already mounted and counted samples from the laboratory's sample stream. If the average result obtained by one counter for the samples in one or more categories is somewhat higher or lower than the average for the other counters, then there may be bias. Instead of determining if the difference is significant, it may be just as simple, and much more informative, to investigate the possible causes. There may be one counter who is doing something obviously different, such as using a different graticule or misadjusting the microscope. At other times, the problem will be more subtle, and it may be necessary to statistically design an experiment to determine the cause.

Intralaboratory s_r must be calculated based on reference samples that have been counted by all the counters in the laboratory. The calculation is the same as for intracounter s_r except that the formula for standard deviation given in Equation 1 cannot be used when there are more than two counters. For three or more counters, use the standard formula:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (6)$$

Then s_r for each sample is calculated as in Equation 2 and pooled as in Equation 3.

The expected level of performance by a competent laboratory may depend on the types of samples. For occupational samples in the range 100–1300 fibers/mm^2, intralaboratory s_r values of 0.17–0.22 have been achieved by laboratories with good quality assurance programs.[12,15] If all the samples evaluated by a laboratory are from low concentration levels, then the s_r estimate may be larger simply due to the higher intrinsic (relative) variability of the samples.

The intralaboratory s_r is used to track the differences between the counters in the laboratory. If intralaboratory s_r changes significantly, the cause should be investigated. Although increases in intralaboratory s_r are of most concern, the reasons for decreases are also of interest. Increased intralaboratory s_r may indicate the presence of bias.

Interlaboratory Bias and RSD

Most analytical methods, including those in the *NIOSH Manual of Analytical Methods*,[3] do not address differences between laboratories. Such differences are considered to be interlaboratory biases, and it is assumed that they are negligible because of the ready availability and widespread use of highly accurate calibration standards and reference materials. Such is not the case for fiber counting. As mentioned in "Variability and Bias" above, calibration standards for fiber counting are not readily available, and the results for fiber counting as performed by human counters are not easily adjusted.

In theory, it should be possible to eliminate fiber count differences among laboratories. In practice, these differences are difficult to eliminate and the reasons for them difficult to identify. Ideally, a reference sample counted by the OSHA laboratory in Salt Lake City, Utah, would be available since that laboratory determines regulatory compliance. In practice, one way a laboratory can approach this ideal is by participating in the PAT program; the OSHA laboratory consistently reports PAT asbestos results close to the reference value. However, PAT samples are laboratory-generated, are not identical, and are not, therefore,

ideal reference samples.

Another way to approach this ideal would be to exchange slides containing mounted field samples among laboratories. This is less than ideal because the OSHA laboratory cannot participate in each of these round robins, and it is logistically difficult for more than six laboratories to participate in such a round robin. Nevertheless, OSHA now requires laboratories counting fibers for compliance purposes to exchange samples from their workload with other laboratories and to analyze and post the results.[16] Each laboratory is required to exchange samples at least twice a year with at least two other laboratories.

A quality assurance program for fiber counting should use both types of reference sample: mounted samples exchanged with other laboratories and the audit samples set quarterly to participants in either the AAR or the PAT program. These samples provide a means of detecting biases. In the case of the PAT program, a laboratory can compare its result with the average result of the reference laboratories, a group of approximately 80 highly qualified laboratories. Note that the analysis of PAT or AAR samples does not fill the OSHA requirement to exchange field samples with other laboratories.

Although these procedures bring laboratories into closer agreement, differences can persist, and NIOSH Method 7400 refers to these differences as interlaboratory s_r. These differences may be considered as residual biases by some, but the differences may not be constant, and finding an assignable cause is difficult. This is not meant to imply that these differences should be accepted; the purpose of the interlaboratory s_r is to quantify differences and to make them known by using them to assign confidence limits to reported fiber counts.

The determination of interlaboratory s_r is based on recounts done by persons in different laboratories. To obtain sufficient data to calculate interlaboratory s_r, a group of laboratories will want to exchange samples at a greater rate than that required by OSHA. A simple plan calls for approximately 20 samples to be counted in each laboratory in the group. The samples will already have been mounted and counted in one of the laboratories. Each laboratory can contribute equally to the number of samples, say five from each of four laboratories. The samples contributed should be representative of the laboratory's workload as far as loading and other characteristics. As the already-mounted samples arrive in a laboratory, each is counted by a different (randomly chosen) counter for that laboratory. The samples are then sent to the next laboratory and the count data sent to the coordinating laboratory.

The calculation of interlaboratory s_r is performed by the coordinating laboratory as follows. For each sample, the results of the participating laboratories should be averaged, and a standard deviation (Equation 6) and s_r (Equation 2) calculated. Note that, to be consistent with NIOSH Method 7400, the calculations are done in total fibers, not fibers/mm^2. Before pooling these interlaboratory s_r (total) values, the Poisson component of variability is removed to obtain the subjective interlaboratory component, $s_{r,s}$, as follows:

$$s_{r,s} = \sqrt{s_r^2\,(\text{total}) - s_r^2\,(\text{Poisson})} = \sqrt{s_r^2\,(\text{total}) - (1/\text{count})} \qquad (7)$$

where:

count = the average of the counts reported by the laboratories for the sample.

These $s_{r,s}$ values are then pooled in the same way as was done for intracounter s_r in Table I, but the result does not need to be converted to the square root scale. Note that the removal of the Poisson component in Equation 7 means that the samples used for determining interlaboratory s_r need not be put into categories based on fiber loading. Some groups of laboratories may wish to put especially difficult samples, e.g., asbestos cement dust, into a separate category and calculate an interlaboratory $s_{r,s}$ just for use with that kind of sample.

The simple plan just given for arriving at interlaboratory $s_{r,s}$ should serve the needs of most groups. There may be questions about how to pick samples, how to assign counters, or whether the data from samples mounted in different laboratories are poolable. If these are a concern, the advice of a statistician should be sought before data are collected or analyzed. Note that it is worthwhile to reduce differences among laboratories[17] since data that are biased will increase the interlaboratory $s_{r,s}$ and negate the advantage of determining it. If a laboratory is not required to exchange samples under the OSHA rule, or if not enough data have been collected since forming the laboratory group, or if there appear to be unresolved problems with the data collected, the conservatively high estimate of variability given in NIOSH Method 7400 can be used, as discussed below.

For most purposes, the interlaboratory $s_{r,s}$, not the intracounter or intralaboratory s_r, is used to calculate confidence limits on reported data. Intracounter s_r may be used for special studies where the purpose is to measure relatively small differences in fiber counts and for which all analyses are performed by one counter. Similarly, when counts for a study are all produced by one laboratory, but not necessarily one counter, it is appropriate to use the intralaboratory s_r. When comparisons are being made between laboratories or when results are to be compared to the PEL or other criterion (REL, TLV®), the interlaboratory $s_{r,s}$ should be used. In general, when comparing a fiber count to the OSHA PEL, one needs to be confident that a fiber count made by any other laboratory, including the OSHA laboratory, will produce a similar result. The following discussion uses interlaboratory $s_{r,s}$ to estimate confidence limits.

When the interlaboratory $s_{r,s}$ has been established in a sample exchange program, it is appropriate to use it to calculate confidence limits for each sample instead of using the graph in NIOSH Method 7400. The following formulae recombine the Poisson component of variability with $s_{r,s}$ to give the 90 percent confidence limits:

$$\text{UCL} = \frac{2x + 2.25 + \sqrt{(2.25 + 2x)^2 - 4\,(1 - 2.25s_{r,s}^2)\,x^2}}{2(1 - 2.25s_{r,s}^2)} \qquad (8)$$

$$\text{LCL} = \frac{2x + 4 - \sqrt{(4 + 2x)^2 - 4(1 - 4s_{r,s}^2)\,x^2}}{2(1 - 4s_{r,s}^2)} \qquad (9)$$

where:

$s_{r,s}$ = subjective interlaboratory s_r
x = total fibers counted on sample
UCL = upper confidence limit
LCL = lower confidence limit

These formulae were derived from the work of Ogden[12] by extrapolating to higher $s_{r,s}$ and to different values of x. The UCL formula is not valid for values of $s_{r,s}$ greater than 0.67, and the LCL formula is not valid for values of $s_{r,s}$ greater than 0.5. The confidence limits calculated for a fiber count provide a range of values within which the mean count of a group of competent laboratories is expected to fall 90 percent of the time.

These confidence limits are in units of total fibers to be consistent with NIOSH Method 7400. Any conversions to concentration units are performed on the confidence limits in the same way as would be done for x, the original fiber count. As an example, if the interlaboratory $s_{r,s}$ is 0.25 and 24 fibers have been counted on a sample, the above equations give 13.8 fibers and 42.8 fibers as the confidence limits. If these fibers were counted in 100 fields of 0.00785 mm^2 on a 25-mm filter and the air volume of the sample was 500 liters, then the confidence limits on the air concentration are 0.014 and 0.042 fibers/cc. Since the primary use of a fiber count is to compare with an exposure limit such as the OSHA PEL, the upper confidence limit is of most interest. For this example, the upper confidence limit on the air concentration is 0.042 fibers/cc.

NIOSH Method 7400 gives an example calculation of confidence limits when 24 fibers, the same as in the example just given, have been counted. Based on the graph in the method, the resulting confidence limits are 0.011 and 0.077 fibers/cc. We can no longer be confident that the fiber concentration, nominally 0.024 fibers/cc, is less than 0.05, 0.06, or even 0.07 fibers/cc. The confidence band is wider because the graph is based on an $s_{r,s}$ of 0.45.[1,15] The graph was produced by substituting 0.45 for $s_{r,s}$ and substituting various values for x in Equations 8 and 9, then plotting UCL and LCL as a percentage of the substituted x. The graph, or the conservatively high estimate of $s_{r,s}$ (0.45), can be used when the data for calculating interlaboratory $s_{r,s}$ are not available. When using the graph in NIOSH Method 7400, note that it gives the percent differences to be added to and subtracted from the original fiber count. These percent differences are relatively constant (+213% and −49%) for fiber counts above 30 because the Poisson component of variability becomes less important. Thus, the confidence limits for 50 fibers counted are 25 and 157 fibers.

A group of laboratories exchanging samples may achieve an interlaboratory $s_{r,s}$ less than the estimate given in NIOSH Method 7400. It is then reasonable for each laboratory in the group to use confidence limits based on that $s_{r,s}$ when reporting their sample results. The lower $s_{r,s}$ indicates that these laboratories have successfully lowered their variability. The laboratories with smaller $s_{r,s}$ values can report results with tighter confidence limits, thereby increasing the number of definitive results (those definitely above or below a given level). It is possible that all the laboratories in a given exchange group, especially a small group, agree well with each other, but that they are all biased. However, participation of laboratories and counters in audit programs such as the PAT program and the AAR reduces the likelihood of that.

Usually, intracounter s_r is about one-half of the interlaboratory s_r.[15] If the intracounter s_r for a given set of samples is a little higher than the historically established s_r, the effect on interlaboratory s_r and sample confidence limits is negligible, particularly if the graph in NIOSH Method 7400 is being used. However, if many of the samples in a set fail the criterion in Equation 5, the confidence limits given by Equations 8 and 9 are not applicable.

Samples are taken from an environment in order to provide an estimate of the airborne concentration in that environment. The above discussion has treated the analytical variability of individual samples at length. Since the environmental variability can often overshadow the analytical variability, strategies have been developed by others to improve the estimates of environmental concentrations.[18] One important strategy is to take multiple samples. If K samples are taken, the relative standard deviation due to environmental variability is reduced by the factor $1/\sqrt{K}$. Multiple samples taken from the same environment and analyzed by a laboratory can also reduce the intralaboratory s_r by the same $1/\sqrt{K}$ factor. However, the largest component of the analytical variability is the between laboratory variability, which is not reduced by the use of multiple samples. Therefore, multiple samples are useful for improving the estimate of airborne concentration from an environment but are not generally useful for reducing the analytical variability. This emphasizes that reducing the interlaboratory s_r through sample exchanges is of primary importance in improving the confidence limits that a laboratory can report for any given sample.

Conclusions and Recommendations

Each laboratory should have a Quality Assurance Coordinator responsible for assessing the quality of fiber count data. The coordinator will include blind reference samples into the laboratory's sample stream, obtain the recount data, and use it to calculate intracounter s_r and intralaboratory s_r. Interlaboratory s_r can also be calculated based on sample exchanges with other laboratories. There are several ways to determine these components of variance, but the simple calculations discussed in this article and outlined in the current version of NIOSH Method 7400 will suffice for most laboratories.

To summarize:

- Calculate a Student's t-statistic for (recount–count) differences in the sample set in order to detect statistically significant bias. If bias exists, determine its cause and correct it.
- Calculate an intracounter s_r (Equations 1–4, Table I) for each counter and each sample category (i.e., loading range, type of interference, etc.) based on recounts by the same counter on recently counted sample slides. Plot s_r as a function of loading if enough data are available. Recounting older samples provides a method of detecting drift in a counter's performance.
- Apply the "Sample Quality Test" (Equation 5, Table II) to data pairs obtained by recounting 10 percent of the samples in each set. The recount is performed by the same counter on slides counted within the last day. If the number of samples "rejected" exceeds the number in Table III, reject the sample set or recount all the samples and apply the test to each sample individually.
- Calculate intralaboratory s_r based on recounts by all counters. Check for changes in s_r and for biases between counters.
- Calculate interlaboratory $s_{r,s}$ based on recounts of mounted field samples already counted in two or more other laboratories (Equation 7). Be alert to bias and, if detected, correct it and recalculate interlaboratory $s_{r,s}$. If recount data are not yet available, an $s_{r,s}$ of 0.45 is assumed.
- Use the interlaboratory $s_{r,s}$ and the total number of fibers counted on a sample to calculate confidence limits (Equations 8 and 9) for that fiber count. If the $s_{r,s}$ of 0.45 has been assumed, use the graph in NIOSH Method 7400. Report these confidence limits with the fiber count.

Intracounter s_r for properly loaded samples that have low background can theoretically approximate a value as low as 0.1. An intralaboratory s_r of 0.17 − 0.22 has been achieved by competent laboratories. NIOSH Method 7400 gives a conservatively high estimate of 0.45 for the subjective component of interlaboratory s_r.

Appendix

In this appendix, some topics mentioned in the main text are

further explained. The topics are: how to test for bias; background on Equation 5 (the "Sample Quality Test"); the distribution of fiber count data and justification for the square root transform; and mention of several alternatives to the use of intracounter s_r and Equation 5 for testing data.

Bias

The Student's t distribution can be used to test for bias in count data. Specifically, we are interested in seeing if the true average difference (recount–count) is equal to zero. If there are k data pairs, the first step is to compute the k differences:

$$d = \sqrt{x_2} - \sqrt{x_1}$$

where:

x_2 = the recount
x_1 = the original count

The reason for using the square root is discussed in the "Distribution of Fiber Count Data" section below. The next step is to calculate

$$t = \frac{\bar{d}}{s_d/\sqrt{k}} \quad (10)$$

where:

$\bar{d}$ = the average of the differences
s_d = the estimated standard deviation of the differences

The value obtained for the t statistic should be compared with the value that is located in the row and column of a Student's t table corresponding to $k-1$ degrees of freedom and the 0.05 significance level (for the typical two-tailed table). If the value obtained in Equation 10 is greater than the table value, then bias can be assumed to be responsible.

Remarks on Equation 5

For a normal random sample of size n, the probability density function has been given for $s/\bar{x}$,[19] where s denotes the sample standard deviation and $\bar{x}$ is the sample mean. Although the function was based on a sample standard deviation defined as:

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (11)$$

it is easy to modify the result for s defined as:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (12)$$

If Prob ($\bar{x} < 0$) is very small, the noncentral t distribution describes the distribution of $\sqrt{n} \cdot \bar{x}/s$. To arrive at Equation 5 in the text, the 95 percent upper confidence limit on the sample s_r was determined as a multiple of the true s_r value.[19,20] This multiplier depends somewhat on the value of the true s_r. However, for $s_r < 0.3$ on the original scale, the multiplier is less than or equal to 2.8, the value used in Equation 5. That is:

$$\Pr\left(\frac{\left(\frac{1}{\sqrt{2}}|y_1 - y_2|\right) \div \bar{y}}{s_r \text{ (pooled, sqrt scale)}} < \frac{2.8}{\sqrt{2}}\right) \geq 0.95 \quad (13)$$

Distribution of Fiber Count Data

Equation 5 assumes that, for fibers/mm² data, $\sqrt{x_2} - \sqrt{x_1}$ data are approximately normally distributed and have an RSD corresponding to that computed for the category in which the average for the two fiber counts falls. The facts associated with Poisson deposition of the fibers and Poisson distribution of the actual counted fibers are as follows: 1) if the actual count is at least 15 fibers in 100 fields, the counts are approximately normal on both the original and square root scales and 2) the absolute standard deviation of the counts is approximately constant on the square root scale. Thus, even if the counts themselves (which reflect both deposition of fibers on the filter and the counting process) are Poisson distributed, square root transformation of those counts does not assure constant RSD. Also, the data used in the test are in fibers/mm², and these do not, in general, follow a Poisson distribution, even if the deposition itself is Poisson. Furthermore, even if the deposition is Poisson, the fact that a particular person counts fibers and uses a particular microscope, introduces variability beyond that expected from the Poisson. These concerns are discussed next: first with regard to normality, then with regard to constant RSD.

The first issue is normality. Assuming Poisson deposition of fibers and Poisson counts of these fibers, what is the degree of non-normality of the fibers/mm² data on the original, square root, and log scales? Application of the stopping rule to samples with different loadings results in variable numbers of fields being counted as well as variable numbers of fibers being counted. When loadings are low, a variable number of fibers in a constant area of the filter (100 fields) is counted. Likewise, when loadings are high, a variable number of fibers in a constant area of the filter (20 fields) is counted. When loadings are intermediate, approximately 100 fibers are counted in a variable area (20–100 fields) of the filter. A fiber density of ≤ 0.8 fibers/field corresponds to low loading since the probability is high that fewer than 100 fibers will be counted in 100 fields. A fiber density of > 8 fibers/field corresponds to high loading since > 100 fibers will have been counted in 20 fields or less. Fiber densities in the range of 0.8 to 8 fibers/field can be taken as the intermediate range.

Thus, in the cases of low and high loadings as just defined, the area of the filter examined is constant, and all the variability is in the number of fibers counted. Since the fibers counted on the filter are assumed to be Poisson, the fiber/mm² measurement is also from a Poisson distribution whose mean is $100 \cdot \mu$ ($20 \cdot \mu$ for high loadings), where μ is the true mean count for one field. For sufficiently high Poisson counts (a count of at least 15 fibers in 100 fields), as stated above, the square roots and the original counts are approximately normally distributed, as are the fiber/mm² values for the low and high loadings. The logarithms of these counts, computer simulations indicate, are not normally distributed for low loadings, but are for high loadings.

For intermediate loadings, the complication of having variable numbers of both fibers and fields counted, makes it difficult to make any statement about the correct transformation to normality of the fiber/mm² values based on theory alone. Computer simulations suggest that the transformation that is best at producing a normal distribution in this loading range depends on the loading. For 1 fiber/field, the untransformed data are not normally distributed, the square root transformed data are marginally normal, and the log transformed data are nearly normal. At 2 fibers/field, only the log transformed data can be said to be normally distributed. At 4 fibers/field, the log transformed data are normal, and the square root transformed data are normal again, but the original data are not normal.

For the entire range under consideration, provided at least 15 fibers are counted, the square root and log transformations in-

duce a normal distribution of the fiber/mm^2 data at most loadings. Thus, for statistical tests that assume normality, such as Equation 5, the result will be accurate on the square root scale for most loadings. However, for loadings of about 2 fibers/field, where the log transformation is more nearly normal than the square root, a test based on the log transform seems more reasonable than one based on the square root. Such a test is also discussed below. The above remarks assume that the counted fibers are approximately Poisson distributed. The authors have not studied the sensitivity to non-normality of the test in Equation 5.

The second issue, after normality, is that of constant RSD within each sample category. For Poisson counts, the square root transform induces approximately constant variance, whatever the loading. Thus, if the categories cover a relatively wide range of loadings and most of the variability is Poisson variability due to counts, a version of Equation 5 based on absolute standard deviation rather than s_r might be useful. This approach is mentioned below. On the other hand, if the sample categories are relatively narrow loading ranges, for counts that have considerable variability in excess of Poisson variability, the relative standard deviation will be slightly more constant than the absolute standard deviation. For moderately excessive variability and narrow bands, the two measures are about equally variable.

Test on the Log Scale

The preceding section argues that fiber count data are normal on the square root scale. However, as mentioned in "Variability and Bias," the distribution has also been described as lognormal in the literature. Thus, it is reasonable to compare the test in Equation 5, which assumes square root normality, with an analogous test based on log normality. A test analogous to Equation 5, but based on the log scale, is:

$$\text{If } |\ln(x_1) - \ln(x_2)| > 2.77 \cdot s_r, \text{ reject the sample} \qquad (14)$$

where:

s_r = the established intracounter s_r on the original scale

Table IV provides a comparison of the power of the tests in Equations 5 and 14, assuming an historical s_r of 0.30 on the original scale. In Table IV, the first column is the ratio of the true RSD of the count–recount data pairs being tested to the established intracounter RSD. The table shows that neither the square root test of Equation 5 (column 2) nor the log test of Equation 14 (column 3) has great power, even when the test being used matches the distribution of the data. For example, when the true RSD of the data pairs being tested is 2.92 times the established intracounter RSD of 0.30, the tests have a 43 percent (square root) and 56 percent (log) chance of accepting the data. The aggregate test using Table III has more power.

The fourth column of Table IV shows the probability of accepting the data when the square root test is used on lognormally distributed data, again assuming an historical RSD of 0.30 on the original scale. The table shows, however, that never is there an appreciable difference in power from the test based on the logs of the data. We conclude that the square root test can be used even when the data are lognormally distributed. However, as stated above, there could be instances, for example, when the true filter deposition is about 2 fibers/field, that use of Equation 14 would be more appropriate.

TABLE IV. Comparison of Outlier Tests Assuming Square Root and Lognormal Distributions

RSD(true)* / RSD(hist)	Probability of Accepting Samples		
	Sq Root Test & Sq Root Dist.	Log Test & Log Dist.	Sq Root Test & Log Dist.
1.0	0.95	0.95	0.95
1.5	0.80	0.83	0.83
2.92	0.43	0.56	0.57
3.74	0.29	0.48	0.51
4.70	0.11	0.42	0.45

*RSD(hist) = 0.3

Test Based on Absolute Standard Deviation

A test can be based on the historical standard deviation instead of the historically established intracounter s_r. The data pairs to be used for determining the standard deviation should be grouped into categories just as was done to determine intracounter s_r. For these calculations, the fibers/mm^2 data are converted to their square roots (y_1 and y_2) immediately since there is no simple way to convert the standard deviation to the square root scale as there is for s_r. The calculation of standard deviation for each sample is similar to Equation 1 ($s = 0.707\ |y_1 - y_2|$), and the standard deviations are pooled as in Equation 3. As an example, the low-range data in Table I yield an s[pooled, square root scale] of 0.929. The data to be tested are also converted to the square root scale:

$$\text{If } |y_1 - y_2| > 2.77 \cdot s \text{ (pooled, sqrt scale), reject sample} \qquad (15)$$

The factor 2.77 provides an upper limit for $|y_1 - y_2|$ that will be exceeded not more than 5 percent of the time. Factors of 3.64 and 4.65 correspond to 1 percent and 0.1 percent upper limits.[21] This test is more powerful (will correctly reject more samples) than the s_r test in Equation 5,[22] but for most cases the power difference is relatively small. The test based on Equation 15 is appropriate when most of the variability of the counts is due to the Poisson distribution of counts.

Test Based on Pooled Sample Set s_r

Another way to use the data when every sample in a set has been recounted is to calculate a pooled s_r for each sample category for comparison to the established intracounter s_r for that category. If it can be established that the RSD corresponding to the pooled s_r is no greater than the RSD corresponding to the established intracounter s_r, then the results can all be reported. If the former RSD is appreciably greater than the established intracounter s_r, it may be necessary to widen the reported confidence limits for the samples. This test has some difficulties. If the samples in the set fall into different loading categories there will be more than one pair of s_r values to compare, and for categories with only a few samples, there will be little power to detect differences in s_r. Also, this test is not as simple or flexible as the sample-by-sample test given in Equation 5.

Test Based on Means

Tests of sample quality could be based on means rather than variability. One such test is a chi-square test that seeks to determine if means μ_1 and μ_2 are equivalent, where μ_1 and μ_2 are the true means as estimated by the count and recount, x_1 and x_2. This test also requires that the intracounter s_r be determined. If a minimum intracounter RSD (Poisson component only) is assumed, the true means of a pair of counts with a true average of 20 would have to differ by 11 or more for the test to have 95 percent power to reject the data. For an average of 40 counts, the difference of the true means would have to be 13. The differences can be decreased to 8.5 and 10 fibers, respectively, if

80 percent power is acceptable. These differences would have to be even greater for larger, more realistic, intracounter RSD. Thus, as with the tests based solely on variability, this test does not have great power to detect differences.

In fact, this is the same test as that given in Equation 15. In using this as a test of means, we are assuming a different alternative, namely, that the true means corresponding to the two counts may differ. In using it as a test of variances, we are testing whether the absolute variances differ. Since the fields counted on a slide are a random selection, it may make more sense to view the alternative hypothesis for the test in Equation 15 as inequality of variances associated with two random samples from the same source.

Acknowledgment

The authors thank those who contributed to the evolution of this article, including the following: M. Attfield, P. Bierbaum, L. Bloomfield, K. Busch, M. Carmel, D. Crane, T. Dinh, L. Doemeny, P. Eller, R. Hartle, R. Zumwalde, and an anonymous reviewer for the journal.

References

1. Carter, J.; Taylor, D.; Baron, P.A.: Fibers, Method 7400, Revision #3: 4/15/89. In: NIOSH Manual of Analytical Methods, 3rd ed. P.M. Eller, Ed. DHHS (NIOSH) Pub. No. 84–100. Cincinnati, OH (1984).
2. International Union of Pure and Applied Chemistry, Analytical Chemistry Division: Compendium of Analytical Nomenclature Definition Rules 1987, 2nd ed., p. 5. H. Freiser and G. Nancollas, Ed. Blackwell Scientific Publishers, Oxford (1987).
3. NIOSH Manual of Analytical Methods, 3rd ed. P.M. Eller, Ed. DHHS (NIOSH) Pub. No. 84–100. Cincinnati, OH (1984).
4. Walton, W.H.: The Nature, Hazards and Assessment of Occupational Exposure to Airborne Asbestos Dust: A Review. Ann. Occup. Hyg. 25(2):203 (1982).
5. Rajhans, G.S.; Sullivan, J.L.: Asbestos Sampling and Analysis, pp. 115–128. Ann Arbor Science Publishers Inc., Ann Arbor, MI (1981).
6. Attfield, M.: Investigation of Stopping Rules in Fiber Counting, and the Development of a New Rule. Ph.D. dissertation (unpublished). West Virginia University, Morgantown, WV (1986).
7. Chesson, J.; Rosenberg, J.: Statistical Evaluation of the Performance of the TEM Clearance Procedure. EPA Final Report Contract PO 7C3071NAST (1987).
8. Schlecht, P.C.; Shulman, S.A.: Performance of Asbestos Fiber Counting Laboratories in the NIOSH Proficiency Analytical Testing (PAT) Program. Am. Ind. Hyg. Assoc. J. 47:259 (1986).
9. Baron, P.A.: Asbestos Analysis—NIOSH Method 7400. Appl. Ind. Hyg. 2:R8 (1987).
10. Cherrie, J.; Jones, A.D.; Johnston, A.M.: The Influence of Fiber Density on the Assessment of Fiber Concentration Using the Membrane Filter Method. Am. Ind. Hyg. Assoc. J. 47:465 (1986).
11. Shenton-Taylor, T.; Ogden, T.L.: Permanence of Membrane Filter Clearing and Mounting Methods for Asbestos Measurement. Microscope 34:161 (1986).
12. Ogden, T.L.: The Reproducibility of Asbestos Counts. Health and Safety Executive Research Paper 18. London (1982).
13. Busch, K.A.; Hornung, R.W.; Smith, R.J. Unbiased Estimates of Coefficients of Variation for Asbestos Counting Determined from Johns-Manville Data. In: Dusts and Disease, pp. 185–197. Pathotox Publishers (1979). Also in: Leidel, N.A.; Bayer, S.G.; Zumwalde, R.D.; Busch, K.A.: USPHS/NIOSH Membrane Filter Method for Evaluating Airborne Asbestos Fibers, Appendix C. DHEW (NIOSH) 79–137. Cincinnati, OH (1979).
14. Brownlee, K.A.: Statistical Theory and Methodology in Science and Engineering, 2nd ed., pp. 290–295. John Wiley and Sons, New York (1965).
15. Baron, P.A.; Shulman, S.A.: Evaluation of the Magiscan Image Analyzer for Asbestos Fiber Counting. Am. Ind. Hyg. Assoc. J. 48:39 (1987).
16. Occupational Safety and Health Administration, U.S. Department of Labor: Occupational Exposure to Asbestos, Tremolite, Anthophyllite, and Actinolite; Final Rules, 29 CFR 1910.1001 and 29 CFR 1926.58. Fed. Reg. 51:22612 (June 20, 1986).
17. Tombes, C.; Calpin, J.A.: Interlaboratory Differences in Fiber Counting in Accordance with NIOSH 7400 Method. Am. Ind. Hyg. Assoc. J. 49:A-695 (1988).
18. Leidell, N.A.; Busch, K.A.; Lynch, J.R.: Occupational Exposure Sampling Strategy Manual. DHEW (NIOSH) Pub. No. 77–173. Cincinnati, OH (1977).
19. McKay, A.T.: Distribution of the Coefficient of Variation and the Extended "t" Distribution. J.R.S.S. A-95:695 (1932).
20. Johnson, N.L.; Kotz, S.: Distribution in Statistics: Continuous Univariate Distributions, I, pp. 75–76. Houghton Mifflin, Boston (1969).
21. Duncan, A.J.: Quality Control and Industrial Statistics, 3rd ed., pp. 383–389, 908. Richard D. Irwin, Inc., Homewood, IL (1965).
22. Hald, A.: Statistical Theory with Engineering Applications, pp. 323–324, 725–726. John Wiley and Sons, New York (1952).

Received 7/26/88; review decision 8/29/88; revision 6/12/89; accepted 6/15/89