Counting Rules for Estimating Concentrations of Long Asbestos Fibers

KENNY S. CRUMP* and D. WAYNE BERMAN

1Department of Mathematics and Statistics, Louisiana Tech University, Ruston, LA 71273, USA; 2Aeolus, Inc., 751 Taft Street, Albany, CA 94706, USA

Received 12 October 2010; in final form 7 April 2011; published online 19 July 2011

Mounting evidence that long asbestos fibers (e.g. >20 or even 40 μm) pose the greatest cancer risk underscores the need for accurate measurement of concentrations of such fibers. These fiber lengths are of the same order of magnitude as the size of openings in the grids (typically ≈90 μm per side) used to analyze asbestos samples by transmission electron microscopy. This means that a substantial proportion of long fibers will cross the edge of a grid opening (GO) and therefore not be completely visible. Counting rules generally deal with such fibers by assigning a length equal to twice the visible length. Using both theoretical and simulation methods, we show that this doubling rule introduces bias into estimates of fiber concentrations and the amount of bias increases with fiber length. We investigate an alternative counting rule that counts only fibers that lie completely within a GO and weights those fibers by the reciprocal of the probability that a fiber of that length lies totally within a GO. This approach does not have the bias inherent in the doubling rule and is essentially unbiased if the stopping rule specifies a fixed number of GOs to be scanned. However, a stopping rule based on successively scanning GOs until a fixed number of fibers have been counted will introduce bias into any counting method, although this bias may typically not be large enough to be of practical concern. We recommend use of the weighted approach as a supplement to use of the doubling rule when estimating concentrations of long fibers, irrespective of the stopping rule employed.

Keywords: asbestos analysis; doubling rule; exposure estimation; fiber counting methods; statistical bias; transmission electron microscopy

INTRODUCTION

Exposure to airborne asbestos fibers is associated with risk of lung cancer and mesothelioma and continues to be a health problem throughout the world. Quantification of the health risk associated with a particular environment requires information on the airborne levels in that environment. These airborne levels typically are determined by passing measured amounts of air through filters and examining the residues on the filters. Traditionally, such filters were analyzed using phase contrast microscopy (PCM). However, PCM has several limitations for this use including (i) inability to distinguish fibers composed of asbestos-related minerals from those of other compositions, (ii) inability to identify fibers thinner than ~0.25 μm, and (iii) inability to distinguish among different types of asbestos (e.g. NIOSH, 1994a,b). Because of these limitations, transmission electron microscopy (TEM) has replaced PCM for many applications. TEM can identify the smallest asbestos fiber and can distinguish asbestos minerals from one another and from non-asbestos materials. This latter feature is important because different types of asbestos (chrysotile and different types of amphiboles) appear to pose unique health risks (Hodgson and Darnton, 2000; ERG, 2003; Berman and Crump, 2008a,b; Berman, 2011).
During TEM analysis of a sample [e.g. International Organization for Standardization (ISO), 1995], a filter through which air has been drawn is first coated with carbon and then one or more grids are positioned over cut portions of the filter. The filter portion is then dissolved leaving the collected debris in the carbon coat over each grid. Each grid contains a mesh of square grid openings (GOs) with sides typically measuring \( \frac{90}{C^2} \text{μm} \) (Fig. 1). During analysis, GOs are systematically scanned in the TEM and all the structures within each GO that meet pre-assigned criteria are recorded. GOs are sequentially processed until a stopping rule is satisfied, which is typically defined either as a minimum number of structures to be counted or as a minimum area of the filter (number of GOs) to be scanned. In some cases, the two rules are combined so that counting stops when a minimum number of GOs are scanned or when the scan of the GO is completed on which the last of a minimum number of fibers is counted [e.g. National Institute for Occupational Safety and Health (NIOSH, 1994b; ISO, 1995)].

Traditionally, PCM analyses accounted for all structures visible by PCM that were \( \geq 5 \text{μm} \) and with an aspect ratio \( \geq 3:1 \) and gave equal weight to all such structures (e.g. NIOSH, 1994a). However, evidence from both animal studies (Stanton and Wrench, 1972; Pott et al., 1974, 1976, 1987; Wagner et al., 1976, 1982, 1985; Stanton et al., 1977, 1981; Bertrand and Pezerat, 1980; Bolton et al., 1982, 1983, 1986; Pott, 1982; Davis et al., 1985, 1986a,b, 1987, 1988; Bonneau et al., 1986a,b; Muhle et al., 1987; Wylie et al., 1987, 1993; Berman et al., 1995) and newer epidemiological studies (Stayner et al., 2008; Berman and Crump, 2008b; Loomis et al., 2010; Berman, 2011) suggest that risk of cancer from asbestos exposure increases with fiber length, and longer fibers (e.g. with lengths \( \geq 20 \text{μm} \)) pose a greater risk than shorter fibers. However, the vast majority of fibers typically identified by TEM are short (e.g. with lengths \( < 5 \text{μm} \)) and the frequency of longer fibers decreases with increasing length. Consequently, the precision with which concentrations of long fibers are measured is often much lower than that for short fibers.

To cost-effectively increase the precision in the counting of long fibers, multiple scans of a filter are sometimes conducted in which counted fibers are limited to those of a minimum length which increases during each successive scan. Such stratified counting has been employed in numerous studies (e.g. Berman et al., 1993; Dement et al., 2008, 2009; Berman, 2010; Loomis et al., 2010) to efficiently estimate length distributions. For example, during TEM analysis of archived filters from a textile mill, Dement et al. (2008) conducted three separate scans. The first scan counted fibers of all lengths, the second scan counted only fibers \( \geq 5 \text{μm} \), and the third scan counted only fibers \( \geq 15 \text{μm} \). The scan for all fibers used a magnification of \( \times 20000 \) and the other two scans used a magnification of \( \times 10000 \). With each scan, complete GOs were sequentially scanned until the number of fibers counted exceeded a minimum number (50 fibers for scans for all fibers, 80 fibers for scans for fibers \( \geq 5 \text{μm} \), and 50 fibers for scans for fibers \( \geq 15 \text{μm} \)). The information recorded for each fiber included, among other characteristics, its length, width, and asbestos mineral type.

Using TEM, if a fiber intersects the grid bar at one of the sides of a GO, only the portion of the fiber inside the GO is visible. Such fibers often are assigned a length equal to twice the length that is visible inside the GO (e.g. ISO, 1995). We evaluate this ‘doubling rule’ mathematically in this paper and show that it provides biased estimates of the concentrations of fibers in specific length categories. The amount of bias can be appreciable for categories of long fibers. As a possible remedy for this problem, we propose an alternative that avoids the bias inherent in the doubling rule.

Asbestos structures occur in more complex forms than simple fibers. They are usually categorized as fibers, bundles (groups of attached fibers all with the same orientation), clusters (groups of fibers with random orientations that are either attached or at least overlain), and matrices (complex
Estimating concentrations of fibers from data on multiple scans

We focus on estimating the mean number of fibers per GO. Such an estimate can be converted into an air concentration by dividing by the amount of air estimated to have passed through an area of the filter the size of a GO.

Suppose we have data from a number of different scans and we wish to estimate the concentration of fibers in some category of fibers represented by \( C \). \( C \) may be defined in terms of fiber type (e.g. chrysotile or various categories of amphibole) in addition to fiber dimension. For example, \( C \) could be all chrysotile fibers \( \geq 10 \mu m \) and thinner than 0.4 \( \mu m \) (a fiber category relevant to a protocol for quantifying risk evaluated in Berman and Crump, 2008b). We divide \( C \) into mutually disjoint subcategories so that each scan counts either all or none of the fibers in each subcategory. We then estimate the concentration of fibers in each subcategory by the total number of fibers counted in that subcategory (in all scans), divided by the number of GOs (in all scans) in which fibers in that subcategory were counted. The total concentration of fibers in \( C \) is estimated by the sum of the concentrations of fibers in the subcategories.

As an example of this procedure, suppose we have the following data from three scans: Scan 1 that counted all fibers, Scan 2 that counted only fibers between 10 and 15 \( \mu m \) and thinner than 0.4 \( \mu m \), and Scan 3 that counted only fibers \( \geq 15 \mu m \). The total concentration of fibers per GO \( \geq 10 \mu m \) is then estimated as the sum of the estimated concentrations in \( C_{10-15} \) and \( C_{\geq 15} \) (5.71 + 3.05 = 8.76).

### Evaluation of the doubling rule

Under the doubling rule (e.g. ISO, 1995), fibers visible in a GO that cross one edge of the GO are assigned a length equal to twice the length that is visible within the GO. Since such fibers could also be counted in a GO containing the other end of the fiber, to avoid double counting, only fibers that cross one of two particular sides of a GO (e.g. the bottom and right side) are counted or else all fibers that cross any one side are counted but given a weight of 1/2 (e.g. NIOSH, 1994b). For the same reason, a fiber that crosses two edges of the GO, so that neither end is visible, is not counted in that GO.

The doubling rule modifies both the distribution of fiber lengths and the proportion of fibers in given length categories. The bias caused by this modification can be particularly important for long fibers. For example, suppose that a particular environment only has fibers with lengths up to 30 \( \mu m \). In this situation, the doubling rule will assign fiber lengths ranging up to 60 \( \mu m \), whereas in fact, there are no fibers \( \geq 30 \mu m \).

The extent of the bias caused by the doubling rule is examined in this section. Suppose one end of a fiber of length \( x \) is placed randomly in a GO with sides of length \( L \) and the fiber is given a random orientation (Fig. A1). Let \( W \) be the length assigned to the fiber by the doubling rule (i.e. \( W = x \) if the fiber doesn’t cross the edge of the GO and otherwise, \( W = 2x \)). The Appendix 1 contains a derivation of the cumulative probability distribution of \( W \) in terms of \( L \) and the true fiber length \( x \) and the true fiber length \( x \) (equation (A9)). Intuitively, it might seem that if the fiber crosses an edge, \( W \) would be uniformly distributed on the interval \([0, 2x]\). Although this would be true if the GO were bounded only on one side, the distribution is not exactly uniform due to the finite size of the GO. Figure 2 shows the exact distribution of \( W \) given the fiber crosses some edge of the GO, for different fiber lengths \( x \) (equation (A9)). A uniform distribution would be represented by a straight line running from the origin to the point...
This figure shows that, although the deviation from uniformity can be considerable for very long fibers, the deviation is small for fiber of lengths up to the length of the side of a GO.

The probability that a fiber of length $x$ placed randomly in a GO will cross some edge of the GO will play an important role in what follows. This probability depends only on the ratio $r = \frac{x}{L}$ of the true fiber length $x$ to the length of the GO side and is given by

$$P(r) = \begin{cases} \frac{1}{2}(r - \frac{1}{4}) & 0 \leq r \leq 1 \\ \frac{1}{2}(\sqrt{\frac{2}{3} + \frac{1}{2} - \sqrt{\frac{2}{3}}} - \arcsin(\frac{1}{2})) & 1 \leq r \leq \sqrt{2} \\ \frac{1}{2} & \sqrt{2} \leq r \end{cases}$$

(1)

where ‘arcsin’ denotes the inverse sine function. This expression is a solution to a special case of the classical Buffon–Laplace needle problem (e.g. Weisstein, 2002) and is derived in the Appendix 1 [equation (A10)] as a special case of the cumulative probability distribution of $W$.

Figure 3 depicts how $P(r)$ increases with increasing $r$. The probability increases almost linearly with the ratio $r = \frac{x}{L}$ and is slightly larger than that ratio for ratios $<0.86$. For example, if a GO is 90 μm on a side (a typical value), a fiber 45 μm in length ($r = 0.5$) has probability of 0.56 of crossing an edge of the GO. If $r = 1$, then $P(r) = \frac{1}{2} \approx 0.5$. Consequently, most of the probability range is covered by the case $0 \leq r \leq 1$, where the probability is simply a linear-quadratic form, $P(r) = (4/\pi)(r - r^2/4)$.

Rather than fixing the true fiber length $x$, we now consider $x$ to be a specific value of the length $X$ of a randomly selected fiber from the distribution $F_W(x)$ of true fiber lengths. The Appendix 1 derives an expression for the cumulative probability distribution $F_W(w)$ of the length $W$ assigned by the doubling rule in terms of $F_X(x)$ and the length $L$ of the side of a GO [equation (A15)]. We now use this equation to study how the probability distribution for fiber lengths assigned under the doubling rule differs from the true underlying distribution of fiber lengths.

We compared $F_W(s)$ and $F_X(s)$ using three log-normal distributions and three gamma distributions for $F_X(s)$ (Johnson et al., 1994). The log-normal distributions have log means of 1 and log standard deviations (SDs) of 0.9, 1.0, and 1.1 (corresponding untransformed means and SDs of $\mu = 4.1$ and $\sigma = 4.6$, $\mu = 4.5$ and $\sigma = 5.8$, and $\mu = 5.0$ and $\sigma = 7.6$). To compare the effect of different distributions with the same general shape, we assumed the same means and SDs for the gamma distribution. All lengths are in microns and the GO side was assumed to be 90 μm in length. The integrals in equation (A15) were evaluated numerically using a Cash–Karp variable step size Runge–Kutta routine with a tolerance of $10^{-10}$ (Press et al., 2007).

Figure 4 contains graphs of the ratio of the proportion of fibers longer than a given length computed using the doubling rule to the true proportion. In all cases, the ratio is slightly $<1.0$ for very short fibers but exceeds 1.0 by increasingly larger amounts with increasing fiber length. This indicates that with these distributions the doubling rule assigns slightly too few short fibers, but as fiber length increases, assigns increasingly too many long fibers. Within a single distribution, the value of the ratio for large fiber lengths appears to be inversely related to the log SD. These ratios are larger for the gamma distribution than for the lognormal. For example, for the proportion of fibers $\geq 40 \mu$m, the case $\mu = 4.1$ and $\sigma = 4.6$ gives a ratio of 1.3 for the lognormal and 3.9 for the gamma. The errors introduced by the doubling rule can be extreme for very long fibers. For example, with gamma-distributed fiber lengths ($\mu = 4.1$ and $\sigma = 4.6$), the expected proportion of fibers $\geq 60 \mu$m obtained using the doubling rule is 26 times the true number (Fig. 4).

These results suggest that errors in counting long fibers resulting from use of the doubling rule would be...
more severe if fiber lengths have gamma distributions than if they have log-normal distributions. However, there is no theoretical basis for selecting one of these distributions over the other, and since these distributions have the same mean and variance, they would be difficult to distinguish based on sampling data. Thus, for any sampled population (whose true distribution is unknown and may be difficult to concisely define), the magnitude of the bias introduced by the doubling rule would be unknown and could be large.

Figure 5 contains graphs of the expected fraction of the fibers assigned lengths longer than a given value under the doubling rule whose lengths were not directly observed but were assigned by doubling the visible length [based on Appendix 1 equation (A16)]. The proportion of fibers assigned lengths ≥ 40 μm provided from fibers that cross an edge of the GO (and consequently whose lengths cannot be observed) ranges between 0.69 and 0.72 for the three log-normal distributions and between 0.69 and 0.89 for the three gamma distributions. Thus, with each of these distributions, a substantial proportion of long fibers identified using the doubling rule will come from fibers that cross an edge of a GO and whose true lengths cannot be observed.

In addition to the bias inherent in the doubling rule, an additional source of bias is the use of stopping rules that depend upon the number of fibers counted (e.g. fixing in advance the number of fibers to be counted as opposed to fixing the number of GOs to be examined). A simple example is provided in the Appendix 1 to show that such stopping rules can induce bias. However, these biases may not be appreciable and such stopping rules can assure that an adequate number of fibers are counted. Thus, the mere fact that such bias exists may not be sufficient reason to eschew stopping rules that depend upon the number of fibers counted. Nevertheless, practitioners should be aware of the potential for such bias. The extent of this bias is examined in specific cases in the simulations presented in the next section.

A weighted estimator that uses only fibers whose total length is visible

In this section, we propose and study the properties of a simple estimator that eliminates the bias inherent in the doubling rule. Consider a single GO. If the lengths of all fibers in the GO (i.e. all
fibers contained completely within the GO plus all other fibers that cross one of two specified GO sides) could be determined, the observed number of fibers in a particular fiber category, $C$ (e.g. $C$ could be amphibole fibers $\geq 20 \mu m$ and thinner than $1.5 \mu m$, which defines a promising metric for risk recently examined by Berman, 2011), would be an unbiased estimate of the average number of fibers in category $C$ per GO. This estimator can be considered a sum over all fibers in the GO in which each fiber in $C$ contributes 1 to the sum and each fiber not in $C$ contributes zero. The proposed estimator is a simple modification to this estimator that considers only fibers that do not cross an edge of a GO and, instead of contributing 1 to the sum, each fiber in $C$ is weighted by the reciprocal of the probability that a fiber of that length would not cross an edge. I.e. the estimator is the sum of terms of the form

$$\frac{1}{1 - P(x_f/L)}$$

where $x_f$ is the length of fiber $f$, $L$ is the length of a side of the GO, $P$ is given by equation (1) and the sum is taken over all fibers $f$ in $C$ that do not cross an edge. Intuitively, each fiber is given an increased weight that accounts for the expected number of other fibers of the same length in the GO that are not included in the estimate because they cross an edge of the GO. If $L = 90 \mu m$, a fiber $5 \mu m$ in length would be given a weight of $\frac{1}{1 - P(5/90)} = 1.07$ and a fiber of length $40 \mu m$ would be given a weight of $\frac{1}{1 - P(40/90)} = 2.01$. The greater weight given to the longer fiber reflects the fact that the fiber is ‘standing in’ for a larger number of fibers of that length that are not included because they cross an edge of a GO. This estimation procedure is closely related to the Miles–Lantuéjoul correction used to correct for edge effects in the random stacking of solid particles (Miles, 1974; Lantuéjoul, 1980).

Since a fiber of length greater than the diagonal of a GO ($\sqrt{2}L$) is sure to cross one GO edge, the denominator in equation (2) will be zero for such a fiber. This does not prevent the estimator from being well-defined in practice because such fibers can never be sampled by this estimation scheme. However, because of this feature, rather than providing unbiased estimates of the proportions of all fibers lying in various length intervals, instead the estimator provides unbiased estimates based on the truncated distribution restricted to fibers whose lengths are $< \sqrt{2}L$. This limitation is mainly of theoretical interest since in practical situations, the proportion of fibers longer than $\sqrt{2}L$ will be extremely small. For example, in the six distributions used as examples in Fig. 3, the proportion of such fibers was $< 0.003\%$ for the log-normal distributions and $< 0.0000000001\%$ ($10^{-11}$) for the gamma distributions. Among 14 samples of 13 diverse asbestos materials (Berman et al., 1995; Berman, 2011), only 2 assigned any structure length $>90 \mu m$. Of these, wet dispersed chrysotile exhibited 0.2\% structures $\geq \sqrt{2} \times 90 = 127 \mu m$ and long chrysotile (a product specifically contrived to increase the percentage of long structures) exhibited 0.03\% structures $>90 \mu m$. Among the structures found in archived samples of textile factory dust (Dement et al., 2008), 0.4\% were assigned lengths $>90 \mu m$. However, these values derived from sampling data are likely overestimates because (i) each sampling protocol utilized stratified sampling designed to preferably identify long structures and (ii) structure lengths were assigned using the doubling rule which likely overestimated the counts of long structures. For example, Dement et al. identified only one structure (0.007\%) $>90 \mu m$ among those structures contained completely within a GO and consequently whose length could be measured.

Table 1 contains the results of a simulation that compares estimates of the average number of fibers per GO obtained using the doubling rule versus use of the weighted estimator based on equation (2) and including only fibers that do not cross an edge of a GO. Following Dement et al. (2008), three scans were simulated (Scan 1 for all fibers, Scan 2 for fibers $\geq 5 \mu m$, and Scan 3 for fibers $\geq 15 \mu m$). The total number of fibers of all lengths on a GO was assumed to have a negative binomial distribution with a mean of 40 and a coefficient of variation of 1. Fiber lengths were assumed to have either a log-normal or gamma distribution. The log-normal distribution was assigned a log mean of 1 and a log SD of 0.9, which results in a mean of 4.076 and a SD of 4.553 for the untransformed log-normal distribution. The same mean and SD were assigned to the (untransformed) gamma distribution. Estimates were made both of the number of fibers per GO $\geq 10 \mu m$ and of the number $\geq 30 \mu m$. GOs were assumed to be $90 \mu m$ on a side. In the first set of simulations (top half of Table 1), GOs were sequentially processed until a fixed minimum number of fibers were counted (50 fibers of all sizes for Scan 1, 80 fibers $\geq 5 \mu m$ for Scan 2, and 50 fibers $\geq 15 \mu m$ for Scan 3).

In the second set of simulations (bottom half of Table 1), the numbers of GOs to be scanned were fixed in advance. To promote comparability between the two sets of simulations, these were set equal to the average number (rounded to the nearest integer) scanned in the
first set of simulations. In each case, 1000 sets of sampling data were generated and analyzed. For data sets analyzed by the doubling rule, a fiber was counted if it did not cross an edge and its length was at least the minimum for the scan or if it crossed the boundary and its observed length was at least one half of the minimum for the scan. For data sets to which the weighted estimator was applied, a fiber was counted if it did not cross an edge and its length was at least the minimum for the scan. Since fibers that crossed an edge were not used, the weighted estimator required a larger number of GOs than the estimator based on the doubling rule to achieve the minimum number of fibers to be counted for a scan.

With the stopping rule based on the number of fibers counted (top half of Table 1), the two estimators performed roughly the same for fibers $\geq 10\,\mu m$ (similar percent biases and SDs). However, for fibers $\geq 30\,\mu m$, the percent biases were substantially larger for the doubling estimator than for the weighted estimator. This difference was most pronounced with the gamma distribution where the bias was 100% for the doubling estimator but only 2% for the weighted estimator. There was no statistical evidence of bias in estimates of fibers $\geq 30\,\mu m$ from the weighted estimator (the 90% confidence intervals contain the true concentration for both the log-normal and the gamma distributions—results not shown). Appreciable differences in the performance of the two estimators as measured by the percent square root mean square error were evident only for fibers $\geq 30\,\mu m$ and then only for gamma-distributed fiber lengths and not for log-normally distributed lengths.

When the stopping rule was defined by a fixed number of GOs (bottom half of Table 1), the percent bias was greatly reduced in estimates of the concentration of fibers $\geq 10\,\mu m$. This indicates that most of the bias in these estimates in the top half of Table 1 was due to the stopping rule being defined by fixed numbers of fibers rather fixed numbers of GOs. However, for fibers $\geq 30\,\mu m$, the reductions in the biases in the doubling estimator were minimal. By comparison, in agreement with the theoretical result that the weighted estimator is essentially unbiased under this stopping rule, there is no statistical evidence of any bias in this estimator (the 90% confidence intervals contain the true concentration for both the log-normal and the gamma distributions—results not shown).

<table>
<thead>
<tr>
<th>Stopping rule based on fixed number of fibers counted</th>
<th>Lognormal</th>
<th></th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>True mean number of fibers/GO</td>
<td>3.0</td>
<td>3.0</td>
<td>0.15</td>
</tr>
<tr>
<td>Average estimate</td>
<td>3.1</td>
<td>3.2</td>
<td>0.19</td>
</tr>
<tr>
<td>% Bias$^a$</td>
<td>5.5</td>
<td>7.3</td>
<td>24</td>
</tr>
<tr>
<td>% Square root MSE$^b$</td>
<td>30</td>
<td>32</td>
<td>50</td>
</tr>
<tr>
<td>Average number of GOs</td>
<td>55</td>
<td>75</td>
<td>55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stopping rule based on fixed number of grid openings examined</th>
<th>Lognormal</th>
<th></th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>True mean number of fibers/GO</td>
<td>3.0</td>
<td>3.0</td>
<td>0.15</td>
</tr>
<tr>
<td>Average estimate</td>
<td>2.9</td>
<td>3.0</td>
<td>0.18</td>
</tr>
<tr>
<td>% Bias$^a$</td>
<td>-0.65</td>
<td>0.056</td>
<td>19</td>
</tr>
<tr>
<td>% Square root MSE$^b$</td>
<td>26</td>
<td>24</td>
<td>46</td>
</tr>
<tr>
<td>Number of GOs (fixed)</td>
<td>54</td>
<td>75</td>
<td>54</td>
</tr>
</tbody>
</table>

$^a$% Bias = |bias|/|true value| $\times$ 100.

$^b$% Square root MSE = [square root of mean square error]/|true value| $\times$ 100.
hypothesize that it is related in some way to the difficulty in enumerating and sizing components of complex structures. Although beyond the scope of the present paper, this suggests that possible biases and uncertainties related to quantification of such structures warrants further attention.

**DISCUSSION**

The evidence cited earlier that cancer risk increases with fiber length underscores the need for accurate determination of concentrations of long fibers. Currently, estimates of fiber concentrations are generally based on the doubling rule, in which the length of a fiber that crosses the edge of a GO is assumed to be twice the visible length. This assumption plays an increasingly important role with increasing fiber length. As shown herein, with a typical GO size of 90 μm on a side, a fiber 45 μm in length has a >50% probability of lying across the edge of a GO. Although the proportion of counted fibers whose lengths are estimated by doubling depends upon the distribution of fiber lengths, it can easily be the case that the majority of counts of long fibers come from doubling the observed lengths of fibers whose true lengths are obscured (Fig. 5). This problem has been given insufficient attention to date as many papers that report fiber concentration data from TEM analysis do not even describe how such fibers were handled.

In this paper, we show that use of the doubling rule introduces bias into estimates of fiber concentrations and the amount of bias increases with increasing fiber length (Fig. 4, Table 1). In the cases simulated (Table 1), the doubling estimator resulted in estimates of the concentration of fibers ≥30 μm that were 20–100% too large. This excess would be expected to increase as the minimum size fiber counted is restricted to even longer fibers.

We show herein that some bias in estimates of fiber concentrations can also result from use of stopping rules that fix a lower bound for the number of fibers to be counted rather than fixing the number of GOs to be scanned. However, this bias may not be serious in practical situations. In the simulations conducted herein, this bias was <10% of the true value in all cases (Table 1). Biases of this magnitude may not be sufficient to outweigh the benefits from stopping rules that assure that sufficient numbers of fibers in various size ranges are counted.

As an alternative to use of the doubling rule, we propose a simple weighted estimator that involves counting only fibers that lie completely within a GO. To account for the undercounting due to omitting fibers that cross the edge of a GO, rather than contributing 1 to a count, each such fiber contributes the reciprocal of the probability that a randomly positioned fiber of that length does not cross an edge of a GO. This weighted estimator avoids the bias introduced by the doubling rule and is essentially unbiased (i.e. if the generally minuscule proportion of fibers >√2 times the length of the edge of a GO are omitted from consideration) if stopping rules that fix the number of GOs are employed.

The specific distributions and related parameter values used herein were chosen to illustrate various points rather than necessarily to represent fiber distributions encountered in practice. In fact, we believe that it is difficult to determine the extent of the bias in the doubling rule from standard statistical analyses of existing sets of data. The right tail of the fiber length distribution is of critical interest when evaluating the bias in the doubling rule. Just because a particular distribution provides an adequate statistical description of an entire data set does not imply that it will also predict concentrations of long fibers with sufficient accuracy. In the examples studied herein, log-normal and gamma distributions with the same mean and variances, and which therefore would be difficult to distinguish in practice, presented very different amounts of bias under the doubling rule.

### Table 2. Comparison of length distributions of fibers and bundles in archived samples of textile factory dust (Demert et al., 2008) estimated using the doubling rule (Dbl.) or the weighted procedure (Wgt.)

| Lengths (μm) | Primary Structures | | Component structures | | |
|-------------|-------------------|---|-------------------|---|
| ≥5 | 1540 | 1863 | 1829 | 1.02 | 4961 | 5864 | 6667 | 0.88 |
| ≥10 | 543 | 766 | 726 | 1.06 | 2792 | 3573 | 4261 | 0.84 |
| ≥20 | 123 | 235 | 205 | 1.15 | 1129 | 1682 | 2178 | 0.77 |
| ≥30 | 36 | 92 | 78 | 1.18 | 460 | 840 | 1193 | 0.70 |
| ≥40 | 15 | 49 | 41 | 1.18 | 205 | 469 | 738 | 0.64 |
| ≥50 | 5 | 27 | 18 | 1.46 | 110 | 309 | 523 | 0.59 |
| ≥60 | 3 | 18 | 12 | 1.45 | 54 | 204 | 362 | 0.56 |

*aCounts of fibers and bundles that did not cross an edge of the GO.*
This problem is compounded by the fact that the doubling rule is typically employed in determining fiber lengths, which introduces bias into a statistical analysis. There is no theoretical basis for assuming a log-normal, gamma, or any other particular parametric distribution for fiber lengths. For these reasons, simulating data from, for example, a log-normal distribution having the mean and variance estimated from actual sampling data would not necessarily provide a true picture of the extent of bias resulting from use of the doubling rule. Because of these problems, we suggest that, rather than fitting parametric models to data, a better method of evaluating the extent of bias introduced by the doubling rule in a particular data set would be to estimate fiber concentrations using both the doubling rule and the weighted estimator and compare the results.

Although it is beyond the scope of this paper to recommend counting rules for analyzing air sampling data, it seems that it should be possible to develop rules optimally designed for the weighted estimator, so that fibers that cross GO edges would not need to be recorded. To achieve comparable numbers of fiber counts would require scanning a larger number of GOs than approaches in which partially obscured fibers are included in the counts. However, this additional effort would be at least partially offset by the fact that such fibers could be quickly eliminated by scanning along the length of any fiber that is encountered and moving on as soon as it is found that the fiber crosses the edge of a GO. Regardless, eliminating what can otherwise be a substantial bias may be well worth the additional time and cost.

Even if fibers that cross an edge of a GO are counted, the weighted estimator can still be applied, provided that the data recording allows such fibers to be identified (which ideally should always be the case). Although we consider that it would be premature to recommend exclusive use of the weighted procedure based on totally visible fibers, we do recommend applying this procedure as a supplement to use of the doubling rule when estimating concentrations of long fibers.

**FUNDING**

The work by K.S.C. was supported in part by a fellowship at Louisiana Tech funded by an appointment to the Research Participation Program of the National Center for Environmental Assessment, US EPA, through the Oak Ridge Institute for Science and Education.

**ACKNOWLEDGEMENTS**—The authors thank the reviewers for their comments that led to substantial improvements to this paper. The views expressed in this article do not necessarily reflect the views or policies of the US EPA. The work by D.W.B. was unfunded; D.W.B. consults for a variety of government and private organizations with competing interests, none of whom were involved in this work. Thus, the authors have no competing interests to declare.

**APPENDIX 1**

Distribution of fiber lengths assigned by the doubling rule

We first derive the distribution of the length assigned by the doubling rule to a fiber of length \( x \) having one end placed at random on a GO with sides of length \( L \) and given a random orientation (i.e. the angle \( \theta \) that the fiber makes with horizontal is uniformly distributed between 0 and \( 2\pi \) radians—see Fig. A1). Let \( R \) denote the event the fiber crosses one of the sides of the GO, \( R_B \) the event that the fiber crosses the bottom side, \( Z \) the fiber length visible inside GO, and \( W \) the length assigned by the doubling rule. It follows that \( W = 2Z \) on \( R \) and \( W = x \) on \( R_c \), where \( R_c \) denotes the complement of \( R \) and represents the event the fiber lies entirely within the GO. With this configuration, we derive the conditional distribution of the assigned length \( W \) given that the fiber crosses an edge, i.e.

\[
\Pr(W \leq w | R) = \Pr(Z \leq w/2 | R) = (\Pr(\{Z \leq w/2\} \cap R)) / (\Pr(R)).
\]

(A1)

It is sufficient to derive an expression for the numerator of this expression because the numerator becomes equal to the denominator as \( w \) becomes large. We can write

![Fig. A1. The assumed orientation of fiber of length \( x \) on a GO with sides of length \( L \).](http://annhyg.oxfordjournals.org/)
\[
\Pr \left( \left\{ Z \leq \frac{w}{2} \right\} \cap R \right) = \frac{1}{2\pi} \int_{0}^{\pi} \Pr \left( \left\{ Z \leq \frac{w}{2} \right\} \cap R|\theta \right) d\theta. \tag{A2}
\]

Since the fiber must cross exactly one of the four sides, by symmetry we can replace \( R \) by \( R_{B} \), which represents the event the fiber crosses the bottom edge of the GO, and multiply the result by 4. Further noting that the resulting integrand is zero for \( \pi \leq \theta \leq 2\pi \) and is symmetric about \( \theta = \frac{\pi}{2} \), we get

\[
\Pr \left( \left\{ Z \leq \frac{w}{2} \right\} \cap R \right) = \frac{4}{\pi} \int_{0}^{\pi/2} \Pr \left( \left\{ Z \leq \frac{w}{2} \right\} \cap R_{B}|\theta \right) d\theta. \tag{A3}
\]

The derivation proceeds by obtaining expressions for the integrand \( \Pr \left( \left\{ Z \leq \frac{w}{2} \right\} \cap R_{B}|\theta \right) \) and performing the integration. There are a number of cases to consider.

We first consider the case \( \frac{\pi}{4} \leq x \leq L \). For a fixed fiber orientation \( \theta \), a fiber of length \( x \) crosses the bottom edge and has a length showing in the GO of \( Z \leq \frac{w}{2} \) if and only if the location of the fiber end is in the shaded region shown in Fig. A2A. The proportion of the entire GO that is shaded is

\[
\Pr \left( \left\{ Z \leq \frac{w}{2} \right\} \cap R_{B}|\theta \right) = \frac{w}{4L} \sin(\theta) \left[ 2 - \frac{w}{2L} \cos(\theta) \right]. \tag{A4}
\]

Applying this expression in equation (A3) and performing the integration yields

\[
\Pr \left( \left\{ Z \leq \frac{w}{2} \right\} \cap R \right) = \frac{w}{\pi L} \left( 2 - \frac{w}{4L} \right), \tag{A5}
\]

for \( \frac{\pi}{4} \leq x \leq L \). If \( x \leq \frac{\pi}{4} \), the same expression holds but with \( w \) replaced by \( 2x \).

We next consider the case \( L \leq x \leq \sqrt{2}L \). There are three ranges to consider for \( \theta \) defined by values of \( \theta \) that are too small for a fiber of length \( x \) to be contained totally within the GO (Case 1), too large for this to occur (Case 3), or otherwise (Case 2).

Case 1: If \( 0 \leq \theta \leq \theta_{1} = \arccos \left( \frac{L}{x} \right) \), then the fiber must cross either the bottom or the right edge of the GO. It crosses the bottom edge whenever its end is in the lower right triangle in Fig. A2B with diagonal of length \( x \). Its length showing in the GO is \( \leq \frac{w}{2} \) whenever its end is in the region below the dotted lines. That region has probability given by expression (A4) provided \( \frac{1}{2} \sin(\theta) \) is smaller than the height of the shaded triangle (i.e. provided \( \frac{1}{2} \sin(\theta) \leq L\tan(\theta) \) or, equivalently, \( \cos(\theta) \geq \frac{2L}{w} \)). Otherwise, the probability is equal to the proportion of the GO contained in the shaded triangle or

\[
\Pr \left( \left\{ Z \leq \frac{w}{2} \right\} \cap R_{B}|\theta \right) = \frac{1}{2} \tan(\theta). \tag{A6}
\]

Case 2: If \( \theta_{1} = \arccos \left( \frac{L}{x} \right) \leq \theta \leq \arcsin \left( \frac{L}{w} \right) = \theta_{2} \), a fiber with an orientation angle in that range does not necessarily cross an edge. So long as \( x \leq L \), \( \Pr \left( \left\{ Z \leq \frac{w}{2} \right\} \cap R_{B}|\theta \right) \) is given by equation (A4). Otherwise, this expression must be modified by replacing \( w \) by \( 2x \).

Case 3: If \( \theta_{2} \leq \arcsin \left( \frac{L}{w} \right) \leq \theta \leq \frac{\pi}{4} \) by reasoning similar to Case 1, equation (A4) holds provided \( \sin(\theta) \leq \frac{2L}{w} \), and

![Fig. A2. Illustration of the description of the calculation of the probability distribution of the length of that portion of a randomly situated fiber of length x that lies within a GO of length L. (A) Case x ≤ L. (B) Case x ≥ L.](image-url)
\[
\text{Pr}\left\{ Z \leq \frac{w}{2} \right\} \cap R_B | \theta = \frac{1}{2} [2 - \cot(\theta)], \quad (A7)
\]
otherwise.

Figure A3 is a plot with \( \theta \) on the vertical axis running from 0 to \( \frac{\pi}{2} \) and \( x \) on the horizontal axis running from 0 to \( x \) that shows the different regions in which \( \text{Pr}\left\{ Z \leq \frac{w}{2} \right\} \cap R_B | \theta \) has a unique definition. It is clear from this plot that if \( \frac{w}{2} \leq L \), then \( \text{Pr}\left\{ Z \leq \frac{w}{2} \right\} \cap R_B | \theta \) is defined by equation (A4) throughout the range of integration and hence, equation (A5) holds in this range. However, if \( L \leq \frac{w}{2} \leq x \), then the integration indicated by equation (A3) must be performed in three parts as follows:

\[
\text{Pr}\left\{ Z \leq \frac{w}{2} \right\} \cap R
\]
\[= \frac{1}{\pi} \int_{0}^{\arccos(2L/w)} \tan(\theta) \, d\theta \]
\[+ \int_{\arccos(2L/w)}^{\arcsin(2L/w)} \frac{w}{2L} \sin(\theta) \left[ 2 - \frac{w}{2L} \cos(\theta) \right] \, d\theta \]
\[+ \int_{\arcsin(2L/w)}^{\frac{\pi}{2}} \left[ 2 - \cot(\theta) \right] \, d\theta. \quad (A8)
\]
\[
= \frac{4}{\pi} \left\{ \frac{1}{2} + \frac{\pi}{2} + \frac{1}{4} \left( \frac{w}{2L} \right)^2 \right. \\
\left. - \sqrt{\left( \frac{w}{2L} \right)^2 - 1} - \arcsin\left( \frac{2L}{w} \right) \right\}.
\]

Other terms in the solution are derived in a similar fashion. The complete solution can be expressed as follows:

\[
H(w, x, L) = \text{Pr}(W \leq w) = \text{Pr}\left\{ \{W \leq w\} \cap R \right\} = \text{Pr}\left\{ \{Z \leq \frac{w}{2}\} \cap R^c \right\} = \\
\text{equation (A5)} \quad \text{for } x \leq L \text{ and } w/2 \leq x; \\
\text{equation (A5) with } w \text{ replaced by } 2x \quad \text{for } x \leq L \text{ and } x \leq w/2; \\
\text{equation (A5)} \quad \text{for } L \leq x \leq \sqrt{2L} \text{ and } w/2 \leq L; \\
\text{equation (A8)} \quad \text{for } L \leq x \leq \sqrt{2L} \text{ and } L \leq w/2 \leq \sqrt{2L}; \\
\text{equation (A9)} \quad \text{for } \sqrt{2L} \leq x \text{ and } L \leq w/2 \leq \sqrt{2L}; \\
\text{for } \sqrt{2L} \leq w/2. \\
\]

The probability that a fiber crosses some side of a GO depends upon the ratio \( r = \frac{x}{L} \) of the fiber length divided by the GO side length and will be denoted by \( P(r) \). As noted earlier, \( P(r) \) can be obtained by making \( w \) large in equation (A9). Specifically, from the second, fifth, and eighth terms in equation (A9), we get

\[
P(r) = \begin{cases} 
\frac{4}{\pi} \left( r - \frac{w}{2} \right) & 0 \leq r \leq 1 \\
\frac{4}{\pi} \left( r + \frac{w}{2} - \sqrt{r^2 - 1} - \arcsin\left( \frac{r}{2} \right) \right) & 1 \leq r \leq \sqrt{2} \\
\frac{1}{\sqrt{2}} & \sqrt{2} \leq r 
\end{cases}
\]

The conditional distribution of \( W \) given that the fiber crosses some side of the GO can now be written as

\[
\text{Pr}(W \leq w | R) = \frac{H(w, x, L)}{P(r)}. \quad (A11)
\]

Next, rather than fixing the true fiber length, we assume that it is random (denoted by \( X \)) with probability density function \( f_X(x) \) and indicate how the probability distribution of \( W \), the length assigned by the doubling rule, can be obtained from \( f_X(x) \) and equations (A9) and (A10). We assume that there is a minimum assigned length \( l_{\min} \) (which can be zero) for a fiber to be included in the count.

The probability a fiber is counted and its assigned length is \( \leq w \) can be written as

\[
\text{Pr}[l_{\min} \leq W \leq w] = \int_{0}^{w} \text{Pr}[l_{\min} \leq W \leq w | X = x] f(x) \, dx \\
= \int_{0}^{w} \text{Pr}[l_{\min} \leq W \leq w | R = x] f(x) \, dx \\
+ \int_{0}^{\infty} \text{Pr}[l_{\min} \leq W \leq w, R^c | X = x] f(x) \, dx. \quad (A12)
\]

The second integral can be rewritten as
and an edge of the GO is given by

$$\int_0^\infty \Pr[l_{\text{min}} \leq W \leq w|R^c, X = x] \Pr[R^c|X = x] f(x) dx.$$  

We note that

$$\Pr[l_{\text{min}} \leq W \leq w|R^c, X = x] = \left\{ \begin{array}{ll} 1 & \text{if } l_{\text{min}} \leq x \leq w \\ 0 & \text{otherwise} \end{array} \right.$$  

and

$$\Pr[R^c|X = x] = 1 - P(x/L),$$  

where $P$ is given by equation (A10). Consequently, the second integral in equation (A12) is equal to

$$\int_0^w [1 - P(x/L)] f(x) dx.$$  

(A13)

Turning to the first integral in equation (A12) we note that

$$\Pr[l_{\text{min}} \leq W \leq w, R|X = x] = H(w, x, L) - H(l_{\text{min}}, x, L),$$  

(A14)

where the function $H$ represents the conditional distribution of $W$ given $X = x$ (equation (A9)). Combining equations (A12), (A13) and (A14) and setting $l_{\text{min}} = 0$, we can write the cumulative distribution function of $W$ as

$$F_W(w) = \Pr(W \leq w) = \int_0^\infty H(w, x, L) f(x) dx + \int_0^w [1 - P(x/L)] f(x) dx.$$  

(A15)

Using these same methods, we can show that the fraction of fibers longer than a posited value $s$ estimated using the doubling rule from fibers that cross an edge of the GO is given by

$$\Pr[W \geq s, R] \Pr[W \geq s] = 1 - \frac{\int_s^\infty [1 - P(x/L)] f(x) dx}{[1 - F_W(s)].}$$  

(A16)

Example to illustrate that stopping rules which depend upon the number of counted fibers can produce biased estimates

It suffices to consider a very simple example. Suppose we wish to estimate the average number of fibers per GO that are $\geq 10\ \mu m$ and, in fact, exactly one half of the fibers satisfy this condition. Further suppose that each GO contains exactly one fiber, so that the true average is 0.5. The counting rule is to count until a single fiber $\geq 10\ \mu m$ is counted and the estimator is the total number of fibers $\geq 10\ \mu m$ counted divided by the number of GOS examined.

In this case, the probability that $n$ GOS are counted is $\frac{1}{n}$, and if $n$ GOS are counted, then $n$ fibers are examined, one of which is $\geq 10\ \mu m$ and so the estimate is $\frac{1}{n}$. Thus, the expected value of this estimate is

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \ln(2) \approx 0.693 \neq 0.5,$$

and consequently, this estimator is biased.

REFERENCES


