# APPENDIX C

# CHARACTERIZATION OF UNCERTAINTY AND BIAS IN ASBESTOS EPIDEMIOLOGICAL DATA

## 1.0 OVERVIEW

Each of the epidemiological studies utilized in the fitting exercise described in this report provide data on the level of cumulative exposure in groups of workers and on the incidence of lung cancer or mesothelioma observed in those workers. However, both the independent variable (level of cumulative exposure) and the dependent variable (occurrence of disease) are subject to uncertainty and potential bias that may arise from numerous sources. This appendix describes the approach that OSWER is proposing for characterizing the uncertainty and potential bias in the reported data.

## 2.0 METHODS FOR CHARACTERIZATION OF UNCERTAINTY AND BIAS IN THE DATA

Uncertainty in a reported data item is attributable to random errors that occur when observations or measurements are collected. Typically, uncertainty extends in both directions around a measured value, and the true value may be either higher or lower than the measured estimate. Bias occurs when the measured data differ in a systematic (rather than random) way from the true values.

One way to characterize uncertainty and/or bias around a measured data item in through the specification of a probability density function (pdf). The density function describes the relative probability that some alternative value is the true value of the measured data item. If there is no bias in the methods used to measure the value, the reported value will generally be located in the central portion of the probability density, which may be either symmetrical or skewed. If there is a known or suspected bias in the method used to collect a data item, the reported value is likely to be located either in the lower or upper part of the density (depending on the direction of the measurement bias).

Specification of a pdf for any uncertain or biased data input term has two elements: the mathematical form of the distribution (e.g., normal, lognormal, uniform, triangular, etc.), and the parameters of the distribution (e.g., mean, standard deviation, minimum, maximum, mode, etc.). The following sections describe the mathematical forms and parameters proposed for specifying

the density distributions for each of the major sources of uncertainty and bias in the data. It is important to emphasize that, because most of the densities used to characterize uncertainty and/or bias in the input data can only be specified at a screening level, the results of any analysis based on these distributions should also be interpreted as screening level, with conclusions about uncertainty being characterized semi-quantitatively ("low", "medium" or high"), rather than as falling within precise numeric ranges.

## 3.0 UNCERTAINTY AND BIAS IN CUMULATIVE EXPOSURE

As discussed in the main text, cumulative exposure (CE) to asbestos is expressed as follows:

Lung Cancer: $CE = C \cdot d_{10}$ (s/cc-yrs)
Mesothelioma: $CE = C \cdot Q$ (s/cc-yrs$^3$)

where:

C = concentration of asbestos is air (s/cc)
$d_{10}$ = exposure duration (yrs) excluding the most recent 10 years of life
Q = cubic function of exposure duration and time since first exposure (yr$^3$), excluding the most recent 10 years of life

There are numerous sources of uncertainty in these measures of cumulative exposure, and this uncertainty can be substantial. The following sections identify the main sources of uncertainty in cumulative exposure estimates, and describe the pdfs proposed to characterize the nature and magnitude of the uncertainty from each source.

For mathematical and programming convenience, most of the pdfs are specified as multipliers of the starting data value (x0), as follows:

$$x \sim x0 \cdot pdf$$

That is, the uncertainty distribution for the random variable x is generated by making random draws from the density distribution specified for x and multiplying each random draw by the value x0. When there is more than one source of uncertainty or bias affecting a data item, the distribution of alternative values is generated as follows:

$$x \sim x0 \cdot \prod pdf_i$$

### 3.1 Uncertainty Due to Sampling and Measurement Error in Concentration Values

In most studies, the assignment of the cumulative exposure value for each person-year of observation is based on a job-exposure matrix (JEM), which describes the average concentration of asbestos in air for a series of different jobs or locations, often divided into a series of time intervals. Each of these values is the average of some number of independent samples obtained in the specified time interval for the specified job or location. As always, the uncertainty around the mean of a set of independent measurements of concentration is a function of a) the number of samples used to computed the mean, b) the between-sample variability, and c) the nature of the underlying distribution. In addition, in the case of asbestos, there is an additional source of variation arising from random Poisson measurement error in the number of particles counted during the microscopic analysis of each sample. The relative importance of the Poisson counting error may be significant if the number of particles counted is small, but tends to diminish as the number of particles counted becomes large. For the purposes of this evaluation, it is assumed that most analyses of air samples collected from asbestos workplaces yielded particle counts sufficiently large that this contribution to uncertainty may be ignored.

The impact of uncertainty in the mean concentration in a work area due to sampling variability on the values of CE10 reported for each person-year of exposure is difficult to evaluate because of the multi-step nature of the procedure used to stratify the data and form exposure groups. Therefore, the possible magnitude of this source of uncertainty was explored using Monte Carlo simulation.

The simulation was performed in several steps. First, a job-exposure matrix (JEM) was established for four work areas (jobs) and for four consecutive time periods at a hypothetical workplace. For each of the 16 cells in this JEM, the values of individual air samples over time and space was assumed to be distributed lognormally with a specified true mean and true standard deviation. From each of these 16 specified distributions, a set of N samples were drawn at random, and the sample mean was calculated to represent the reported (observed) value for that cell in the JEM. The second step was to simulate the individual exposure histories of a series of hypothetical workers. The cohort was defined as all workers who worked at least 1 year between year a and year b, with the epidemiological study being performed in year c. The age and date of first exposure, the job worked, and the duration of exposure were all treated as random variables. The cumulative exposure (CE10) for each person-year was based either on the true (ideal) value or on the measured (with error) value of asbestos in air for the appropriate cell of the JEM. The vital status of each worker in each year of life was simulated in two steps. First, survival from all causes of death (other than workplace asbestos exposure) was simulated

by performing a random draw from a Bernoulli distribution B[1,p(i)], where p(i) is the all-cause probability of death in year i of life. The value of p was based on 1970 all-cause death rates for males in the United States. Second, if the simulated worker did not die from a non-workplace cause in year i, then the outcome for death from lung cancer in that year was simulated by a second draw from a Bernoulli distribution with probability equal to the risk of death from lung cancer in that year. This probability was calculated by computing the relative risk of lung cancer based on the true level of simulated cumulative exposure to asbestos and multiplying the relative risk by the baseline risk of lung cancer death in that year of life. The final step was to group the person-years of observation into exposure groups based on either the true or the measured CE10 values, and compare the results by calculating the ratio of the KL values estimated using the "measured" CE10 values divided by the KL estimated using the ideal (true) CE10 values.

Some example results are shown in Figure C-1. As seen, if the inter-sample variability in workplace air samples is relatively low (GSD = 2), the distribution of the ratio values is centered approximately on 1, with the width of the distribution tending to increase as sample size decreases. For higher inter-sample variability (GSD = 4), the distribution becomes left-shifted, indicating that the effect of measurement error in the exposure metric is to tend to decrease the estimate of potency (KL).

Unfortunately, none of the published studies provide sufficient detail on the raw data to allow an estimation of sample number and inter-sample variability used in the JEM. In the absence of data, a default assumption is made that the number of samples used in each cell of the JEM is unlikely to be much larger than 10, and that the inter-sample variability is unlikely to be smaller than a GSD of 2. Therefore, the distribution of Panel C of Figure C-1 is treated as the default. For convenience, this distribution was modeled as:

pdf(sampling error) ~ *triangular*(0.6, 1, 1.5)

### 3.2 Use of the Mid-Point of the Range as the Mean

In most studies, an exposure group is formed by combining the data for person-years of observation that fall within a specified range of values. For example, in a lung cancer study, one exposure group might be defined as all person-years of observations where the value of CE10 is between 10 and 50 f/cc-yrs. In these cases, the midpoint of the exposure range is usually taken as the point estimate (PE) of the average exposure for the group:

PE(average exposure) = Mid = (Min + Max) / 2

This approach will yield a reliable estimate of the mean exposure if the individual values of exposure that fall within the range for the group are distributed approximately at random (uniformly) within the range, but may be unreliable if individual values within the range are not distributed uniformly. For example, if the density of observations increases for increasing exposure, the mean will be somewhat closer to the upper bound than the lower bound of the range. Conversely, if the density of observations decreases as exposure increases, the mean will be closer to the lower bound than the upper bound of the range. In the absence of the raw data, the distribution of observations within the range is not known, so the true mean value for the group can not be computed with confidence from the bounds of the group. Therefore, for the purposes of this analysis, in cases where the point estimate of cumulative exposure for a group is based on the mid-point of the exposure range, uncertainty in this estimate of the point estimate is characterized as follows:

pdf(Exposure, bounded) ~ TRIANGULAR(Min/Mid, 1, Max/Mid)

### 3.3 Point Estimate Values for Groups with Unbounded Upper Range

In some cases, an exposure group is characterized only as being above some specified value (e.g., CE10 > 50 s/cc-yrs). In the absence of any additional data or assumptions, it would not be possible to specify an estimate of the mean exposure for such an unbounded group. However, it is considered likely that the (unreported) upper bound for the group will usually not be higher than about three-times the lower bound, because if there were many values higher than this, a new exposure group would likely have been established by the researchers. Thus, the upper bound in this case is assumed to be three-times the lower bound. Moreover, it is considered likely that sample density will tend to decrease as exposure increases, so it is expected that the mean will likely be somewhat closer to the lower bound than the (assumed) upper bound. Based on this, the point estimate for the group is assumed to be 1/3 of the distance from the lower bound to the (assumed) upper bound of the bin. Given these assumptions, the uncertainty in the mean of an exposure group with only a lower bound reported is modeled as:

pdf(Exposure, unbounded) ~ TRIANGULAR(1, 5/3, 3)

### 3.4 Uncertainty in Conversion from mppcf to PCM f/cc

In a number of studies, data on asbestos levels in the workplace are based on measurements of total airborne dust particles (reported in units of mppcf) rather than on direct measurements using filter-based PCM analysis. In these cases, the concentration in units of PCM f/cc is estimated from the particle data using a conversion factor (CF):

$$C(\text{PCM f/cc}) = C(\text{mppcf}) \cdot CF$$

The value of CF at a location may be determined by measuring the concentration values in a set of identical (paired) or similar (concurrent) samples by both techniques, and finding the ratio of the measurements. However, at some locations, no paired or concurrent measurements are available, so the conversion must be based on observations from other sites. The approach for characterizing uncertainty in the conversion factor for each of these two situations is presented below.

Site-Specific Conversion Factor is Available

The uncertainty in a site-specific conversion factor is a function of the number of paired or concurrent samples collected at the site, and the magnitude of between-sample variability in the ratio of the measured concentration values. This type of site-specific conversion factor is available at a number of workplaces considered in this evaluation. The uncertainty in each of these site-specific CF values is described in the study-specific descriptions (see Appendix A).

No Site-Specific Conversion Factor is Available

Table C-1 summarizes average conversion factors that have been observed at different sites. As seen, the conversion factor ranges substantially, both between and often within a site. Based on this, most investigators conclude that no single conversion factor is appropriate for all locations. However, most values tend to fall in the 1-10 range, and a value of about 3 is often assumed as a default when site-specific data were not available. However, because of the wide variability between sites, it is evident that the use of this default factor is associated with substantial uncertainty. For the purposes of this effort, the uncertainty associated with the use of an assumed (default) conversion factor of 3.0 is given by:

$$\text{pdf(Conversion factor)} \sim \text{TRIANGULAR}(0.33, 1, 3.33)$$

## 3.5 Uncertainty in Representativeness of the Measured Concentration Values

*Personal vs Stationary Monitors*

Ideally, the concentration values used to estimate the cumulative exposure for each worker would be based on the mean of a series of personal air monitor samples worn by each worker. However, in many cases, estimates of exposure are based on mean values for stationary air samplers located in various areas of the workplace rather than personal air measurements. This introduces uncertainty in exposure estimates because stationary air samplers measure

concentration values that may be different that what workers inhale (e.g., see Doll and Peto 1985). This is most likely to be an issue in cases where a worker is actively disturbing asbestos (e.g., pouring bags of asbestos into a hopper), because the concentration of asbestos in the breathing zone of the worker may be (much) higher than at some nearby stationary monitor. However, the converse may also occur. For example, if the stationary air monitors were purposely placed in areas of maximum airborne levels of asbestos, then the stationary air monitor values may tend to overestimate exposures of workers who do not work in that immediate area.

Table C-2 summarizes some data on the magnitude of the differences between paired or concurrent personal and stationary air samples for asbestos or dust that have been noted in the literature. As seen, most average ratios are relative small (between 1 and 2.5), but a few may be in the range of 4-10, and occasionally a value may also be lower than 1. Based on the data in Table C-2, the pdf selected to characterize uncertainty in cumulative exposure for a group due to the use of data from stationary area monitors is:

pdf(Personal vs stationary) = BETA(2,20,0.9,10)

This beta distribution ranges from 0.9 to 10, and has 88% of its mass between 1.0 and 2.5, with a mean of 1.72. The distribution is shown in Figure C-2. Note that, to the extent that actual personal exposures were higher than estimated based on stationary air measurements, this adjustment will tend to decrease potency estimates for asbestos.

*Temporal Representativeness*

In some cases, concentration values of asbestos in air at a workplace tend to decrease over time due to addition or improvement in dust control systems in various areas of the workplace. However, because many workplaces did not have a regular schedule for collection of airborne dust or asbestos levels, especially prior to the 1960s, data used to calculate exposures for workers exposed prior to about 1960 may be based on estimated values that are extrapolations across time. For the purposes of this evaluation, the uncertainty due to use of concentration data that are extrapolated over time is characterized as follows:

pdf(temporal representativeness) = TRIANGULAR(1-a, 1, 1+a)

where the value of a is selected to reflect the relative degree to which extrapolated concentration values were used to estimate cumulative exposure of workers:

| Data Description | Uncertainty | Value of a |
|---|---|---|
| Data are available over most of the exposure interval; use of extrapolated data is minor | Low | 0.1 |
| Data are available at intermittent times during the exposure interval; use of extrapolated data is moderate | Medium | 0.2 |
| Data are available only for a few times during the exposure interval; use of extrapolated data is predominant | High | 0.5 |

### 3.6 Use of Data Reported as CE Rather than CE10

As discussed in the main text, the relative risk of lung cancer at any specified age is believed to be a linear function of cumulative exposure up to that age, ignoring any exposures that occurred in the most recent 10 years of life. This measure of cumulative exposure is referred to as CE10.

Some epidemiological studies report observed and expected cases of lung cancer using cumulative exposure without adjusting for the 10-year lag. This is referred to as CE. When CE is used as the measure of cumulative exposure, both the measure of exposure and the measure of response are changed compared to when CE10 is used.

In order to investigate the magnitude of the effect of using CE rather than CE10 as the metric of exposure, a number of Monte Carlo experiments were performed in which exposure histories for a number of workers (100) were simulated. For each worker, the exposure concentration, age at first exposure, and duration of exposure were all treated as random variables. Each worker was followed either until death (simulated based on national mortality statistics for male smokers) or until the time of the epidemiological study (whichever came first). The relative risk (RR) of lung cancer for each person year of observation was calculated based on the assumption that RR was a function of CE10, not CE. The simulated data were then used to form 4 exposure groups based either on CE10 or CE, selecting the bin cutoffs to yield approximately equal numbers in each bin. For the CE10 approach, all person years from the first 10 years of exposure (CE10 = 0) were excluded. For the CE approach, two alternative strategies were assessed. In the first strategy, all person years were included. In the second approach, the first 10 years of observation were excluded.

Figure C-3 shows several example results. In all cases, when groups are formed by grouping person-years of observation based on CE rather than CE10, the within-study exposure-response curve remains linear, but the slope of the line (an estimate of the potency for that particular atmosphere) is reduced. That is, when CE is used, the magnitude of the response is

underestimated compared to when CE10 is used, and this can result in an underestimation of bin-specific potency factors.

The magnitude of the effect depends on a number of attributes of the study. Most important are the average length of follow-up and the average exposure duration, with the bias being higher for short follow-up than for long follow-up (compare Panels A and B). If the first 10 years of person-years of observation are excluded, the bias is diminished (Panel C).

Because the magnitude of the bias associated with the use of CE rather than CE10 depends on the details of each study design, the same Monte Carlo approach described above was used to estimate a bias correction factor (BCF) for each study, based on the attributes of the study. The BCF was calculated by finding the ratio of the slope of the study-specific exposure-response curves based on the CE approach compared to the slope based on the CE10 approach:

$$BCF = Slope(CE) / Slope(CE10)$$

This BCF is independent of assumed concentration level and assumed KL, and is only weakly dependent on the between-worker variability in exposure concentration. The BCF is used to adjust reported CE values to yield an estimated CE10 value as follows:

$$CE10(estimated) = CE \cdot BCF$$

A similar approach is used to estimate the BCF for studies where data were grouped by person rather than by person years (e.g., U.S. retirees).

Table C-3 shows the study attributes used in the Monte Carlo simulations for each study. variability in exposure duration was modeled as a beta distribution, bounded by a minimum of 1 year and a maximum of the reported or estimated maximum duration. The shape parameters were selected to yield a mean exposure duration similar to the reported or assumed value. The variability in concentration was assumed to be characterized by a lognormal distribution with a GSD of 2.5. Because age at first exposure, exposure duration, exposure concentration, and length of follow-up are all random variables in each simulation, the BCF for each study is a distribution rather than a constant. Figure C-4 shows one example. For simplicity, each study-specific BCF was modeled as a triangular distribution, with parameter values derived from the $5^{th}$, $50^{th}$, and $95^{th}$ percentiles of the study-specific Monte Carlo simulation:

$$BCF \sim TRI(5th, 50th, 95th)$$

### 3.7 Uncertainty in Cumulative Exposure Estimates for Mesothelioma Studies

Recall that, for mesothelioma, cumulative exposure is expressed in terms of C·Q, where:

C = Exposure concentration (s/cc)
$Q = (T - 10)^3 – (T- 10 -d)^3$

and:

T = time since first exposure (yrs)
d = exposure duration (yrs)

In the ideal case, exposure-response data for mesothelioma studies would be based on person-years of exposure grouped according to the value of C·Q for each person year of observation. However, there are no studies that report mesothelioma exposure-response data in this format. Therefore, in order for a study to be employed in the quantitative model fitting, the value of C·Q must be estimated for each group based on the data reported in the study.

Peto et al (1985) reported mesothelioma incidence as a bi-variate function of T and d. Using the mid-points of the bins for T and d, these data allow computation of the average value of Q for each group. The value of C·Q for each group is then estimated by multiplying the group-specific values of Q by the study-wide average value of C.

For other mesothelioma studies utilized in this effort, incidence is reported as a mono-variate function of time since first exposure (T). In this case, the average value of Q ($\overline{Q}$) for a bin bounded between T = a and T = b may be estimated from $\overline{T}$ = (a+b)/2 and the average exposure duration ($\overline{d}$), as follows:

$$\overline{Q} \approx q(\overline{T}, \overline{d})$$

where:

$\overline{Q}$ = Average value of Q for person-years in the bin
$\overline{T}$ = Average value of T for person-years in the bin
$\overline{d}$ = Average value of duration for person-years in the bin

$$q(\overline{T}, \overline{d}) = \begin{cases} 0 & \overline{T} \le 10 \\ (\overline{T} - 10)^3 & 10 < \overline{T} \le 10 + \overline{d} \\ (\overline{T} - 10)^3 - (\overline{T} - 10 - \overline{d})^3 & \overline{T} > 10 + \overline{d} \end{cases}$$

However, this approach is subject to error because Q is not a linear function of T. That is:

$$q(\overline{T}, \overline{d}) \neq \overline{q}(a, b, \overline{d})$$

The error caused by this approximation is to underestimate $\overline{Q}$, which tends to lead to an overestimation of KM. The magnitude of the error is generally small if the bin width is narrow, but can become substantial if the bin width is wide, especially if data for T < 10 are included in the bin (a < 10).

In order to investigate whether or not these approximation techniques result in significant uncertainty or bias compared to the case where results are grouped based on actual values of C·Q, a Monte Carlo simulation approach was used, similar to that described above for assessing the effect of CE vs CE10 for lung cancer. That is, the detailed exposure histories of a series of 100 workers was simulated, treating the age at first exposure (T), exposure duration (d), and exposure concentration (C) as random independent variables. Each worker was observed either until death (simulated using all-cause death rates for male smokers), or until the time of observation. For each person year of observation, the expected incidence of mesothelioma was calculated from the value of C·Q. The results were then grouped in two different ways. In case 1 (ideal), each person-year was grouped according to the value of C·Q. In the second case (approximated), the results were grouped according to time since first exposure, and the value of C·Q for each group was estimated as described above. In both cases, the data were fit to a linear model with zero intercept, and the ratio R of the slopes of the lines (the KM values) was computed as:

R = Slope based on Case 2 (approximation) / Slope based on Case 1 (ideal data)

Figure C-5 shows an example result. As seen, the effect of these approximation methods results in a slightly skewed uncertainty distribution centered close to 1.0. The width of the uncertainty distribution depends on a number of study-specific attributes, including exposure duration, length of follow-up, and variability in concentration value. Assuming that variability in the yearly average exposure concentration for each worker is moderate (modeled as a lognormal distribution with a GSD of 2.5), then variability in the ratio R is generally characterized by a 5$^{th}$ percentile value between about 0.7 and 0.85, and a 95$^{th}$ percentile value of about 1.3 to 1.5. Based on this, the following pdf is selected to approximate the uncertainty due to the use of these approximation methods:

pdf(approximation of C·Q) ~ TRI(0.75, 1, 1.4)

### 3.8 Uncertainty in Mesothelioma Studies by McDonald et al.

In three mesothelioma studies by McDonald et al. (1982, 1983, 1984), the authors reported the total number of mesothelioma cases, but the number of person years of observation were not reported. Therefore, incidence of mesothelioma can not be calculated directly. However, total (all-cause) mortality was reported for the cohort, stratified according to age at death. These data can be used to estimate the person years of observation for each group using the procedure described in Appendix A, Attachment A-1. The values of C·Q for each group can be estimated using an approach similar to that described above, where the value of T (time since first exposure) is estimated as the midpoint of the age bin minus the average age at first exposure.

In order to investigate whether or not these approximation techniques result in significant uncertainty or bias in the data compared to the case where results are grouped based on actual person-years and actual values of C·Q, a Monte Carlo simulation approach was used, similar to that described above. That is, the detailed exposure histories of a series of 100 workers was simulated, treating the age at first exposure (T), exposure duration (d), and exposure concentration (C) as random independent variables. Each simulated worker was observed either until death from mesothelioma or from other causes, or until the time of observation. For each person-year of observation, the expected incidence was calculated from the value of C·Q. The results were then grouped in two different ways. In case 1 (ideal), person-years were grouped according to the value of C·Q. In the second case (mimicking the data reported in the three McDonald reports), the person-years were grouped according to age at observation. The number of all-cause deaths was then used to estimate the number of person years of observation for each group (see Attachment 1), and the value of C·Q was estimated as described above. In both cases, the data were fit to a linear model with zero intercept, and the ratio R of the slopes of the lines (the KM values) was computed as:

R = Slope based on Case 2 (approximation) / Slope based on Case 1 (ideal data)

Figure C-6 shows the results. Panel A shows the error introduced by using the estimated rather than the true values of C·Q. Panel B shows the error introduced by using the estimated rather than the true number of person years. Panel C shows the combined error when both approximations are used (as is the case with the actual studies). As seen, both approximations introduce uncertainty, with a slight left shift for estimation of C·Q, a slight right shift for estimation of person years, and a distribution centered approximately on 1 when the two approximations are combined. Based on this, the following pdf is selected to approximate the uncertainty due to the use of these approximations for the three mesothelioma studies by McDonald et al. (1982, 1983, 1984):

pdf(combined McDonald approximations) ~ TRIANGULAR(0.4, 1, 2.5)

## 4.0 UNCERTAINTY IN RESPONSE

As discussed in the main text, error in the response variable (observed number of cases in each group) is assumed to be Poisson distributed, and this variability is accounted for by use of the Poisson MLE fitting strategy. However, diagnosis error may also contribute to the error in the reported number of cases. For example, in the absence of detailed histopathological confirmation, mesothelioma cases may sometimes be mis-classified as lung cancer, which would lead to a tendency for over-estimation of the reported incidence of lung cancer and underestimation of the reported incidence of mesothelioma.

For the purposes of this analysis, the effect of potential misdiagnosis is ignored. This is because the magnitude of the error, if any, is unknown, and because the error is likely to be small compared to the magnitude of the Poisson variation. For example, using the data from the cohort of New Jersey insulation manufacturers (Seidman et al. 1986), the total number of lung cancer cases is 111 and the total number of mesothelioma cases is 17. Assume that the misdiagnosis rate for mesothelioma is about 10% (i.e., 10% of all true mesothelioma cases are diagnosed as lung cancer). This would mean that about 2 cases of mesothelioma were reported as lung cancer, and the correct values would be 109 for lung cancer and 19 for mesothelioma. However, the 95% Poisson confidence interval around a value of 109 is from 92 to 128, and the interval around a value of 19 is from 12 to 28. Thus, assuming that mesothelioma diagnosis error is not substantially higher than 10%, the consequence of ignoring the diagnosis error is expected to be minimal.

## 5.0 ERROR IN THE CALCULATION OF BIN-SPECIFIC CONCENTRATIONS

### 5.1 Uncertainty in Fraction Amphibole

Estimates of the amphibole content of each workplace atmosphere were based on information provided from each published study. As discussed in Section 5.3.2, available studies may be classified into three groups, as follows:

- Chrysotile Only
- Chrysotile plus Amphibole
- Amphibole Only

The strategy for characterizing uncertainty in fraction amphibole for each case is described below.

*Chrysotile Only Studies*

Some epidemiology studies evaluate workplaces where only chrysotile asbestos was used. In these cases, it is expected that fraction amphibole is approximately zero. However, chrysotile asbestos may contain trace levels of amphibole asbestos. For example, Addison and Davies (1990) used XRD to measure the amount of amphibole in 81 different samples of chrysotile asbestos. The results are summarized below.

Sample Results as Reported by Addison and Davies (1990)

| Number of samples | Tremolite content as percentage of sample | | | |
|---|---|---|---|---|
| | Number with detectable tremolite | Average in detects | Range (where detected) | Detection limit range |
| 81 | 28 | 0.09 | 0.1-0.6 | 0.01-0.06 |

Source: Addison and Davies (1990) Table 7.

As seen, tremolite was detected in 28 of the 81 samples (35%), with a detection limit of 0.01% - 0.06%. In the samples where tremolite was detected, the average was 0.09%, and the maximum was 0.6%. Based on these data, the point estimate for fraction amphibole in "pure" chrysotile may be derived as the count-weighted average tremolite concentration, treating non-detects at ½ the average detection limit (0.035%):

$$\text{Point Est. (fraction amphibole)} = (28 \cdot 0.09\% + 53 \cdot 0.035\%) / 81 = 0.054\%$$

The uncertainty around this point estimate may be reasonably characterized by a lognormal distribution with a mean of 0.054 and a standard deviation of 0.1. This distribution would yield an expected detection frequency of about 40% based on an average detection limit of 0.035%, and the maximum observed value (0.6%) corresponds to the $99.5^{th}$ percentile.

*Mixed Chrysotile and Amphibole Studies*

In some studies, the description of the workplace and its operations makes clear that both chrysotile and amphibole were used in the workplace. The choice of the fraction amphibole term and the characterization of the uncertainty about the term is study-specific, based on the level of information available to estimate the term. Thus was generally characterized as a triangular pdf, as follows:

$$\text{pdf(fraction amphibole)} \sim \textit{triangular}(\text{lb, be, ub})$$

*Amphibole Only Studies*

Some epidemiological studies are performed in workplaces where only amphibole is stated to be present. No data were located to indicate that chrysotile occurs as a trace contaminant of amphibole asbestos, so the fraction amphibole term was assumed to be 100% in these cases.

## 5.2 Uncertainty in Particle Size Data

As discussed in Section 5.3.2, the concentration of asbestos in each bin is estimated from the reported data but multiplying by a factor (referred to as k[b]) that is the ratio of the fraction of fibers in bin "b" to the fraction of fibers that are PCM(E). The data needed to calculated the k[b] terms are derived from studies of particle size distribution that typically include several thousand fibers (see Appendix B), so there is very little uncertainty in these terms, and they were treated as constants. However, because the particle size data are based on samples collected in workplaces that are not identical to the environment where the health effects data were collected, there is uncertainty as to the relevance of the available particle size distributions to the actual particle size distributions. For screening purposes, this uncertainty in data relevance was also modeled as a triangular distribution:

$$pdf(k[i]) \sim triangular(1\text{-}x, 1.0, 1+x)$$

The assignment of the parameter x that characterizes the uncertainty in the extrapolation of the particle size ratio data is very subjective. For the purposes of this effort, each extrapolation was ranked as having low, low, medium, or high uncertainty based on a consideration of how nearly the setting where the particle size data were collected matches the setting where the epidemiological data were collected, especially with regard to three variable:

- The type of industry
- The type of operations being performed
- For amphibole, the form of the amphibole (amosite, crocidolite, tremolite, etc.)

In cases where more than one type of amphibole is used, emphasis is placed on matching the primary (most common) type. The following tables summarizes how this information was used to make uncertainty level assignments:

Level of Uncertainty in Extrapolation of Particle Size Data

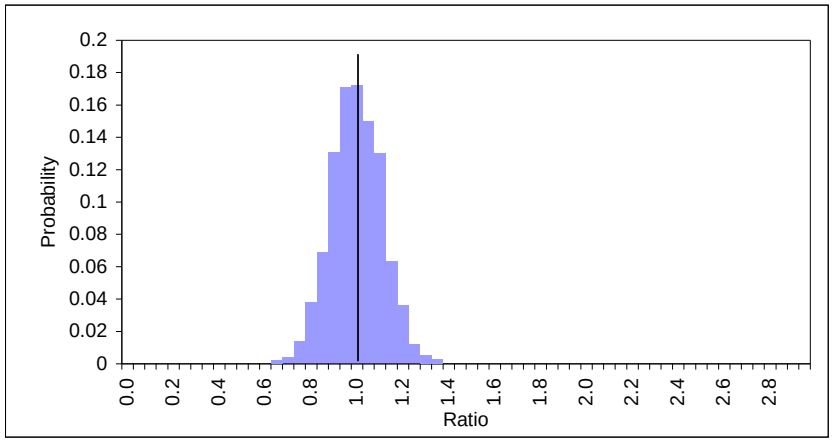
| Uncertainty Level | Chrysotile | Amphibole | x | pdf |
|---|---|---|---|---|
| Low | The industry is the same | The industry is the same<br>The primary type of amphibole is the same | 0.2 | *triangular*(0.8, 1, 1.2) |
| Medium | The industry is different | The industry is the same<br>The primary type of amphibole is different | 0.5 | *triangular*(0.5, 1, 1.5) |
| | | The industry is different<br>The primary type of amphibole is the same | | |
| High | Not used | The types is different<br>The industry is different | 0.8 | *triangular*(0.2, 1, 1.8) |

## 6.0 SUMMARY

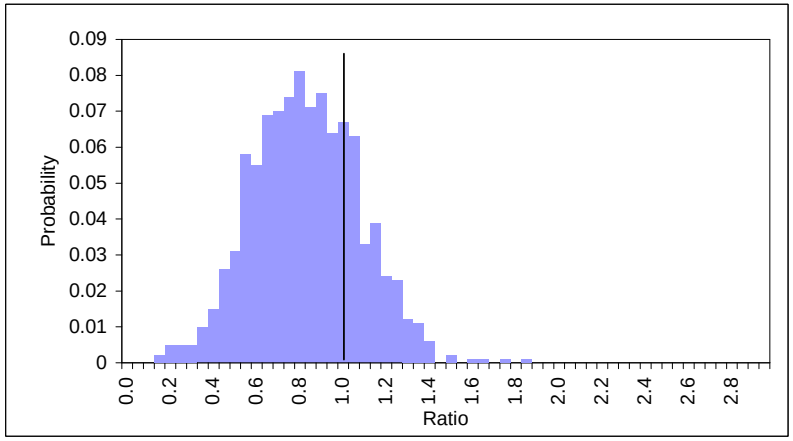
Table C-4 summarizes the general forms of the probability densities used to characterize the uncertainty and/or bias in data items used in the fitting process. The densities selected for use in each study are presented in Appendix A.

## FIGURE C-1 EFFECT OF SAMPLING ERROR

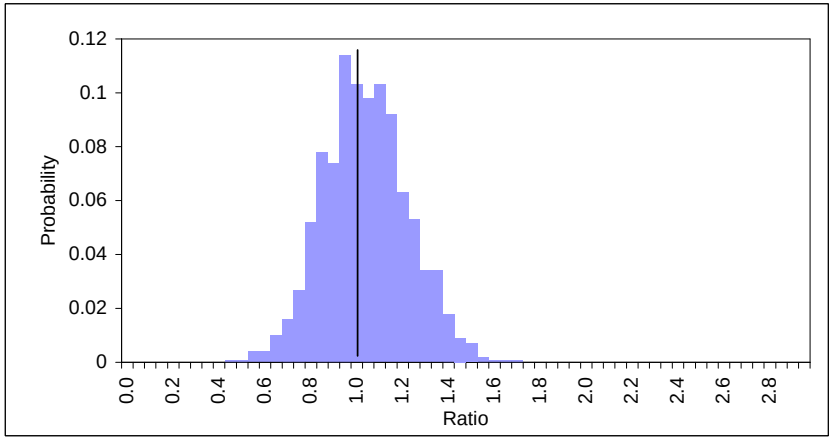
Panel A: GSD = 2 N = 30




Panel B: GSD = 4 N = 30




Panel C: GSD = 2 N = 10




Panel D: GSD = 4 N = 10

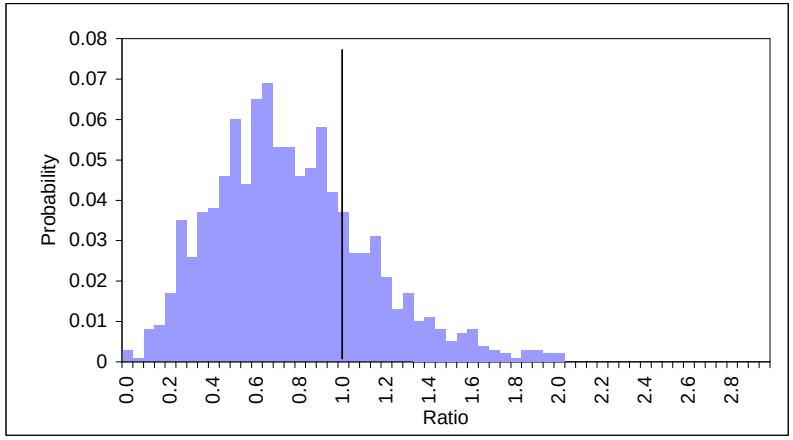
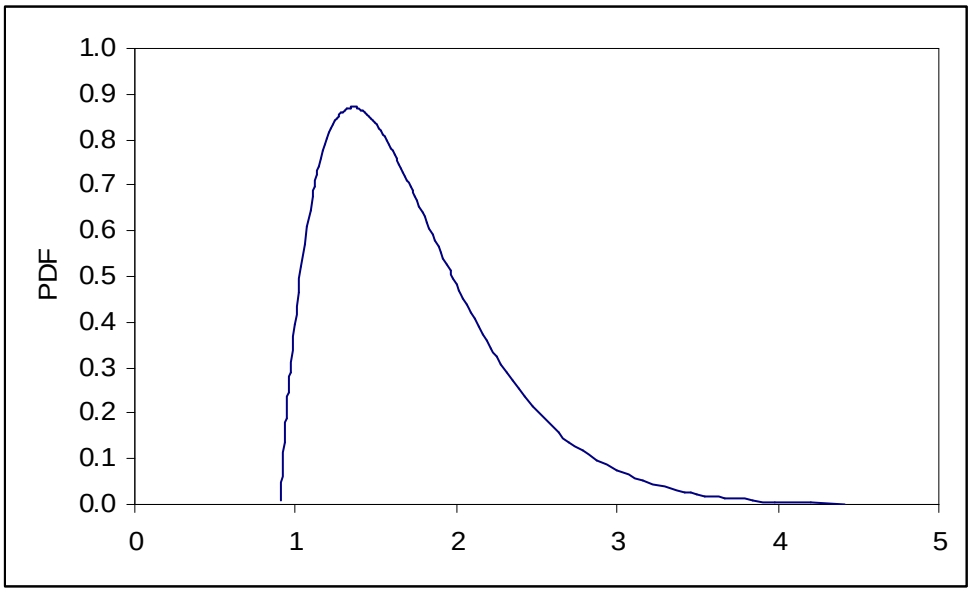
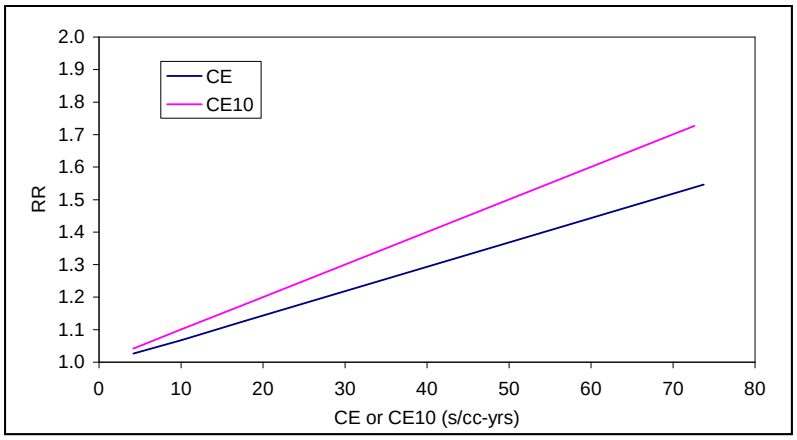


**FIGURE C-2**
**BETA DISTRIBUTION USED TO ACCOUNT FOR BIAS AND UNCERTAINTY IN USE OF STATIONARY AIR MONITORS**

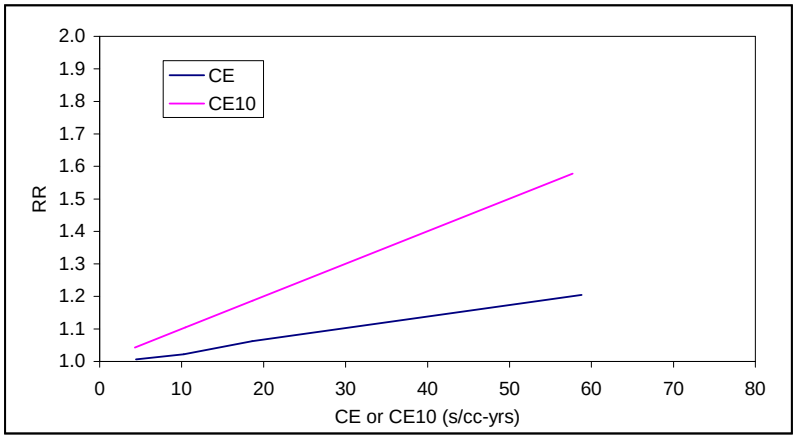


## FIGURE C-3
## EFFECT OF USING CE RATHER THAN CE10 IN LUNG CANCER STUDIES

Panel A: Long Followup




Panel B: Short Followup




Panel C: Short Followup, 10 PYAR Excluded

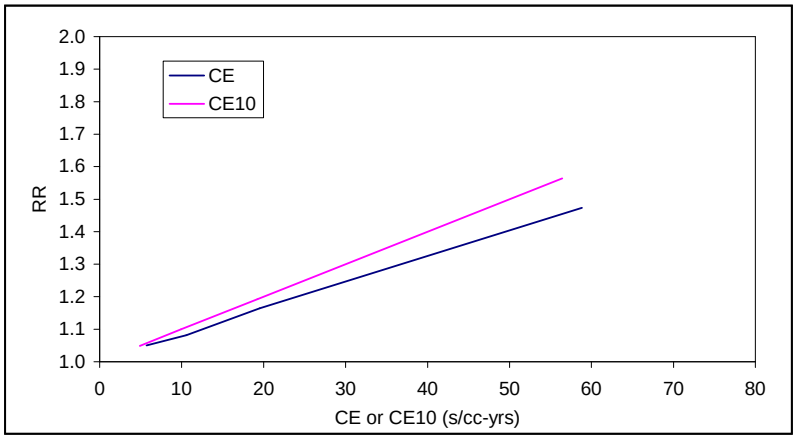
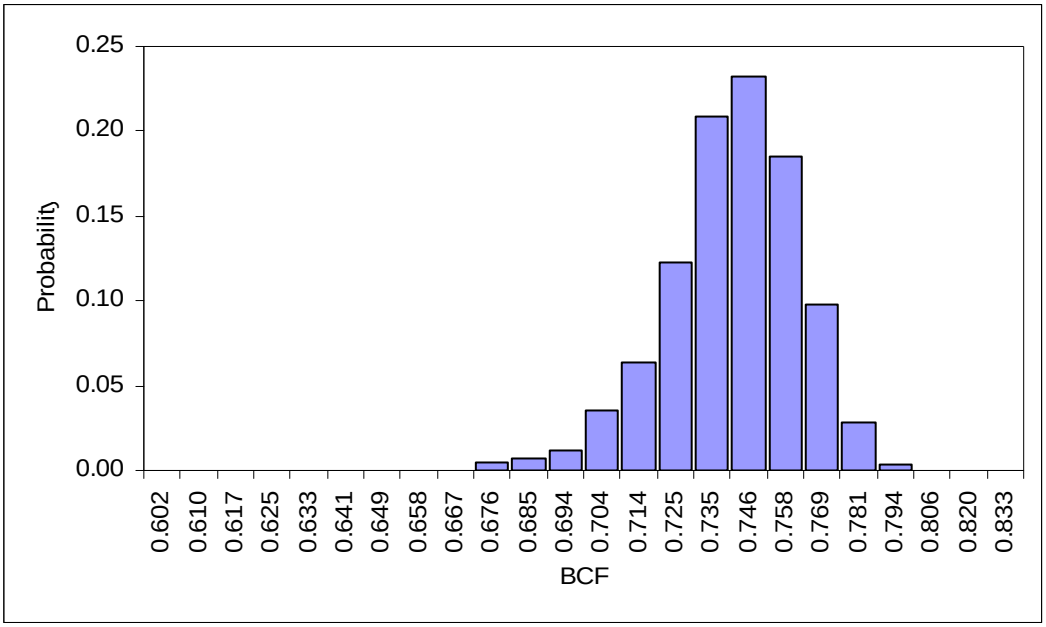


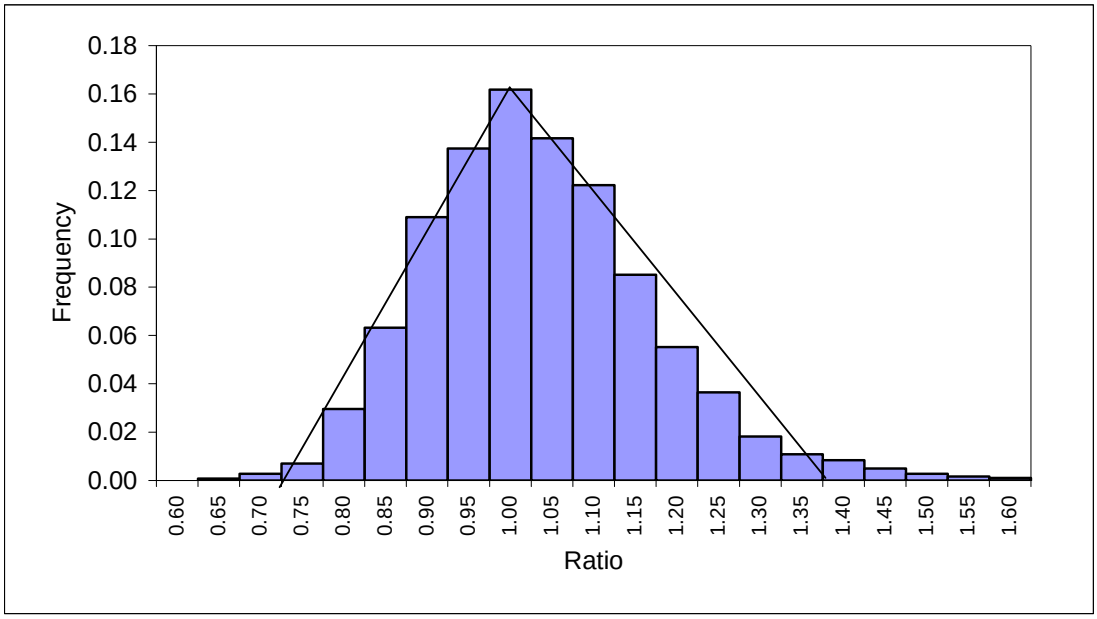
**FIGURE C-4**
**EXAMPLE DISTRIBTION OF BCF TO ACCOUNT FOR**
**USE OF CE RATHER THAN CE10 IN LUNG CANCER STUDIES**




Simulation parameters used in this example:

Duration ~ *beta*(2,4,2,20) (mean = 8.0)
Concentration ~ *lognormal*(10, 15)
Cohort definition: Must work 1 year between 1930 and 1960
Year of study: 1990 (average follow-up = 31.5 years)
PYAR Excluded = 0

**FIGURE C-5**
**EXAMPLE UNCERTAINTY ASSOCIATED WITH**
**APPROXIMATION OF C·Q TERM IN MESOTHELIOMA STUDIES**




Simulation parameters used in this example:

Duration ~ *beta*(2,8,2,20) (mean = 5.6)

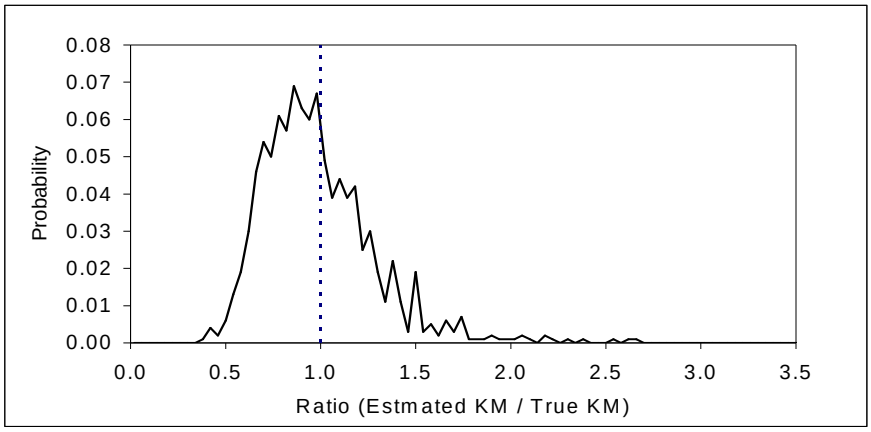
Concentration ~ *lognormal*(3, 3.4)

Cohort definition: Must work 1 year between 1930 and 1960

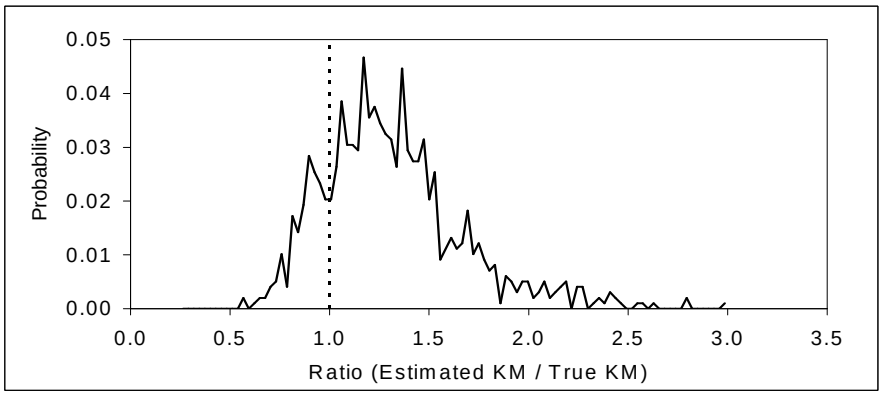
Year of study: 2000 (average follow-up = 35 years)

**FIGURE C-6**
**UNCERTAINTY ASSOCIATED WITH APPROXIMATIONS REQUIRED TO UTILIZE MESOTHELIOMA STUDIES BY McDONALD ET AL.**

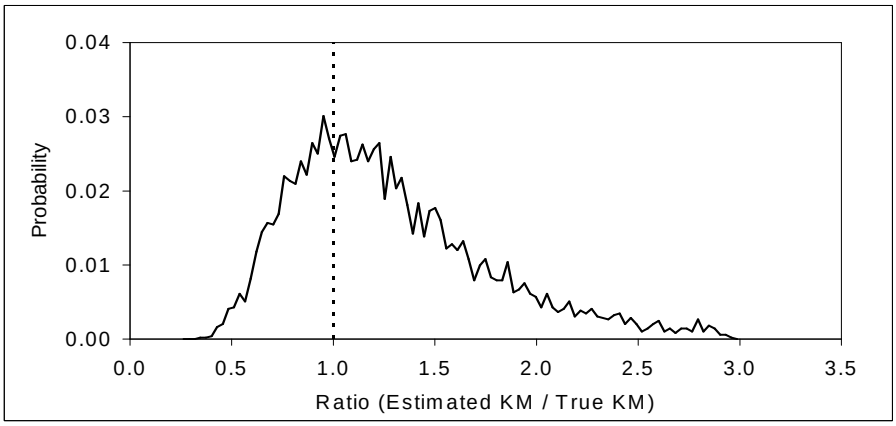
Panel A: Effect of Estimating Cumulative Exposure




Panel B: Effect of Estimating Person-Years of Observation




Panel C: Effect of Combined Approximations



# TABLE C-1
# SUMMARY OF CONVERSION FACTORS
# REPORTED IN THE LITERATURE

| Industry | Operation | Asbestos Types | N | CF (f/cc per mppcf) | Notes | Reference |
|---|---|---|---|---|---|---|
| Four Textile Plants | Fiber Preparation | Chrysotile | | 5.2 | Too few fibers > 5 um were counted to yield reliable ratios. However, authors stated that 10 total f = 6 f > 5 um. Thys, values shown are based on ratios for all fibers multiplied by 0.6. | Ayer et al. 1965 |
| | Carding | | | 7.6 | | |
| | Spinning | | | 5 | | |
| | Twisting | | | 3.9 | | |
| | Weaving | | | 7.5 | | |
| | All | | 230 | 5.6 | | |
| South Carolina Textile Plant | All operations | Chrysotile | 120 | 2.9 | paired samples | Dement et al. 1983 |
| | All ops except fiber prep. | | 986 | 2.5 | concurrent samples | |
| | Fiber preparation | | | 7.8 | | |
| Quebec mining and milling | overall mean Mine A | Chrysotile | 28 | 4.5 | Low correlation coeffient (0.32) | Gibbs and LaChance 1974 |
| | overall mean Mine B | | 18 | 11.4 | | |
| | overall mean Mine C | | 18 | 21.9 | | |
| | overall mean Mine D | | 11 | 5.9 | | |
| | overall mean Mine E | | 12 | 1.7 | | |
| | All Underground | | 1 | 1.7 | | |
| | All Open Pit - Drill | | 7 | 5.3 | | |
| | All Open Pit - Shovel | | 7 | 4.6 | | |
| | All Dryer | | 11 | 9.5 | | |
| | All Crushers | | 17 | 5.3 | | |
| | All Mill Rock Screening | | 12 | 14.2 | | |
| | All Mill Fiber Screening | | 14 | 11 | | |
| | All Mill Bagging | | 11 | 10.4 | | |
| | All Mill Storage | | 7 | 8.1 | | |
| | 7 mines | | 623 | mppcf factor<br>0.1 2.3<br>1.0 11<br>10 52 | Non-linear relationship given by: s/cc = 10.97·(mppcf)^0.68 Uncertainty range is wide (100-fold) | Dagbert (1976) as summarized in Doll and Peto (1985) |
| | Thetford and Asbestos mines | | 10205 | 3.46 (0.3 - 30) | Mean and range for all jobs; values for individual jobs not reported | Liddell et al. 1984 |
| New Orleans asbestos cement products | forming | Mainly chrysotile, some crocidolite and amosite | 23 | 0.63 | correlation coefficient = 0.18 | Hammad et al. 1979 |
| | mixing | | 27 | 1.3 | correlation coefficient = 0.91 | |
| | shingle finishing | | 14 | 1.1 | correlation coefficient = 0.31 | |
| | corrugated finishing | | 23 | 2.5 | correlation coefficient = 0.47 | |
| | panel plant | | 15 | 1.5 | correlation coefficient = 0.31 | |
| | All | | 102 | 1.4 | correlation coefficient = 0.57 | |
| Swedish cement plant | Milling | Mainly chrysotile, some crocidolite and amosite | | 0.45 | | Albin et al. 1990 |
| | Mixing | | | 0.10 | | |
| | Machine line | | | 0.15 | | |
| | Sawing | | | 0.19 | | |
| Montana vermiculite mine and mill | Mine and dry mill | Amphibole | 336 impinger and 81 filter (not paired) | 4.0 | Data from 1967-1971; ratios from other time intervals ranged from 1.2 to 11.5. | Amandus et al. 1987 |

**TABLE C-2**
**COMPARISON OF CONCENTRATION VALUES MEASURED BY PERSONAL AND STATIONARY AIR MONITORS**

| Reference | Setting | Endpoint | Ratio (Personal / Stationary) | |
|---|---|---|---|---|
| | | | N | Average |
| Sakai et al (2006) | Brake shop (location 1) | Asbestos | 37 | 9.26 |
| | Brake shop (location 2) | Asbestos | 29 | 1.1 |
| | Brake shop (location 3) | Asbestos | 21 | 1.27 |
| Lange et al. (2000) | Asbestos abatement (location 1) | Asbestos | 9 | 2.5 |
| | Asbestos abatement (location 2) | Asbestos | 13 | 2.5 |
| | Asbestos abatement (location 3) | Asbestos | 12+ | 2 |
| Ferro et al. (2004b) | Residential living space | $PM_{2.5}$ | | 1.4 |
| | | $PM_5$ | | 1.6 |
| Ferro et al. (2004b) | Residential living space | $PM_5$ | | 3.8 |
| Lange et al. (1996) | Asbestos abatement | Asbestos | 42 | 0.97 |

## TABLE C-3. BIAS CORRECTION FACTORS FOR LUNG CANCER STUDIES BASED ON CE RATHER THAN CE10

| Cohort | Reference | Cohort Definition | | Date of Observation | Duration (years) | | PYAR Excluded | BCF | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Start Year | End Year | | Range | Mean | | 5th | 50th | 95th |
| British Friction Product Workers | Berry and Newhouse 1983 | 1941 | 1977 | 1979 | 1 to >30 | 5 (assumed) | 0 | 0.53 | 0.61 | 0.67 |
| U.S. Retirees | Henderson and Enterline 1979 | 1916 (retired in 1941) | 1967 (retired) | 1973 | 3-51 | 25 | 0 | 0.67 | 0.74 | 0.81 |
| Ontario Cement Manufacturers | Finkelstein 1984 | (1948) | 1960 | 1980 | 1-20 (assumed) | 6.67 | 20 | 0.98 | 0.99 | 1.00 |
| Quebec Miners | McDonald et al. 1993 | 1891 (birth) | 1920 (birth) | 1989 | 2.5-49 | 7 (assumed) | 20 | 0.99 | 1.00 | 1.00 |
| Connecticut Friction Product Workers | McDonald et al. 1984 | 1938 | 1959 | 1977 | 0.1-45 | 8.04 | 20 | 0.99 | 1.00 | 1.00 |
| Italian Miners | Piolatto et al. 1990 | 1946 | 1987 | 1987 | 1-??? | 5.94 | 0 | 0.56 | 0.63 | 0.69 |
| New Jersey Insulation Manufacturers | Seidman et al. 1986 | 1941 | 1945 | 1982 | 0.05-14 (most < 2 years) | 0.94 | 5 | 0.79 | 0.83 | 0.85 |
| Swedish Cement Manufacturers | Albin et al. 1990 | 1907 | 1977 | 1986 | 0.25-47 | 15 (assumed) | 20 | 0.86 | 0.91 | 0.94 |
| Libby, MT Miners | McDonald et al. 2004 | (1923) | 1963 | 1998 | 3-16 (based on means for different jobs) (min. = 1 year) | 8.7 | 10 | 0.84 | 0.86 | 0.88 |
| Australian Miners | de Klerk et al. 1989 | 1943 | 1966 | 1980 | < 0.1 to > 5 | 1 | 0 | 0.53 | 0.62 | 0.68 |
| Belgium Cement Manufacturers | Lacquet et al. 1980 | 1963 | 1977 | 1977 | 1-??? | 7 (assumed) | 0 | 0.22 | 0.30 | 0.36 |
| Austrian Cement Manufacturers | Neuberger and Kundi 1990 | 1950 | 1981 | 1986 | 3-??? | 7 (assumed) | 0 | 0.54 | 0.61 | 0.67 |
| Chinese Workers | Yano et al. 2001 | (1939) | 1972 | 1996 | SD ? 7 | 24.6 | 0 | 0.70 | 0.76 | 0.82 |

## TABLE C-4. SUMMARY OF PROBABILITY DENSITIES

| Data Category | Uncertainty Source | Probability Density Function | Parameter Values |
|---|---|---|---|
| Exposure Response | Sampling and measurement error in values used to compute average exposure concentration | triangular(a, b, c) | a = 0.6<br>b = 1.0<br>c = 1.5 |
| | Use of the mid-point of a bin | triangular(a, b, c) | a = lb<br>b =mid<br>c =ub |
| | Un-bounded bin | triangular(a, b, c) | a = lb<br>b = 5/3·lb<br>c = 3·lb |
| | Conversion from mppcf to PCM | Study-specific (when available) | |
| | | Default = triangular(a,b,c) | a = 1<br>b = 3<br>c = 10 |
| | Personal vs Stationary monitors | beta(p1, p2, min, max) | p1 = 2<br>p2 = 20<br>min = 0.8<br>max = 10 |
| | Temporal representativeness | triangular(1-a, 1, 1+a) | a = 0.1 (low uncertainty)<br>a = 0.2 (medium uncertainty)<br>a = 0.5 (high uncertainty) |
| | Use of CE rather than CE10 in lung cancer studies | Study specific (see Table C-3) | |
| | Approximation of C·Q term in mesothelioma studies | triangular(a,b,c) | a = 0.75<br>b = 1.0<br>c = 1.4 |
| | Estimation of person years and C·Q in mesothelioma studies by McDonald et al | triangular(a, b, c) | a = 0.4<br>b = 1.0<br>c = 2.5 |
| Composition of the atmosphere | Fraction amphibole | Chrysotile only: lognormal(m,s)<br>Mixed: triangular(lb, be, ub)<br>Amphibole only: triangular(1,1,1) | m = 0.00054, s = 0.001<br>lb, be, ub based on info. from study |
| | Use of k[b] from TEM data sets to calculate bin-specific concentrations | triangular(1-x, 1, 1+x) | x = 0.2 (low uncertainty)<br>x = 0.5 (medium uncertainty)<br>x = 0.8 (high uncertainty) |

lb = lower bound
mid = midpoint
ub = upper bound