Draft, May 2010
Do Not Cite or Quote
Development Support Document

Appendices for

Section 4.2 Carcinogenic Potential

Arsenic and Inorganic Arsenic Compounds

CAS Registry Numbers:
7440-38-2 (Arsenic)

Prepared by

Neera Erraguntla, Ph.D.
Roberta L. Grant, Ph.D.
Toxicology Division

Chief Engineer's Office

TEXAS COMMISSION ON ENVIRONMENTAL QUALITY

TABLE OF APPENDICES

Appendix A. Lung Cancer Mortality/Incidence Rates and Survival Probabilities

	US Total Population 2000-2003	Texas Statewide 2001-2005	US Total Population 1975-2005	Texas Statewide 2001-2005
	Total Lung Cancer Mortality Rates per 100,000 [1]	Total Lung Cancer Mortality Rates per 100,000 [2]	Total Lung Cancer Incidence Rates per 100,000 [3]	Total Lung Cancer Incidence Rates per 100,000 [4]
Years	Rate	Rate	Rate	Rate
00	0.0	0.0	0.0	0.0
01-04	0.0	0.0	0.0	0.0
05-09	0.0	0.0	0.0	0.0
10-14	0.0	0.0	0.0	0.0
15-19	0.0	0.0	0.1	0.1
20-24	0.1	0.1	0.3	0.3
25-29	0.2	0.2	0.5	0.5
30-34	0.6	0.4	1.1	1.2
35-39	2.5	1.6	3.6	3.0
40-44	8.8	7.9	10.9	12.2
45-49	20.6	18.6	25.5	28.0
50-54	40.9	36.7	51.5	54.1
55-59	81.5	75.1	102.3	107.2
60-64	148.8	143.8	184.9	199.2
65-69	229.3	225.0	283.7	307.9
70-74	315.0	312.4	378.8	403.0
75-79	373.3	376.1	433.9	456.2
80-84	376.4	384.1	408.6	427.4
85+	300.3	294.8	294.9	289.6

[1] Appendix E. United States Lung Cancer Mortality Rates. US Total Population (Table XV-7, SEER Cancer Statistics Review 1975-2005) Total Lung Cancer Mortality Rates per 100,000.

[2] Age-specific lung cancer (C34) mortality rates. Prepared by the Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry. Data Request # 08240 08/12/2008 Source: Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Mortality, 1990-2005, created 03-31-08, SEER Pop-Adj, SEER*Prep 2.4.

[3] Table XV-7, SEER Cancer Statistics Review 1975-2005 Surveillance, Epidemiology, and End Results database.

[4] Age-specific lung cancer (C340:C349) incidence rates. Prepared by the Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry. Data Request # 08240 08/12/2008 Source: Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence, 1995-2005, NPCR-CSS Sub 01-31-08, SEER Pop-Adj, SEER*Prep 2.4.0

2004 US All Life Tables [1]		2005 Total Texas Population Life Tables [2]	
Age	**Survival**	**Age**	**Survival**
0	1	0	1
1	0.9932	1	0.99348
5	0.99202	5	0.99227
10	0.99129	10	0.99149
15	0.99036	15	0.99052
20	0.98709	20	0.98739
25	0.98246	25	0.9828
30	0.97776	30	0.97823
35	0.9725	35	0.97305
40	0.96517	40	0.9661
45	0.95406	45	0.95449
50	0.93735	50	0.93756
55	0.91357	55	0.91315
60	0.88038	60	0.87949
65	0.83114	65	0.82873
70	0.76191	70	0.75979
75	0.66605	75+	0.66292
80	0.53925		
85	0.38329		

[1] Arias, E., United States Life Tables, 2004. National Vital Statistics Reports. 2007. 56(9): 3, Table B. Available from http://www.cdc.gov/nchs/data/nvsr/nvsr56/nvsr56_09.pdf

[2] Table 24, Appendix C. Texas Life Table, last update: 8/12/08

Appendix B. Linear Multiplicative Relative Risk Model (Crump and Allen 1985)

Robert L. Sielken Jr., Ph.D.
Ciriaco Valdez-Flores, Ph.D., P.E.
Sielken & Associates Consulting Inc
3833 Texas Avenue, Suite, 230, Bryan, TX 77802
Tel: 979-846-5175; Fax: 979-846-2671; Email: SielkenAssoc@aol.com

TCEQ, Austin, TX
December 17, 2007

B.1 Adjustments for Possible Differences Between the Population Background Cancer Rate and the Cohort's Cancer Rate in the Relative Risk Model

A multiplicative relative risk model that uses reference population background cancer rates to fit the cohort's observed cancer rates should adjust for the possibility of discrepancies between the background cancer rates in the reference population and the background cancer rates in the cohort.

Crump and Allen (1985) discuss the relative risk model with a factor that accounts for the possibility of different background rates in an epidemiological cohort and its reference population. This factor may adjust for issues like the healthy worker effect, the difference between internally and externally derived background cancer rates, covariate effects not explicitly incorporated in the summary epidemiological data, etc. For example, the multiplicative relative risk model with no adjustment for differences in background rates can be extended from

$$E(O_j) = E_{oj} \times (1 + \beta \times d_j)$$

to

$$E(O_j) = \alpha \times E_{oj} \times (1 + \beta \times d_j)$$

where the α term adjusts for any possible difference between the population's background cancer rates and the cohort's observed cancer rates in unexposed workers.

In the equations above the variables are:

$E(O_j)$ = expected number of lung cancer deaths for exposure group j predicted by the model;

E_{oj} = expected number of background lung cancer deaths for exposure group j based on the reference population background cancer rates;

β = multiplicative factor by which background risk increases with cumulative exposure;

d_j = cumulative exposure for exposure group j;

α = multiplicative factor that accounts for differences in cancer mortality background rates between the study cohort and the reference population.

B.2 Estimating the Slope Parameter, β, in the Relative Risk Model Adjusting for Differences in Background Rates

Poisson regression is a standard modeling technique in epidemiological studies. Poisson regression relies on the assumption that the number of cancer deaths in a dose group follows a Poisson distribution with mean equal to the expected number of cancer deaths and uses the maximum likelihood estimation procedure for the estimation of the parameters α and β in the model.

The Poisson distribution that describes probabilistically the number of cancers observed in a group is given by:

$$P(x) = \lambda^x \times e^{-\lambda} / x!,$$

where P(x) is the probability of observing x cancers, x is the number of cancer deaths actually observed, x! = x (x-1) (x-2) … 1, and λ is the expected number of cancers in the group. Thus, for dose group j, $x_j=O_j$ and $\lambda_j= E(O_j) = \alpha \times E_{oj} \times (1 + \beta \times d_j)$. That is, for each group j of person-years with average dose d_j, the observed number of cancer deaths in the dose interval (O_j) follows a Poisson distribution with parameter $\lambda_j= E(O_j) = \alpha \times E_{oj} \times (1 + \beta \times d_j)$ and the likelihood of observing O_j cancer deaths is given by,

$$P(O_j) = \lambda_j^{Oj} \times e^{-\lambda j} / O_j!.$$

The likelihood (L) is given by the product of the likelihoods of observing the number of cancer deaths in each dose group. That is,

$$L = P(O_1) \times P(O_2) \times \ldots.$$

or, equivalently,

$$L = (\lambda_1^{O1} \times e^{-\lambda 1} / O_1!) \times (\lambda_2^{O2} \times e^{-\lambda 2} / O_2!) \times \ldots.$$

where O_j is the number of cancer cases observed for the person-years with cumulative exposures equal to d_i. Substituting the value of λ_j by $\alpha \times E_{oj} \times (1 + \beta \times d_j)$ in the equation above, the likelihood is expressed as follows:

$$L = \prod [\alpha \times E_{oj} \times (1 + \beta \times d_j)]^{Oj} \times \exp\{-[\alpha \times E_{oj} \times (1 + \beta \times d_j)]\} / O_j!$$

where the symbol $\prod$ indicates that it is the product over all dose groups j=1,2,… and exp{.} is the base of the natural logarithm (e) raised to the power in the braces.

The maximum likelihood estimates of α and β can then be obtained by selecting the values of α and β that maximize the value of L. Finding the values of α and β that maximize the value of the likelihood L cannot be determined using a close-form solution because there are two variables. However, any routine that can maximize non-linear functions of more than one variable can be used to calculate the maximum likelihood estimates of α and β.

The parameters α and β that maximize the likelihood function given above also maximize the logarithm of the likelihood because the logarithm is a monotone function. The logarithm of the likelihood function (LL) given above is,

$$LL = \sum \{ O_j \times \ln[\alpha \times E_{oj} \times (1 + \beta \times d_j)] - [\alpha \times E_{oj} \times (1 + \beta \times d_j)] - \ln(O_j!) \}$$

where the symbol $\sum$ indicates that it is the sum over all dose groups j=1,2,… and ln(x) is the natural logarithm of x. The LL function can also be written as,

$$LL = \sum \{ O_j \times \ln(\alpha) + O_j \times \ln(E_{oj}) + O_j \times \ln(1 + \beta \times d_j) - [\alpha \times E_{oj} \times (1 + \beta \times d_j)] - \ln(O_j!) \}.$$

Note that the terms $O_j \times \ln(E_{oj})$ and $\ln(O_j!)$ on the equation above do not depend on the values of α and β, and hence, the values of α and β that maximize the LL also maximize the following simplified LL function:

$$LL = \sum \{ O_j \times \ln(\alpha) + O_j \times \ln(1 + \beta \times d_j) - [\alpha \times E_{oj} \times (1 + \beta \times d_j)] \}.$$

Finally, the maximum likelihood estimates of α and β can also be obtained by solving for α and β in the following system of equations:

$$\frac{\partial LL}{\partial \alpha} = \sum \{ O_j/\alpha - E_{oj} \times (1 + \beta \times d_j) \} = 0$$

$$\frac{\partial LL}{\partial \beta} = \sum \{ (O_j \times d_j) / (1 + \beta \times d_j) - \alpha \times E_{oj} \times d_j \} = 0$$

where $\partial LL/\partial\alpha$ and $\partial LL/\partial\beta$ are the partial derivatives of the logarithm of the likelihood with respect to α and β, respectively.

B.3 Estimating the Asymptotic Variance for the Slope Parameter in the Relative Risk Model

The system of equations of the partial derivatives of the logarithm of the likelihood given in the previous section can be used to estimate the asymptotic variance of the maximum likelihood estimates of α and β. The variance-covariance matrix of the parameters α and β is approximated by

$$\text{Cov}(\alpha,\beta) = -\begin{pmatrix} \partial^2 LL/\partial\alpha^2 & \partial^2 LL/\partial\alpha\partial\beta \\ \partial^2 LL/\partial\alpha\partial\beta & \partial^2 LL/\partial\beta^2 \end{pmatrix}^{-1}$$

where $[.]^{-1}$ is the inverse of the matrix, $\partial^2 LL/\partial\alpha^2$ is the second partial derivative of the logarithm of the likelihood with respect to α, $\partial^2 LL/\partial\beta^2$ is the second partial derivative of the logarithm of the likelihood with respect to β, and $\partial^2 LL/\partial\alpha\partial\beta$ is the partial derivative of the logarithm of the likelihood with respect to α and β. The approximation of the covariance is then given by

$$\text{Cov}(\alpha,\beta) = -\begin{pmatrix} \partial^2 LL/\partial\beta^2 & -\partial^2 LL/\partial\alpha\partial\beta \\ -\partial^2 LL/\partial\alpha\partial\beta & \partial^2 LL/\partial\alpha^2 \end{pmatrix} / \text{Determinant}$$

where

$$\text{Determinant} = 1 / [\ \partial^2 LL/\partial\alpha^2 \times \partial^2 LL/\partial\beta^2 - (\partial^2 LL/\partial\alpha\partial\beta)^2\ ]$$

The second-order derivatives used for the estimation of the variance-covariance matrix are:

$$\frac{\partial^2 LL}{\partial\alpha^2} = \sum -O_j/\alpha^2$$

$$\frac{\partial^2 LL}{}$$

$$\frac{\partial^2 LL}{\partial \beta^2} = \sum -(O_j \times d_j^2) / (1 + \beta \times d_j)^2$$

$$\frac{\partial^2 LL}{\partial \alpha \partial \beta} = \sum -E_{oj} \times d_j$$

A better asymptotic variance calls for substituting the variance-covariance matrix of α and β by the expected value of the above matrix. That is, by replacing the observed number of cancer deaths in a dose group j (O_j) by its expected value (i.e., $E(O_j) = \alpha \times E_{oj} \times (1 + \beta \times d_j)$). After substituting O_i by $\alpha \times E_{oj} \times (1 + \beta \times d_j)$ in the second-order derivatives and the variance-covariance matrix given above and some simplification, the better approximation of $Cov(\alpha,\beta)$ is given by:

$$Cov(\alpha,\beta) = \begin{pmatrix} \sum E_{oj} \times (1 + \beta \times d_j)/\alpha & \sum E_{oj} \times d_j \\ \sum E_{oj} \times d_j & \alpha \times \sum (E_{oj} \times d_j^2) / (1 + \beta \times d_j) \end{pmatrix}^{-1}$$

The determinant for the matrix is

$$\text{Determinant} = [\sum E_{oj} \times (1 + \beta \times d_j)] \times [\sum (E_{oj} \times d_j^2) / (1 + \beta \times d_j)] - (\sum E_{oj} \times d_j)^2$$

and the variance of the maximum likelihood estimate of α is

$$var(\alpha) = [\alpha \times \sum (E_{oj} \times d_j^2) / (1 + \beta \times d_j)] / \text{Determinant},$$

while the variance of the maximum likelihood estimate of β is

$$var(\beta) = [\sum E_{oj} \times (1 + \beta \times d_j)/\alpha] / \text{Determinant},$$

and the standard errors (SE) of the estimated parameters are the square root of their respective variances.

B.4 References

Crump, KS and BC Allen, 1985. Methods of Quantitative Risk Assessment Using Occupational Studies. *The Am Stat* **39**: 442-450.

Appendix C. Analyses of the Tacoma Smelter (Enterline et al. 1995)

Robert L. Sielken Jr., Ph.D., and Ciriaco Valdez Flores, Ph.D., P.E.
Sielken & Associates Consulting Inc.
3833 Texas Avenue, Suite 230, Bryan, TX 77802
Tel: 979-846-5175; Fax: 979-846-2671;
Email: SielkenAssoc@aol.com

June 15, 2009

C.1 Adjusting for the Difference between the Reference Population and Cohort Background Rates in the Multiplicative Relative Risk Model

Viren and Silvers (1999) found that the model that fit the Enterline et al. (1995) data best (i.e., the lowest AIC) is the following multiplicative relative risk linear model with intercept ($\beta_1\beta_2$):

$$\lambda_t = E_t \times (b_1 + b_2 \times d_j)$$

The standard parameterization of the multiplicative relative risk linear model with intercept, used more often and readily usable for excess risk estimation, is (Crump and Allen 1985):

$$E(O_j) = \alpha \times E_{oj} \times (1 + \beta \times d_j)$$

where $\lambda_t = E(O_j)$ and $E_t = E_{oj}$. Thus, the α in the standard multiplicative relative risk linear model with intercept is equal to b_1 in the Viren and Silvers' linear – with intercept ($\beta_1\beta_2$) model. Similarly, the β in the standard multiplicative relative risk linear model with intercept is equal to b_2/b_1 in the Viren and Silvers' linear – with intercept ($\beta_1\beta_2$) model. By replacing $\alpha \times E_{oj}$ by a target population's background risks, the standard multiplicative relative risk linear model can be used to estimate excess risks for a target population with background risks different than those of the cohort.

Appendix B describes the methodology to determine the maximum likelihood estimates and corresponding variances of the parameters in the standard multiplicative relative risk model with intercept.

C.2 Adjusting for Year of First Hire in the Multiplicative Relative Risk Model

Viren and Silvers (1999) used Enterline et al. (1995) epidemiological data to fit a multiplicative relative risk model discussed in Section C.1. The Enterline et al. data included information on the

first year of hire (< 1940 or ≥ 1940). The multiplicative relative linear risk model with intercept used by Viren and Silvers can be extended to adjust for the first year of hire. The model can be adjusted for the first year of hire using a nonparametric covariate effect. The advantage of using a nonparametric effect adjustment as opposed to a functional effect adjustment is that the nonparametric adjustment does not restrict the effect to have any specified functional form.

The multiplicative relative risk linear model with **no** adjustment for year of first hire

$$E(O_j) = \alpha \times E_{oj} \times (1 + \beta \times d_j)$$

can be extended to **adjust** for year of first hire as follows:

$$E(O_{kj}) = \mathbf{h} \times \alpha \times E_{kj} \times (1 + \beta \times d_{kj})$$

where all the parameters are as described in Appendix B; namely,

$E(O_{kj})$ = expected number of lung cancer deaths for exposure group j predicted by the model in the group of workers first hired before 1940 (k=1) or first hired in or after 1940 (k=2);

E_{kj} = expected number of background lung cancer deaths for exposure group j based on the reference population background cancer rates in the group of workers first hired before 1940 (k=1) or first hired in or after 1940 (k=2);

β = multiplicative factor by which the background risk increases with cumulative exposure;

d_{kj} = cumulative exposure for exposure group j rates in the group of workers first hired before 1940 (k=1) or first hired in or after 1940 (k=2);

α = multiplicative factor that accounts for differences in cancer mortality background rates between the study cohort and the reference population

h = multiplicative factor for the effect of year of hire

The effect of year of hire, h, is fixed to 1 for workers first hired before 1940 and is estimated to be a number greater than zero for workers first hired in or after 1940 – an estimate of h>1 implies that workers first hired in or after 1940 have a background rate of lung cancer greater than workers first hired before 1940 while an estimate of h<1 implies the opposite.

C.3 Estimating the Slope Parameter, β, in the Relative Risk Model Adjusting for Differences in Background Rates and Year of First Hire

Poisson regression is a standard modeling technique in epidemiological studies. Poisson regression relies on the assumption that the number of cancer deaths in a dose group follows a Poisson distribution with mean equal to the expected number of cancer deaths and uses the maximum likelihood estimation procedure for the estimation of the parameters α and β in the model.

The Poisson distribution that describes probabilistically the number of cancers observed in a group is given by:

$$P(x) = \lambda^x \times e^{-\lambda} / x!,$$

where P(x) is the probability of observing x cancers, x is the number of cancer deaths actually observed, x! = x (x-1) (x-2) … 1, and λ is the expected number of cancers in the group. Thus, for dose group j and the k-th group of workers first hired before 1940 or after 1939, $x_{kj}=O_{kj}$ and $\lambda_{kj}= E(O_{kj}) = h \times \alpha \times E_{kj} \times (1 + \beta \times d_{kj})$. That is, for each group j of person-years in the k-th group of workers with average dose d_{kj}, the observed number of cancer deaths in the dose interval (O_{kj}) follows a Poisson distribution with parameter $\lambda_{kj}= E(O_{kj}) = h \times \alpha \times E_{kj} \times (1 + \beta \times d_{kj})$ and the likelihood of this is given by,

$$P(O_{kj}) = \lambda_{kj}^{Okj} \times e^{-\lambda kj} / O_{kj}!.$$

The likelihood (L) is given by the product of the likelihoods of observing the number of cancer deaths in each dose group. That is,

$$L = P(O_{11}) \times P(O_{12}) \times \ldots P(O_{21}) \times P(O_{22}) \times \ldots.$$

or, equivalently,

$$L = (\lambda_{11}^{O11} \times e^{-\lambda 11} / O_{11}!) \times (\lambda_{12}^{O12} \times e^{-\lambda 12} / O_{12}!) \times \ldots.(\lambda_{21}^{O21} \times e^{-\lambda 21} / O_{21}!) \times (\lambda_{22}^{O22} \times e^{-\lambda 22} / O_{22}!) \times \ldots$$

where O_{kj} is the number of cancer cases observed for the person-years with cumulative exposures equal to d_{ki} for workers of the k-th group of year of first hire. Substituting the value of λ_{kj} by $h \times \alpha \times E_{kj} \times (1 + \beta \times d_{kj})$ in the equation above, the likelihood is expressed as follows:

$$L = \prod [h \times \alpha \times E_{kj} \times (1 + \beta \times d_{kj})]^{Okj} \times \exp\{-[h \times \alpha \times E_{kj} \times (1 + \beta \times d_{kj})]\} / O_{kj}!$$

where the symbol $\prod$ indicates that it is the product over all combinations of groups of first hire (k=1,2) and dose groups j=1,2,…, and exp{.} is the base of the natural logarithm (e) raised to the power in the braces.

The maximum likelihood estimates of h, α and β can then be obtained by selecting the values of h, α and β that maximize the value of L. Finding the values of h, α and β that maximize the value of the likelihood L cannot be determined using a close-form solution because there are three variables. However, any routine that can maximize non-linear functions of more than one variable can be used to calculate the maximum likelihood estimates of h, α and β.

The values of h, α and β that maximize the likelihood function given above also maximize the logarithm of the likelihood because the logarithm is a monotone function. The logarithm of the likelihood (LL) of the function given above is,

$$LL = \sum \{ O_{kj} \times \ln[h \times \alpha \times E_{kj} \times (1 + \beta \times d_{kj})] - [h \times \alpha \times E_{kj} \times (1 + \beta \times d_{kj})] - \ln(O_{kj}!) \}$$

where the symbol $\sum$ indicates that it is the sum over all combinations of groups of first hire (k=1,2) and all dose groups j=1,2,…, and ln(x) is the natural logarithm of x. The LL function can also be written as,

$$LL = \sum \{ O_{kj} \times \ln(h) + O_{kj} \times \ln(\alpha) + O_{kj} \times \ln(E_{kj}) + O_{kj} \times \ln(1 + \beta \times d_{kj}) - [h \times \alpha \times E_{kj} \times (1 + \beta \times d_{kj})] - \ln(O_{kj}!) \}.$$

Note that the terms $O_{kj} \times \ln(E_{kj})$ and $\ln(O_{kj}!)$ in the equation above do not depend on the values of h, α or β, and hence, the values of h, α and β that maximize the LL also maximize the following simplified LL function:

$$LL = \sum \{ O_{kj} \times \ln(h) + O_{kj} \times \ln(\alpha) + O_{kj} \times \ln(1 + \beta \times d_{kj}) - [h \times \alpha \times E_{kj} \times (1 + \beta \times d_{kj})] \}.$$

Finally, the maximum likelihood estimates of h, α and β can also be estimated by solving for h, α and β in the following system of equations:

$$\frac{\partial LL}{\partial \alpha} = \sum \{ O_{kj}/\alpha - h \times E_{kj} \times (1 + \beta \times d_{kj}) \} = 0$$

$$\frac{\partial LL}{\partial \beta} = \sum \{ (O_{kj} \times d_{kj}) / (1 + \beta \times d_j) - h \times \alpha \times E_{kj} \times d_{kj} \} = 0$$

$$\frac{\partial LL}{\partial \beta} = \sum \{ O_{2j}/h - \alpha \times E_{2j} \times (1 + \beta \times d_{2j}) \} = 0$$

∂ h

where $\partial LL/\partial h$, $\partial LL/\partial \alpha$ and $\partial LL/\partial \beta$ are the partial derivatives of the logarithm of the likelihood with respect to h, α and β, respectively. Note that the parameter h, for the year of hire, is being estimated for groups of person-years of workers first hired in or after 1940 and is a fixed value of 1 for workers first hired before 1940. Thus, the summation for $\partial LL/\partial h$ is only over workers first hired in or after 1940.

C.4 Estimating the Asymptotic Variance for the Slope Parameter in the Relative Risk Model

The system of equations of the partial derivatives of the logarithm of the likelihood given in the previous section can be used to estimate the asymptotic variance of the maximum likelihood estimates of h, α and β. The variance-covariance matrix of the parameters h, α and β is approximated by

$$\mathrm{Cov}(h,\alpha,\beta) = -\begin{pmatrix} \partial^2 LL/\partial\alpha^2 & \partial^2 LL/\partial\alpha\partial\beta & \partial^2 LL/\partial\alpha\partial h \\ \partial^2 LL/\partial\alpha\partial\beta & \partial^2 LL/\partial\beta^2 & \partial^2 LL/\partial\beta\partial h \\ \partial^2 LL/\partial\alpha\partial h & \partial^2 LL/\partial\beta\partial h & \partial^2 LL/\partial h^2 \end{pmatrix}^{-1}$$

where $[.]^{-1}$ is the inverse of the matrix, $\partial^2 LL/\partial h^2$ is the second partial derivative of the logarithm of the likelihood with respect to h,$\partial^2 LL/\partial \alpha^2$ is the second partial derivative of the logarithm of the likelihood with respect to α, $\partial^2 LL/\partial \beta^2$ is the second partial derivative of the logarithm of the likelihood with respect to β, $\partial^2 LL/\partial h\partial \alpha$ is the partial derivative of the logarithm of the likelihood with respect to h and α, $\partial^2 LL/\partial h\partial \beta$ is the partial derivative of the logarithm of the likelihood with respect to h and β, and $\partial^2 LL/\partial \alpha\partial \beta$ is the partial derivative of the logarithm of the likelihood with respect to α and β.

The second-order derivatives used for the estimation of the variance-covariance matrix are:

$$\frac{\partial^2 LL}{\partial \alpha^2} = \sum -O_{kj}/\alpha^2$$

$$\frac{\partial^2 LL}{\partial \beta^2} = \sum -(O_{kj} \times d_{kj}^2) / (1 + \beta \times d_{kj})^2$$

$$\frac{\partial^2 LL}{\partial h^2} = \sum -O_{2j}/h^2$$

$$\frac{\partial^2 LL}{\partial \alpha \partial \beta} = \sum -h \times E_{kj} \times d_{kj}$$

$$\frac{\partial^2 LL}{\partial \alpha \partial h} = \sum -E_{2j} \times (1 + \beta \times d_{2j})$$

$$\frac{\partial^2 LL}{\partial \beta \partial h} = \sum -\alpha \times E_{2j} \times d_{2j}$$

Note that the parameter h, for the year of hire, is being estimated for workers first hired in or after 1940 and is a fixed value of 1 for workers first hired before 1940. Thus, the summations for $\partial^2 LL/\partial h^2$, $\partial^2 LL/\partial \alpha \partial h$, and $\partial^2 LL/\partial \beta \partial h$ are only over groups of person-years of workers first hired in or after 1940.

A better asymptotic variance calls for substituting the variance-covariance matrix of h, α and β by the expected value of the above matrix. That is, by replacing the observed number of cancer deaths in a dose group j (O_j) by its expected value (i.e., $E(O_j) = h \times \alpha \times E_{oj} \times (1 + \beta \times d_j)$). After

substituting O_i by $h \times \alpha \times E_{oj} \times (1 + \beta \times d_j)$ in the second-order derivatives and the variance-covariance matrix given above and some simplification, the better approximation of Cov(h,α,β) is given by:

$$\mathrm{Cov}(h,\alpha,\beta) = \begin{pmatrix} \sum h\times E_{kj}\times(1+\beta\times d_{kj})/\alpha & \sum h\times E_{kj}\times d_{kj} & \sum E_{2j}\times(1+\beta\times d_{2j}) \\ \sum h\times E_{kj}\times d_{kj} & h\times\alpha\times\sum (E_{kj}\times d_{kj}^2)/(1+\beta\times d_{kj}) & \sum \alpha\times E_{2j}\times d_{2j} \\ \sum E_{2j}\times(1+\beta\times d_{2j}) & \sum \alpha\times E_{2j}\times d_{2j} & \sum \alpha\times E_{2j}\times(1+\beta\times d_{2j})/h \end{pmatrix}^{-1}$$

The element on the first row and first column of the Cov(h,α,β) matrix is the variance for the estimate of the intercept (α). The element on the second row and second column of the Cov(h,α,β) matrix is the variance for the estimate of the slope (β).The element on the third row and third column of the Cov(h,α,β) matrix is the variance for the estimate of the year of hire effect (h). The standard errors (SE) of the estimated parameters h, α and β are the square root of their respective variances.

Although there is no simple close-form inverse for a three by three matrix, the matrix can be easily inverted in most spreadsheet programs like Excel.

C.5 Beta (β), SE, and 95% (LCL and UCL) β Values (Enterline et al. 1995)

Beta (β), Standard Error (SE), and 95% Lower Confidence Limit (LCL) and Upper Confidence Limit (UCL) β Values (Enterline et al. 1995) [a]				
	$O = \alpha \times E \times (1 + \beta \times d)$			
	Intercept (α)	**β (MLE) ± SE**	**β (95% LCL) [c]**	**β (95% UCL) [d]**
All workers adjusting for year of hire (h = 1.38[b])	1.46	3.15E-05 ± 1.48E-05	7.17E-06	5.59E-05
All workers with no adjustment	1.81[e]	2.13E-05 [e] ± 1.13E-05	2.64E-06	3.99E-05
Workers hired < 1940	1.43[f]	3.44E-05 [f] ± 1.89E-05	3.29E-06	6.56E-05
Workers hired 1940+	2.05[g]	2.67E-05 [g] ± 2.33E-05	-1.17E-05	6.51E-05

[a] Units are in ERR per $\mu g/m^3$-yr.

[b] the background lung cancer mortality rate for workers hired 1940+ is 1.38-fold higher than the background lung cancer mortality rate for workers first hired < 1940

[c] 95% LCL = β - (1.645 x SE) for a standard normal distribution.

[d] 95% UCL = β + (1.645 x SE) for a standard normal distribution.

[e] intercept = 1.68 and potency/intercept = 2.14E-05 (Table 3 in Viren and Silvers 1999)

[f] intercept = 1.43 and potency/intercept = 3.44E-05 (Table 5 in Viren and Silvers 1999)

[g] intercept = 2.05 and potency/intercept = 2.68E-05 (no association, regression didn't achieve statistical significance at P < 0.01 based on the corresponding likelihood ratio statistic (Table 5 in Viren and Silvers 1999))

C.6 References

Crump, KS and BC Allen, 1985. Methods of Quantitative Risk Assessment Using Occupational Studies. *The Am Stat* **39**: 442-450.

Enterline, PE, R Day, and GM Marsh, 1995. Cancers Related to Exposure to Arsenic at a Copper Smelter. *Occupational and Environmental Medicine* **52**:28-32.

Viren, J and A Silvers, (1999). Nonlinearity in the Lung Cancer Dose-Response for Airborne Arsenic: Apparent Confounding by Year of Hire in Evaluating Lung Cancer Risks from Arsenic Exposure in Tacoma Smelter Workers. *Regulatory Toxicology and Pharmacology* **30**: 117-129.

Appendix D. URFs and 10^{-5}-Risk Air Concentrations using United States Lung Cancer Mortality Rates and Survival Probabilities

Table D-1. URFs and 10^{-5}-Risk Air Concentrations (Enterline et al. 1995)[a]

	Background Rates	β (MLE) URF 10^{-5}-Risk Air Concentration	β (95% LCL) URF 10^{-5}-Risk Air Concentration	β (95% UCL) URF 10^{-5}-Risk Air Concentration
All workers adjusting for year of hire	US	1.25E-04/ µg/m^3 0.0799 µg/m^3	2.85E-05/ µg/m^3 0.351 µg/m^3	2.22E-04/ µg/m^3 0.0450 µg/m^3
All workers with no adjustment	US	8.47E-05/ µg/m^3 0.118 µg/m^3	1.05E-05/ µg/m^3 0.953 µg/m^3	1.59E-04/ µg/m^3 0.0630 µg/m^3
Workers hired < 1940	US	1.37E-04/ µg/m^3 0.0731 µg/m^3	1.31E-05/ µg/m^3 0.765 µg/m^3	2.61E-04/ µg/m^3 0.0383 µg/m^3

[a]URFs based on the parameter estimates given in Table C-5

Table D-2. URFs and 10^{-5}-Risk Air Concentration (Lubin et al. 2000; 2008)[a]

	Background Rates	β (MLE) URF 10^{-5}-Risk Air Concentration	β (95% LCL) URF 10^{-5}-Risk Air Concentration	β (95% UCL) URF 10^{-5}-Risk Air Concentration
Lubin et al. (2000) Restricted sub-cohort	US	8.07E-04/ µg/m^3 0.0124 µg/m^3	1.05E-04/ µg/m^3 0.0953 µg/m^3	1.51E-03/ µg/m^3 0.00664 µg/m^3
Lubin et al. (2008) Full cohort	US	2.28E-04/ µg/m^3 0.0437 µg/m^3	1.23E-04/ µg/m^3 0.0811 µg/m^3	3.34E-04/ µg/m^3 0.0299 µg/m^3

[a]URFs based on the parameter estimates given in Table E-2

Table D-3. URFs and 10^{-5}-Risk Air Concentration (Järup et al. 1989)[a]

	Background Rates	β (MLE) URF 10^{-5}-Risk Air Concentration	β (95% LCL) URF 10^{-5}-Risk Air Concentration	β (95% UCL) URF 10^{-5}-Risk Air Concentration
All workers adjusting for year of hire (h=1.19)	US	1.16E-04 / μg/m^3 0.0861 μg/m^3	9.18E-06 / μg/m^3 1.09 μg/m^3	2.23E-04/ μg/m^3 0.0448 μg/m^3
Total Cohort	US	9.46E-05/ μg/m^3 0.106 μg/m^3	3.49E-05/ μg/m^3 0.286 μg/m^3	1.55E-04/ μg/m^3 0.0647 μg/m^3
First hired < 1940	US	1.04E-04/ μg/m^3 0.0960 μg/m^3	1.59E-05/ μg/m^3 0.629 μg/m^3	1.92E-04/ μg/m^3 0.0520 μg/m^3
First hired 1940+	US	2.45E-04/ μg/m^3 0.0408 μg/m^3	NA	6.32E-04/ μg/m^3 0.0158 μg/m^3

[a]URFs based on the parameter estimates given in Table F-1
NA, not available as the 95%LCL β value was negative, suggesting zero risk, calculation of an air concentration at 1 in 100,000 excess risk was not possible.

Table D-4. URFs and 10^{-5}-Risk Air Concentration Estimates Based on Weighted Cumulative Exposure (Jones et al. 2007)[a]

Extrapolation assumption for exposures prior to 1972	Background Rates	β (MLE) URF 10^{-5} Risk Air Concentration	β (95% LCL) URF 10^{-5} Risk Air Concentration	β (95% UCL) URF 10^{-5} Risk Air Concentration
Scenario A	US	1.27E-03 / μg/m^3 0.00790 μg/m^3	NA	2.60E-03 / μg/m^3 0.00384 μg/m^3
Scenario B	US	7.46E-04 / μg/m^3 0.0134 μg/m^3	NA	1.67E-03 / μg/m^3 0.00599 μg/m^3
Scenario C	US	8.62E-04 / μg/m^3 0.0116 μg/m^3	NA	1.78E-03 / μg/m^3 0.00561 μg/m^3

[a]URFs based on the parameter estimates given in Table G-1
NA, not available as the 95%LCL β value was negative, suggesting zero risk, calculation of an 10^{-5} risk air concentration was not possible.

Table D-5. Preferred URFs and 10^{-5}-Risk Air Concentrations from All Studies Based on U.S. Rates

Study And Person-years (PY)	Back-ground Rates	β (MLE) URF 10^{-5}-Risk Air Concentration	β (95% LCL) URF 10^{-5}-Risk Air Concentration	β (95% UCL) URF 10^{-5}-Risk Air Concentration
Enterline et al. (1995) All workers adjusting for year of hire 84,916 PY	US	1.25E-04/ μg/m^3 0.0799 μg/m^3	2.85E-05/ μg/m^3 0.351 μg/m^3	2.22E-04/ μg/m^3 0.0450 μg/m^3
Lubin et al. (2008) Full cohort 256,850 PY	US	2.28E-04/ μg/m^3 0.0437 μg/m^3	1.23E-04/ μg/m^3 0.0811 μg/m^3	3.34E-04/ μg/m^3 0.0299 μg/m^3
Järup et al. (1989) All workers adjusting for year of hire 127,189 PY	US	1.16E-04 / μg/m^3 0.0861 μg/m^3	9.18E-06 / μg/m^3 1.09 μg/m^3	2.23E-04/ μg/m^3 0.0448 μg/m^3
Jones et al. (2007) Scenario B 35,942 PY	US	7.46E-04 / μg/m^3 0.0134 μg/m^3	NA	1.67E-03 / μg/m^3 0.00599 μg/m^3

NA, not available as the 95% LCL β value was negative, suggesting zero risk, calculation of an air concentration at 1 in 100,000 excess risk was not possible.

A weighted URF based on US lung cancer mortality rates and survival probabilities using PY was calculated:

Final URF (risk per μg/m^3)

$$= \frac{[(URF_1 \times weight_1) + (URF_2 \times weight_2) + (URF_3 \times weight_3) + (URF_4 \times weight_4)]}{[weight_1 + weight_2 + weight_3 + weight_4)]}$$

$$= \frac{[(1.25E\text{-}04 \times 84{,}916) + (2.28E\text{-}04 \times 256{,}850) + (1.16E\text{-}04 \times 127{,}189) + (7.46E\text{-}04 \times 35{,}942)]}{[84{,}916 + 256{,}850 + 127{,}189 + 35{,}942)]}$$

= 2.19E-04 per μg/m^3

The URF based on US lung cancer mortality rates and survival probabilites is 2.19E-04 per $\mu g/m^3$. The URF is 2.2E-04 per $\mu g/m^3$ and the resulting air concentration at a 1 in 100,000 excess lung cancer risk is 0.045 $\mu g/m^3$ (rounded to two significant figures).

Appendix E. Analyses of the Anaconda Smelter in Montana (Lubin et al. 2000; 2008)

Robert L. Sielken Jr., Ph.D., and Ciriaco Valdez Flores, Ph.D., P.E.
Sielken & Associates Consulting Inc.
3833 Texas Avenue, Suite 230, Bryan, TX 77802
Tel: 979-846-5175; Fax: 979-846-2671;
Email: SielkenAssoc@aol.com

June 24, 2009

E.1 Concentration as an Effect-Modification Factor

The dose-response relationship used by Lubin et al. (2008) uses concentration as an effect-modification factor rather than as a covariate. A covariate effect is generally used to account for differences in background hazard rates of different groups of person-years. An effect-modification factor, on the other hand, is used to model how the excess hazard rate changes due to the effect-modification factor. The covariate effects are normally excluded in the estimation of excess risks and the background risks of a target population are used instead. The effect-modification factors, on the other hand, are kept in the estimation of excess risks because they describe how the risk changes with these factors. One can think of these effect-modifying factors as part of the dose metric. The usual dose metric in dose-response models for epidemiological data is cumulative exposure. Lubin et al. (2008), however, used a dose metric that is equal to the cumulative exposure multiplied by the average concentration over the exposure period raised to a power.

It would be incorrect to not include the effect-modification factor in the estimation of excess risks. Thus, as long as the effect-modification factor (concentration in the Lubin et al. 2008 models) is correctly accounted for in the estimation of excess risks, the average exposure concentration and the cumulative exposure are not confounded in the dose-response relationship. In other words, the dose metric used in the estimation of excess risks has to be the same as the dose metric used in the estimation of the model parameters.

As an example, parameter estimates of multiplicative relative risk models with cumulative exposure lagged x number of years as the dose metric are often published. Excess risks based on these models can be appropriately calculated only if the same dose metric is used (i.e., cumulative exposure lagged x number of years).

The exposure concentration in the Lubin et al. (2008) models is an effect-modification factor. This factor is part of the dose metric and cannot be excluded whenever excess risks are to be estimated. The effect-modification factor (exposure concentration) can be used to provide a measure of uncertainty by fixing the concentration at levels well above the average

environmental exposures – i.e., assuming that the dose metric is cumulative exposure and that the modification-factor affects the slope of the relative risk model. Assuming average concentration larger than the environmental concentrations in the estimation of excess risks results in an overestimation of the slope and, therefore, in health protective risk estimates. On the contrary, assuming average concentration less than the environmental concentrations in the estimation of excess risks results in an underestimation of the slope and, therefore, in less health protective risk estimates.

E.2 Estimates Based on the Lubin et al. (2008) Paper Compared to the Lubin et al. (2000) Paper

In the Lubin et al. 2000 paper, the multiplicative relative risk models were fit to a restricted data set that included only "current workers and former workers last exposed over 50 years." That is, more than 50% of the person-years of follow-up and more than 40% (194 of the 446) of the respiratory cancers were not included in the estimation of the relative risk model.

Lubin et al. (2008) analyzed both, the full cohort and a restricted subset of the cohort. The restricted sub-cohort included only "current workers, recent former (< 5 years) workers, and workers with last employment at ≥ 50 years of age." This restricted sub-cohort is slightly larger than the restricted sub-cohort used in the Lubin et al. 2000 paper (261 respiratory cancers versus 252 respiratory cancers in the 2000 paper). Still the 2008 restricted sub-cohort excludes approximately 44% of the person years and 185 or 41% of the respiratory cancer deaths.

In the 2008 paper, Lubin et al. considered only the cumulative doses that weighted with λ=0.1 the exposures in the jobs with high arsenic concentrations. Lubin et al. (2000, 2008) concluded that the weight of 0.1 on the exposures in jobs with high concentrations of arsenic is more appropriate because workers in those jobs used protective equipment. Furthermore, using the weight of 0.1 on high-exposure jobs resulted in: 1) rate ratios that conformed to a linear dose-response relationship with cumulative exposure to arsenic and 2) steeper estimates of the slopes, which imply more health-protective excess risks of respiratory cancer deaths.

E.3 Model of Full Cohort Using the Multiplicative Relative Risk Model and Cumulative Exposure

Table 2 in Lubin et al. (2008) lists the mean cumulative exposure to arsenic (mg/m^3-yr), the number of respiratory cancer deaths and the standardized mortality ratios (SMRs) for six cumulative exposure intervals for the full cohort. The SMRs for respiratory cancers adjusted for calendar period and country of birth are more appropriate than the unadjusted SMRs also listed in Table 2. The adjusted SMRs include the effects of possible fluctuations of background respiratory cancer mortality rates in different calendar years and different countries of birth. The relevant data extracted from Table 2 of the Lubin et al. (2008) paper are:

Table E-1. Observed, Expected and Standard Mortality Rates (SMRs) from Table 2 in Lubin et al. (2008)				
Cumulative exposure interval ($\mu g/m^3$-yr)	Mean Exposure ($\mu g/m^3$-yr)	Observed number of respiratory cancer deaths	Expected[1] number of respiratory cancer deaths	SMR
< 750	470	62	73.81	0.84
750-2,000	1,240	96	75.00	1.28
2,000-5,000	3,430	74	68.52	1.08
5,000-10,000	7,270	83	74.77	1.11
10,000-15,000	11,900	84	50.00	1.68
≥15,000	21,900	47	20.00	2.35

[1]Expected = Observed / SMR

Using the data in the table above, the multiplicative relative risk model proposed by Crump and Allen (1985) with a factor that accounts for the possibility of different background rates in an epidemiological cohort and its reference population can be used. That is, the same model used in the Tacoma study; namely,

$$E(O_j) = \alpha \times E_{oj} \times (1 + \beta \times d_j)$$

where the α term adjusts for any possible difference between the population's background cancer rates and the cohort's observed cancer rates in unexposed workers.

In the equation above the variables are:

$E(O_j)$ = expected number of respiratory cancer deaths for exposure group j predicted by the model;

E_{oj} = expected number of background respiratory cancer deaths for exposure group j based on the reference population background cancer rates;

β = multiplicative factor by which background risk increases with cumulative exposure;

d_j = cumulative exposure for exposure group j;

α = multiplicative factor that accounts for differences in cancer mortality background rates between the study cohort and the reference population.

The maximum likelihood parameter estimates of the multiplicative linear rate ratio model and the 95% LCL and 95% UCL on the slope are:

α = 9.42E-01

β = 5.75E-05 per $\mu g/m^3$-yr

SE = 1.61E-05

β(95% LCL) = 5.75E-05 - 1.645×1.61E-05} = 3.10E-05 per $\mu g/m^3$-yr

β(95% UCL) = 5.75E-05 + 1.645×1.61E-05} = 8.40E-05 per $\mu g/m^3$-yr

Lubin et al. (2000, 2008), however, focused in the results obtained from the restricted sub-cohort as opposed to the results based on the full cohort. The main reason for focusing in the restricted sub-cohort was to minimize the effects of unmeasured exposures "because there was no information on exposures after the workers left the smelter."

In generating the cumulative exposures for the full cohort, Lubin et al. (2008) assumed that workers were not exposed to arsenic after they left the smelter. This is a standard assumption made in epidemiological studies and, by assuming zero exposure when there might have been non-zero exposures, results in an underestimation of cumulative exposures. Underestimation of actual cumulative exposures results in overestimation of the slope in a multiplicative relative risk model and, consequently, in more health protective risk estimates. Thus, the slope for the multiplicative relative risk model based on the full cohort derived here is probably greater than the slope that would have been obtained if exposures for workers that had left the smelter were assumed to be greater than zero.

E.4 Models in Lubin et al. (2008)

The objective of the Lubin et al. (2008) paper was to evaluate the shape of the dose response relationship between respiratory cancer mortality and cumulative exposure to arsenic and the modification of this relationship by the average exposure concentration. There are two ways of interpreting Lubin et al. (2008) models:

1) Interpretation 1: Lubin et al. (2008) estimated the multiplicative relative risk linear model but instead of assuming a slope (β) that is a constant, they assumed that the slope is a function of the average arsenic concentration (c). The function of the average arsenic concentration for the slope of the linear relative risk model that Lubin et al. used is:

$$\beta(c) = \beta \times c^{\varphi}$$

where φ models the effect that the concentration has on the excess risk per unit of cumulative exposure and is estimated from the data. That is, the relative risk is given by the following

$$RR = 1 + \beta \times c^{\varphi} \times CumExp$$

where CumExp is the cumulative exposure to arsenic. Lubin et al. went beyond the adjustment of the slope by the functional form shown above, and also considered nonparametric modifications of the slope by age and time since last exposure as well as nonparametric effects of exposure concentrations on the slope.

2) Interpretation 2: Lubin et al. (2008) estimated the multiplicative relative risk linear model assuming a constant slope (β) but the dose metric was the product of the cumulative exposure and the average arsenic concentration (c) raised to a power. That is, the dose metric is given by the following relation

$$\text{Dose Metric} = \text{CumExp} \times c^{\varphi}$$

where CumExp is the cumulative exposure to arsenic and φ models the effect that the concentration has on the cumulative exposure and is estimated from the data. That is, the relative risk is given by the following

$$RR = 1 + \beta \times c^{\varphi} \times \text{CumExp}$$

Lubin et al. went beyond defining the dose metric by the functional form shown above and also considered nonparametric effects of age and time since last exposure modifying the cumulative exposure.

The second interpretation of the Lubin et al. (2008) model is how BEIR IV (BEIR. *Health Effects and of Exposure to Radon (BEIR VI)*. Washington, DC: National Academy Press, 1999) and Jones et al. (Jones, S.R., P. Atkin, C. Holroyd, E. Lutman, J. Vives i Batlle, R. Wakeford and P. Walker (2007). Lung Cancer Mortality at a UK Tin Smelter. *Occupational Medicine*, **57**:238-245) applied these models for exposures to radon and arsenic, respectively.

E.4.1 Slope estimates for person-years with exposures to different average concentrations

Before estimating the parameters of the multiplicative relative risk model with the slope being a function of the average arsenic concentration, age and times since last exposure, (or with a dose metric that is a function of cumulative exposure, average arsenic concentration, age and times since last exposure) Lubin et al. fit the standard multiplicative relative risk model with cumulative exposure as the dose metric to four subsets of the full cohort of workers. The four subsets and the corresponding estimates of the slopes are (see Table 2 and Figure 1 in Lubin et al. (2008)):

1) person-years with exposures to mean arsenic concentration equal to 290 $\mu g/m^3$ (i.e., low-exposure jobs)

$$\beta = 1.6\text{E-}05 \text{ per } \mu g/m^3\text{-yr}$$

95% CI = (-5.0E-06 to 4.1E-05)

Standard Error (back calculated) = 1.17E-05

β(95% LCL) = -3.23E-06 per μg/m^3-yr

β(95% UCL) = 3.52E-05 per μg/m^3-yr

2) person-years with exposures to mean arsenic concentration of 300-400 μg/m^3

β = 6.7E-05 per μg/m^3-yr

95% CI = (2.4E-05 to 1.19E-04)

Standard Error (back calculated) = 2.41E-05

β(95% LCL) = 2.73E-05 per μg/m^3-yr

β(95% UCL) = 1.07E-04 per μg/m^3-yr

3) person-years with exposures to mean arsenic concentration of 400-500 μg/m^3

β = 7.7E-05 per μg/m^3-yr

95% CI = (1.7E-05 to 1.59E-04)

Standard Error (back calculated) = 3.58E-05

β(95% LCL) = 1.81E-05 per μg/m^3-yr

β(95% UCL) = 1.36E-04 per μg/m^3-yr

4) person-years with exposures to mean arsenic concentration ≥ 500 μg/m^3

β = 7.2E-05 per μg/m^3-yr

95% CI = (4.3E-05 to 1.07E-04)

Standard Error (back calculated) =1.63E-05

β(95% LCL) = 4.53E-05 per μg/m^3-yr

β(95% UCL) = 9.87E-05 per μg/m^3-yr

Lubin et al. observed that the first group, with the lowest average concentration of 290 μg/m^3, had the smallest slope β and that the slope increased with increasing concentration (except for the fourth group which had a slope slightly smaller than the third group). Figure 1 in Lubin et al. (2008) is included here for convenience.

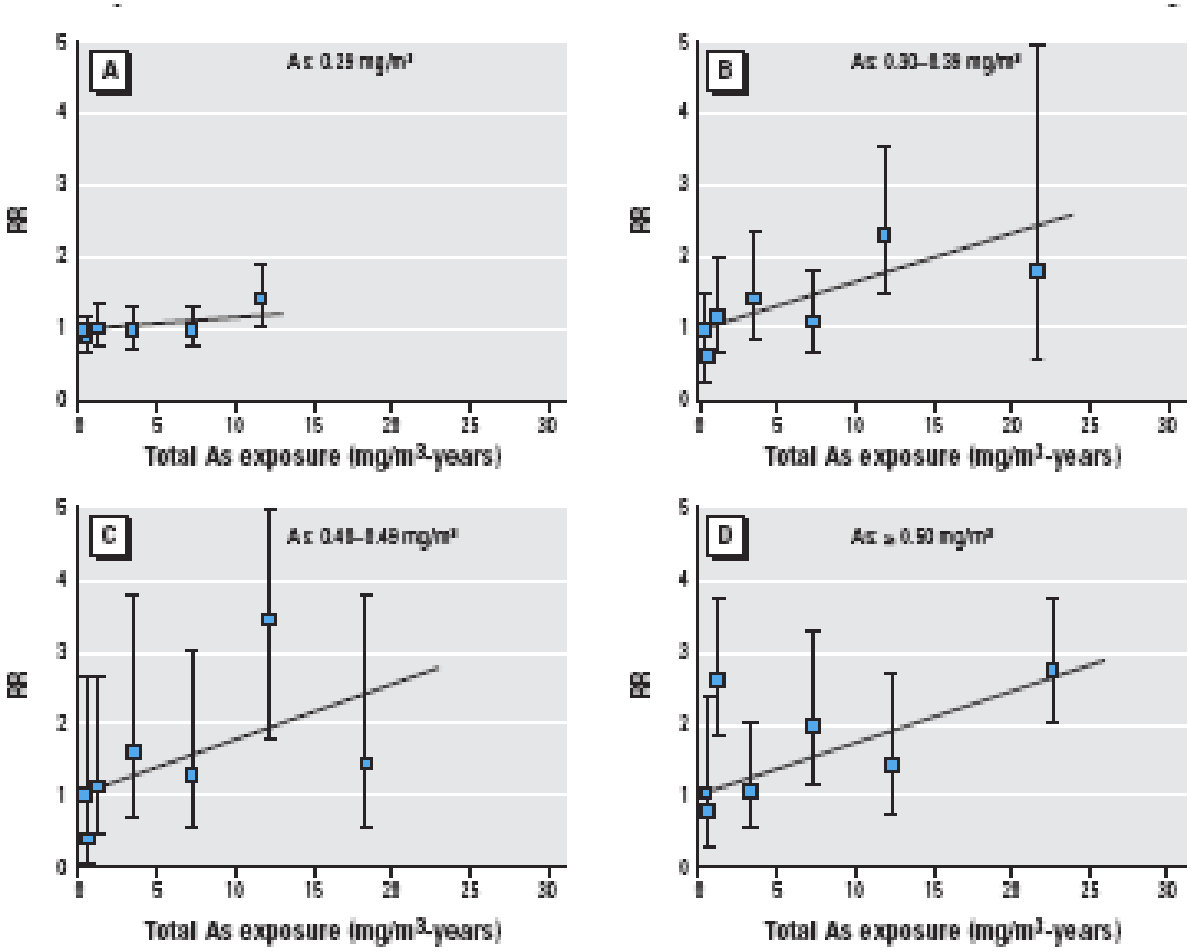


Figure 1. RRs of respiratory cancer mortality by categories of cumulative arsenic exposure (mg/m^3-years) and arsenic concentration (mg/m^3) relative to U.S. mortality rates for white males, adjusted to nonexposed workers, and fitted linear ERR models for cumulative arsenic exposure: 0.29 mg/m^3 (*A*), 0.30–0.39 mg/m^3 (*B*), 0.40–0.49 mg/m^3 (*C*), and ≥ 0.50 mg/m^3 (*D*). Estimates of the ERR per mg/m^3-year and 95% CIs for the four concentration categories were as follows: A, 0.016 (–0.005 to 0.041); B, 0.067 (0.024 to 0.119); C, 0.077 (0.017 to 0.159); D, 0.072 (0.043 to 0.107).

E.4.2 Slope estimates for the full cohort using the standard multiplicative relative risk model

In Figure 2 of Lubin et al. (2008) the dotted line is the slope of the standard multiplicative relative risk model for the cohort that includes all the workers in the study. The slope (β) is equal to

4.756E-05 per μg/m^3-yr.

Lubin et al. report neither a confidence interval nor a standard error for the estimate of the slope. An annotated version of Figure 2 in Lubin et al. (2008) is included here for convenience.

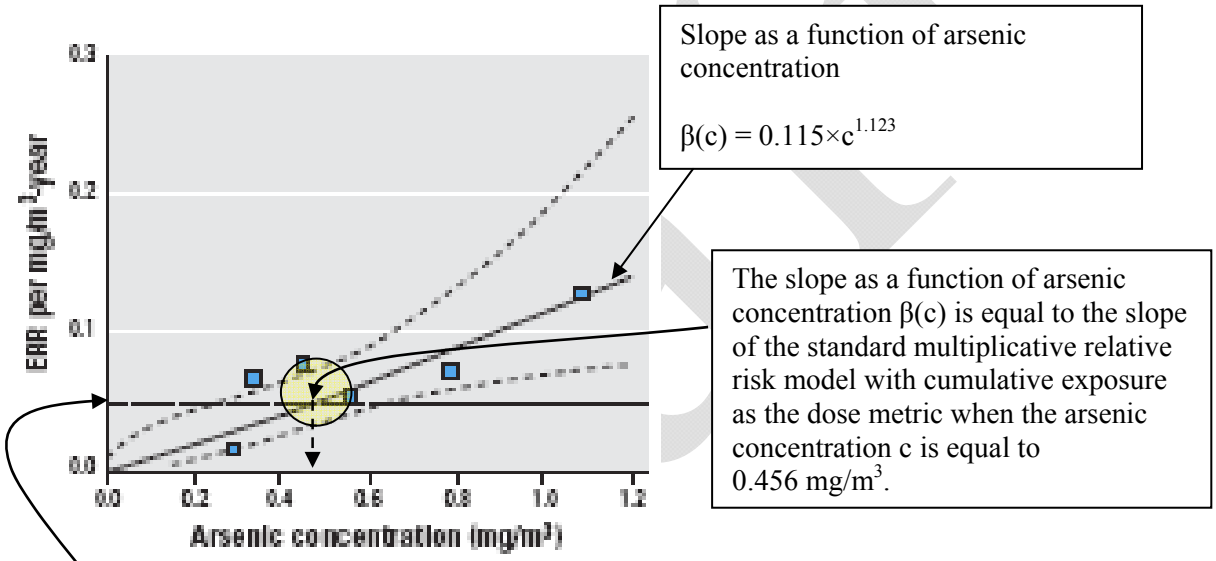


Figure 2. Estimates of ERR per mg/m^3-year based on a linear RR model within six categories of arsenic concentration (square symbols), fitted model 2 (solid line), its pointwise, two-sided, Wald 95% CI (dashed lines), and model 2 omitting variation with concentration (dotted line; ERR/mg/m^3/year = 0.04756).

Note that the slope estimated by Lubin et al. (2008) for the full cohort and using the standard multiplicative relative risk model with cumulative exposure as the dose metric (4.756E-05 per μg/m^3-yr) is different than the slope estimated from the data in their Table 2 (5.75E-05 per μg/m^3-yr). This difference is because the slope estimated using the data in Table 2 is adjusted using external background hazard rates (i.e., SMRs) whereas Lubin et al. (2008) adjusted the slope using cohort-specific background rates that can be obtained only when the data are available.

E.4.3 Slope estimates as a parametric function of average exposure concentration

Table 3 of Lubin et al. (2008) lists the slopes of the relative risk model as a function of the exposure concentration. The slope functions are shown for both, the full cohort and the restricted sub-cohort. The results are as follows:

1) full cohort

$\beta(c) = 0.115 \times c^{1.123}$ per mg/m^3-yr

MLE and 95% CI: 0.115 (0.07-0.19) and 1.123 (0.41-1.84)

2) restricted sub-cohort

$\beta(c) = 0.083 \times c^{0.822}$ per mg/m^3-yr

MLE and 95% CI: 0.083 (0.04-0.15) and 0.822 (0.01-1.63) (note: footnote c in Table 3 of Lubin et al. 2008 incorrectly lists 0.63 instead of 1.63)

The slopes, β (c), are rate of increase in the relative risk per mg/m^3-yr and the concentration c is in units of mg/m^3. Even though the variance for β and φ could be inferred from their confidence intervals, upper and lower confidence limits on the slope $\beta(c)$ cannot be estimated without knowing the covariance between β and φ.

In addition to the concentration-dependent slope given above for the full cohort and the restricted sub-cohort, there are other six definitions of slope for the full cohort and for the restricted sub-cohort given in Table 3 of Lubin et al. (2008). Namely;

1) full cohort

   a) T1: β(c, time since last exposure (TSLE)) = $0.120 \times c^{1.153} \times \theta_{TSLE}$ per mg/m^3-yr where TSLE is time since last exposure and θ_{TSLE} is 1.00, 0.83, or 1.20 for TSLE<5, 5≤TSLE<15, and 15≤TSLE, respectively.

   b) T2: $\beta(c,TSLE) = 0.115 \times c^{\square TSLE}$ per mg/m^3-yr where TSLE is time since last exposure and $\square_{TSLE}$ is 0.923, 1.278, or 2.077 for TSLE<5, 5≤TSLE<15, and 15≤TSLE, respectively.

   c) T3: $\beta(c,TSLE) = 0.095 \times c^{\square TSLE} \times \theta_{TSLE}$ per mg/m^3-yr where TSLE is time since last exposure, $\square_{TSLE}$ is 0.723, 1.095, or 2.661 for TSLE<5, 5≤TSLE<15, and 15≤TSLE, respectively, and θ_{TSLE} is 1.00, 1.02, or 2.52 for TSLE<5, 5≤TSLE<15, and 15≤TSLE, respectively.

d) A1: $\beta(c, Age) = 0.153 \times c^{1.175} \times \theta_{Age}$ per mg/m^3-yr where θ_{Age} is 1.00, 0.88, or 0.52 for Age<60, 60≤Age<70, and 70≤Age, respectively.

e) A2: $\beta(c, Age) = 0.115 \times c^{\square Age}$ per mg/m^3-yr where $\square_{Age}$ is 1.285, 1.012, or 1.187 for Age<60, 60≤Age<70, and 70≤Age, respectively.

f) A3: $\beta(c, Age) = 0.200 \times c^{\square Age} \times \theta_{Age}$ per mg/m^3-yr where $\square_{Age}$ is 1.830, 1.153, or 0.077 for Age<60, 60≤Age<70, and 70≤Age, respectively, and θ_{Age} is 1.00, 0.67, or 0.20 for Age<60, 60≤Age<70, and 70≤Age, respectively.

2) restricted sub-cohort

a) T1-R: $\beta(c,TSLE) = 0.102 \times c^{0.848} \times \theta_{TSLE}$ per mg/m^3-yr where TSLE is time since last exposure and θ_{TSLE} is 1.00, 0.75, or 0.18 for TSLE<5, 5≤TSLE<15, and 15≤TSLE, respectively.

b) T2-R: $\beta(c,TSLE) = 0.085 \times c^{\square TSLE}$ per mg/m^3-yr where TSLE is time since last exposure and $\square_{TSLE}$ is 0.632, 1.111, or 3.486 for TSLE<5, 5≤TSLE<15, and 15≤TSLE, respectively.

c) T3-R: $\beta(c,TSLE) = 0.095 \times c^{\square TSLE} \times \theta_{TSLE}$ per mg/m^3-yr where TSLE is time since last exposure, $\square_{TSLE}$ is 0.739, 1.240, or 17.53 for TSLE<5, 5≤TSLE<15, and 15≤TSLE, respectively, and θ_{TSLE} is 1.00, 0.99, or 0.11 for TSLE<5, 5≤TSLE<15, and 15≤TSLE, respectively.

d) A1-R: $\beta(c, Age) = 0.088 \times c^{0.878} \times \theta_{Age}$ per mg/m^3-yr where θ_{Age} is 1.00, 0.88, or 0.66 for Age<60, 60≤Age<70, and 70≤Age, respectively.

e) A2-R: $\beta(c, Age) = 0.082 \times c^{\square Age}$ per mg/m^3-yr where $\square_{Age}$ is 1.118, 0.813, or 0.678 for Age<60, 60≤Age<70, and 70≤Age, respectively.

f) A3-R: $\beta(c, Age) = 0.156 \times c^{\square Age} \times \theta_{Age}$ per mg/m^3-yr where $\square_{Age}$ is 1.724, 1.001, or -0.281 for Age<60, 60≤Age<70, and 70≤Age, respectively, and θ_{Age} is 1.00, 0.64, or 0.20 for Age<60, 60≤Age<70, and 70≤Age, respectively.

None of the models using additional parameters (Age or TSLE) to adjust the slope fitted the data statistically significantly better than the models where the slope depended only on the concentration raised to a power. That is, the introduction of TSLE or Age as effect-modification factors do not improve the model fit to the observed data.

E.5 Slope Estimates at Specific Average Concentrations - Sensitivity Analyses

Since the slope for the cumulative exposure of the multiplicative relative risk model is dependent on the average exposure concentration, the slope at some specific concentrations may be of interest. The following items 1 and 2 are slopes at specific arsenic concentrations followed by items 3 to 6 with average arsenic concentrations that make the concentration-dependent slope equal to the other estimates of the slope of the relative risk model:

1) slope for the full cohort at the mean airborne arsenic concentration for the full cohort (0.35 mg/m^3 in Table 1 of Lubin et al. 2008)

$$0.115 \times 0.35^{1.123} \text{ per } mg/m^3\text{-yr} \times 0.001 \text{ } \mu g/mg = 3.54E\text{-}05 \text{ per } \mu g/m^3\text{-yr}$$

2) slope for the restricted sub-cohort at the mean airborne arsenic concentration for the restricted sub-cohort (0.36 mg/m^3 in Table 1 of Lubin et al. 2008)

$$0.083 \times 0.36^{0.822} \text{ per } mg/m^3\text{-yr} \times 0.001 \text{ } \mu g/mg = 3.58E\text{-}05 \text{ per } \mu g/m^3\text{-yr}$$

3) average arsenic concentration at which the slope (5.75E-05 per $\mu g/m^3$-yr) estimated from the data in Table 2 of Lubin et al. (2008) is equal to the slope $\beta(c)$ based for the full cohort. This concentration can be calculated by solving for c in the following equation:

$$0.115 \times c^{1.123} \text{ per } mg/m^3\text{-yr} = 0.0575 \text{ per } mg/m^3\text{-yr}$$

which implies

$$c = (0.0575/0.115)^{(1/1.123)} = 0.539 \text{ } mg/m^3$$

That is, the concentration-dependent slope $\beta(c)$ based on the full cohort is equal to the constant slope β estimated for the full cohort when the concentration is equal to 0.539 mg/m^3. The concentration-dependent slope $\beta(c)$ based on the full cohort is less (greater) than to the constant slope β when the average arsenic concentration is less (greater) than 0.539 mg/m^3. This also means that using the slope 5.75E-05 per $\mu g/m^3$-yr with concentrations below 539 $\mu g/m^3$ results in higher risk estimates (more health protective risk estimates) than using the concentration-dependent slope.

4) average arsenic concentration at which the slope for the full cohort (4.756E-05 per $\mu g/m^3$-yr from Figure 2 of Lubin et al. (2008)) is equal to the slope $\beta(c)$ based on the full cohort. This concentration can be calculated by solving for c in the following equation:

$$0.115 \times c^{1.123} \text{ per } mg/m^3\text{-yr} = 0.04756 \text{ per } mg/m^3\text{-yr}$$

which implies

$$c = (0.04756/0.115)^{(1/1.123)} = 0.456 \text{ } mg/m^3$$

That is, the concentration-dependent slope β(c) based on the full cohort is equal to the constant slope β when the average arsenic concentration is equal to 0.456 mg/m^3. The concentration-dependent slope β(c) based on the full cohort is less (greater) than to the constant slope β when the average arsenic concentration is less (greater) than 0.456 mg/m^3. This also means that using the slope 4.756E-05 per μg/m^3-yr with concentrations below 456 μg/m^3 results in higher risk estimates than using the concentration-dependent slope.

5) average arsenic concentration at which the slope for the restricted sub-cohort (2.1E-04 per μg/m^3-yr) in the Lubin et al. (2000) paper is equal to the slope β(c) based on the restricted sub-cohort. This concentration can be calculated by solving for c in the following equation:

$$0.083\times c^{0.822} \text{ per mg/m}^3\text{-yr} = 0.21 \text{ per mg/m}^3\text{-yr}$$

which implies

$$c = (0.21/0.083)^{(1/0.822)} = 3.09 \text{ mg/m}^3$$

That is, the concentration-dependent slope β(c) based on the restricted sub-cohort is equal to the constant slope β of 0.21 per mg/m^3-yr when the average arsenic concentration is equal to 3.09 mg/m^3. The concentration-dependent slope β(c) based on the restricted sub-cohort is less (greater) than to the constant slope β when the average arsenic concentration is less (greater) than 3.09 mg/m^3. This also means that using the slope 2.10E-04 per μg/m^3-yr with concentrations below 3090 μg/m^3 results in higher risk estimates than using the concentration-dependent slope.

6) average arsenic concentration at which the concentration-dependent slope for the full cohort ($0.115\times c^{1.123}$) equals the concentration-dependent slope for the restricted sub-cohort ($0.083\times c^{0.822}$). This average arsenic concentration can be calculated by solving for c in the following equation:

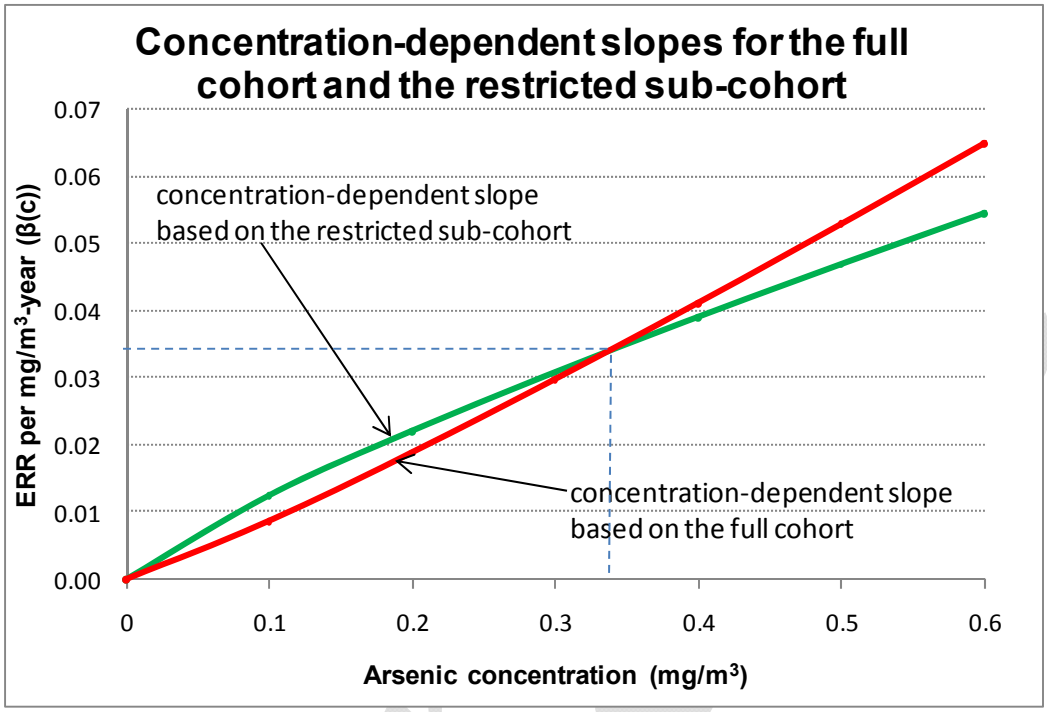
$$0.115\times c^{1.123} \text{ per mg/m}^3\text{-yr} = 0.083\times c^{0.822} \text{ per mg/m}^3\text{-yr}$$

which implies

$$c = (0.083/0.115)^{(1/(1.123-.822))} = 0.338 \text{ mg/m}^3$$

Thus, for the average arsenic concentration c=0.338 mg/m^3, the β(c) based on the full cohort and the β(c) based on the restricted sub-cohort are equal to 0.0338 per mg/m^3-yr. The β(c) based on the full cohort is less than the β(c) based on the restricted sub-cohort for arsenic concentrations c < 0.338 mg/m^3. The β(c) based on the full cohort is greater than the β(c) based on the restricted sub-cohort for arsenic concentrations c>0.338

mg/m^3. The following figure shows the concentration-dependent slopes for the full cohort and the restricted sub-cohort.



Summary table of the results given above: (NOTE: estimates of the average arsenic concentrations for an added risk of 1 in 100,000 are well below 0.3 mg/m^3 which is equivalent to an environmental arsenic concentration (24 hrs a day, 365 days a year) of approximately 0.1 mg/m^3. That implies that any of the slopes listed below would be conservative -- i.e., predict more health-protective excess risks.)

3

c (mg/m^3-yr)	β(c) per $\mu g/m^3$-yr	Comments
0.35	3.54E-05 $(0.115 \times c^{1.123})$	Slope at mean exposure concentration in full cohort using Lubin et al. β(c) derived from full cohort. The slope β(c) is smaller at concentrations less than 0.35 mg/m^3.
0.36	3.58E-05 $(0.083 \times c^{0.822})$	Slope at mean exposure concentration in restricted sub-cohort using Lubin et al. β(c) derived from restricted sub-cohort. The slope β(c) is smaller at concentrations less than 0.36 mg/m^3.
0.54	5.75E-05 $(0.115 \times c^{1.123})$	Slope estimated from full cohort data in Table 2 is equal to the slope β(c) derived from the full cohort at a concentration of 0.54 mg/m^3. The slope β(c) is smaller at concentrations less than 0.54 mg/m^3.
0.46	4.76E-05 $(0.115 \times c^{1.123})$	Slope reported for full cohort in Figure 2 is equal to the slope β(c) derived from the full cohort at a concentration of 0.46 mg/m^3. The slope β(c) is smaller at concentrations less than 0.46 mg/m^3.
3.09	2.10E-04 $(0.083 \times c^{0.822})$	Slope reported for restricted sub-cohort in Lubin et al. (2000) is equal to the Lubin et al. slope β(c) derived from the restricted sub-cohort at a concentration of 3.09 mg/m^3. The slope β(c) is smaller at concentrations smaller than 3.09 mg/m^3.
0.34	3.40E-05 $(0.115 \times c^{1.123} = 0.083 \times c^{0.822})$	The slope β(c) derived from the full cohort is equal to the slope β(c) derived from the restricted sub-cohort at a concentration of 0.34 mg/m^3. The slope β(c) for the full cohort is smaller at concentrations less than 0.34 mg/m^3 and greater at concentrations greater than 0.34 mg/m^3.

E.6 Summary of Maximum Likelihood Estimates of the Slope and 95% Confidence Limits

The Lubin et al. (2008) model based on the full cohort and the Lubin et al. (2008) model based on the restricted sub-cohort of workers exposed to an average arsenic concentration of 0.29 mg/m^3 were fit using the standard multiplicative relative risk model. These two models seem to be the most defensible for environmental risk assessment purposes. The first model is based on all the data and parallels the estimation procedures used for the Tacoma and Swedish cohorts. The second estimate is based on low arsenic occupational concentration exposures which are more similar to the environmental concentration exposures of the general population. In addition, these two estimates are in the range of the estimates obtained with the Tacoma and Swedish cohorts.

Table E-2 summarizes the maximum likelihood estimates of the slope as well as the corresponding 95% lower and upper confidence limits.

3

Table E-2. Estimates of β (MLE), SE, β (95% LCL) and β (95% UCL) (Lubin et al. 2000; 2008) [a]			
	β (MLE) ± SE per μg/m^3-yr	**β (95% LCL)** [b] per μg/m^3-yr	**β (95% UCL)** [b] per μg/m^3-yr
Lubin et al. (2000) [c] (restricted sub-cohort)	2.03E-04 ± 9.48E-05	2.64E-05	3.79E-04
Lubin et al. (2008) [d] (full cohort)	**5.75E-05 ± 1.61E-05**	**3.10E-05**	**8.40E-05**
Lubin et al. (2008) [e] (full cohort) 290 μg/m^3	**1.6E-05 ± 1.17E-05** [f]	**-3.23E-06**	**3.52E-05**
Lubin et al. (2008) [e] (full cohort) 300-390 μg/m^3	6.7E-05 ± 2.41E-05 [g]	2.73E-05	1.07E-04
Lubin et al. (2008) [e] (full cohort) 400-490 μg/m^3	7.7E-05 ± 3.58E-05 [h]	1.81E-05	1.36E-04
Lubin et al. (2008) [e] (full cohort) >500 μg/m^3	7.2E-05 ± 1.63E-05 [i]	4.53E-05	9.87E-05
Other alternatives include the use of the concentration-dependent slopes derived by Lubin et al. 2008 and approximate 95% lower and upper confidence limits			
Lubin et al. (2008) [j] (full cohort)	$\exp\{\ln(0.115)\pm0.255\}\times c^{1.123\ k}$ per mg/m^3-yr	$0.0756\times c^{1.123\ n}$ per mg/m^3-yr	$0.175\times c^{1.123\ n}$ per mg/m^3-yr
Lubin et al. (2008) [l] (restricted sub-cohort)	$\exp\{\ln(0.083)\pm0.335\}\times c^{0.822\ m}$ per mg/m^3-yr	$0.0478\times c^{0.822\ n}$ per mg/m^3-yr	$0.144\times c^{0.822\ n}$ per mg/m^3-yr

4 [a] Units are in ERR per μg/m^3-yr and cumulative exposure estimates with a weight of 0.1 in heavy
5 exposure areas
6 [b] 95% LCL = β - (1.645 × SE) for a standard normal distribution; 95% UCL = β + (1.645 × SE) for a
7 standard normal distribution
8 [c] Linear model fit to the rate ratios in Table 4 of Lubin et al. (2000) with weight λ=0.1 using least squares
9 regression with a multiplicative intercept. Lubin et al. (2000) estimates are 2.1E-04 (95% CI: 1.0E-05,
10 4.6E-04) – page 558.
11 [d] Maximum likelihood estimate of the slope and its SE for the multiplicative linear relative risk model
12 based on the full cohort data in Table 2 of Lubin et al. 2008
13 [e] Estimates of the ERR per μg/m^3-yr of respiratory cancer mortality by categories of cumulative arsenic
14 exposure (μg/m^3-yr), (from Model 1, Figure 1 of Lubin et al. 2008)
15 [f] The average SE was back-calculated from 95% confidence intervals of -5.00E-06, 4.10E-05 per μg/m^3-
16 yr based on the following equation: confidence interval = β ± (1.96 x SE)
17 [g] The average SE was back-calculated from 95% confidence intervals of 2.40E-05, 1.19E-04 per μg/m^3-yr
18 based on the following equation: confidence interval = β ± (1.96 x SE)
19 [h] The average SE was back-calculated from 95% confidence intervals of 1.70E-05, 1.59E-04 per μg/m^3-yr
20 based on the following equation: confidence interval = β ± (1.96 x SE)
21 [i] The average SE was back-calculated from 95% confidence intervals of 4.30E-05, 1.07E-04 per μg/m^3-yr
22 based on the following equation: confidence interval = β ± (1.96 x SE)
23 [j] Estimate of the ERRs of respiratory cancer mortality with concentration-dependent slope based on the
24 full cohort, (Model B0 in Table 3 of Lubin et al. 2008)

[k] The average SE was back-calculated from the 95% confidence interval (ln(0.07), ln(0.19)), based on the following equation: confidence interval = ln(β) ± (1.96 x SE)

[l] Estimate of the ERRs of respiratory cancer mortality with concentration-dependent slope based on the restricted sub-cohort, (Model B0-R in Table 3 of Lubin et al. 2008)

[m] The average SE was back-calculated from the 95% confidence interval (ln(0.04), ln(0.15)), based on the following equation: confidence interval = ln(β) ± (1.96 x SE)

[n] 95% lower and upper confidence limits assuming the power of the concentration is a constant with zero variability

(This bounds are an approximation because in general the SE for the β parameter would not suffice for this model because a full variance/covariance matrix is required since it is a multiparameter (i.e., at least the two parameters β and □ were estimated) model (2-3-09 email from Dr. Lubin))

E.7 References

Lubin, JH, LM Pottern, BJ Stone, and JF Fraumeni, Jr. (2000). Respiratory Cancer in a Cohort of Copper Smelter Workers: Results from More than 50 Years of Follow-up. *American Journal of Epidemiology*, **151**:554-565.

Lubin, JH, LE Moore, JF Fraumeni, Jr, and KP Cantor (2008). Respiratory Cancer and Inhaled Inorganic Arsenic in Copper Smelter Workers: A Linear Relationship with Cumulative Exposure that Increases with Concentration. *Environmental Health Perspectives*, **116**:1661-1665.

Appendix F. Analyses of the Copper Smelter in Sweden (Järup et al. 1989)

Robert L. Sielken Jr., Ph.D., and Ciriaco Valdez Flores, Ph.D., P.E.
Sielken & Associates Consulting Inc.
3833 Texas Avenue, Suite 230, Bryan, TX 77802
Tel: 979-846-5175; Fax: 979-846-2671;
Email: SielkenAssoc@aol.com

June 18, 2009

F.1 New Analysis Adjusting for Year of First Hire

The slope of the multiplicative relative risk linear model for the total cohort but adjusting for the first year of hire is shown in Table F.1. This new analysis parallels the analyses done for the Tacoma cohort (see Appendix C). In fact, the data structure for the Ronnskar cohort is so similar to the data structure of the Tacoma cohort that the model descriptions can be essentially the same. The results of the entire cohort adjusting for the year of first hire is the most defensible result because it is based on more data than the separate analyses based on subsets of the cohort and adjusts for the effect of potential differences in exposure concentrations with calendar year by using a nonparametric estimate for the effect of year of hire.

Table F-1. Estimates of β (MLE), SE, β (95% LCL) and β (95% UCL) (Järup et al. 1989) [a]

	Intercept (α)	β (MLE) ± SE	β (95% LCL) [b]	β (95% UCL) [c]
All workers adjusting for year of hire (h = 1.19[d])	2.37	2.92E-05 ± 1.63E-05	2.31E-06	5.61E-05
All workers with no adjustment	2.67	2.38E-05 ± 9.14E-06	8.79E-06	3.89E-05
Workers hired < 1940	2.48	2.62E-05 ± 1.35E-05	4.00E-06	4.84E-05
Workers hired 1940+	2.60	6.17E-05 ± 5.92E-05	-3.57E-05	1.59E-04

[a] Units are in ERR per $\mu g/m^3$-yr.
[b] 95% LCL = β - (1.645 x SE) for a standard normal distribution.
[c] 95% UCL = β + (1.645 x SE) for a standard normal distribution.
[d] the background lung cancer mortality rate for workers hired 1940+ is 1.19-fold higher than the background lung cancer mortality rate for workers first hired < 1940

F.2 Consistency of Conclusions in Järup et al. (1989) and Lubin et al. (2008)

Two conclusions in Järup et al. (1989) are consistent with a conclusion in Lubin et al. (2008). Namely, Järup et al. indicate that their "data suggest that arsenic concentration is more important than duration of exposure for the risk of developing lung cancer." In addition, Järup et al. indicate that they "did not find a clear dose-response relationship in the low exposure categories." These two statements in Järup et al. (1989) are consistent with Lubin et al. (2008) conclusion that their results suggested a "direct concentration effect on the exposure-response relationship, indicating that for a fixed level of cumulative arsenic exposure, inhalation of higher concentrations of arsenic over shorter durations was more deleterious than inhalation of lower concentrations over longer durations."

F.3 Uncertainty Analysis

The data in Järup et al. do not include the average cumulative exposure for each of the cumulative dose categories. Viren and Silvers (1994) used the midpoints of the dose ranges in fitting the models to the Järup et al. data. Here, we also used the midpoints of the dose ranges in fitting the models. The midpoints of the dose ranges are good approximations to the average cumulative exposure for the person-years in the dose ranges. However, the last dose range (cumulative exposures greater than 100 mg/m^3-yr) is unbounded and Viren and Silvers "assumed that the median exposure in this group was 25% greater than the lower bound of the given interval." The estimation for the midpoint for the highest, unbounded, cumulative exposure range is always controversial, unless it is based on actual data. Oftentimes reviewers are uncertain of the influence that the value of the midpoint for the highest dose range may have on the estimates of the model parameters. One analysis that helps in satisfying the uncertainty that the specific value for the highest dose range may have introduced in the estimates of the parameters is to evaluate the same dose response model without the data on the highest dose range. Table F-2 shows the parameter estimates based on the Järup et al. data after removing the person years in the highest dose range.

Table F-2. Estimates of β (MLE), SE, β (95% LCL) and β (95% UCL) (Järup et al. 1989) [a] excluding the highest (> 100,000 µg/m^3-yrs) cumulative exposure range				
	Intercept (α)	**β (MLE) ± SE**	**β (95% LCL)** [b]	**β (95% UCL)** [c]
All workers adjusting for year of hire (h = 1.17[d])	2.40	2.75E-05 ± 2.11E-05	-7.17E-06	6.22E-05
All workers with no adjustment	2.71	2.15E-05 ± 1.13E-05	2.88E-06	4.01E-05
Workers hired < 1940	2.57	2.25E-05 ± 1.63E-05	-4.28E-06	4.93E-05
Workers hired 1940+	2.60	6.17E-05 ± 5.92E-05	-3.57E-05	1.59E-04

[a] Units are in ERR per µg/m^3-yrs.
[b] 95% LCL = β - (1.645 x SE) for a standard normal distribution.
[c] 95% UCL = β + (1.645 x SE) for a standard normal distribution.
[d] the background lung cancer mortality rate for workers hired 1940+ is 1.17-fold higher than the background lung cancer mortality rate for workers first hired < 1940

The maximum likelihood estimates and corresponding lower and upper confidence limits are very similar whether or not the highest exposure group of person-years is included in the estimation. The conclusions of the uncertainty analysis of including/excluding the highest cumulative exposure group of person years can be summarized as follows:

1) The estimates for "Workers hired 1940+" do not change because there were no person-years in the highest cumulative exposure group.
2) Maximum likelihood estimates are slightly smaller when the person years in the highest cumulative exposure range are excluded.
3) Standard errors of the estimated slope are slightly larger when the person years in the highest cumulative exposure range are excluded. The standard errors were expected to be larger here because the estimates are based on fewer observations.
4) The 95% upper confidence limits on the slope were slightly larger when the person years in the highest cumulative exposure range are excluded. This is not surprising because standard errors (as expected) were larger.
5) The intercepts (α) are slightly larger when the person years in the highest cumulative exposure range are excluded.
6) The likelihood of the data that excludes the highest cumulative exposure range using the models fit to the data that excludes the highest cumulative exposure range was compared to the likelihood of the data that excludes the highest cumulative exposure range using the models fit to the data that include all dose ranges. They are essentially equal;

indicating that the model fit to the data that includes all the dose ranges is as good as the model fit to the data that excludes the highest dose range. Table F-3 shows these results.

Table F-3 . Logarithm of the likelihood of observing the data that excludes the highest cumulative exposure range

	Maximum Likelihood Estimates based on the data without the highest cumulative exposure range	**Logarithm of the Likelihood of observing the data without the highest cumulative exposure range**	**Maximum Likelihood Estimates based on all cumulative exposure ranges**	**Logarithm of the Likelihood of observing the data without the highest cumulative exposure range**
All workers adjusting for year of hire	h=1.17 α=2.40 β=2.75E-05	129.450	h=1.19 α=2.37 β=2.92E-05	129.444
All workers with no adjustment	α=2.71 β=2.15E-05	25.069	α=2.67 β=2.38E-05	25.045
Workers hired < 1940	α=2.57 β=2.25E-05	5.239	α=2.48 β=2.62E-05	5.239
Workers hired 1940+	α=2.60 β=6.17E-05	20.182	α=2.60 β=6.17E-05	20.152

7) The likelihood of the data that includes all dose ranges using the models fit to the data that excludes the highest dose group was compared to the likelihood of the data that includes all dose ranges using the models fit to the data that include all dose ranges. They are essentially equal; indicating that the model fit to the data that excludes the highest dose range is as good as the model fit to the data that include all the dose ranges. Table F-4 shows these results.

Table F-4 . Logarithm of the likelihood of observing the data that includes all the dose ranges

	Maximum Likelihood Estimates based on the data without the highest cumulative exposure range	**Logarithm of the Likelihood of observing the data that includes all the dose ranges**	**Maximum Likelihood Estimates based on all cumulative exposure ranges**	**Logarithm of the Likelihood of observing the data that includes all the dose ranges**
All workers adjusting for year of hire	h=1.17 α=2.40 β=2.75E-05	147.245	h=1.19 α=2.37 β=2.92E-05	147.257
All workers with no adjustment	α=2.71 β=2.15E-05	42.146	α=2.67 β=2.38E-05	42.191
Workers hired < 1940	α=2.57 β=2.25E-05	5.239	α=2.48 β=2.62E-05	5.239
Workers hired 1940+	α=2.60 β=6.17E-05	37.230	α=2.60 β=6.17E-05	37.295

8) The parameters obtained using all the dose ranges are not statistically significantly different than the parameters obtained from the data that excludes the highest dose range. The parameter estimates using all the dose ranges are preferable than the parameter estimates based on the data that excludes the highest dose range because the former are more precise (i.e., have smaller standard errors) and because they rely on more data.

F.4 References

Järup, L, G Pershagen, and S Wall (1989). Cumulative Arsenic Exposure and Lung Cancer in Smelter Workers: A Dose-Response Study. *American Journal of Industrial Medicine*, **15**:31-41,

Viren, J and A Silvers (1994). Unit Risk Estimates for Airborne Arsenic Exposure: An Updated View Based on Recent Data from Two Copper Smelter Cohorts. *Regulatory Toxicology and Pharmacology*, **20**:125-138.

Appendix G. Analyses of the Humberside, UK Tin Smelter (Jones et al. 2007)

Robert L. Sielken Jr., Ph.D., and Ciriaco Valdez Flores, Ph.D., P.E.
Sielken & Associates Consulting Inc.
3833 Texas Avenue, Suite 230, Bryan, TX 77802
Tel: 979-846-5175; Fax: 979-846-2671;
Email: SielkenAssoc@aol.com

July 8, 2009

G.1 Review of Jones et al. (2007) Study

Jones et al. (2007) analyze a cohort of 1426 male workers employed for at least one year between November 1, 1967 and July 28, 1995 that were followed-up through the end of 2001. Jones et al. focus their analyses on the dose response relationship between lung cancer and exposures to arsenic, cadmium, antimony, lead, and polonium-210 with the purpose of identifying the cause or causes of the excess lung cancer deaths observed. This excess of lung cancers in the same cohort had been previously reported by Brinks et al. (2005).

G.2 Exposure Concentrations

Jones et al. used the measurements of numerous air samples to estimate the concentrations of the different agents at the smelter. These measurements were recorded for the period 1972 to 1991. In addition, Jones et al. used the work history for each cohort member to calculate the exposure profiles of each worker. The measurements of air concentrations for jobs that started before calendar years 1972 were not available (there were work histories starting in 1937). Jones et al. extrapolated exposures concentrations to years prior to 1972 using three alternative extrapolation assumptions. The following figures illustrate a hypothetical example of the three alternative extrapolation exposure scenarios used by Jones et al.

Figure G-1. Exposure concentration extrapolation example using Scenario A
"Constant back-extrapolation in each process area, as the mean of the levels in the three earliest years for which data were available."

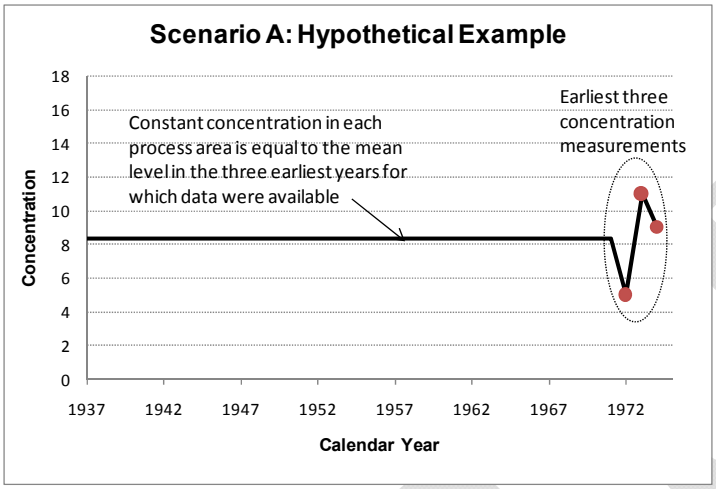


Figure G-2. Exposure concentration extrapolation example using Scenario B
"Back-extrapolation in each process area on a linear increasing trend from a baseline value, to values 2-fold higher in the early 1940s, based on a weak trend seen in per-caput average exposure levels over the period 1972-91."

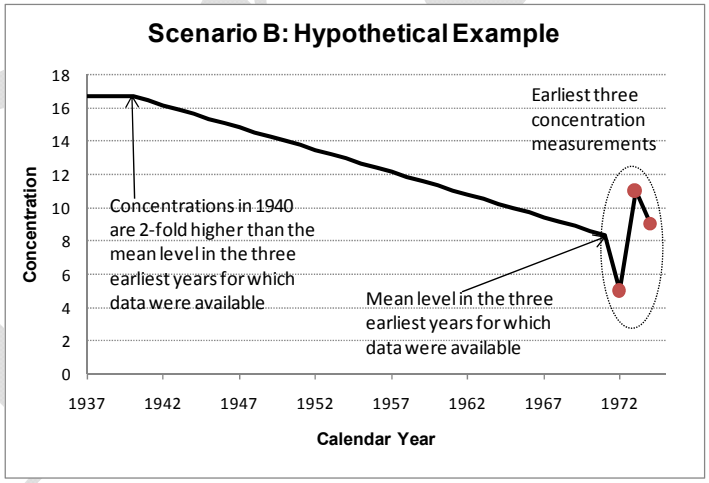
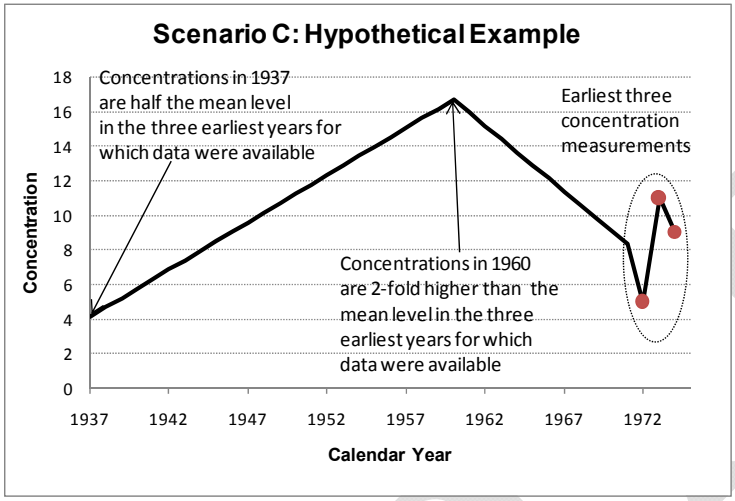


Figure G-3. Exposure concentration extrapolation example using Scenario C
"Back-extrapolation in each process area from a baseline value to values 2-fold higher in 1960, subsequently, declining linearly to values one-half of the baseline in 1937."



The cumulative exposure or area under the curve (AUC) up to the earliest year with concentration measurements for each of the three scenarios can be calculated as a function of the average concentration of the earliest three concentration measurements (AvgC). Thus, the extrapolated cumulative exposures through the end of 1971 are as follows (assuming all of the scenarios extrapolate back to 1937):

$AUC_{ScenarioA} = (1972\text{-}1937)\times AvgC$

$AUC_{ScenarioB} = (1940\text{-}1937)\times 2\times AvgC + (1972\text{-}1940)\times(3/2)\times AvgC$

$AUC_{ScenarioC} = (1960\text{-}1937)\times(5/4)\times AvgC + (1972\text{-}1960)\times(3/2)\times AvgC$

It can be shown, after some algebra, that $AUC_{ScenarioB} > AUC_{ScenarioC} > AUC_{ScenarioA}$. That is, extrapolating concentrations using Scenario B results in the largest cumulative exposures, followed by the cumulative exposures estimated using Scenario C, and the smallest cumulative exposures of the three scenarios are predicted using Scenario A.

G.3 Modeling

Jones et al. fit Poisson regression models to the number of lung cancer deaths split into quintiles of the distribution of the dose metric among the lung cancer decedents. The weighted average of the dose metric in each dose interval was used in fitting the relative risk linear dose response model with additive intercept.
Jones et al. fitted the dose response model using the following two different dose metrics for each of the five agents (arsenic, cadmium, antimony, lead, and polonium-210):

1) Cumulative exposure
2) Weighted cumulative exposure

The cumulative exposure dose metric is in units of concentration-year (e.g., mg/m^3-yr). The weighted cumulative exposure dose metric is an exposure that is modified by other factors that weight the effect that the concentration might have on lung cancer. Jones et al. suggest using a weighted cumulative exposure dose metric that diminishes the risk of lung cancer with the time since exposure and the age of the worker. They indicate that Binks et al. (2005) "found evidence of diminution of lung cancer risk with time since exposure." The weights used by Jones et al. to calculate the weighted cumulative exposure were taken from the "exposure-age-concentration model" in BEIR VI (Tables 3-3 and A-4). These weights were initially derived from dose-response models for exposures to radon progeny. The weighted cumulative exposure used by Jones et al. is as follows:

Weighted Cumulative Exposure at age n = $\square_n \times \sum_{i=1 \text{ to } n} C_i \times \theta_{n-i}$

where C_i is the exposure concentration at age i,

$\square_{age}$ = 1 if age < 50 years
= $4.8 - 0.105 \times age + 0.000575 \times age^2$ if 50 years ≤ age < 80 years
= 0.09 if age ≥ 80 years

and, defining tse (time since exposure) as age n minus i in the above equation,

θ_{tse} = 0 if tse < 5 years
= 1 if 5 years ≤ tse < 10 years
= $1.17 - 0.0145 \times tse - 0.00025 \times tse^2$ if 10 years ≤ tse < 30 years
= 0.51 if tse ≥ 30 years

Jones et al. smoothed the step function for φ_{age} and θ_{tse} specified in Tables 3-3 and A-4 in BEIR VI. The following two figures (similar to Figure 1 in Jones et al.) show the step functions for the weights and the smoothed functions used by Jones et al.

Figure G-4. BEIR VI weighting factor for attained age (γ_{age}) and the smoothed function used by Jones at al. (2007)

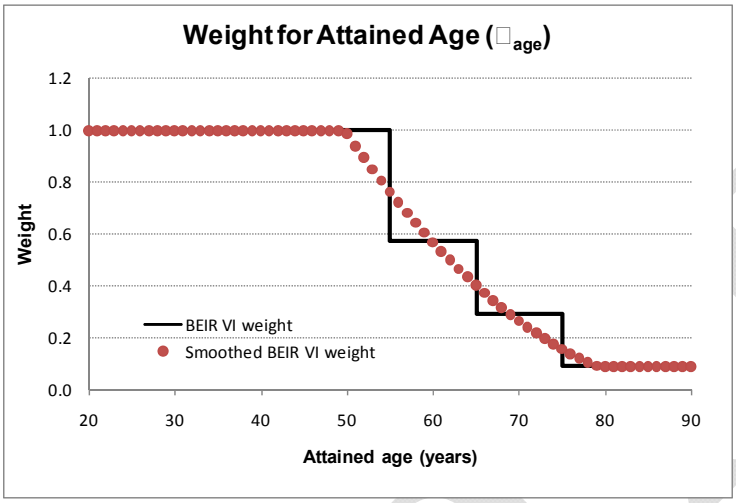
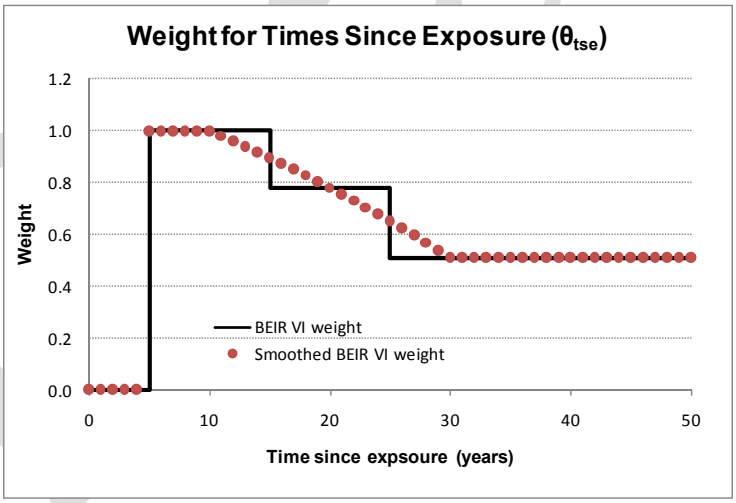


Figure G-5. BEIR VI weighting factor for the time since exposure (θ_{tse}) and the smoothed function used by Jones at al.



Jones et al. indicate that fitting the models using the smoothed function and the step-function version of the weights result in approximately the same estimates. The weights given in Tables 3-3 and A-4 of the BEIR VI report, however, included another weighting factor that was ignored by Jones et al. That is, the weighted cumulative exposure dose metric used in BEIR VI is equal to

$$\text{BEIR VI Weighted Cumulative Exposure at age n} = \gamma_z \times \gamma_n \times \sum_{i=1 \text{ to } n} C_i \times \theta_{n-i}$$

where all the components are identical to the equation for the weighted cumulative exposure specified in Jones et al. with the exception of γ_z. The variable γ_z is the effect of the exposure rate. In Tables 3-3 and A-4 of the BEIR VI report the definition of γ_z is as follows:

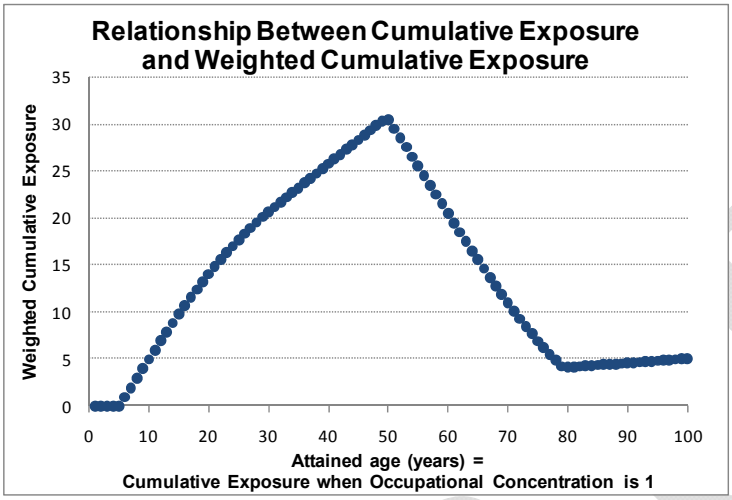
γ_z = 1.00		if exposure rate < 0.5 WL
= 0.49		if 0.5 WL ≤exposure rate < 1.0 WL
= 0.37		if 1.0 WL ≤exposure rate < 3.0 WL
= 0.32		if 3.0 WL ≤exposure rate < 5.0 WL
= 0.17		if 5.0 WL ≤exposure rate < 15.0 WL
= 0.11		if exposure rate ≥ 15.0 WL

Jones et al. could not use this weighting factor directly because the exposure rates in BEIR VI are for units of radon progeny concentrations in WL, which may be very different to the units of average concentrations of the five agents analyzed by Jones et al. However, Jones et al. could have fit this weighting parameter using a linear approximation if they believe the mechanism of the five agents in causing lung cancer is similar to the mechanism of radon progeny in causing lung cancer.

Jones et al. could also have used the "exposure-age-duration model" proposed by BEIR VI in Tables 3-3 and A-4. The modifying effects for age attained and time since exposure were very similar to those in the "exposure-age-concentration model" also proposed in BEIR VI and used by Jones et al. The weights for the "duration of exposure" under the "Exposure-age-duration model" given in Table 3-3 and A-4 of BEIR VI could have been used because they depend on time and not on specific concentrations of radon progeny.

For a fixed concentration or exposure rate, the weighted cumulative exposure used by Jones et al. (an also proposed in the "exposure-age-concentration model" in BEIR VI) is zero for the first 5 years of exposure, then increases for the next 45 years followed by a decrease for the next 30 years, to slowly increase 80 years after the first exposure. Figure G-6 shows the weighted cumulative exposure as a function of age using the smoothed weights derived by Jones et al.

Figure G-6. Weighted cumulative exposure using the smoothed weights in Jones et al. for a concentration of 1



Although BEIR VI used the following multiplicative relative risk model with a multiplicative intercept and the weighted cumulative exposure to radon progeny,

$$E(O_j) = \alpha \times E_{oj} \times (1 + \beta \times d_j)$$

Jones et al. used a multiplicative relative risk model with additive intercept given by

$$E(O_j) = E_{oj} \times (\alpha + \beta_a \times d_j)$$

where the α term adjusts for any possible differences between the population's background cancer rates and the cohort's observed cancer rates in unexposed workers.

In the equation above the variables are:

$E(O_j)$ = expected number of lung cancer deaths for exposure group j predicted by the model;

E_{oj} = expected number of background lung cancer deaths for exposure group j based on the reference population background cancer rates;

β = multiplicative factor by which the cohort's background risk increases with cumulative exposure;

β_a = multiplicative factor by which the reference population's background risk increases with cumulative exposure;

d_j = cumulative exposure (weighted or unweighted) for exposure group j;

α = multiplicative factor that accounts for differences in cancer mortality background rates between the study cohort and the reference population.

The interpretations of slope parameters β (in the multiplicative relative risk model with multiplicative intercept) and β_a (in the multiplicative relative risk model with additive intercept) are different. The interpretation of the intercept (α), however, is the same in both models.

G.4 Results

Table 3 in the Jones et al. (2007) lists the maximum likelihood estimates of the additive intercept and slope for the relative risk model along with a p-value for trend and the logarithm of the maximum likelihood. The table shows the results for both unweighted cumulative exposure and the weighted cumulative exposure for each of the five agents analyzed and for each of the three exposure scenarios considered.

Arsenic has the largest logarithm of the maximum likelihood (i.e., fits the data the best) when the unweighted cumulative exposure is used as the dose metric, regardless of which exposure scenario is used. However, when the weighted cumulative exposure is used as the dose metric, antimony (Sb) has the largest logarithm of the maximum likelihood for all three exposure scenarios.

Weighted cumulative exposures to antimony, arsenic and lead were statistically significantly associated with lung cancer mortality for the three exposure scenarios. Exposure to these three agents, however, are highly correlated and Jones et al. acknowledged that "the data alone do not permit unambiguous attribution of causality to arsenic exposure, antimony exposure, lead exposure or a combination of the three." Although the likelihood of the data is largest when weighted cumulative exposure to antimony is used, the difference in likelihood between using antimony versus arsenic or lead is not statistically significant. Jones et al. concluded that arsenic exposure is the cause for the increased lung cancer mortality because there is evidence from other studies that exposures to arsenic increase lung cancer mortality, and because there is no strong historical evidence of a relationship between antimony or lead exposure and lung cancer.

G.5 Modeling Comparison of Jones et al. (2007) and Lubin et al. (2008)

The objective of the Lubin et al. (2008) paper was to evaluate the shape of the dose-response relationship between respiratory cancer mortality and cumulative exposure to arsenic and the modification of this relationship by the average exposure concentration (Appendix E). Similarly, the objective of the Jones et al. (2007) paper was to "investigate the relationship between lung cancer mortality and quantitative measures of exposure." There are two ways of interpreting Lubin et al. and Jones et al. models. (Although Lubin et al. used the multiplicative relative risk model with a multiplicative intercept and Jones et al. used the multiplicative relative risk model

with an additive intercept, the interpretations of the slopes given below are still applicable regardless of which model is used.)

1) Interpretation 1: Lubin et al. (2008) (Appendix E) estimated the multiplicative relative risk linear model but instead of assuming a slope (β) that is a constant, they assumed that the slope is a function of the average arsenic concentration (c). The function of the average arsenic concentration for the slope of the linear relative risk model that Lubin et al. used is:

$$\beta(c) = \beta \times c^{\square}$$

where □ is another parameter that is estimated from the data. That is, the relative risk is given by the following

$$RR = 1 + \beta \times c^{\square} \times CumExp$$

where CumExp is the cumulative exposure to arsenic. Lubin et al. went beyond the adjustment of the slope by the functional form shown above, and also considered nonparametric modifications of the slope by age and time since last exposure as well as nonparametric effects of exposure concentrations on the slope.

Similar to Lubin et al., Jones et al. (2007) estimated the relative risk linear model with additive intercept but instead of assuming a slope (β_a) that is a constant, they assumed that the slope is a function of the weighted time since exposure and the age. The function of the weighted time since exposure and the age for the slope of the linear relative risk model with additive intercept that Jones et al. used is:

$$\beta_a(c) = \beta_a \times \square_{age} \times f \times (\sum \theta_i / n)$$

where φ_{age} and θ_i are parameters estimated from epidemiological studies of workers exposed to radon progeny and reported in BEIR VI, f is calculated for each individual worker using the following function

$$f = [\ \sum \theta_i \times C_i\ ] / [CumExp \times (\sum \theta_i / n)\ ].$$

To be precise, the slope of the relative risk model with additive intercept depends on the age, the time since exposure, the concentration of exposure and the cumulative exposure. That is, the relative risk is given by the following

$$RR = 1 + \{\ [\ \beta_a \times \square_{age} \times f \times (\sum \theta_i / n)] \times CumExp\ \} / \alpha$$

Interpretation 2: Lubin et al. (2008) (Appendix E) estimated the multiplicative relative risk linear model assuming a constant slope (β) but the dose metric was the product of the cumulative exposure and the average arsenic concentration (c) to a power. That is, the dose metric is given by the following relation

$$\text{Dose Metric} = \text{CumExp} \times c^{\square}$$

where CumExp is the cumulative exposure to arsenic and φ is another parameter that is estimated from the data. That is, the relative risk is given by the following

$$RR = 1 + \beta \times c^{\square} \times \text{CumExp}$$

Lubin et al. went beyond defining the dose metric by the functional form shown above and also considered nonparametric effects of age and time since last exposure modifying the cumulative exposure.

Similar to Lubin et al., Jones et al. (2007) estimated the relative risk linear model with additive slope assuming a constant slope (β_a) but the dose metric was the product of a parameter that depended on the age of the workers and a sum of the arsenic concentrations multiplied by a weight that depended on the time since exposure. That is, the dose metric is given by the following relation

$$\text{Dose Metric} = \square_{age} \times \sum \theta_i \times C_i$$

where φ_{age} and θ_i are parameters estimated from epidemiological studies of workers exposed to radon progeny and reported in BEIR VI. That is the added risk is given by the following

$$RR = 1 + \{ \beta_a \times \square_{age} \times \sum \theta_i \times C_i \} / \alpha$$

The second interpretation of the Lubin et al. (2008) and Jones et al. (2007) models is how BEIR VI (BEIR. *Health Effects and of Exposure to Radon (BEIR VI)*. Washington, DC: National Academy Press, 1999) used these weights.

G.6 Data for Dose Response Modeling

Jones et al. (2007) present the maximum likelihood estimates and 90% confidence intervals of the intercept and the slope for the relative risk model with additive intercept under the three extrapolation exposure scenarios (A, B, and C), the two dose metrics (cumulative exposure and weighted cumulative exposure), and the five agents analyzed (lead, antimony, arsenic, cadmium, and polonium-210). The standard errors for the intercept and the slope of the relative risk model with additive intercept could be obtained separately from their respective 90% confidence intervals. The maximum likelihood estimate of the slope corresponding to the multiplicative relative risk model can also be easily obtained from the maximum likelihood estimates of the intercept and the slope of the relative risk model with additive intercept given in Table 3 of Jones et al. (2007). That is, if the maximum likelihood estimate of the intercept (α) and the slope (β_a) for the following relative risk model with additive intercept

$$E(O_j) = E_{oj} \times (\alpha + \beta_a \times d_j)$$

are known, then, the maximum likelihood estimates of the intercept (α) and the slope (β) for the multiplicative relative risk model

$$E(O_j) = \alpha \times E_{oj} \times (1 + \beta \times d_j)$$

are,

$$\alpha = \alpha$$

and

$$\beta = \beta_a / \alpha.$$

Although the maximum likelihood estimates of the multiplicative relative risk model can be obtained from the maximum likelihood estimates of the relative risk model with additive intercept, the standard error for the slope of the multiplicative relative risk model with multiplicative intercept cannot be estimated from the standard errors for the parameters of the relative risk model with additive intercept. (The estimates and the standard errors of the estimates are identical for the intercept of both models). Dr. Steve Jones sent an electronic mail message showing the calculation of the maximum likelihood estimates of the slope for the relative risk parameter from the maximum likelihood estimates for the relative risk model with additive intercept for the weighted cumulative exposure to arsenic using the three exposure extrapolation scenarios. Dr. Jones' calculated slopes for the multiplicative relative risk model with multiplicative intercept are

Scenario A: 1.35/1.25 = 1.08 per mg/m^3-yr = 0.00108 per μg/m^3-yr (occupational)

Scenario B: 0.85/1.33 = 0.64 per mg/m^3-yr = 0.00064 per μg/m^3-yr (occupational)

Scenario C: 0.95/1.27 = 0.75 per mg/m^3-yr = 0.00075 per μg/m^3-yr (occupational)

Similar maximum likelihood estimates of the slope from the maximum likelihood estimates for the relative risk model with additive intercept for the unweighted cumulative exposure to arsenic using the three exposure extrapolation scenarios can be obtained as follows:

Scenario A: 0.09/1.53 = 0.0588 per mg/m^3-yr = 5.88E-05 per μg/m^3-yr (occupational)

Scenario B: 0.038/1.58 = 0.0241 per mg/m^3-yr = 2.41 E-05 per μg/m^3-yr (occupational)

Scenario C: 0.06/1.55 = 0.0387 per mg/m^3-yr = 3.87 E-05 per μg/m^3-yr (occupational)

Neither Dr. Jones calculations for the weighted cumulative exposure to arsenic nor the calculations given above for the unweighted cumulative exposure to arsenic provide standard

errors for the estimates, because there is no sufficient information in the Jones et al. (2007) paper to infer these standard errors.

One way to estimate the standard errors of the slope for the multiplicative relative risk linear model using the weighted cumulative exposure to arsenic and the three extrapolation exposure scenarios is using the data given in Table 4 of Jones et al. (2007) (Table 4 is reproduced below). Table 4 shows observed and expected number of lung cancer deaths in the cohort for each interval of weighted cumulative exposure assuming exposure extrapolation scenarios A, B and C. Jones et al. also show in Table 4 the mean weighted cumulative exposure for each of the intervals defined therein. Using the data in Table 4, then the parameters and corresponding standard errors of a multiplicative relative risk model for each exposure scenario can be estimated using Poisson regression. The slope and standard errors of the estimates are:

Scenario A: MLE (SE) = 0.001099 (0.000703) per $\mu g/m^3$-yr (occupational)

Scenario B: MLE (SE) = 0.000649 (0.000490) per $\mu g/m^3$-yr (occupational)

Scenario C: MLE (SE) = 0.000748 (0.000489) per $\mu g/m^3$-yr (occupational)

The maximum likelihood estimates of the slope of the multiplicative relative risk model using Poisson regression on the observed and expected number of lung cancer deaths given in Table 4 of Jones et al. are essentially equal to the slope estimates that Dr. Jones back-calculated from the estimates in Table 3 of Jones et al. (2007) for the relative risk model with additive intercept and weighted cumulative exposure to arsenic. However, significance (p-values) of the slopes of the multiplicative relative risk model are greater than 0.10 (0.12, 0.18 and 0.13) using the Wald's test for significance as opposed to the much smaller p-values (0.012, 0.053, and 0.013) for the slopes of the relative risk model with additive intercept reported by Jones et al. in their Table 3.

Table 4. Regression results: weighted cumulative exposures to arsenic

Exposure range (mg year m^{-3} As)	Mean exposure (mg year m^{-3} As)	Observed	Expected	Observed/expected	CI[a]
Scenario A					
0.0 to ≤0.042	0.0088	13	11.1	1.17	0.62–2.00
>0.042 to ≤0.11	0.075	12	6.8	1.78	0.92–3.11
>0.11 to ≤0.29	0.19	12	8.8	1.36	0.70–2.38
>0.29 to ≤0.62	0.43	12	7.4	1.61	0.83–2.82
>0.62	1.2	13	4.2	3.11	1.66–5.32
Scenario B					
0.0 to ≤0.045	0.0097	13	10.4	1.25	0.67–2.14
>0.045 to ≤0.12	0.081	12	6.7	1.78	0.92–3.12
>0.12 to ≤0.32	0.21	12	8.8	1.37	0.71–2.39
>0.32 to ≤0.71	0.48	12	7.5	1.60	0.83–2.79
>0.71	1.4	13	4.9	2.67	1.42–4.57
Scenario C					
0.0 to ≤0.044	0.0093	13	10.2	1.28	0.68–2.19
>0.044 to ≤0.12	0.083	12	7.0	1.72	0.89–3.00
>0.12 to ≤0.35	0.23	12	9.3	1.29	0.67–2.25
>0.4 to ≤0.84	0.54	12	7.8	1.55	0.80–2.71
>0.84	1.7	13	4.1	3.17	1.69–5.43

G.7 Discussion of Exposure Scenarios for Extrapolation

As discussed in Jones et al. (2007) paper and in the summary given above, the slopes developed for weighted and unweighted cumulative exposures to arsenic are based on three different assumptions in extrapolating exposure concentrations for years before 1972. As a result, the parameter estimates are different, depending on the extrapolation assumption made. As discussed above, Scenario A (which assumes a constant concentration equal to the average concentration of the earliest three years of data) is probably an underestimate of the actual concentrations of arsenic. Scenario B (which assumes a concentration in the early 1940s that is twice the average concentration of the earliest three years of data) is probably a more realistic estimate of the concentrations of arsenic in the early years. Scenario C (which compromises between Scenario A and Scenario B) is an estimate of the concentrations of arsenic in the early years that incorporates production volume in the early years as a possible explanation of arsenic concentration levels.

In most epidemiological studies, exposure concentrations tend to be larger in the early years and decrease to smaller exposure concentrations in later years (similar to the Scenario B assumption). Exposure concentrations like those described in Scenario C are not very common in epidemiological studies. Production increases usually come with larger facilities or newer technologies that tend to dilute exposure concentrations and, therefore, tend to not necessarily increase the exposure concentration as assumed in Scenario C.

G.8 Calculation of Excess Risks

Calculating excess risks from the Jones et al. (2007) models, after converting them to multiplicative relative risk models with multiplicative intercept, can be accomplished using the BEIR IV methodology when the dose metric is cumulative exposures. Similar methodology can be used when the dose metric is weighted cumulative exposure. The weighted cumulative exposure for a constant exposure rate can be easily calculated either using the macros that Jones sent via e-mail or by setting up a recursive calculation in Excel. That is, the weighted cumulative exposure for a constant exposure rate is the product of three components: 1) a sum of weights, 2) an effect of age, and 3) the constant exposure concentration. The sum of weights at a given age ($cumW_{age}$) can be calculated as follows:

1) sum of weights:

$$cumW_{age} = \sum_{i=1 \text{ to age}} \theta_{age-i}$$
$$= cumW_{age-1} + \theta_{age-1}$$

where $cumW_0 = 0$ and θ_{tse} (tse = age-i in the equation above) is

$$\theta_{tse} = 0 \quad \text{if tse} < 5 \text{ years}$$
$$= 1 \quad \text{if } 5 \text{ years} \le \text{tse} < 10 \text{ years}$$
$$= 1.17 - 0.0145 \times \text{tse} - 0.00025 \times \text{tse}^2 \quad \text{if } 10 \text{ years} \le \text{tse} < 30 \text{ years}$$
$$= 0.51 \quad \text{if tse} \ge 30 \text{ years}$$

2) effect of age

The effect of age ($\square_{age}$) is given by

$$\square_{age} = 1 \quad \text{if age} < 50 \text{ years}$$
$$= 4.8 - 0.105 \times \text{age} + 0.000575 \times \text{age}^2 \quad \text{if } 50 \text{ years} \leq \text{age} < 80 \text{ years}$$
$$= 0.09 \quad \text{if age} \geq 80 \text{ years}$$

3) the constant exposure concentration

The constant exposure concentration (c) is the same for each of the years at risk.

Thus, the weighted cumulative exposure at a specific age ($wCumExp_{age}$) is the product of the three components which is equal to

$$wCumExp_{age} = cumW_{age} \times \square_{age} \times c.$$

G.9 Summary of Parameter Estimates

The following table summarizes the maximum likelihood estimates and 95% lower and upper confidence limits on the slope of the multiplicative relative risk model with multiplicative intercept.

Table G-1. Estimates of β (MLE), SE, β (95% LCL) and β (95% UCL) (Jones et al. 2007) [a]

Extrapolation assumption for exposures prior to 1972	β (MLE) ± SE per μg/m^3-yr	β (95% LCL) [b] per μg/m^3-yr	β (95% UCL) [b] per μg/m^3-yr
Estimates based on unweighted cumulative exposure			
Scenario A	5.88E-05 ± SE [c]	na[d]	na
Scenario B	2.41E-05 ± SE	na	na
Scenario C	3.87E-05 ± SE	na	na
Estimates based on weighted cumulative exposure [e]			
Scenario A	1.10E-03 ± 7.03E-04	-5.86E-05	2.26E-03
Scenario B	6.49E-04 ± 4.90E-04	-1.56E-04	1.45E-03
Scenario C	7.48E-04 ± 4.89E-04	-5.67E-05	1.55E-03

[a] Units are in ERR per μg/m^3-yrs

[b] 95% LCL = β - (1.645 × SE) for a standard normal distribution; 95% UCL = β + (1.645 × SE) for a standard normal distribution

[c] The standard error for the slope of the multiplicative relative risk linear model with multiplicative intercept cannot be back-calculated from the standard error of the parameters of the relative risk linear model with additive intercept reported by Jones et al. (2007)

[d] The confidence limit on the estimate of the slope cannot be calculated because the standard error is not available

[e] Multiplicative relative risk linear model with multiplicative intercept fit using Poisson regression to the observed numbers of lung cancer in Table 4 of Jones et al. (2007) with weighted cumulative exposure

G.10 Summary of Concentration Exposures and Unit Risk Factors for Texas Rates

The following table shows the risk specific environmental concentrations of arsenic for an extra risk of lung cancer mortality of 1 in 100,000 at age 70 years when the Texas lung cancer background mortality rates and the Texas competing risks given in Appendix A are used. The values for the estimates based on the unweighted cumulative exposures to arsenic were obtained using the standard BEIR IV life-table calculations provided by Sielken & Associates. The values for the estimates based on the weighted cumulative exposures to arsenic were obtained using a modified version of the BEIR IV life-table calculations whereby the weighted (instead of the unweighted) cumulative exposures are used in calculating the extra risk at age 70 years.

Table G-2. URFs and 10-5- Extra Risk Environmental Air Concentrations (Jones et al. 2007) using Texas Lung Cancer Mortality Background Rates and Competing Risks				
Extrapolation assumption for exposures prior to 1972	**β (MLE) URF** **10^{-5} Risk Air Concentration**	**β (95% LCL) URF** **10^{-5} Risk Air Concentration**	**β (95% UCL) URF** **10^{-5} Risk Air Concentration**	**Ratio:** **URF (95% UCL) to URF(MLE)**
Estimates based on **unweighted** cumulative exposure				
Scenario A	2.23E-04 / μg/m^3 4.48E-02 μg/m^3	na[a]	na	na
Scenario B	9.17E-05 / μg/m^3 1.09E-01 μg/m^3	na	na	na
Scenario C	1.47E-04 / μg/m^3 6.81E-02 μg/m^3	na	na	na
Estimates based on **weighted** cumulative exposure [b]				
Scenario A	1.19E-03 / μg/m^3 8.39E-03 μg/m^3	Na [c]	2.45E-03 / μg/m^3 4.08E-03 μg/m^3	2.1
Scenario B	7.04E-04 / μg/m^3 1.42E-02 μg/m^3	na	1.57E-03 / μg/m^3 6.36E-03 μg/m^3	2.2
Scenario C	8.13E-04 / μg/m^3 1.23E-02 μg/m^3	na	1.68E-03 / μg/m^3 5.95E-03 μg/m^3	2.1

[a] The LCL's and UCL's based on the unweighted cumulative exposure to arsenic could not be estimated because the standard error on the estimates of the slope for the multiplicative relative risk model with multiplicative intercept was not available for any of the exposure extrapolation assumptions

[b] Risk air concentrations and URF based on the maximum likelihood estimates and the standard errors on the slopes for the multiplicative relative risk model with multiplicative intercept using the three extrapolation assumptions and weighted cumulative exposure to arsenic from Table 4 of Jones et al. (2007)

[c] The 95% LCLs on the slope were negative for the weighted cumulative exposure to arsenic and the three exposure extrapolation assumptions, suggesting zero risk, and calculations of an air concentration at 1 in 100,000 extra risk was not possible

G.11 Summary of Concentration Exposures and Unit Risk Factors for US Rates

The following table shows the risk specific environmental concentrations of arsenic for an extra risk of lung cancer mortality of 1 in 100,000 at age 70 years when the US lung cancer background mortality rates and the US competing risks given in Appendix A are used. The values for the estimates based on the unweighted cumulative exposures to arsenic were obtained using the standard BEIR IV life-table calculations provided by Sielken & Associates. The values for the estimates based on the weighted cumulative exposures to arsenic were obtained using a modified version of the BEIR IV life-table calculations whereby the weighted (instead of the unweighted) cumulative exposures are used in calculating the extra risk at age 70 years.

Table G-3. URFs and 10^{-5}- Extra Risk Environmental Air Concentrations (Jones et al. 2007) Using US Lung Cancer Mortality Background Rates and Competing Risks

Extrapolation assumption for exposures prior to 1972	β (MLE) URF 10^{-5} Risk Air Concentration	β (95% LCL) URF 10^{-5} Risk Air Concentration	β (95% UCL) URF 10^{-5} Risk Air Concentration	Ratio: URF (95% UCL) to URF(MLE)
Estimates based on **unweighted** cumulative exposure				
Scenario A	2.34E-04 / μg/m^3 4.28E-02 μg/m^3	na [a]	na	na
Scenario B	9.62E-05 / μg/m^3 1.04E-01 μg/m^3	na	na	na
Scenario C	1.54E-04 / μg/m^3 6.50E-02 μg/m^3	na	na	na
Estimates based on **weighted** cumulative exposure [b]				
Scenario A	1.27E-03 / μg/m^3 7.90E-03 μg/m^3	na[c]	2.60E-03 / μg/m^3 3.84E-03 μg/m^3	2.1
Scenario B	7.46E-04 / μg/m^3 1.34E-02 μg/m^3	na	1.67E-03 / μg/m^3 5.99E-03 μg/m^3	2.2
Scenario C	8.62E-04 / μg/m^3 1.16E-02 μg/m^3	na	1.78E-03 / μg/m^3 5.61E-03 μg/m^3	2.1

[a] The LCL's and UCL's based on the unweighted cumulative exposure to arsenic could not be estimated because the standard error on the estimates of the slope for the multiplicative relative risk model with multiplicative intercept was not available for any of the exposure extrapolation assumptions

[b] Risk air concentrations and URF based on the maximum likelihood estimates and the standard errors on the slopes for the multiplicative relative risk model with multiplicative intercept using the three

extrapolation assumptions and weighted cumulative exposure to arsenic from Table 4 of Jones et al. (2007)

[c] The 95% LCLs on the slope were negative for the weighted cumulative exposure to arsenic and the three exposure extrapolation assumptions, suggesting zero risk, and calculations of an air concentration at 1 in 100,000 extra risk was not possible

G.12 References

Binks, K, R Doll, M Gillies, C Holroyd, S R Jones, D McGeoghegan, L Scott, R Wakeford, and P Walker (2005). Mortality experience of male workers at a UK tin smelter. *Occupational Medicine*, **55**:215-226.

Jones, S R, P Atkin, C Holroyd, E Lutman, J Vives i Batlle, R Wakewford, and P Walker (2007). Lung cancer mortality at a UK tin smelter. *Occupational Medicine, **57**:238-245.*