# 25 Weight of Evidence Analysis of Lung Cancer in Colorado Plateau Uranium Miners

*Peter G. Groër*[1], *Carlos A. de B. Pereira*[2], *and Phillip W. Wallace*[1]

[1] Center for Epidemiologic Research; Oak Ridge Associated Universities, PO Box 117, Oak Ridge, Tennessee 37831-0117, USA;
[2] IME, Universidade de Sao Paulo, C.P. 20570-01498, Sao Paulo, Brazil.

This paper describes a Bayesian contingency table analysis using 'weight of evidence' (Good, 1983). From the data on the Colorado Plataeau uranium miners we estimated an average exposure rate, $w$, in WLM per year by dividing the Total Working Level Months (TWLM) for each miner by $y$, the total time spent underground in either hard rock or uranium mines. TWLM is the sum of the WLM accumulated in hard rock mining and in uranium mining. Recent analyses of these data (Whittemore and McMillan, 1983; Hornung and Meinhardt, 1987; BEIR-IV 1988) did not include WLM from hard rock mining in the analysis. We give the posterior expectation of the weight of evidence $E(\omega)$ provided by $w$ and $y$ in favour of death from lung cancer as compared with death from another cause and the posterior variance of $\omega$. $E(\omega)$ and var $(\omega)$ are given for three subcohorts comprised by cigarette smokers, ex-smokers, and non-smokers. First exposure to $^{222}$Rn-daughters occurs, of course, immediately after birth. We did not add WLM from exposure in homes and the environment to TWLM. This non-occupational exposure accumulates at a rate of about 0.2–0.4 WLM per year in the USA. The year exposure rate may be somewhat higher in uranium mining areas. For the anlaysis, it was assumed that TWLM, the random WLM accumulated by each miner during his underground work experience represents his total $^{222}$Rn-daughter exposure.

## *Weight of evidence:*

Good (1983, 1985) describes the interesting and long history, meaning and derivation of 'weight of evidence,' $\omega$. He was also the first one (Good, 1967) to apply $\omega$ to the analysis of lung cancer. Other authors who used this concept are Jeffreys (1939) and Kemeny and Oppenheim (1952). In more elaborate notation, $\omega(H{:}E)$ is the weight of evidence provided by the evidence E in favour of an hypothesis H as compared with $\overline{H}$, the negation of H. Additional background information, B, that might be available can be included in writing: $\omega(H{:}E\ B)$. Subsequently we will drop B for brevity and most of the time we will only use $\omega$. Expressions for $E(\omega)$ and $\text{var}(\omega)$ were derived by Pereira and Pericchi (in press). Before we develop the detail of their estimation procedure, we will first introduce some notation and take an exploratory look at the data.

The indicator variable L indicates the following states for an individual uranium miner: L=0 alive, L=1 died from lung cancer, L=2 died from a cause other than lung cancer. The indicator S indicates non-smoker (S=0), smoker (S=1), and ex-smoker (S=2). All miners considered in the contingency table analysis stopped uranium mining before 1969. Thus, we could be reasonably sure that TWLM for each of these miners included all of their occupational exposures in WLM. WLM values were last updated in 1969 for this data set. Therefore, miners who continued work beyond 1969 accumulated additional. WLM.

In the sequel, we will study the effect of $w$ and $y$ defined earlier, on the relative frequency of lung cancer in the subcohorts of miners defined by S. We chose to compare L=1 with L=2 since the subgroup with L=0 could still die from lung cancer or have a latent and so far undiagnosed lung cancer. There are 253 miners with L =1 and 972 miners with L=2.

The parameters that are estimable from the sample are:

$P_1(w,y) = Pr(w,y \mid L=1)$

$$P_1(w,y) = \Pr(w, y \mid L=1) \text{ and } P_2(w,y) = \Pr(w, y \mid L=2)$$

However, we are interested in:

$$\pi 1(w,y) = \Pr(L=1 \mid w, y, L \neq O) \text{ and}$$

$$\pi 2(w,y) = Pr(L=2 \mid w, y, L \neq O).$$

Since we are only using information on deceased miners, all probabilities are conditioned on $L \neq 0$. Subsequently we omit this conditioning event for brevity.

Before observations of $(w,y)$ the odds in favour of L=1 are defined as:

$$o = \Pr(L=1)/\Pr(L=2)$$

Observation of $(w, y)$ changes the odds to:

$$0=\pi_1/\pi_2$$

To evaluate the ability of the evidence $(w, y)$ to predict L=1, or L=2, we must focus on the change from o to O. Good (1985) proved that the weight of evidence should be an increasing function of r=0/o, the Bayes factor. It is not difficult to prove that $r=p_1/p_2$, i.e. r is the likelihood ratio. Since intuitively we would like weights of evidence to be additive we can, for instance, choose $\omega=\ln r$, see (Good, 1985 for more detail). $\omega$ is the main parameter of interest in this anlaysis.

To estimate $p_1$ and $p_2$ we have to specify prior distributions for these parameters. We chose $p_1 \sim Be(a, b)$ $p_2 \sim Be(a', b')$ with $a=a'=0.1$ and $b=b'=1.5$. Be(a,b) indicates the beta distribution with parameters a and b. We assume $p_1$ and $p_2$ to be independent. Since $p_1$ and $p_2$ depend on $(w,y)$ the parameters a,b,a', and b' depend also on $(w, y)$. Let $f_1=f_1(w, y)$, $f_2=f_1(w, y)$ be the observed frequency of lung cancers and deaths from other causes in the cell defined by $(w, y)$ respectively and let $n_1$, $n_2$ be the total number of lung cancers and deaths from other causes respectively.

*A posteriori*, after observation of $(f_1, n_1)$ and $(f_2, n_2)$ we find:

$$P_1|(f_1,n_1) \sim BE(A,B) \text{ and } p_2|(f_2,n_2) \sim Be(A', B')$$

where $A=a+f_1$, $B=b+n_1-f_1$ and A', B' are defined analogously. Pereira and Pericchi (1988) show that the posterior expectation and variance of r are given by:

$$\begin{aligned} E(r|f, n) &= A(A'+DB'-1) / [(A+B)(A'-1)] \text{ for } A'>1 \\ Var(r|n) &= E(r|f, n)\,[(A+1)(A'+B'-2) / \\ &\quad [(A+B)\,A'-2)]-E(r|f, n)] \text{ for } A'>2, \end{aligned} \qquad (1)$$

where f and n are shorthand for $f_1$, $f_2$ and $n_1$, $f_2$ respectively. $E(\omega)$ and $var(\omega)$ are, of course, functions of $(w, y)$. In the sequel we will, in fact, consider sixteen different classes defined by different pairs of $(w, y)$. We define the class boundaries in terms of $w$ and $y$ and assign indices to the categories as follows:

| | | |
|---|---|---|
| $w$ | $\leq 5$ (WLM/year) | (.,1) |
| 5 | $\leq w < 15$ (WLM/year) | (1.,2) |
| 15 | $\leq w < 50$ (WLM/year) | (.,3) |
| 50 | $\leq w$ (WLM/year) | (.,4) |

Similarly we assign categories (i, .=1,4) to $y \leq 1$, $1 > y \leq 5$, $5 < y \leq 10$ and $y > 10$ respectively. Therefore, class (2,3) corresponds to $1 < y \leq 5$ and

$15 < w \leq 50$. The sixteen different classes are thus defined by the pairs (1,1), . . . , (1,4), . . . , (4,1), . . . , (4,4), respectively.

Pereira and Pericchi (1988) give the following expressions for $E(\omega)$, and var $(\omega)$:

$$E(\omega|f, n) = \Psi(A) - \Psi(A+B) - [\Psi(A') - \Psi(A'+B')]$$

and

$$\text{var}(\omega|f, n) = \Psi'(A) - \Psi'(A+B) + \Psi'(A') - \Psi'(A'+B') \tag{2}$$

where $\Psi$ and $\Psi'$ are the digamma and trigamma functions, defined correspondingly as the first and second derivative of the natural logarithm of the gamma function. To prove equations (1) and (2) two lemmas about beta random quantities are needed. These lemmas are stated without proof in the Appendix.

## *Results*

The distribution of lung cancer deaths and deaths from other causes over the sixteen categories $(y_i, w_j)$ in non-smokers, smokers, ex -smokers, and all miners is shown in Tables 25.1–25.8. Application of the formulas for the posterior expectation and variance of $\omega$ given in the previous section yields the results shown in Tables 25.9–25.12. Each entry represents the posterior expectation for $\omega$ for the cell characterized by $y_i$ and $w_j$. The parenthesized entry is the posterior standard deviation for the corresponding $E(\omega)$.

**Table 25.1 Deaths from lung cancer in non-smokers (S=0, L=1).**

| *Y/W* | *1* | *2* | *3* | *4* | *Total* |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 5 | 5 |
| 4 | 0 | 0 | 6 | 7 | 13 |
| Total | 0 | 0 | 6 | 12 | 18 |

**Table 25.2 Deaths from other causes in non-smokers (S=0, L=1).**

| *Y/W* | *1* | *2* | *3* | *4* | *Total* |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 3 | 13 | 17 |
| 2 | 1 | 3 | 7 | 34 | 45 |
| 3 | 0 | 1 | 7 | 17 | 25 |
| 4 | 0 | 14 | 18 | 29 | 61 |
| Total | 1 | 19 | 35 | 93 | 148 |

**Table 25.3 Deaths from lung cancer in smokers (S=1, L=1).**

| *Y/W* | *1* | *2* | *3* | *4* | *Total* |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 3 | 4 |
| 2 | 0 | 1 | 0 | 22 | 23 |
| 3 | 0 | 1 | 2 | 41 | 44 |
| 4 | 5 | 8 | 31 | 91 | 135 |
| Total | 5 | 10 | 34 | 157 | 206 |

**Table 25.4 Deaths from other causes in smokers (S=1, L=2).**

| *Y/W* | *1* | *2* | *3* | *4* | *Total* |
|---|---|---|---|---|---|
| 1 | 0 | 4 | 11 | 31 | 46 |
| 2 | 4 | 13 | 22 | 162 | 201 |
| 3 | 2 | 16 | 32 | 125 | 175 |
| 4 | 12 | 67 | 106 | 147 | 332 |
| Total | 18 | 100 | 171 | 465 | 754 |

**Table 25.5 Deaths from lung cancer in ex-smokers (S=2, L=1).**

| *Y/W* | *1* | *2* | *3* | *4* | *Total* |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1 | 1 |
| 2 | 0 | 0 | 0 | 5 | 5 |
| 3 | 0 | 0 | 0 | 3 | 3 |
| 4 | 0 | 1 | 5 | 14 | 20 |
| Total | 0 | 1 | 5 | 23 | 29 |

**Table 25.6 Deaths from other causes in ex-smokers (S=2, L=2).**

| *Y/W* | *1* | *2* | *3* | *4* | *Total* |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 1 | 2 |
| 2 | 0 | 0 | 1 | 9 | 10 |
| 3 | 0 | 0 | 3 | 18 | 21 |
| 4 | 1 | 1 | 22 | 13 | 37 |
| Total | 1 | 1 | 27 | 41 | 70 |

**Table 25.7 Deaths from Lung cancer in all miners (L=1).**

| *Y/W* | *1* | *2* | *3* | *4* | *Total* |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 4 | 5 |
| 2 | 0 | 1 | 0 | 27 | 28 |
| 3 | 0 | 1 | 2 | 49 | 52 |
| 4 | 5 | 9 | 42 | 112 | 168 |
| Total | 5 | 11 | 45 | 192 | 253 |

**Table 25.8 Deaths from other causes in all miners (L=2).**

| *Y/W* | *1* | *2* | *3* | *4* | *Total* |
|---|---|---|---|---|---|
| 1 | 0 | 5 | 15 | 45 | 65 |
| 2 | 5 | 16 | 30 | 205 | 256 |
| 3 | 2 | 17 | 42 | 160 | 221 |
| 4 | 13 | 82 | 146 | 189 | 430 |
| Total | 20 | 120 | 233 | 599 | 972 |

**Table 25.9 Weight of evidence for lung cancer in non-smokers: posterior expectation and (Posterior standard deviation).**

| *Y/W* | *1* | *2* | *3* | *4* | *Marginal* |
|---|---|---|---|---|---|
| 1 | 2.0548 | −7.9452 | −9.3305 | −10.9029 | −3.3340 |
| | (14.2411) | (10.1394) | (10.0873) | (10.0724) | (2.6773) |
| 2 | −7.9519 | −9.3305 | −10.2570 | −11.8835 | −4.3110 |
| | (10.1394) | (10.0873) | (10.0760) | (10.0700) | (2.6905) |
| 3 | 2.0548 | −7.9452 | −10.2570 | 0.7733 | 0.4308 |
| | (14.2411) | (10.1394) | (10.0760) | (0.4666) | (0.4295) |
| 4 | 2.0548 | −10.9792 | 0.9109 | 0.5894 | 0.5031 |
| | (14.2411) | (10.0721) | (0.4194) | (0.3567) | (0.1867) |
| Marginal | 0.4452 | −3.4459 | 0.2784 | 0.0001 | |
| | (2.8708) | (2.6961) | (0.3724) | (0.1889) | |

**Table 25.10 Weight of evidence for lung cancer in smokers: posterior expectation and (Posterior standard deviation).**

| *Y/W* | *1* | *2* | *3* | *4* | *Marginal* |
|---|---|---|---|---|---|
| 1 | 1.2936 | −10.4142 | −1.4913 | −1.1659 | −1.1691 |
| | (14.2429) | (10.0848) | (1.2335) | (0.6377) | (0.5203) |
| 2 | −10.4142 | −1.6640 | −12.2029 | −0.7187 | −0.8779 |
| | (10.0848) | (1.2274) | (10.0734) | (0.2153) | (0.2061) |
| 3 | −9.6154 | −1.8775 | −1.6742 | 0.1723 | −0.0886 |
| | (10.1012) | (1.2212) | (0.7952) | (0.1628) | (0.1495) |
| 4 | 0.3703 | −0.8762 | 0.0550 | 0.8124 | 0.3934 |
| | (0.5446) | (0.3745) | (0.1897) | (0.1082) | (0.0654) |
| Marginal | −0.0003 | −1.0176 | −0.3240 | 0.2874 | |
| | (0.5031) | (0.3237) | (0.1709) | (0.0487) | |

**Table 25.11 Weight of evidence for lung cancer in ex-smokers: posterior expectation and (Posterior standard deviation).**

| *Y/W* | *1* | *2* | *3* | *4* | *Marginal* |
|---|---|---|---|---|---|
| 1 | 0.8595 | 0.8615 | −9.1405 | 0.8595 | 0.1472 |
| | (14.2415) | (14.2415) | (10.1400) | (1.6791) | (1.2220) |
| 2 | 0.8595 | 0.8595 | −9.1405 | 0.2352 | 0.1575 |
| | (14.2415) | (14.2415) | (10.1400) | (0.5342) | (0.5070) |
| 3 | 0.8595 | 0.8595 | −10.5258 | −1.0470 | −1.1108 |
| | (14.2415) | (14.2415) | (10.0879) | (0.6242) | (0.5849) |
| 4 | −9.1405 | 0.8595 | −0.6852 | 0.9359 | 0.2421 |
| | (10.1400) | (1.6791) | (0.4643) | (0.3249) | (0.1734) |
| Marginal | −1.6405 | 0.8595 | −0.8417 | 0.2796 | |
| | (2.8729) | (1.4154) | (0.4396) | (0.1443) | |

Table 25.12 **Weight of evidence for lung cancer in all miners: posterior expectation and (Posterior standard deviation).**

| *Y/W* | *1* | *2* | *3* | *4* | *Marginal* |
|---|---|---|---|---|---|
| 1 | 1.3428 | −10.6090 | −1.7622 | −1.1709 | −1.2391 |
| | (14.2429) | (10.0819) | (1.2234) | (0.5416) | (0.4624) |
| 2 | −10.6090 | −1.8284 | −12.4688 | −0.6973 | −0.8734 |
| | (10.0819) | (1.2216) | (10.0728) | (0.1934) | (0.1865) |
| 3 | −9.5663 | −1.8906 | −1.9000 | 0.1537 | −0.1056 |
| | (10.1012) | (1.2200) | (0.7911) | (0.1476) | (0.1372) |
| 4 | 0.3368 | −0.9068 | 0.9000 | 0.8180 | 0.4026 |
| | (0.5392) | (0.3511) | (0.1606) | (0.0959) | (0.0576) |
| Marginal | −0.0571 | −1.0548 | −0.3034 | 0.2046 | |
| | (0.4987) | (0.3083) | (0.1470) | (0.0438) | |

Inspection of Tables 25.9–25.12 shows that positive weights of evidence with a coefficient of variation less than one occur only in the lower right-hand corner of these Tables. This indicates that only large exposure rates and long work periods underground increase the final odds for lung cancer as compared with other causes of deaths in uranium miners. The positive weights of evidence in the upper left-hand corner of the Table reflect the fact that there are zero observations in these cells and that the sample size is larger for deaths from other causes. Since zero observations out of a smaller total of lung cancer deaths is weaker evidence than zero deaths from other from other causes out of a greater total of deaths from other causes, the expectation of the weight of evidence is positive. Note, however, the large standard deviation which indicates the uncertainty about $\omega$.

## *Critical comments:*

1. TWLM, the total WLM accumulated by each miner, is a rough estimate of the total exposure accumulated by each miner. Uncertainty about TWLM and, therefore, also about $w$ causes classification errors. The analysis given above assumes that TWLM and, therefore, $w$ are precisely known.
2. The Colorado Plateau Uranium miners were obviously also exposed to external and interal $\gamma$ radiation. External $\gamma$ doses to the lung are not available for these miners. Internal $\gamma$ doses from inhaled $^{222}$Rn-daughters could be estimated, but these doses are certainly negligible compared with the $\alpha$-doses. In the weight of evidence analysis presented here effects of external $\gamma$ radiation are 'hidden' and implicitly attributed to TWLM. The same is true for other exposures, to exhaust fumes or ore dust for example.
3. If the cohort is not subdivided into three subcohorts (smokers, non

-smokers and ex-smokers), it is possible to estimate the weights of evidence for smoking and Rn-daughter exposure separately. In the same spirit Good (1967) separated weights of evidence for lung cancer provided by different factors like, for example urbanity and pipe and cigar smoking. Results given in Pereira and Pericchi (1988) can be applied to estimate weights of evidence for more than one source of evidence.

4. Weights of evidence (e.g. smoking, TWLM) can be ordered using a divergence measure as shown by Pereira and Pericchi (1988). This would shed light on the question whether it is more effective to stop smoking or reduce the Rn-daughter concentrations in mines to lower the odds on lung cancer.
5. The analysis presented above considers only data from *deceased* miners ($L \neq O$)! Roughly speaking a positive weight of evidence results where $P_1 > P_2$. Then $r > 1$ and $w = \ln r > O$. This happens only for large $\omega$ and $y$ as one would expect since many different causes of death make up the category defined by $L=2$. To investigate the question of an increased lung cancer risk data for 'unexposed' miners are needed. Tables analogous to Tables 25.1 25.3 25.5, and 25.7 have to be constructed. With such Tables, $\omega$ can be used to investigate to which extent $(w, y)$ tends to cause lung cancer (see Good, 1961.)

## *Acknowledgments*

This report concerns work done under Contract No. DE-AC05-760R00033 between the Department of Energy, Office of Energy Research, and Oak Ridge Associated Universities. Additional support for this research was provided by the mining Safety and Health Administration under Interagency Agreement No. 0881-0881-A1 with the United States Department of Energy. Robert Roscoe and Richard Hornung of the National Institute for Occupational Safety and Health provided tapes with the data on the Colorado Plateau uranium miners and answered patiently numerous questions about the data during the course of this investigation. Messrs Goodwin and Sexauer of MSHA encouraged this study from start to finish.

## *References*

BEIR (Committee on the Biological Effects of Ionizing Radiation) (1988) *Health Risks of Radon and Other Internally Deposited Alpha-Emitters*, (BEIR-IV Report), Washington, DC, National Academy Press.

Good, I. J. (1961) A causal calculus. *British Journal for the Philosophy of Science*, **11**, 305–318; **12**, 43–51; **13**, (1962), 88.