BAYES THEOREM and RADON RISK

W. E. Hobbs, Ph.D.
Radon Reduction & Research
Santa Barbara, CA

INTRODUCTION

The health risk of carcinogens in the environment is managed through the application of a single parameter, the unit risk. This is effectively the probability that an average person will contract cancer and die as the result of a seventy-year exposure to a given substance at a constant concentration of one unit. The unit of exposure depends on the specific carcinogen. Generally, the unit of concentration for an organic substance is $\mu g/m^3$. Since radioactive substances are found in such minuscule mass concentrations, their concentrations are given in decay rate per unit volume; common units are either the currie per liter or the bequerrel per cubic meter (appropriately named for the two discoverers of radioactivity).

It is assumed that a period of exposure will result in an absorbed dose. For an organic substance, the risk of cancer death is assumed proportional to the mass of the substance absorbed. While for a radioactivity, the risk is proportional to the amount of nuclear energy absorbed. For radiation the amount of energy absorbed (the dose) may be calculated. An example of a unit for absorbed dose is the rad(tissue) or, for radon, the working-level month (WLM). Exposure to a substance at a given concentration will result in a dose rate, the time derivative of absorption. This rate depends on various variables such as body size and activity level.

Summarizing mathematically, if D is the cumulative absorbed dose and E(t) is the time-dependent exposure, then we write

$$D = k \int_0^\tau E(t)dt$$

where we use k for an average man (it might be better to use k as either for a specific individual or even specific organ, e.g., the lungs). Using the mean-value of exposure <E>,

$$D = k \langle E \rangle T.$$

Since the cancer risk C is assumed proportional to dose, we write

$$C = \lambda D = \lambda k \langle E \rangle T.$$

The unit risk factor is the product λkT where T may be defined as 70 years, a lifetime.

There are two averages: over the susceptibility of different individuals (for k) and over the instantaneous ambient concentration of the offending substance (<E>). This model highlights some assumptions which are made for carcinogens. The model says that risk does not go away.

Once a person has accumulated a given dose, he has a risk for getting cancer which remains even if he receives no further dose. Other effects are believed to fade away with time. The model is a simplification because there may be rate-specific or age-specific effects. An example of a time effect is the latency period for cancer. In studies of cancer the dose received during the final 5-10 years before death may not be included in analyzing the risk. Another assumption is that risk has no threshold. Any dose carries some risk, an ultimate limit would be the risk associated with the absorption of a single organic molecule or a single pulse of energy from the decay of a radioactive atom. This is an unnecessary abstraction since for radon-related exposures, we all have millions of radioactive atoms decay in our lungs during our lives. This leads to the concept of an "acceptable risk." A common acceptable risk is one-in-a-million for cancers from man-made environmental pollutants. Since the background death rate due to cancer exceeds 20%, this is a negligible risk. The extrapolation to low risk levels may involve large uncertainties.

In this paper I will review a technique for combining studies of cancer risk. This will be applied to the miner studies used in BEIR IV and the subsequent study of residential radon exposure in Sweden.

MATHEMATICAL METHODS

The excess relative risk (y=ERR) is the additional risk of cancer death caused by some specific agent. It assumes that a given population has a basic cancer rate which is a fundamental characteristic of that population without any of the given agent. The ERR is the number of cancer deaths divided by the zero-dose number minus one ($D/D_o - 1$). It is dimensionless. (The relative risk does not subtract one.) I focus on radon and its associated radiation dose as measured in units of the working-level month (x=dose in WLM). The WLM is a measure of the potential absorption of alpha-particles energy from radon decay products (RDP). It is 1.5×10^5 MeV/L of alpha energy from the two polonium decay products and corresponds to a average lung dose from RDP in secular equilibrium with radon at a concentration 100 pCi/L for a period of 170 hours. This unit has technically nothing to do with the radon itself, but the RDP in air, of course, would not occur except for the inert radon emitted from the soil. The WLM reflects essentially all the nuclear energy absorbed in the lungs.

In this discussion, the dose x corresponds to the product k<E> above and y to the excess relative lung cancer risk. The model is
$$y = \kappa x$$
where κ is the true unit risk with units of probability (percent) per WLM. The parameter κ is for the human population and is fundamentally unknowable, but the objective is to find an estimate with some confidence (uncertainty). The estimate is designated k and its uncertainty σ. If studies are performed and find a value for the y_i=ERR(%) for an average dose x_i=WLM, then the estimated value for κ will be $k_i=y_i/x_i$ with the individual points having some characteristic spread σ_i about the best estimate.

If the unknown factors and errors contribute to the uncertainty in an additive fashion, then our distribution is written as a gaussian

$$F(x,y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[\frac{-(y/x-k)^2}{2\sigma^2}\right]$$

Inside $\pm\sigma$ of k approximately 68% of the data (y_i/x_i) will fall. Here exposure and probability of cancer may refer to either the individuals or groups within the study. A group may refer to a subset of individuals that have approximately the same exposure, a cohort.

Consider the situation where the research begins with some existing prior knowledge of the relationship between exposure and cancer risk. This might occur because of prior studies with animals or prior epidemiology studies. In other words, we already have an estimate for the unit risk k and its uncertainty σ. An epidemiology study is assumed to be an independent simple random sample from the human population. A simple way to combine the analyses is Bayes Theorem. We have a prior probability distribution $f(k,\sigma)$, the results of some additional analysis (x_i,y_i), and we wish to calculate a new and better posterior distribution using all this information. The posterior distribution for k and σ may be written

$$f[(k,\sigma)\,|\,(x,y)] = \frac{F[(x,y)\,|\,(k,\sigma)]f(k,\sigma)}{\int F[(x,y)\,|\,(k,\sigma)]f(k,\sigma)\,dk\,d\sigma}$$

The right-hand side is the product of the prior estimate for the unit risk f, and the probability of getting current results given that these prior estimate is valid F. The denominator is simply a constant which normalizes the distribution (makes the total probability distribution equal to one). The left-hand side is simply the posterior best estimate for the unit risk distribution given the results of the current experiment. This can be recognized as the differential form of Bayes Theorem as presented in elementary probability and statistics courses.

If we have no idea of what k is, then f is taken as a constant function (all values of k are equally probable). In this situation, the best fit posterior k will be given by the results of the experiment alone, for example by least squares. Of course, the benefit of Bayes theorem occurs when we have some reason to suspect a prior distribution for k (i.e., not constant).

APPLICATIONS, MINERS & RESIDENTIAL

The results of two calculations are presented in this paper. First, the four studies of uranium miners from the BEIR IV study[1] will be combined. Second, this analysis will be extended to the residential cancer risk reported from Sweden[2].

When there are several studies to combine, each of the results is assumed to be jointly independent. Then the distribution of combined data, given k and σ, is simply written as the product of the individual probability functions:

$$\prod_{1\le i\le n}F(x_i,y_i\,|\,k,\sigma) = \frac{1}{\sigma^n(2\pi)^{n/2}}\exp\left[-\sum\frac{(y_i/x_i-k)^2}{2\sigma^2}\right]$$

As you can see from this expression, for a constant prior distribution (complete ignorance) the probability distribution function will be dominated by values that minimize the sum of the squares, classical least squares. Four studies of uranium miners were part of the BEIR IV review conducted in 1985-6. All these studies show statistical significance at the 95% confidence level; i.e., 95% probability that k>0. (The number of cancer deaths in the defined samples has continued to increase as the study continues to this day.)

The unit risk for the four BEIR IV studies varried considerably, from 0.42 to 2.21 %/WLM. As noted above the most probable slope is equal to the estimate found by weighting the results according to the inverse of their variances. The result is k_{max}=0.73 %/WLM. The distributions are shown in the Figure 1.

The variance in the posteror distribution is small, smaller than any of the individual distributions. This is mathematically correct as one can see by considering two simple distributions. Suppose they were both broad (large variance), but only intersected in a small region of parameter space. The Bayes combination of these distributions would only exist in the region of intersection and, thus, would have small variance. In this case the geometric standard diviation of the studies varies between 26% and 69%, but the combination had a geometric stanadard diviation of 25%. While mathematically correct, this indicates a possible shortcoming. Our approach is to treat each of these studies as a simple random sample of the human population. The miners have distinct characteristics, mainly smoking and breathing deeply, which limit the validity of this generalization. Here it appears that the various miner cohorts are also distinct from one another. The worldwide population of uranium miners is not homogeneous with respect to susceptibility for lung cancer. The variance from the BEIR IV analysis is considerably larger than that shown here. BEIR IV gives a conservative unit risk of 1%/WLM with a factor of 2 uncertainty for the 95% confidence interval. This corresponds to a geometric standard diviation of about 40%., approximately the same as the Swedish study shown.

Even though the calculated variance in the unit risk is too small, this BEIR IV distribution is used as the prior distribution and it is combined with the results of the Swedish residential radon risk study. This study involved a large population (1281 deaths) and much longer period of exposure (1950-70). So, even though the exposure levels were smaller than the uranium mines, the cumulative dose was still significant. (The average dose in this study exceeded the average dose for the dead miners at the Beaverlodge, Canada.) There is a large amount of scatter in the data and this is reflected in the large variance; the Swedish study was also statistically significant at the 95% confidence level, 0.01% < k < 0.22%. The results are shown in Figure 2. Because of this, the experiment results are rather flat (weak information) and the prior and posterior distributions are approximately the same. The posterior has a k_{max}=0.71 %/WLM, even though the data had k_{max}=0.56 %/WLM The small scatter in the prior means that we have more confidence in this distribution: "We believe it more."

DISCUSSION

The linear model of carcinogenesis is very appealing because of its simplicity. The physical concept of radon and cancer envisions a single alpha-particle doing the damage to initiate the cancer.

Clearly, at the high exposures there is an effect. At doses above 20 WLM, there is almost always a statistically significant increase in the number of cancer deaths. The relation between cancer and exposure is not statistically significant below some exposure level. Of the eleven studies[3] of miners, radon and cancer, seven have a lowest data point which shows a benefit from radon exposure (the ERR is less than zero, fewer cancers than expected for no radon exposure). The possibility of a significant threshold for radon exposure inducing cancer cannot be eliminated by the various studies. Maybe radon, at low doses does play some beneficial role in our health.
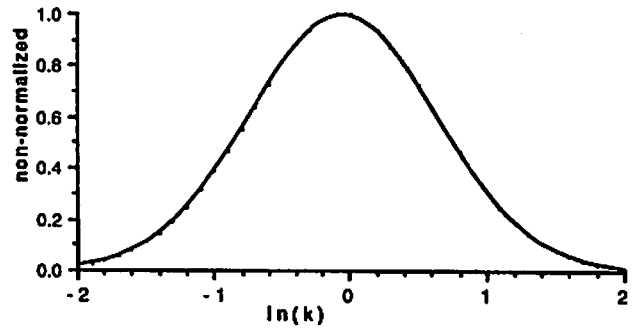
REFERENCES

1,   Jacob I. Fabrikant & others, "Health Risks of Radon and Other Internally Deposited Alpha-Emitters BEIR IV," Committee on the Biological Effects of Ionizing Radiation, National Research Council, National Academy Press, Washington, DC, 1988.

2.   Gögan Pershagen & others, "Residential Radon Exposure and Lung Cancer in Sweden," The New England Journal of Medicine, Vol. 330, No. 3, p. 159.

3.   Jay H. Jubin & others, "Radon and Lung Cancer Risk: A Joint Analysis of 11 Underground Miners Studies," National Institutes of Healt, National Cancer Institute, NIH Publication No, 94-3644, January 1994.
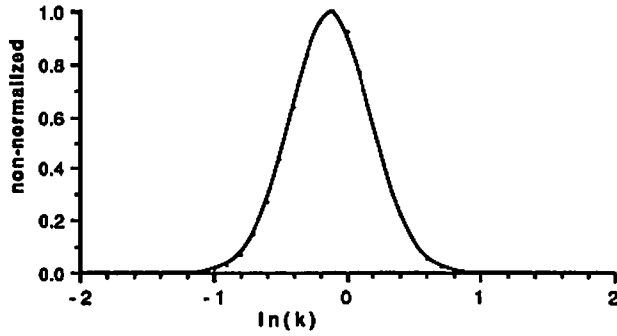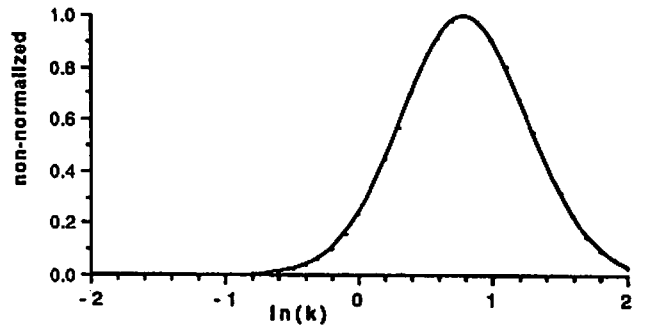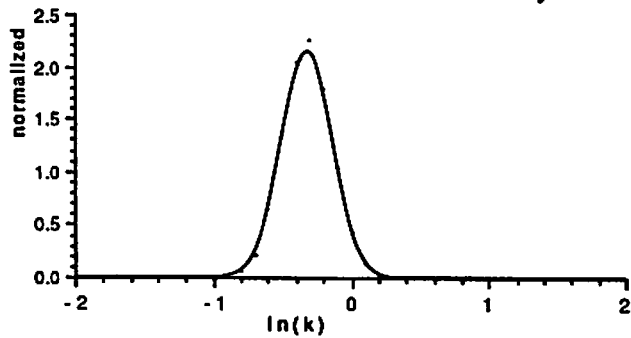
Data from Colorado
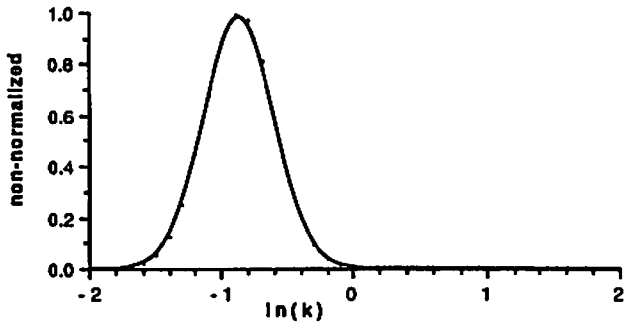
Data from Sweden

Data from Ontario

Data from Beaverlodge

Data from BEIR IV - all combined w/ Bayes

Data from Bayes

Figure showing probability density vs ln(k), with curves for posterior (w/ dots), BEIR IV prior (line), and residential data.