Accounting for Bias and Measurement Error in Occupational Studies[1]

George Ostrouchov
Computational Mathematics and Statistics Section, Computer Science and Mathematics Division,
Oak Ridge National Laboratory

This work is motivated by the need to adjust for dose bias and uncertainty in epidemiologic dose-response analyses. Typically, epidemiological studies of the effects of external penetrating radiation on worker health have relied on recorded annual doses to the individuals in the population. At Oak Ridge, these annual doses were obtained by adding up recorded weekly readings. In statistical analyses, these dose values have been treated as though they are known exactly, although everyone recognizes that there is uncertainty due to measurement error and bias. It is usually assumed that the measurement errors "average out" and that the bias is small. A recent study of Oak Ridge workers [1] used a preliminary dose adjustment procedure and found an upward bias in dose-response coefficients and likelihood ratio test statistics. This analysis was based on a crude adjustment for missing dose and did not consider measurement and other dosimetry errors.

Although our goal is to account for bias and uncertainty in occupational risk estimates for ionizing radiation, we find that the necessary first step is an adjustment for bias and quantification of uncertainty in dose estimates. So far, this is where most of our effort has concentrated [2, 3]. We describe our results in radiation dose estimation and comment on how the bias corrected dose estimates that include quantification of uncertainty can be used in risk estimation.

Among occupational studies based on historical data, occupational radiation risk estimation is relatively "data rich." However, the data was collected for compliance rather than individual dose estimation. Consequently the bias can be substantial. Studies have shown that there was a systematic underestimation of doses for ORNL workers from 1945 to 1955 [2, 3]. The first study [2] concentrated on dose estimation from film badge data and the second study [3] provided a systematic way of combining pocket meter data with film badge data for a better dose estimate. The results show that both bias and uncertainty vary widely between individuals and are poorly correlated with recorded annual dose. This suggests that the additional information contained in daily and weekly dosimetry records is needed for effective bias adjustment and quantification of uncertainty.

The dose estimate proposed for each individual is a probability distribution. This is the most general description of uncertainty and can be reduced to other descriptions of uncertainty. A non-parametric probability distribution estimate, consisting of many (say 100) density points, can be reduced to a more concise description such as the five points of a boxplot, or to a few parameters of an assumed parametric distribution (such as a normal or a lognormal distribution). Each reduction is a loss of information and a gain in simplicity. These can be computed for an individual or for any cohort of individuals. Such generality allows the dose estimates to be useful for many purposes, including adjustment for dose uncertainty in epidemiologic dose-response analyses by methods yet to be developed.

---

[1] An invited presentation at the ASA Confrence on Radiation and Health, San Diego, June 14-17, 1998, to appear in the January 1999 issue of *Radiation Research*. Research was supported in part by the Office of Epidemiology and Health Surveillance, Environment, Safety and Health, U. S. Department of Energy under contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp., and by grant R01 OH12956 from the National Institute for Occupational Safety and Health of the Centers for Disease Control and Prevention.

Our methodology is based on Bayesian estimation of "true dose" from available dose measurements in the form of a probability distribution. Bayesian dose distributions from individual measurements are combined with convolution computations to obtain dose distribution estimates for longer periods.

The Bayesian statistical approach estimates the unobserved quantities (true doses) given the values of the observed ones (recorded doses). A relationship between the true dose and the recorded dose in the form of a conditional probability distribution is the key element of the method. We begin by defining $P(x)$ to be the true dose distribution that concerns one individual in one measurement period. The key component for implementing our approach is the conditional probability distribution $P(z|x)$. In effect, $P(z|x)$ is the answer to the question: "If the true dose is x, what is the probability that the recorded value is z?" This is determined by careful consideration of the properties of the measuring device (in this case the film badge or the pocket meter and the system used in reading and recording its dose). A necessary component is some information on the calibration error of the measuring device as well as recording practice. For the ORNL data, we assume a lognormal calibration error whose parameters are estimated from historical information. The rounding and censoring practices as well as use practices known from historical ORNL documents are included in the model of $P(z|x)$.

Note that $P(z|x)$ is a function of two variables, namely x and z, and it is constructed by specifying a distribution on z for each possible (fixed) value of x. After specifying $P(z|x)$ for all possible z and x, it is used as a function of x for each observed z. This is the "likelihood" of x for the observed z and is denoted by $L(x|z)$. Bayes's theorem then combines the likelihood $L(x|z)$ with prior distribution $P(x)$ to get $P(x|z)$, the posterior distribution of the true dose x given the recorded measurement z.

After obtaining the posterior dose distributions for each measurement period (in our case, a day or a week), the distributions must be "added" to compute a cumulative dose for a longer period. We compute yearly cumulative dose distributions, but other periods (such as a quarter to correspond to measurement periods in later years) may be used. Conditional on the recorded doses, the posterior distributions are independent and their convolutions can be efficiently computed with the discrete Fourier transform.

Because posterior distributions for individual measurements are often not symmetric about the recorded measurement (in the ORNL case, especially the zero recorded doses), the cumulative distribution uncovers the bias in the added recorded doses.

When dose distributions for individuals are available, how can they be used in dose risk estimation? If we only wish to correct for bias in recorded doses, we can use the medians of the dose distributions and proceed with traditional dose-response analyses. Including dose uncertainty is more difficult. A Monte Carlo solution is possible but is also very computationally intensive. Other approaches that extend the dose estimation Bayesian methodology into risk estimation can probably be developed. For example, some simple parametric assumptions about the dose distributions may carry through some known dose-response estimation methods. These may also require substantial computation, but less than the Monte Carlo solution.

A Monte Carlo simulation is more complex than it at first appears. The collection of risk estimates computed from samples generated by the individual dose distributions only accounts for the uncertainty in the data. We also need to account for error in the dose-response model. In traditional dose-estimation, the model error is usually expressed by a confidence interval. The

Monte Carlo simulation must also account for the variability expressed by these intervals.

The Bayesian methodology developed in [2, 3] can be used to quantify bias and measurement error in any situation, where it is possible to build a data generation model. In particular, the application of this methodology to dosimetry data at other facilities and for other time periods may require relatively small modifications to the data generation models. The distribution convolution methodology remains the same.

References

[1] E. L. Frome, D. L. Cragle, J. Watkins, S. Wing, C. Shy, W. Tankersley, and C. M. West. A mortality study of employees of the nuclear industry in Oak Ridge, Tennessee. *Radiat. Res.*, 148:64–80, 1997. (also ORNL-6785 Technical Report).

[2] Toby J. Mitchell, George Ostrouchov, Edward L. Frome, and George D. Kerr. A method for estimating occupational radiation dose to individuals, using weekly dosimetry data. *Radiation Research*, 147:195–207, 1997.

[3] George Ostrouchov, Edward L. Frome, and George D. Kerr. Dose estimation from daily and weekly dosimetry data. Technical Report ORNL-????, ORNL, Oak Ridge, TN 37830, 1998.