Uncertainty Analysis of Multiple Epidemiological Studies Using Frequency Distributions of Relative Risks

Alexander Shlyakhter

Department of Physics and Harvard Center for Risk Analysis
Harvard University, Cambridge, Ma 02138 USA

Abstract

A new format for presenting uncertainty in the results of multiple epidemiologic studies of the same outcome is suggested. A set of 95% confidence intervals for relative risk, *RR*, is transformed to a frequency distribution of the normalized deviations, *ln(RR)/SE(ln(RR))*, from the null value *ln(RR)*=0 (*RR*=1). I assume that deviations from *RR*=1 are due to unaccounted residual biases and compare the distribution of these deviations with the distributions of the actual errors in physical measurements where the true values have subsequently become known, and the incidence of large errors can be estimated. Comparison of these distributions can, by analogy, help to understand how convincing is the evidence of elevated risk in observational studies.

1 Introduction

Epidemiological results are often presented in terms of a Risk Ratio (*RR*). The uncertainty in sampling makes *RR* uncertain, and for large samples, the estimate of *ln(RR)* follows a normal (Gaussian) distribution [1]. Uncertainty in the results of epidemiologic studies is commonly reported as 95 percent confidence intervals (95% CI) for the relative risk, *RR*, which represent uncertainty in the value of *RR*. These 95% CI account only for one component of uncertainty, namely the random errors caused by the finite number of subjects. Non-statistical errors in observational studies become steadily more important as epidemiologists deal with smaller risk ratios. Although investigators always try to minimize the uncertainties caused by various possible biases (coming from such sources as selection, differential misclassification, and confounding), they cannot quantify the effect of the residual biases on uncertainty in the results. Such biases can be viewed as analogues of systematic uncertainties in physical measurements.

Uncertainties associated with random errors in physical experiments are usually controlled by increasing the number of independent measurements until random errors become comparable to the estimated systematic errors. Random and systematic uncertainties are then combined into a "combined standard uncertainty" which serves as the basis for calculating intervals corresponding to the required level of confidence. Even so, the history of physical measurements demonstrates a strong tendency for researchers to underestimate uncertainties in their results (see refs. [2,3] and section 2 below).

On the other hand, in observational studies, sources of bias and their probable effect on the value of *RR* are discussed only qualitatively. The relative magnitude of random and systematic errors can differ from study to study; one can expect that systematic errors are more important for large studies when 95% CI are narrow.

In this paper, I propose a new graphical procedure for presenting the results of epidemiologic studies which may help in evaluating the strength of evidence provided by multiple epidemiologic studies of the same outcome.

0-8186-7126-2/95 $4.00 © 1995 IEEE
Proceedings of ISUMA-NAFIPS '95

The quality of data available to researchers in observational studies is generally lower than the quality of data used in experimental science [4]. The frequent occurrence of contradictory results in case-control studies of the same outcome [5] suggests that residual biases (such as differential misclassification and confounding) in observational studies may be even more widespread than are the unsuspected systematic errors in physical measurements. I assume that the true risk is not elevated and consider the observed *RR* values as deviations from this assumed "true" value. I then use an analogy and compare the distribution of *RR* values with the distribution of the *actual* deviations from the true values in physical measurements. If the two distributions are similar, one can argue that the reported elevated values of risk can be attributed to some residual biases.

2 Distribution of errors in physical measurements

The long record of measurements of elementary particle properties has prompted several early studies of the temporal evolution of errors. Shlyakhter *et al.* [2,3] expanded these original studies by following trends in several data sets. A convenient measure of the deviation of "old" values from the "true" values is $x=(A_{new} - A_{old})/\Delta_{old}$, with A_{new} the new ("true") value, A_{old} the previously measured value, and Δ_{old} the old standard error.

In Figure 1, I present the results of the analysis of the data set of masses and lifetimes of elementary particles [6] and the data set of neutron scattering lengths [7]. Cumulative frequency distributions of x for each dataset are shown together with the cumulative normal (Gaussian) curve. On a logarithmic scale, this cumulative curve may look unfamiliar to the readers accustomed to the bell-shaped probability density curve of the normal distribution. To verify that this is the same distribution one can compare the upper percentiles with the values tabulated for the normal distribution. For example, the upper 97.5 percentile corresponds to $x=1.96$ as it should. Gaussian curve obviously underestimates the probability of large deviations: instead of the 2.5 percent predicted by the normal distribution, there is a 15 to 30 percent chance of $x>2$ for the empirical probability distributions. These distributions also suggest that there is a 1 percent chance of $x>5$, while the normal distribution predicts the value $3 \cdot 10^{-7}$, about 30 thousand times less. A better fit to the data at large values of x is obtained with a compound parametrization which has one additional parameter, u, the relative uncertainty of the old standard error [2,3]. For $u=0$, compound distribution is reduced to the cumulative normal distribution; for $u=1$, it gives an exponential distribution which is close to a straight line on the semi-logarithmic graph of the cumulative probability, $S(x)$, *vs.* the number of standard deviations, x.

3 Distribution of the results of multiple studies

I suggest the use of a similar format to consider the significance of multiple epidemiological results of the same outcome. Instead of the distribution of actual errors I consider frequency distribution of the normalized risk ratios, $x=ln(RR)/SE(ln(RR))$. In the particular examples that I present below, the question that is being posed by the epidemiologists is whether or not there is an association between an environmental cause and a health effect. Alternatively stated, is the true *RR* value different from unity? In order to answer this question graphically, I plot the cumulative probability of the deviation of $x=ln(RR)/SE(ln(RR))$ from zero.

For example, in the study of the risk of lung cancer among nonsmoking U.S. women married to smokers [8] (shown as study #6 in Figure 2) *RR* of 1.32 was reported. The 95% CI reported was 1.03-1.68. The 95% CI of $ln(RR)$ in that study is 0.030-0.519. $SE(ln(RR))=(0.519-0.030)/2/1.96=0.125$. The middle of the confidence interval, $ln(RR)=0.274$, is $x=0.274/0.125=2.20$ standard deviations away from the postulated true value, $ln(RR)=0.0$.

Results of 31 studies of lung cancer among females exposed to environmental tobacco smoke (ETS) US EPA [9] are presented in Figure 2 (see references in Table 2 in the recent analysis of the same studies by Gross [10]). Each study reported the relative risk for lung cancer in nonsmoking women married to smokers *vs.* nonsmokers married to other nonsmokers. Summary value of the risk ratio obtained by pooling the results of $RR=1.19$ and 95%CI is 1.01-1.38. Assuming that the true risk was not elevated, I calculated the x value for each study as described above.

In order to understand whether the distribution of *RR* values is evidence for an effect, I *assume* that there is *no association*. Then the true relative risk is $RR=1.0$ and in Figure 3 I plot the probability distribution of the normalized deviations, $ln(RR)/SE(ln(RR))$, from zero. Presentation is similar to Figure 1.

In Figure 3 I also plot the risk ratio, with 95% CI for each of 31 occupational studies of the possible association of leukemia with occupations including exposure to electromagnetic fields (EMF) compiled in ref. [11]. If all of these studies are averaged assuming that only statistical sampling errors are important, then RR=1.20 (95%CI 1.15-1.26). Also shown in Figure 3 are the Gaussian curve ($u=0$) and the "physical constants" curve ($u=1$). The observed distributions of x are very similar: both have longer tails than the normal distribution and are better described by the curve $u=1$. This suggests that these two datasets provide equally weak evidence of cancer risk associated with exposure to ETS and EMF because the observed tails of the distributions could be due to residual biases in the same way as the observed distribution of unsuspected errors in physical measurements shown in Figure 1. The difference between the attitudes concerning these hypothesized risk factors may lie in biological plausibility: while there is no doubt that some components of tobacco smoke can cause lung cancer, the possibility of biological effects caused by very weak electric and magnetic fields is doubtful [11].

5 Summary

This study provides one more way of looking at the results of multiple epidemiological studies. I use an analogy with physical measurements to illustrate the well-known fact that uncertainty in epidemiologic studies is larger than the range estimated by the 95%CI. Measurements of fundamental physical constants are much better defined than epidemiologic studies, yet experts there are overconfident about their accuracy [2,3]. There is every reason to expect that observational studies with slightly elevated risk ratios incorporate at least as many unaccounted errors as the datasets of physical measurements. The occurrence of false-positive findings in case-control studies with misclassification of exposure for a small fraction of subjects is illustrated by Monte Carlo simulation by Shlyakhter and Wilson in an accompanying paper [12]. Therefore I suggest that a prudent analyst should consider as inconclusive the evidence derived from the sets of observational studies with distributions of $ln(RR)/SE(ln(RR))$ which fall off similarly to the distribution of errors in physical measurements (described by the curve $u=1$).

This procedure is illustrated using studies of the association of leukemia with occupational exposure to electromagnetic fields and of lung cancer among nonsmoking females whose husbands smoke. An interesting observation is that both sets of studies are equally inconclusive.

One can hedge against possible biases in the individual epidemiologic studies by inflating the commonly reported 95% CI for relative risk using the observed probabilities of covering the null value in the studies where the true risk is not elevated. For a weakly positive finding, even a small inflation of the confidence interval can push the lower bound below one and make the conclusion much less convincing [13].

Acknowledgements

This research was partially funded by the Biological Effects Branch of the U.S. Department of Air Force through the contract F33615-92-C-0602 and by a number of gifts to Harvard University. I am deeply grateful to D.Burmaster, A.Feinstein, J.D.Graham, M.MacLure, J.Robins, K.Rothman, E.Washburn, and especially to R.Wilson for discussions and comments.
E-mail address: shlyakhter@physics.harvard.edu

References

[1] Rothman K.J. *Modern Epidemiology*, Little, Brown & Co., Boston, 1986.

[2] Shlyakhter A.I. "Uncertainty Estimates in Scientific Models: Lessons from Trends in Physical Measurements, Population and Energy Projections," in *Uncertainty Modelling and Analysis: Theory and Applications*, editors: B.M. Ayyub and M.M.Gupta, p.477-496. Elsevier Science B.V., 1994

[3] Shlyakhter A.I. An improved framework for uncertainty analysis: accounting for unsuspected errors. *Risk Analysis* 14:441-447, 1994.

[4] Feinstein A.R. "Scientific standards in epidemiologic studies of the menace of daily life," *Science* 242:1257-1263 (1988)

[5] Mayes LC., Horwitz RI, and Feinstein AR. "A collection of 56 topics with contradictory results in case-control research." *Int. J. Epidemiol.* 1988; 17:680-685.

[6] Database maintained by the Lawrence Berkeley Laboratory Particle Data Group.

[7] Koester L, Rauch H, and Seymann E. Neutron scattering lengths: a survey of experimental data and methods. *Atomic Data and Nucl. Data Tables* 1991;49:65-120.

[8] Fontham, E.T.H. et al. (1991) Lung cancer in nonsmoking women. *Cancer Epidemiology, Biomarkers and Prevention:* 1:35-43.

[9] Environmental Protection Agency. *Respiratory Health Effects of Passive Smoking: Lung Cancer and Other Disorders.* US EPA 600/6-90/006F, Washington, DC.

[10] Gross A.J. "The Risk of Lung Cancer in Nonsmokers in the United States and its Reported Association with Environmental Tobacco Smoke," *J. Clin. Epidemiol.* 48:587-598 (1995).

[11] *Health effects of low-frequency electric and magnetic fields*, Report of the Oak Ridge Associated Universities Panel for the Committee on Interagency Radiation Research and Policy Coordination ORAU 92-F8, Washington, D.C., 1992 (Table V-10).

[12] Shlyakhter A.I. and R.Wilson "Monte Carlo Simulation of Uncertainties in Epidemiological Studies: An Example of False-Positive Findings due to Misclassification," these Proceedings.

[13] Shlyakhter A.I., I.A. Shlyakhter, C.L. Broido, and R. Wilson "Estimating uncertainty in physical measurements and observational studies: lessons from trends in nuclear data," *Proceedings of ISUMA'93 the Second International Symposium on Uncertainty Modeling and Analysis*, University of Maryland, College Park, Maryland, April 25-28, 1993, p. 310-317, IEEE Computer Soc. Press, Los Alamitos, California, 1993.

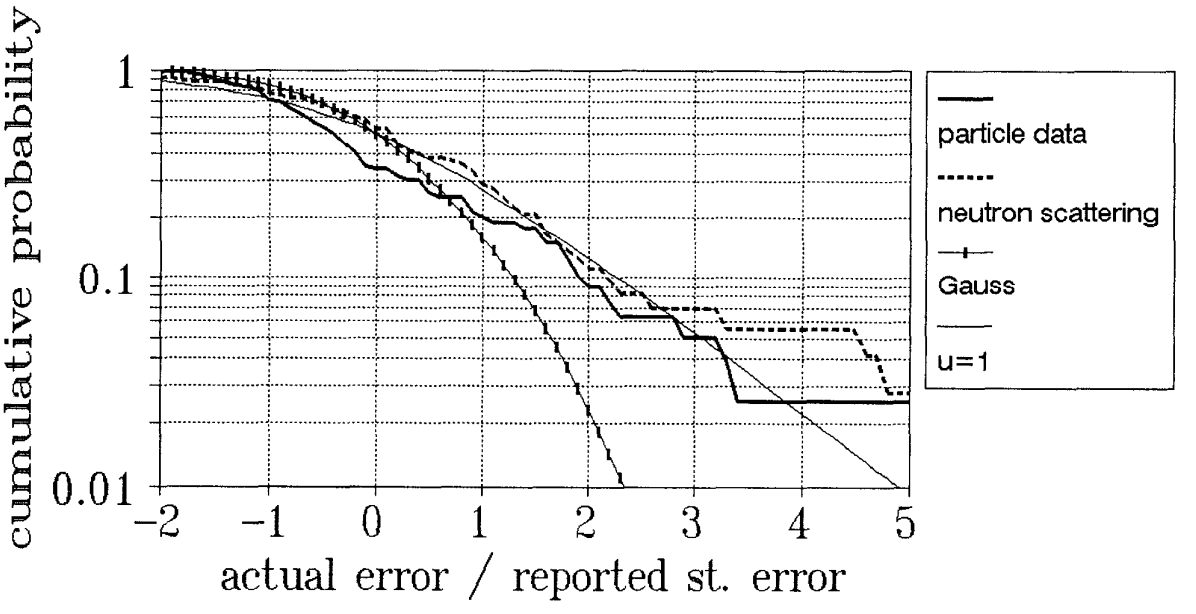

Figure 1. Probability of unexpected results in physical measurements. The plots show the cumulative probability, $S(x) = \int_x^\infty p(t)dt$, that old results *(A)* are at least *x* standard deviations *(Δ)* away from the most recent value *(a)*; $x = (a - A)/\Delta$ as defined in the text. The cumulative probability distributions of *x* are shown for two data sets: particle data [6] and neutron scattering lengths [7]. Also plotted is a cumulative normal distribution, $\mathrm{erfc}(x/\sqrt{2})$ (thin solid line with markers), and compound exponential distribution [2,3] with parameter $u = 1$ (solid line). Note that the scale is logarithmic and the normal distribution is cumulative so that it looks differently from the usual Gaussian bell-shape curve.

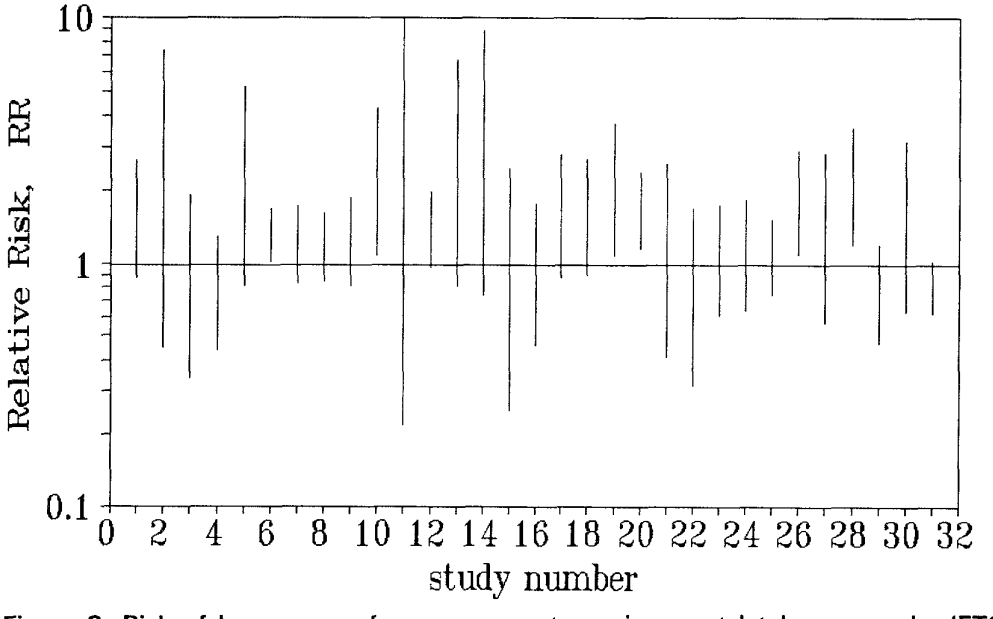

Figure 2. Risk of lung cancer from exposure to environmental tobacco smoke (ETS). Data for 31 studies analyzed by US EPA [9] and Gross [10] are shown.

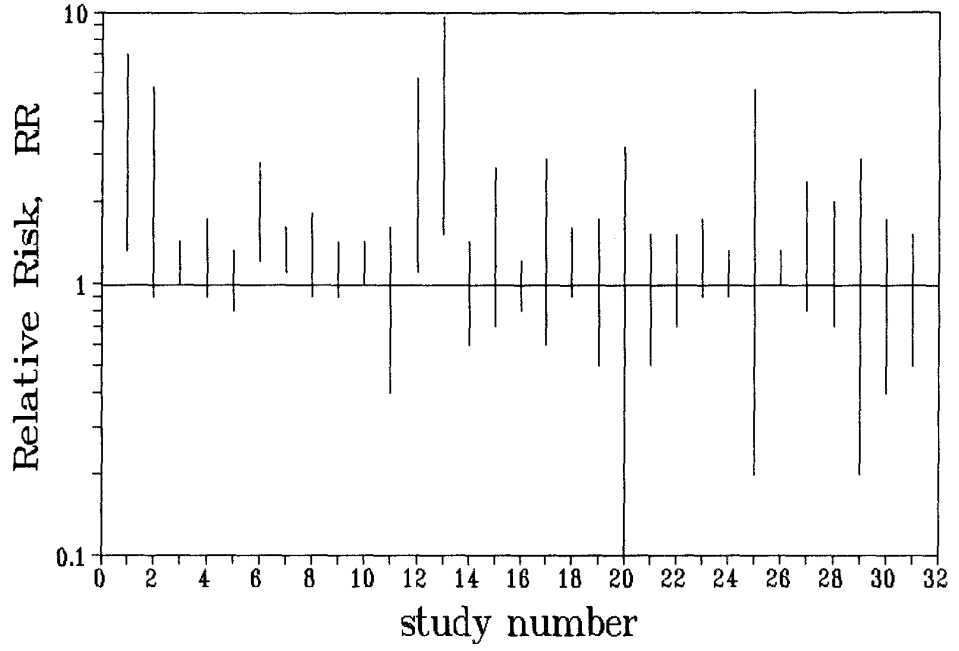

Figure 3. Risks of leukemia from occupational exposure to electromagnetic fields (EMF). Data for 31 studies compiled in [11], in which combined *RR* for all leukemia types was reported, are shown.

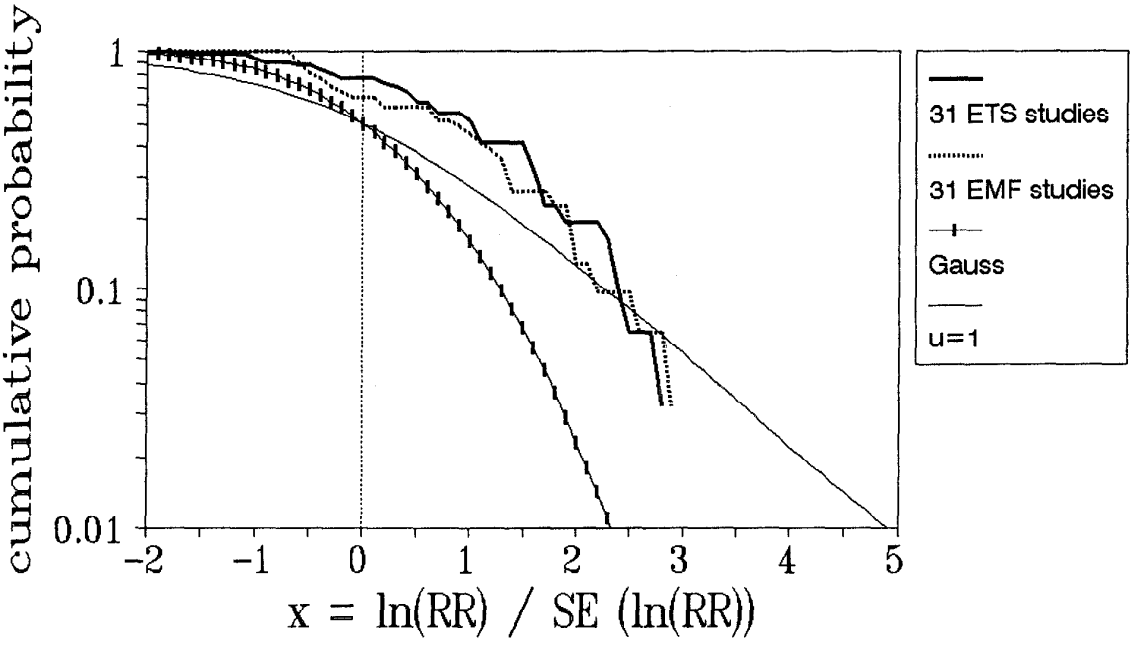

Figure 4. Data from Figures 2,3 presented in a different format. Cumulative distribution of the normalized deviations of *ln(RR)/SE(ln(RR)* from zero is shown for both ETS and EMF datasets together with the curves for $u=0$ (normal distribution) and $u=1$. Presentation is similar to Figure 1.