Environmental Health Perspectives
Vol. 42, pp. 15-21, 1981

Statistical Limitations in Relation to Sample Size

by Charles E. Land*

The statistical difficulties of estimating cancer risks from low doses of a carcinogen are illustrated by examples from radiation carcinogenesis. Although more is known about dose-response relationships for ionizing radiation than for any other environmental carcinogen, estimates of cancer risk from low radiation doses have been extremely controversial; disagreements by factors of 100 or more are not uncommon. Direct estimation, based on data from populations exposed to low doses, is usually impracticable because of sample size requirements. Curve-fitting analyses, by which higher dose data determine lower dose risk estimates, require simple dose-response models if the estimates are to be statistically stable. The current level of knowledge about biological mechanisms of carcinogenesis does not usually permit the confident assumption of a simple model, however; thus frequently the choice is between unstable risk estimates obtained using general models and statistically stable estimates whose stability depends on arbitrary model assumptions.

Introduction

The purpose of this paper is to illustrate some of the statistical difficulties of estimating cancer risks from low doses of a carcinogen, that is, from dose levels producing excess risks that are small relative to normal risk. The illustration is by examples from radiation carcinogenesis. More is known about dose-response relationships for ionizing radiation than for any other environmental carcinogen, and models commonly used in curve fitting have widely accepted radiobiological interpretations. Nevertheless, estimates of cancer risk from low doses of ionizing radiation tend to be extremely controversial; disagreements by factors of 100 or more are common.

Direct Estimation: Hypothesis Testing and Point Estimation

Pochin (*1*) has discussed the difficulties of estimating the increased health risk to populations in areas of unusually high levels of background radiation. These difficulties follow from the necessity of estimating excess risk as the difference between the observed risk in a population exposed to higher-than-usual radiation levels and that in a population exposed to usual levels. In general, the difference is much smaller than the risks in the two populations; thus, changes in the dose difference between the two populations can double or triple the difference in risk between them without having a noticeable effect on the overall risk in the more heavily exposed population.

Example 1

The evidence for a linear dose-response relationship for female breast cancer induced by exposure to sparsely ionizing radiation, like x-rays or γ-rays, is strong (*2*). The 1972 BEIR report estimate of 6 excess cases per million women exposed per year of observation for risk, following a minimal latent period of 10 years after exposure, for each rad to breast tissue (*3*) still seems appropriate for women exposed after the age of 20, although not for women exposed at younger ages (*2*). Consider an idealized experiment in which half of a sample of N women

*Environmental Epidemiology Branch, National Cancer Institute, Bethesda, Maryland 20205.

receive a single mammographic examination resulting in 1 rad average tissue dose to both breasts. Suppose the exposed and nonexposed women are otherwise comparable and suppose, for simplicity, that all were 35 years old at the time of exposure, and that followup information with respect to breast cancer incidence is available for 20 years following exposure for each woman. Ignoring the first 10 years, we might expect to see 60 excess cancers per million exposed women, in addition to the 19100 breast cancers normally seen per million U.S. women of that age in a 10-year period (*4*).

The numbers of breast cancers observed in the exposed and nonexposed women can be assumed to be independent Poisson random variables with means equal to $N/2$ times 19160 per million for the exposed and times 19100 per million for the nonexposed. The estimated yearly excess risk due to radiation, obtained as the difference between the observed rates in the two populations, has mean $D = 6 \times 10^{-6}$ and standard deviation $S = [(19160 + 19100) \times 10^{-6}/(N/2)]^{1/2}/10 = 0.02766/N^{1/2}$. For simplicity, S will be assumed known, but because we are considering only very large values of N this is not misleading; the usual estimate of S itself has standard deviation inversely proportional to N. For N greater than 10,000, the estimate D has approximately a normal distribution. Finally, we ignore the small difference between the above value for S and that corresponding to the null hypothesis of no excess risk, $S = 0.02764N^{1/2}$. Accordingly, the calculations given below are based on normal approximations to the distributions of the estimate D, with mean 6×10^{-6} and standard deviation S, and the test statistic $T = D/S$, with mean $D/S = 0.000217N^{1/2}$ and unit standard deviation.

Under these assumptions we can calculate the approximate statistical power of the level 0.05 test of the hypothesis of no radiation effect on breast cancer risk against the alternative that risk increases with increasing dose, and the probability of a negative estimate of risk, both shown as functions of N in the left-hand panel of Figure 1. Power is low for N less than 100 million (it is greater than 50% only for N greater than 60 million), and the chance of a negative estimate of risk is high when power is low. A negative estimate should not be interpreted as evidence that no radiation effect exists, but such an interpretation is often made, nevertheless.

Even when power is low, the chance of obtaining an estimate that is significantly greater than zero is at least 5%. The minimum value of a statistically significant estimate is graphed in the right-hand panel of Figure 1, and the curve above it is the average value to be expected given statistical significance. For sample sizes corresponding to low power, statistically significant estimates are necessarily too high: for $N = 1$ million, power is only a little above 5%, the probability of a negative estimate is nearly 50%, and the average statistically significant estimate is about 55 per million per

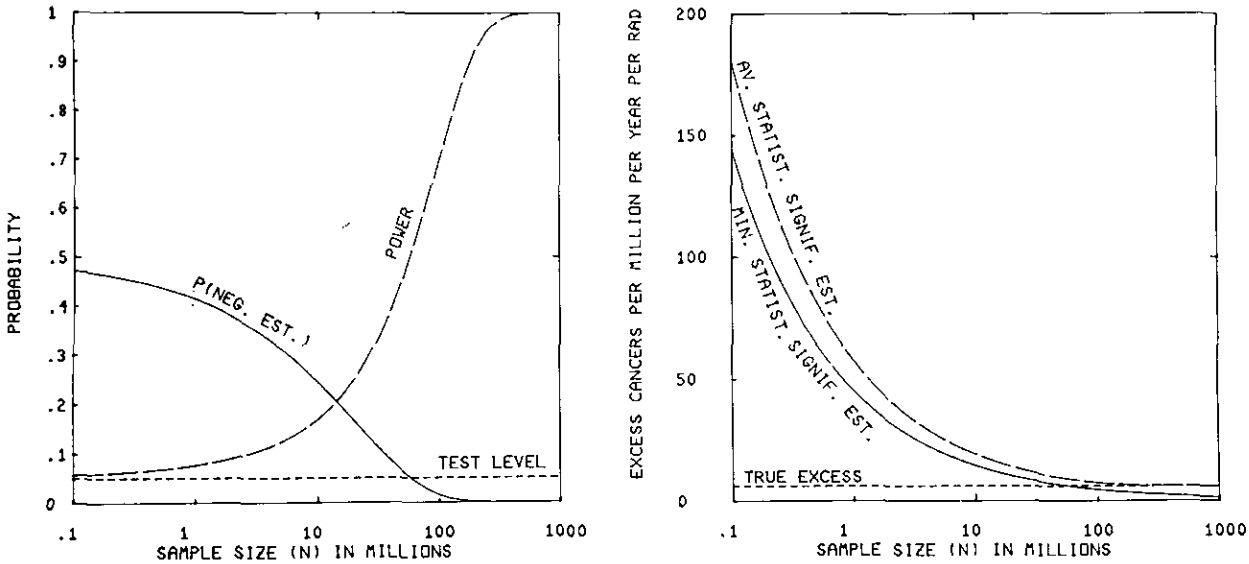

FIGURE 1. Example: Hypothetical 20-year follow-up study of breast cancer incidence among N women, half of them exposed and half not exposed at age 35 to a breast-tissue dose of 1 rad. Assumed excess risk among the exposed is six breast cancers per million women per year after a 10-year minimum latency period. Statistical power, the probability of a negative risk estimate, and the minimum and average risk estimates given statistical significance at level 0.05 are plotted as functions of sample size N. Adapted from Land (*11*).

year, or 9 times the true excess. For $N = 10$ million, power is 17%, the chance of a negative estimate is 25%, and the average significant estimate is 3.2 times the true excess, while for $N = 100$ million power is near 1, negative estimates are unlikely, and statistical significance imposes no appreciable bias.

If all risk estimates received equal attention, and if studies of large populations exposed to low doses of carcinogens were easy to do, the situation illustrated in the right-hand panel of Figure 1 would present no problem, at least in the long run. Unfortunately, estimates of an effect often are considered uninteresting if unaccompanied by evidence that the effect in fact exists, and it is a commonplace among scientists that "positive" studies, those in which the null hypothesis of no exposure effect is rejected, are more likely to be reported and published than "negative," and therefore inconclusive, studies. Large studies involve great effort and expense and for that reason are unlikely to go unreported, but many possible effects tend to be investigated using the same body of data, and it is the statistically significant estimates that receive the most attention. A case in point is the various analyses of the mortality data on workers at the Hanford Plutonium Works, collected by Dr. Mancuso and analyzed first by Stewart and and Kneale (*5*) and, later, by others (*6-9*). It seems fair to say that there has been more attention paid to the two cancers for which everyone has found a statistically significant association with dose—pancreatic cancer and multiple myeloma—than to other cancers, including leukemia, for which no association was found. The point estimates for the two statistically significant sites were very high, even though these cancers, unlike leukemia, are not among those most frequently associated with radiation exposure.

Confidence Intervals. The curves presented in Figure 1, other than the power curve, highlight a common fallacy in the use of statistical methods which can be summed up as a tendency to use only part of the information available from an analysis, either from a desire to make a point or confirm a bias, or from a kind of impatience or mental laziness which leads us to reduce information to a single number. In other words, reporting (or noticing) only point estimates and whether or not the estimates are significantly greater than zero can create an illusion of precision where no precision exists. A strategy based on confidence interval estimation is less likely to be misleading but, perhaps because a confidence interval emphasizes statistical uncertainty while a single number suggests precision, this strategy is too seldom employed.

In the example of Figure 1 the event of rejecting the null hypothesis corresponds to the event that a right-infinite, one-sided, level 0.95 confidence interval for D does not contain zero. The probability of this event, therefore, is given by the power function shown in Figure 1. The probability that the true excess risk will be excluded from the interval is 0.05, regardless of sample size, and the probability that any given larger value is not contained in the interval is a decreasing function of sample size, with an upper limit of 0.05 (Table 1). Exclusion probabilities for positive values smaller than D follow the pattern of the power function, increasing with increasing sample size. In the example, one-sided, left-infinite confidence intervals for D are symmetric with right-infinite intervals in the sense that the probability of exclusion of a value $D + E$ from a left-infinite interval is the same as that for the value $D - E$ from a right-infinite interval of the same confidence level. As can be seen from Table 1, the confidence interval approach is less subject to the problems highlighted in the right-hand panel of Figure 1. For example, for a sample of 10 million women, there is a 17% chance that a statistically significant point estimate of risk will be obtained, and if this happens the estimate can be expected to be 3.2 times as large as the true risk. The probability that the true value will be excluded from the 95% right-infinite confidence interval is only 5%, however, and the chances of excluding all values less than twice the

Table 1. Example 1: Probability of excluding certain multiples of the true parameter value from one-sided level 0.95 confidence intervals for various sample sizes.

	Right-infinite orientation			Left-infinite orientation		
Multiple	1×10^6	10×10^6	100×10^6	1×10^6	10×10^6	100×10^6
0	0.08	0.17	0.70	0.03	0.01	0.00007
1/2	0.06	0.10	0.29	0.04	0.02	0.003
1/4	0.07	0.13	0.49	0.035	0.015	0.0005
1	0.05	0.05	0.05	0.05	0.05	0.05
2	0.03	0.01	0.00007	0.08	0.17	0.70
4	0.01	0.0001	0.000000	0.10	0.66	0.999

true value is only 1%. These probabilities correspond to conditional probabilities, given statistical significance, of 0.29 and 0.06, respectively. The probability of obtaining a negative point estimate of risk is 25%, but the chance that estimates greater than half the true risk will be excluded from a *one-sided, left-infinite confidence interval of level* 0.95 is only 2.3%, and the conditional probability of this, given a negative point estimate, is only 9%.

Sample Size as a Function of Dose. If instead of a 1-rad dose to breast tissue in Example 1 a larger dose were used, the excess risk in the exposed women would be larger, and the power function and other functions graphed in Figure 1 would be shifted. Assuming linearity, or proportionality between dose and excess risk, the effect on power of increasing dose by a given factor for a fixed sample size is approximately the same as that of increasing sample size by the square of that same factor while keeping dose constant. In the example, this relationship holds approximately over the dose range 1-100 rad (Fig. 2). In other words, if 100 million women must be studied to estimate the excess breast cancer risk of 1 rad, only 10 thousand need be studied to estimate the effect of 100 rad.

Curve Fitting. If linearity, or any other simple rule by which low-dose cancer risk can be inferred from high-dose risk, were known to hold in any given case, it clearly would be more efficient to learn about low-dose risk by studying populations exposed to high doses as opposed to low doses. Because the sample size requirements for direct estimation of low-dose risk are so enormous, this is true anyway, but in the absence of knowledge about the shape of the dose-response model there must always be uncertainty about low-dose risk estimates obtained by extrapolation from high-dose data. Even in the case of radiation carcinogenesis, for which radiobiological theory suggests dose-response curves limited, at least for sparsely ionizing radiation of no more than 200 rads or so, to linear-quadratic forms with nonnegative coefficients for dose and the square of dose, differences in the choice of dose-response model can make large differences with respect to estimated risks from low-dose exposures.

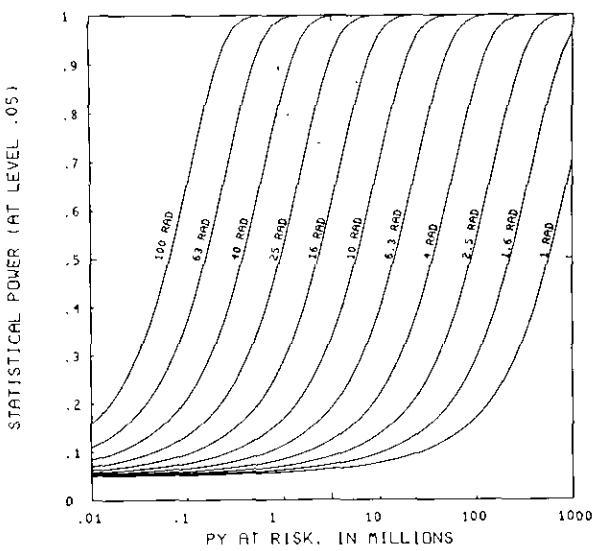

FIGURE 2. Extensions of the example in Figure 1. Power as a function of person-years at risk (PY = N times average years followup after the minimal latency period) for increasingly large doses to breast tissue.

Example 2

The leukemia incidence data from the life-span study sample of survivors of the Nagasaki A-bomb for 1950-1971 (*10*) constitute the most useful existing information on dose-response relationships for leukemia induced by sparsely ionizing radiation. These data yield very different estimates of excess risk at low doses when fitted to a general linear-quadratic dose-response model or to pure linear or pure quadratic models, yet the fitted curves do not differ markedly in their closeness of fit to the data. That is, the chi-square values for lack of fit do not indicate that any one of these models fits the data better than any of the others. This lack of discrimination among competing models is ascribable to lack of statistical power at low doses, as illustrated in the following discussion.

Table 2 gives average radiation doses to bone marrow, person-years at risk for grouped data covering the period 1950-1971 and parameter estimates from regression analyses of age-adjusted rates (*11*). These analyses assumed linear-quadratic, linear and pure quadratic dose-response models. In the present analysis, we assume each of these models, and for each, the estimated parameter values are assumed to be true. For each assumed dose-response function, we consider the statistical properties of curve-fitting analyses using different dose-response models. In particular, statistical power is calculated for level 0.05 hypothesis tests of the coefficients of dose and dose-squared, against positive alternatives. These calculations are based on normal approximations to the distributions of the parameter estimates, assuming the observed numbers of leukemias in each dose group to have the covariance structure of independent Poisson variates. The dose distribution in Table 2 is assumed, but the person-years at risk are uniformly multiplied by factors

Table 2. Example 2: Assumed person-years at risk, dose values, and dose-response functions for power calculations in curve-fitting example.

		Risk/year		
Dose d, rad	PY at risk	Linear-quadratic model	Linear model	Pure quadratic model
0	214,222	$0.0000333 + 0.000001d + 0.00000001d^2$	$0.00029 + 0.0000025d$	$0.000035 + 0.000000016d^2$
2.15	128,288			
11.8	71,676			
38.9	25,643			
79.0	27,355			
132	14,714			
186	5,415			
286	6,981			

between 0.1 and 10 in order to show the dependence of power on sample size. Dependence on average dose level is illustrated by parallel calculations in which all dose values are assumed to be reduced by one tenth. Table 3 summarizes the findings for analyses assuming the linear-quadratic model.

The linear-quadratic model analyses indicate a strong dependence of power on the true parameter value. The power function for tests of the linear coefficient of dose is largest when the linear model is assumed because the assumed linear coefficient is largest according to this model, and it is least when the pure quadratic model, with zero linear coefficient, is assumed. When the linear-quadratic model is assumed, the power for tests of the linear coefficient is high only after the numbers of person-years have been increased by a factor of nearly 10, which explains why analyses of the original data did not discriminate well between the linear-quadratic and pure quadratic models. Similarly, power is low for tests of the dose-squared coefficient unless the assumed sample size is increased, explaining the lack of discrimination between the linear-quadratic and linear models. The values for the reduced dose levels illustrate the formidable sample size requirements for complex curve-fitting analyses of low-dose data.

Power calculations for the linear coefficient of dose, assuming a linear model analysis, are shown in Table 4, and those for the quadratic coefficient of dose, assuming a pure quadratic model analysis, are shown in Table 5. An important difference between these calculations and those in Table 3 is that the linear-quadratic model is a general one, including the linear and pure quadratic models as

Table 3. Power calculations for linear-quadratic model analyses of Example 2, assuming dose values, person-years at risk, and dose-response functions shown in Table 2. Values correspond to multiples of the person-year array in Table 2, and to the given dose array and the array divided by 10.

	Dose			Dose/10		
	Lin-quad	Linear	Pure quad	Lin-quad	Linear	Pure quad
Power for the linear coefficient of dose						
Coeff. × 1,000,000	1	2.5	0	1	2.5	0
PY multiplier (power of 10)						
−1	0.089	0.171	0.050	0.057	0.066	0.050
−0.5	0.132	0.342	0.050	0.062	0.081	0.050
0	0.239	0.710	0.050	0.073	0.114	0.050
0.5	0.508	0.988	0.050	0.096	0.194	0.050
1	0.906	1.000	0.050	0.149	0.400	0.050
Power for the quadratic coefficient of dose						
Coeff. × 1,000,000	0.01	0	0.016	0.01	0	0.016
PY multiplier (power of 10)						
−1	0.118	0.050	0.118	0.051	0.050	0.052
−0.5	0.205	0.050	0.367	0.052	0.050	0.054
0	0.426	0.050	0.751	0.054	0.050	0.058
0.5	0.828	0.050	0.994	0.058	0.050	0.065
1	0.998	1.050	1.000	0.065	0.050	0.079

Table 4. Power calculations for linear model analyses of Example 2, assuming dose values, person-years at risk, and dose-response functions shown in Table 2. Values correspond to multiples of the person-year array in Table 2, and to the given dose array and the array divided by 10.

	Dose			Dose/10		
	Lin-quad	Linear	Pure quad	Lin-quad	Linear	Pure quad
Coeff. × 1,000,000	2.34	2.5	2.14	1.13	2.5	0.21
PY multiplier (power of 10)						
−1	0.336	0.384	0.296	0.069	0.091	0.054
−0.5	0.701	0.775	0.629	0.087	0.137	0.057
0	0.987	0.996	0.969	0.127	0.252	0.062
0.5	1.000	1.000	1.000	0.226	0.537	0.073
1	1.000	1.000	1.000	0.477	0.926	0.096

Table 5. Power calculations for pure-quadratic model analyses of Example 2, assuming dose values, person-years at risk, and dose-response functions shown in Table 2. Values correspond to multiples of the person-year array in Table 2, and to the given dose array and the array divided by 10.

	Dose			Dose/10		
	Lin-quad	Linear	Pure quad	Lin-quad	Linear	Pure quad
Coeff. × 1,000,000	0.0175	0.0187	0.016	0.0847	0.187	0.016
PY multiplier (power of 10)						
−1	0.426	0.497	0.372	0.081	0.116	0.056
−0.5	0.829	0.897	0.367	0.114	0.200	0.061
0	0.999	1.000	0.994	0.194	0.414	0.071
0.5	1.000	1.000	1.000	0.399	0.814	0.092
1	1.000	1.000	1.000	0.796	0.998	0.139

special cases. Thus no bias is introduced by doing a linear-quadratic model analysis of data when the true dose-response relationship is linear or pure quadratic, although, as can be seen from a comparison of the tables, there will be a loss of power from using an unnecessarily general model. Using a linear model to analyze data corresponding to a nonlinear dose response does introduce bias, however. In such a case, the linear model analysis estimates the average excess risk over the range of doses represented, but unlike the linear coefficient in a linear-quadratic model analysis (assuming the true dose response is no more complicated), this value cannot be interpreted as the excess risk per rad at low dose levels. Thus the value to be estimated by a linear model analysis depends not only on the true model but also on the dose distribution of the data; for example, the linear-quadratic dose response with linear coefficient equal to 1 per million, and quadratic coefficient equal to 0.01 per million, corresponds to an average excess per rad of 2.34 per million over the dose distribution in Table 2, but only 1.13 per million over the dose distribution scaled down by a factor of 10. Similar considerations apply to linear model analyses of pure quadratic dose-response data, and to pure-quadratic model analyses of linear and linear-quadratic data.

Perhaps the most surprising thing about Tables 4 and 5 is that power, using linear and pure quadratic model analyses, should appear to depend so strongly on the value of the parameter to be estimated and so little on whether or not the model assumed in the analysis is the same as that generating the data. In other words, lack of fit appears to have little to do with power. The second noteworthy observation, which has already been made, is that the protection against bias obtained through use of a more general model has a cost in reduced power.

Summary

There are formidable statistical difficulties associated with refined estimation of risk from exposure to carcinogens at low dose levels. These difficulties are unlikely to be overcome by sample size expansion or by curve fitting, unless it can be established independently that the dose-response relationship is a particularly simple one. Research into the biological mechanisms of carcinogenesis would appear to be an essential part of the estima-

tion process, by which plausible models can be derived. In the case of radiation carcinogenesis, radiobiological theory suggests that linear model analyses, confined to doses under a few hundred rads to low-LET radiation, may give credible upper limits of risk at low doses, in the form of confidence limits. Although more refined solutions may eventually appear, the concept of upper limits based on conservative, simple models is a useful one, adequate for many purposes.

REFERENCES

1. Pochin, E. E. Problems involved in detecting increased malignancy rates in areas of high natural radiation background. Health Phys. 31: 148-151 (1976).
2. Land, C. E., Boice, J. D., Jr., Shore, R. E., Norman, J. E. Jr., and Tokunaga, M. Breast cancer risk from low-dose exposures to ionizing radiation: Results of an analysis in parallel of data from three different exposed populations. J. Natl. Cancer Inst. 65: 353-356 (1980).
3. National Academy of Sciences Advisory Committee on the Biological Effects of Ionizing Radiation. The Effects on Populations of Exposure to Low Levels of Ionizing Radiation. National Academy of Sciences-National Research Council, Washington, D.C., 1972.
4. Waterhouse, J., Muir, C., Correa, P., and Powell, J. Cancer Incidence in Five Continents, Vol. III. International Agency for Research on Cancer, Lyon, 1976.
5. Mancuso, T. F., Stewart, A., and Kneale, G. Radiation exposures of Hanford workers dying from cancer and other causes. Health Phys. 33: 369-384 (1977).
6. Marks, S., Gilbert, E. S., and Breitenstein, B. D. Cancer mortality in Hanford workers. In: Late Biological Effects of Ionizing Radiation, Vol. I, International Atomic Energy Agency, Vienna, 1978, pp. 369-386.
7. Gilbert, E. S., and Marks, S. An analysis of the mortality of workers in a nuclear facility. Radiat. Res. 79: 122-148 (1979).
8. Hutchison, G. B., MacMahon, B., Jablon, S., and Land, C. E. Review of report by Mancuso, Stewart, and Kneale of radiation exposure of Hanford workers. Health Phys. 37: 207-220 (1979).
9. Gofman, J. W. The question of radiation causation of cancer in Hanford workers. Health Phys. 37: 617-639 (1979).
10. Ichimaru, M., Ishimaru, T., and Belsky, J. L. Incidence of leukemia in atomic bomb survivors belonging to a fixed cohort in Hiroshima and Nagasaki, 1950-71. Radiation dose, years after exposure, age at exposure, and type of leukemia. Japan. Radiat. Res. 19: 262-282 (1978).
11. Land, C. E. Cancer risk from low doses of ionizing radiation. Science 209: 1197-1203 (1980).