1971

AMERICAN JOURNAL OF EPIDEMIOLOGY Vol. 114, No. 5

POWER CONSIDERATIONS IN EPIDEMIOLOGIC STUDIES OF VINYL CHLORIDE WORKERS

JAMES J. BEAUMONT[1] AND NORMAN E. BRESLOW[2]

Beaumont, J. J. (NIOSH, Cincinnati, OH 45226) and N. E. Breslow. Power considerations in epidemiologic studies of vinyl chloride workers. *Am J Epidemiol* 1981;114:725-34.

Nine retrospective mortality studies of workers exposed to vinyl chloride were reviewed to determine whether differences in their hypothesis testing results might be due to differences in statistical power. Where possible, the power of each study was calculated for cancer of the lung, brain and liver. When power was taken *into* consideration, the results for liver and brain cancer were found to be consistent with an etiologic role for vinyl chloride. For lung cancer, the data were not consistent with an etiologic role, in that two studies with very high power yielded negative results.

epidemiologic methods; respiratory tract neoplasms; brain neoplasms; liver neoplasms; vinyl chloride

The nine mortality studies of workers exposed to vinyl chloride that have been completed to date disagree in their hypothesis testing results for some causes of death, and for those causes it is difficult to draw conclusions about excess risk (1–12). The studies that are negative for a cause of death are especially difficult to interpret, since they may or may not have had the statistical power to detect an excess risk if, in fact, one existed. To aid in the interpretation of the studies, their statistical power with respect to three sites of cancer—liver, brain and lung—was calculated, and their hypothesis testing results, positive or negative, were considered in the context of power.

METHODS

Power calculations

Statistical power is the probability of not overlooking an excess risk, i.e., of not making a Type II statistical error. To aid in the interpretation of this frequently confusing subject, a brief review of Type I and Type II errors follows.

Investigators are most familiar with Type I errors. If the true state of nature is the null hypothesis (no increase in risk), then a Type I error is wrongly rejecting the null hypothesis and declaring that there is an increase in risk. The probability of a Type I error is known as alpha, or more commonly, the "level of significance." For example, when an investigator finds an excess and declares it "significant at the 0.05 level," in a sense what he or she is saying is that there is less than a 5 per cent chance of having made a Type I error.

Type II errors are less familiar but,

Received for publication November 14, 1980, and in final form April 13, 1981.

[1] Industrywide Studies Branch F-8, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, OH 45226 (address for reprint requests).

[2] Dept. of Biostatistics, School of Public Health and Community Medicine, U. of Washington, Seattle, WA 98195.

The authors thank Jay Lubin, Richard Waxweiler, Thomas Fischbach, Noel Weiss and William Halperin for their valuable advice. Their special thanks go to Philip Landrigan for encouraging this approach.

This paper was presented at the Conference to Revaluate the Toxicity of Vinyl Chloride, Polyvinyl Chloride and Structural Analogues, Rockville, MD, March 20–21, 1980.

nevertheless, very important (13). Underlying this concept is the fact that when the true state of nature is the alternative hypothesis (increased risk), the research goal is to detect it (i.e., reject the null hypothesis). Suppose one has specified an alpha level (0.05, for example) for testing a null hypothesis of no increased risk for a particular exposure. If an exposure is harmful with a relative risk of R, then not rejecting the null hypothesis (i.e., wrongly accepting the exposure as harmless) is called a Type II error. The probability of a Type II error is usually denoted beta (β). Conversely, the probability of correctly rejecting the null hypothesis and therefore of detecting the excess risk is called the power and is equal to $1 - \beta$. Thus, power quantifies the ability of a particular study to detect an excess risk that truly exists. It is intuitively clear that with a fixed amount of data there is a greater likelihood of detecting a large excess risk. Similarly, an increase in the amount of data increases the chance of observing a given risk, i.e., increases the power.

The approximate power $1 - \beta$ of the vinyl chloride studies to detect a relative risk R at the alpha level of significance was calculated from the following formula, which uses the fact that the square root transformation stabilizes the variance of the Poisson distribution (see the statistical appendix):

$$Z_{1-\beta} = Z_{\alpha} - 2(\sqrt{R} - 1)(\sqrt{E})$$

Here Z_{α} denotes the upper 100_{α} percentile of the standard normal distribution and E the expected number of cancer deaths based on general population rates. This approximation agrees well with exact power calculations based on Poisson probabilities made by Cutler et al. (14, 15). When the discreteness of the exact test based upon Poisson probabilities is accounted for, the approximate and exact power curves are virtually identical.

A family of power curves based upon the approximation is shown in figure 1, where each curve is for a different relative risk. It can be seen that as the assumed relative risk increases for a given number of expected deaths, the power also increases. Similarly, as the expected number of deaths increases for a given relative risk, the power also increases.

Application to vinyl chloride literature

The assumed relative risks for the power calculations were the median (approximate) standardized mortality ratios reported for each cancer site in the vinyl chloride literature. Power calculations are usually performed with a somewhat arbitrarily chosen relative risk, but since relative risk information was available from the vinyl chloride studies, it was thought best to make use of these data. Separate assumptions were made for analyses considering all person-years at risk and for analyses considering only person-years after a minimum latency (time since first exposure), because in the former case the risk was diluted by the 10 to 20 years that are often required for cancer to develop after exposure to a carcinogen. The median reported mortality ratio was not used in one instance: for overall lung cancer, where the median was 1.03, essentially no excess risk. Since some risk needs to be assumed to calculate power, the arithmetic mean (approximately 1.5) was used.

The statistical powers of mortality studies of workers exposed to vinyl chloride were calculated with respect to three sites of cancer: liver, brain and lung. These three sites were chosen because they have been the subject of most of the discussion of causality in the vinyl chloride literature. The powers of the studies were subsequently plotted on power curves for two purposes: to show the variability in powers, and to relate positive and negative findings to power. For the purposes of this review, a positive finding was defined as an excess risk for a

POWER (%)

FIGURE 1.
ratio studies

particular
cance at t
Poisson te
It was
was carci
should be
power sho
the studi
general, t
exposure
of R acro
assumptic
cause, as
differed ir

The hi
vinyl chlc
in table 1
tions vari
study by

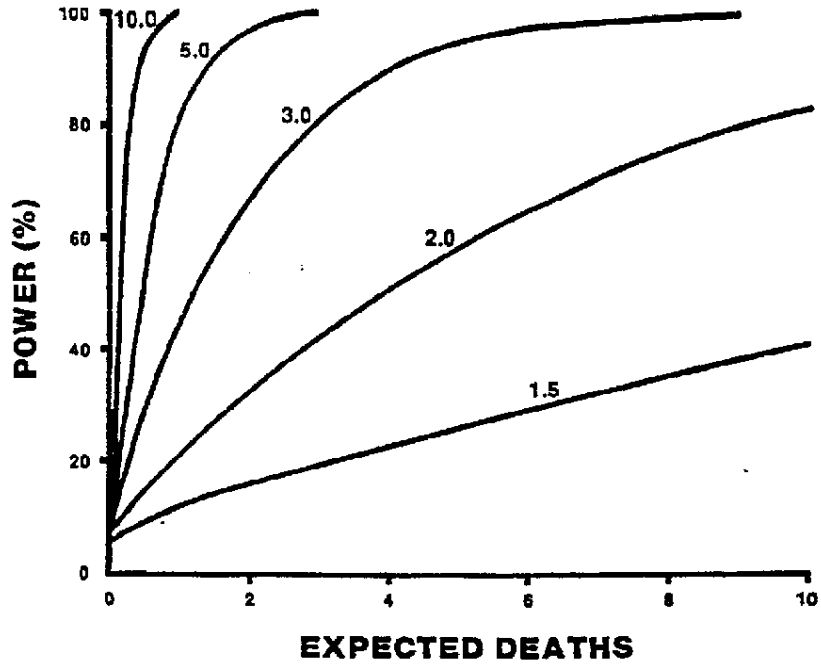

FIGURE. 1. Approximate power curves for various assumed relative risks in standardized mortality ratio studies (for one-sided comparisons with alpha = 0.05).

particular cancer with statistical significance at the 0.05 level, using a one-sided Poisson test.

It was expected that if vinyl chloride was carcinogenic for an organ site, there should be a pattern: the studies with high power should, in general, be positive, and the studies with low power should, in general, be negative. This assumed that exposure would result in a constant value of R across the various populations. The assumption was somewhat tenuous because, as discussed below, the populations differed in a number of respects.

RESULTS

The historical prospective studies of vinyl chloride exposed workers are listed in table 1. The sizes of the study populations varied considerably, from 255 in the study by Nicholson et al. (3) to 9677 in the Equitable Environmental Health study (6, 7). The reports were not entirely independent; for example, many of the workers in the study by Ott et al. (10) were included in the Equitable Environmental Health study, and there was an overlap of about 800 workers in the studies by Waxweiler et al. (4) and Waxweiler (5).

It should be noted that the studies were dissimilar in many other ways. As can be seen in table 1, a minimum exposure of 1–5 years was required in some studies, while others required only one day. Some studies reported findings based upon all person-years at risk, some included person-years at risk only after a minimum time since first exposure, and some reported both types of analysis. The control (standard) populations used were different: four of the five American studies

TABLE 1

Historical prospective (standardized mortality ratio) studies of vinyl chloride exposed workers

Study	Year	Cohort size	Exposure minimum (years)	Person-years considered		Cancer ratios reported		
				All	Minimum latency	Lung	Brain	Liver
Fox and Collier (1, 2)	1976, 1977	7117	>0	X	X	X	X	X
Nicholson et al. (3)	1975	255	5		X			
Waxweiler et al. (4)	1976	1294	5		X	X	X	X
Waxweiler (5)*	1978	4806	>0	X	X	X	X	X†
Equitable Environmental Health (6, 7)	1974, 1978	9677	1	X	X	X	X	
Reinl et al. (8)	1978	7021	>0	X		X	X	X
Buffler et al. (9)	1979	464	>0	X	X	X		
Ott et al. (10)	1976	522	>0	X		X		
Byren et al. (11)	1976	750	>0	X	X	X	X	X
Duck et al. (12)	1976	2120	>0	X		X		

* One of four plants in 1976 study.

† Liver cancer not reported (Waxweiler, personal communication, 1980).

used United States rates, the fifth used Texas rates; and the British, Swedish and German studies used their respective national rates. There were also minor differences in the International Classification of Diseases codes included for the specific cancers. Finally, there were probably differences in age composition which could be important if an effect were dependent upon age. While it was felt that none of these dissimilarities was important enough to prevent comparison of statistical powers, it should be noted that factors other than power could have had a bearing on the ability of the studies to detect excess risk.

Liver cancer was reported in the four studies listed in table 2 and graphically presented in figure 2. When the results were tested for statistical significance with a one-sided Poisson test, it was found that three of the four studies showed significant excesses (4, 5, 8, 11). The powers were all very high, despite the fact that the expected numbers of deaths were in all instances small. The small expected numbers would normally have led to low statistical powers, but the high estimated relative risks had a large effect on the calculations. The powers of the studies are shown in two columns in table 2 and on two lines in figure 2 because the calculations assumed different relative risks for analyses considering all person-years at risk and for analyses considering only person-years after a minimum latency (relative risks of 5 and 10, respectively, were assumed for liver cancer).

The results for brain cancer (table 3) were more variable, in that three of five studies had statistically significant findings (5–7, 11), and that the powers ranged from 12 per cent to a maximum of approximately 80 per cent. In figure 3, the range of powers is presented graphically. It can be seen that, in general, the studies with high power had positive findings and those with low power had negative findings.

Of the eight studies of lung cancer, only

TABLE 2

Liver cancer: results and powers of epidemiologic studies of vinyl chloride

Author	Observed deaths	Expected deaths	Mortality ratio	Excess $p < 0.057$	Power (%) If $RR = 5.0$	Power (%) If $RR = 10.0$
Fox and Collier (1, 2)	4	1.6	2.44	–	93	
Waxweiler (5)	10	2.3	4.27	+	98	
(10+ latency)	8	0.5	15.09	+		92
Waxweiler et al. (4) (15+ latency, 5+ exposure)	7	0.4	16.08	+		86
Byren et al. (11)	4	1.0	4.13	+	80	
(10+ latency)	4	0.7	5.89	+		98
Reinl et al. (8)	12	0.9	15.23	+	76	

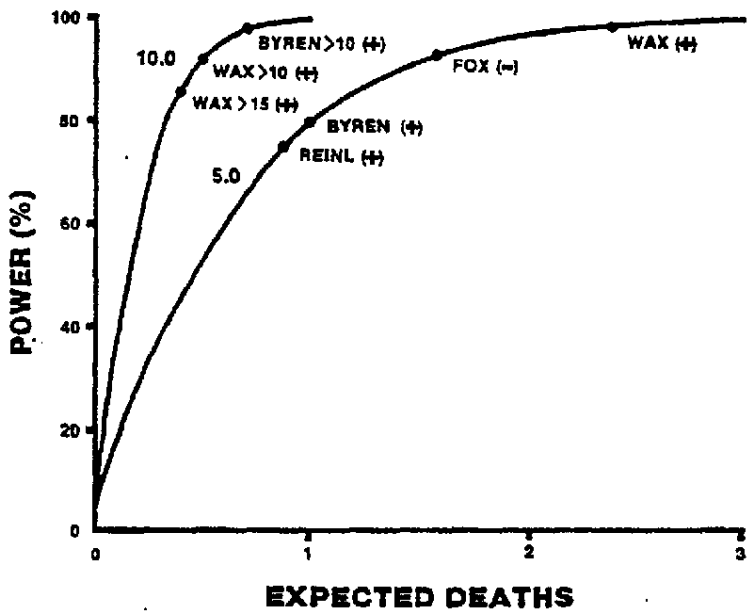

FIGURE 2. Approximate powers of vinyl chloride studies reporting results for liver cancer: 10.0 = assumed relative risk after a minimum latency, 5.0 = assumed relative risk overall. REINL = Reinl et al. (8); BYREN = Byren et al. (11); FOX = Fox and Collier (1, 2); WAX = Waxweiler et al. (4) and Waxweiler (5).

two yielded statistically significant results (4, 5, 9) (table 4 and figure 4). Again, there was a wide range of powers, from 14 per cent to almost 100 per cent, but here the studies with very high power were negative, an unlikely occurrence if vinyl chloride is indeed a human lung carcinogen.

DISCUSSION

The data regarding liver cancer in vinyl chloride exposed workers suggested

TABLE 3

Brain cancer: results and powers of epidemiologic studies of vinyl chloride

Author	Observed deaths	Expected deaths	Mortality ratio	Excess $p < 0.05$?	Power (%)	
					If RR = 2.0	If RR = 3.0
Fox and Collier (1, 2)	2	3.7	0.55	-	48	
Waxweiler (5)	9	4.3	2.09	+	53	
(10+ latency)	8	3.1	2.56	+		82
Waxweiler et al. (4)						
(15+ latency, 5+ exposure)	3	0.9	3.29	-		40
Equitable Environmental Health (6, 7) (1+ exposure)	12	5.9	2.03	+	64	
Byrne et al. (11)	2	0.3	6.12	+	12	
Reinl et al. (8)	2	1.3	1.62	-	24	

strongly that the chemical is carcinogenic for the liver. Three of the four studies reported statistically significant excess risks (4, 5, 8, 11), and the nonsignificant study (1, 2) reported a small excess (4 observed, 1.6 expected). The statistical powers of the studies were high (all above 75 per cent) due to the high observed relative risk. Even without knowledge of the animal studies, and without knowledge that angiosarcoma of the liver is extremely rare, one might conclude from these data that vinyl chloride is carcinogenic for the liver.

The powers of the studies that reported brain cancer ranged from 12 per cent to 80 per cent. The expected trend for a causative association was seen, in that the studies with high power were statistically significant and the studies with low power were, with one exception, nonsignificant. The most reasonable interpretation, therefore, is that the data are consistent with an etiologic hypothesis for vinyl chloride and cancer of the brain.

Only two of eight studies reporting lung cancer results showed significant excesses (4, 5, 9). While some of the negative results could be explained by low statistical power, two studies that were negative had very high power (1, 2, 6, 7). The lack of trend is evidence that vinyl chloride may not be a lung carcinogen, or that the actual relative risks were much lower than the assumed values of 1.5 and 2.0, in which case the powers were also much lower. It should be noted that one of the two negative studies with high power (by Fox and Collier (1, 2)) was also negative for liver cancer. Since it is generally accepted that vinyl chloride is a liver carcinogen, it may be that the exposures in that study were minimal. The other negative study with high power (Equitable) did not report liver cancer results, but did find a significant excess of brain cancer, which may constitute evidence of substantive expos

Examinat "positive" and "nega-

FIGURE 3
relative ris
REINL = 1
EQUIT = 1

tive" hy
light of
several
dence of
For exan
relative
bining t
bers fror
help of
which s
combine
for the l
the pres
relative
also a s
homoge
Reinl (8
bined r
cancer i

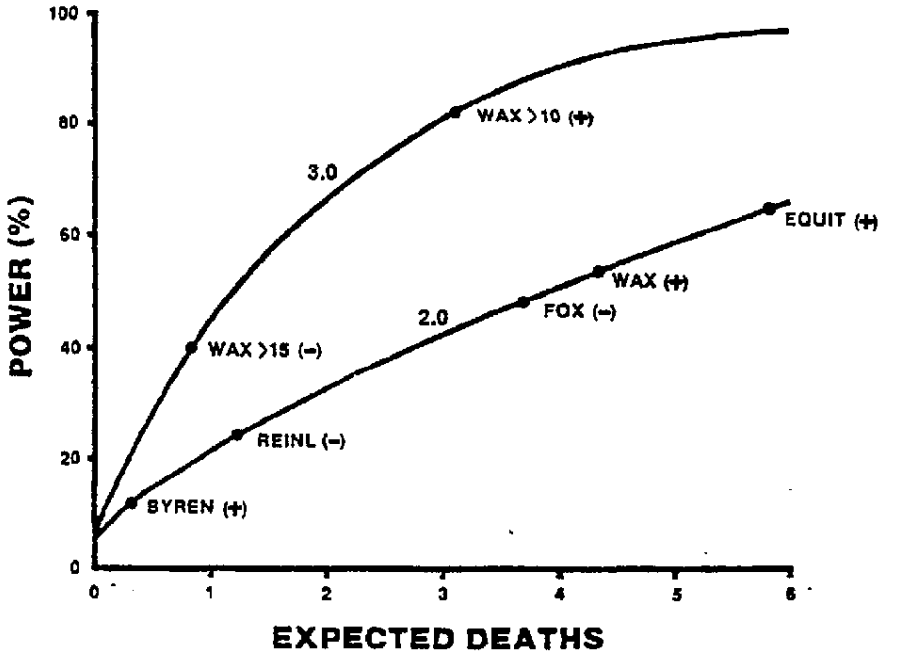

FIGURE 3. Approximate powers of vinyl chloride studies reporting results for brain cancer: 3.0 = assumed relative risk after a minimum latency, 2.0 = assumed relative risk overall. BYREN = Byren et al. (11); REINL = Reinl et al. (8); FOX = Fox and Collier (1, 2); WAX = Waxweiler et al. (4) and Waxweiler (5); EQUIT = Equitable Environmental Health (6, 7).

tive" hypothesis testing results in the light of statistical power is only one of several approaches to reviewing the evidence of carcinogenicity for a substance. For example, one can compute a combined relative risk for each cancer site (by combining the observed and expected numbers from the studies), and then, with the help of homogeneity testing, determine which studies are compatible with the combined risk estimate (see the Appendix for the homogeneity test). Liver cancer in the present review has a combined overall relative risk of 5.17 ($p < 0.00001$), and also a significant ($p = 0.002$) result in homogeneity testing, largely due to the Reinl (8) relative risk of 15.23. The combined relative risk for overall brain cancer is 1.74 ($p < 0.01$); here the results are more homogeneous ($p = 0.10$), although Byren (11) is somewhat of an outlier with a relative risk of 6.12. Finally, the combined relative risk for overall lung cancer is 1.06 (nonsignificant), with some evidence (homogeneity $p = 0.06$) that the results from Waxweiler (4, 5) and Buffler (9) are out of step with the others (relative risks of 1.49 and 2.89, respectively).

Close examination of the individual studies can also be helpful in searching for reasons for differing results. For example, the expected number of deaths from liver cancer in the Byren study (11) is unusually large relative to the expected numbers for brain and lung. Detailed examination of this anomaly might be informative. There are other consid-

TABLE 4

Lung cancer: results and powers of epidemiologic studies of vinyl chloride

Study	Observed deaths	Expected deaths	Mortality ratio	Excess $p < 0.05$?	Power (%)	
					If RR = 1.5	If RR = 2.0
Fox and Collier (1, 2)	46	51.2	0.90	–	94	
(15+ latency)	28	26.0	1.08	–		100
Waxweiler (5)	42	28.2	1.49	+	77	
(10+ latency)	39	24.9	1.56	+		99
Waxweiler et al. (4)						
(15+ latency, 5+ exposure)	11	5.7	1.94	+		63
Equitable Environmental Health						
(6, 7) (1+ exposure)	45	44.3	1.02	–	91	
(15+ latency, 1+ exposure)	41	39.2	1.05	–		100
Buffler et al. (9)	5	1.7	2.89	+	14	
(5+ latency, 2.29+ exposure)	4	1.0	3.81	+		21
Byren et al. (11)	3	1.8	1.68	–	15	
Duck et al. (12)	16	15.5	1.03	–	55	
Heinl et al. (8)	22	24.6	0.95	–	76	
Ott et al. (10)	4	5.2	0.77	–	27	

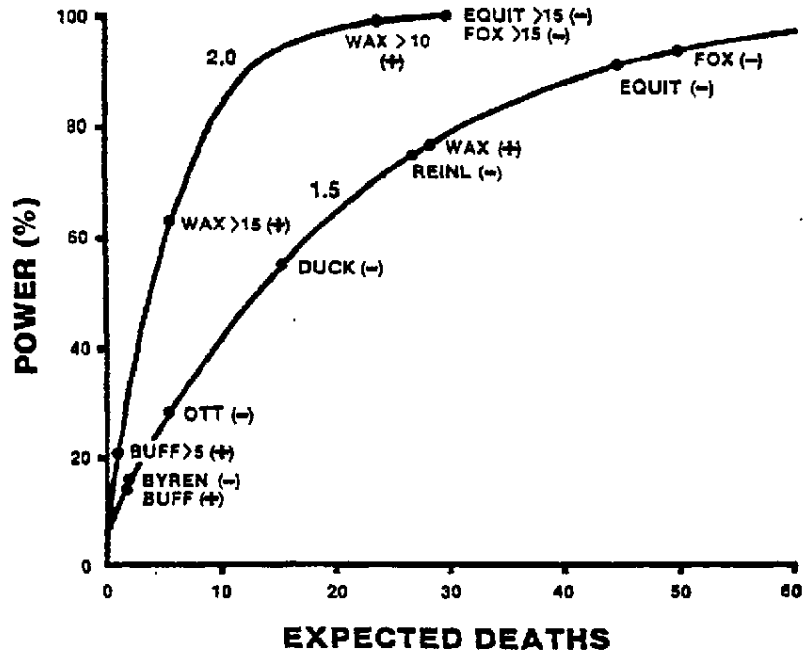

FIGURE 4. Approximate powers of vinyl chloride studies reporting results for lung cancer: 2.0 = assumed relative risk after a minimum latency, 1.5 = assumed relative risk overall. BUFF = Buffler et al. (9); BYREN = Byren et al. (11); OTT = Ott et al. (10); DUCK = Duck et al. (12); REINL = Reinl et al. (8); WAX = Waxweiler et al. (4, 5); EQUIT = Equitable Environmental Health (6, 7); FOX = Fox and Collier (1, 2).

erations, such as dose-response, degree of exposure, concomitant exposures, and confidence limits for risk ratios, that also need to be kept in mind. While consideration of statistical power cannot provide a complete answer, it is one more useful way of looking at epidemiologic evidence.

REFERENCES

1. Fox AJ, Collier PF. Low mortality rates in industrial cohort studies due to selection for work and survival in the industry. Br J Prev Soc Med 1976;30:225-30.
2. Fox AJ, Collier PF. Mortality experience of workers exposed to vinyl chloride monomer in the manufacture of polyvinyl chloride in Great Britain. Br J Ind Med 1977;34:1-10.
3. Nicholson WJ, Hammond EC, Seidman H, et al. Mortality experience of a cohort of vinyl chloride-polyvinyl chloride workers. Ann NY Acad Sci 1975;246:225-30.
4. Waxweiler RJ, Stringer W, Wagoner JK, et al. Neoplastic risk among workers exposed to vinyl chloride. Ann NY Acad Sci 1976;271:40-8.
5. Waxweiler RJ. An epidemiologic investigation of lung cancer in a multixenobiotic environment. Ph.D. Dissertation. Chapel Hill, NC: University of North Carolina, 1978.
6. Tabershaw IR, Gaffey WR. Mortality study of workers in the manufacture of vinyl chloride and its polymers. J Occup Med 1974;16:509-18.
7. Equitable Environmental Health, Inc. Epidemiologic study of vinyl chloride workers. Final report submitted to Manufacturing Chemists Association, 1978.
8. Reinl W, Weber H, Greiser E. Diseases caused by vinyl chloride. Special report of the State Industrial Physician, Dusseldorf, West Germany, 1978.
9. Buffler PA, Wood S, Clayton E, et al. Mortality experience of workers in a vinyl chloride monomer production plant. J Occup Med 1979;21:195-203.
10. Ott MG, Langner RR, Holder BB. Vinyl chloride exposure in a controlled industrial environment. Arch Environ Health 1975;30:333-9.
11. Byren D, Engholm G, Englund A, et al. Mortality and cancer morbidity in a group of Swedish

VCM and PVC production workers. Environ Health Perspect 1976;17:167-70.
12. Duck BW, Carter JT, Coobes EJ. Mortality study of workers in a polyvinylchloride production plant. Lancet 1975;2:1197-9.
13. Freiman JA, Chalmers TC, Smith H, et al. The importance of beta, the Type II error, and sample size in the design and interpretation of the randomized control trial. N Engl J Med 1978;299:690-94.
14. Cutler SJ, Schneiderman MA, Greenhouse SW. Some statistical considerations in the study of cancer in industry. Am J Public Health 1954;44:1159-66.
15. Molina EC. Poisson's exponential binomial limit. Huntington, NY: Robert E. Krieger Publishing Company, 1973.
16. Armitage P. Statistical methods in medical research. Oxford: Blackwell Scientific Publications, 1971.

Appendix

The approximate power formula assumes that the observed number of deaths from the cause of interest follows a Poisson distribution with mean RE, where E is the expected number based on standard population rates and R is the ratio of rates (relative risk) for the study as opposed to the standard population. It follows that for large values of RE, the square root of the observed deaths is approximately normally distributed with mean $\sqrt{RE}$ and variance ¼ (16). Now, if a random quantity has a normal distribution with mean μ and variance σ^2, its power $1 - \beta$ to reject the null hypothesis $\mu = \mu_0$ at the α level of significance is given by

$$Z_{1-\beta} = Z_\alpha - \frac{\mu - \mu_0}{\sigma}$$

Substitution of $\mu = \sqrt{RE}$, $\mu_0 = \sqrt{E}$, and $\sigma = ½$ into this expression yields the required result.

The homogeneity test for the combined relative risk uses the chi-square distribution with $n - 1$ degrees of freedom. The equation is

$$\chi^2_{(n-1)} = \sum_{i=1}^{n} \frac{(O_i - \hat{\theta}E_i)^2}{\hat{\theta}E_i}$$

where n = the number of studies,
O_i = the observed deaths from the ith study,
E_i = the expected deaths from the ith study,
and $\hat{\theta}$ = $\sum O_i / \sum E_i$, the combined relative risk estimate.