Item ID Number 01699

Author Carroll, Ray .J.

Corporate Author

Report/Article Title Typescript: Report #2, False Negatives and their Effect on Power of Statistical Tests of Differences in Disease Rates, August 1982

Journal/Book Title

Year 0000

Month/Day

Color ☐

Number of Images 4

Descripton Notes See Item 1773 for a letter from Carroll that mentions this report.

REPORT #2

FALSE NEGATIVES AND THEIR EFFECT ON POWER OF STATISTICAL TESTS OF DIFFERENCES IN DISEASE RATES

R.J. Carroll

August 1982

In REPORT #1, especially Appendix #3, I calculated statistical power for tests to see if disease rates are the same for exposed and non-exposed soldiers. I concluded that, if the disease rate in the non-exposed population was at all substantial (say 5%), then samples of 6,000 for each group would give very high statistical power for detecting a relative risk due to exposure of 2.0 or more, even if there was substantial misclassification of exposed into the non-exposed group (false negatives). In this report, I will study the effect of false negatives on statistical power when the disease rate in the non-exposed population is fairly small.

To review terminology, recall that

p_1 = disease probability for truly exposed
q_1 = disease probability for reported exposed = p_1
p_2 = disease probability for truly not exposed
q_2 = disease probability for reported not exposed
$r_{true} = p_1/p_2$ = true relative risk
$r_{obs} = q_1/q_2$ = observed relative risk
δ = fraction of false negatives
N_1 = sample size for reported exposed
N_2 = sample size for reported not exposed.

Before constructing a test for equality of exposed and not exposed, we need two concepts. The *level* or *Type 1 error* α of a test is the probability (usually α = .05 or 5%) that we will conclude that the exposed have different disease probabilities than do the non-exposed, but in reality the two are not different.

The *power* of a test $(1 - \beta)$ is the probability we will conclude that the exposed differ from the non-exposed when they really do differ. Generally, we want small Type I error but high power; UCLA chooses $\alpha = .01$, $1-\beta = .95$.

One point not made clear in the UCLA protocol is that they compute sample size based upon a one-sided test and not the usual and more conservative two-sided test. In this particular instance I think what they have done is justifiable; we can hardly expect exposure to agent orange to have improved health status.

If we assume equal sample sizes so that $N_1 = N_2 = N$, then based upon UCLA's arcsin transformation, the statistical power of the usual test is given by the formula

$$1 - \Phi(z_{1-\alpha} - \sqrt{\frac{N}{2}}(2\arcsin\sqrt{q_1} - 2\arcsin\sqrt{q_2}))$$

where $z_{1-\alpha}$ is the $(1-\alpha)100\%$ percentile of a standard normal distribution function Φ and

$$q_1 = p_1$$

$$q_2 = \delta p_1 + (1-\delta)p_2 \quad \text{(Report \#1, Appendix \#1).}$$

If we want to detect a true relative risk r_{true} of at least r_{detect}, we will then need a minimum sample size of at least

$$N = \frac{2(z_{1-\beta} + z_{1-\alpha})^2}{(2\arcsin\sqrt{p_2 r_{detect}} - 2\arcsin\sqrt{\delta p_2 r_{detect} + (1-\delta)p_2})^2}$$

if we want power of $1-\beta$. Note that this is exactly the formula on page 108 of the UCLA protocal as long as there are no false negatives. Some idea of the effect of false negatives can be learned from Table #1. Figure #1 shows this effect even more dramatically. Note that if we desire a Type I error of $\alpha = .01$ and power $1-\beta = .95$ to detect a true relative risk of 2.0 or more, then we need about 4,500 soldiers in each group if there are no false negatives; we will need over 7,600 soldiers in each group if there are 20% false negatives.

This report shows that for diseases which are fairly rare, the effect of misclassification can be severe.

TABLE #1

EFFECT OF FALSE NEGATIVES ON CALCULATION OF SAMPLE SIZE

α	$1-\beta$	r_{detect}	p_2	δ	UCLA N	Actual N
.01	.95	2.0	.005	0.0	9,124	9,124
				.1	9,124	11,721
				.2	9,124	15,397
				.3	9,124	20,828
				.4	9,124	29,305
.01	.95	2.0	.01	0.0	4,528	4,528
				.1	4,528	5,815
				.2	4,528	7,637
				.3	4,528	10,328
				.4	4,528	14,527
.01	.95	2.0	.02	0.0	2,230	2,230
				.1	2,230	2,863
				.2	2,230	3,757
				.3	2,230	5,078
				.4	2,230	7,138

FIGURE #1

Sample Sizes for $\alpha = .01$, $1-\beta = .95$, $\delta = .20$, $F_{true} = 2.($