11901 Samuel Drive, Garden Grove, CA 92840-2546

Item ID Number 01698

Author Carroll, Ray .J.

Corporate Author

Report/Article Title Typescript: Report #1 (with typographical corrections), False Negatives and their Effect on Estimates of the Risk of Exposure to Agent Orange, August 1982

Journal/Book Title

Year 0000

Month/Day

Color ☐

Number of Images 9

Descripton Notes

REPORT #1
(with typographical corrections)

FALSE NEGATIVES AND THEIR EFFECT ON ESTIMATES OF THE RISK OF EXPOSURE TO AGENT ORANGE

R.J. Carroll

August 1982

I am going to assume that a large random sample is taken from individuals known to be at risk (e.g. combat troops) and that another large random sample is taken from those thought not to be at risk (rear-echelon troops). As explained to me, this is not really possible because only battalions and not individuals can easily be sampled for their exposure levels.

I will also assume that all those reported to have been exposed actually were exposed: no false positives. A certain fraction δ are false negatives, i.e., if $\delta = .25$, 25% of those thought not to be at risk (exposed) actually were. These false negatives do have an effect on estimating risk of disease due to exposure.

Let p_1 = true probability of disease for exposed individuals, and let p_2 = true probability of disease for non-exposed individuals. The relative risk is defined to be

$$r_{true} = p_1/p_2 .$$

In practice we might get the following table:

	Reported Exposed	Reported Not Exposed
"sick"	N_E	N_{NE}
"well"	N_1-N_E	N_2-N_{NE}
	N_1	N_2

The sample sizes N_1 and N_2 of those reported exposed and reported not exposed (respectively) have been fixed in advance. N_E represents the number of "sick" individuals among those reported exposed, while N_{NE} is the number of "sick" individuals among those reported not exposed. The observed relative risk can be computed by

$$\hat{r}_{observed} = \frac{(N_E/N_1)}{(N_{NE}/N_2)}$$

Because of the false negatives, the observed relative risk estimates not the true value r_{true} but rather

$$\hat{r}_{observed} \approx \frac{r_{true}}{\delta\, r_{true} + (1-\delta)}$$

For example, if the true relative risk is 2.0 (exposed are twice as likely to be "sick" as non-exposed) and if we have a false negative rate of 25%, then we would report a relative risk of only

$$\hat{r}_{observed} \approx \frac{2.}{(.25)(2.) + (1-.25)} = 1.6 \ .$$

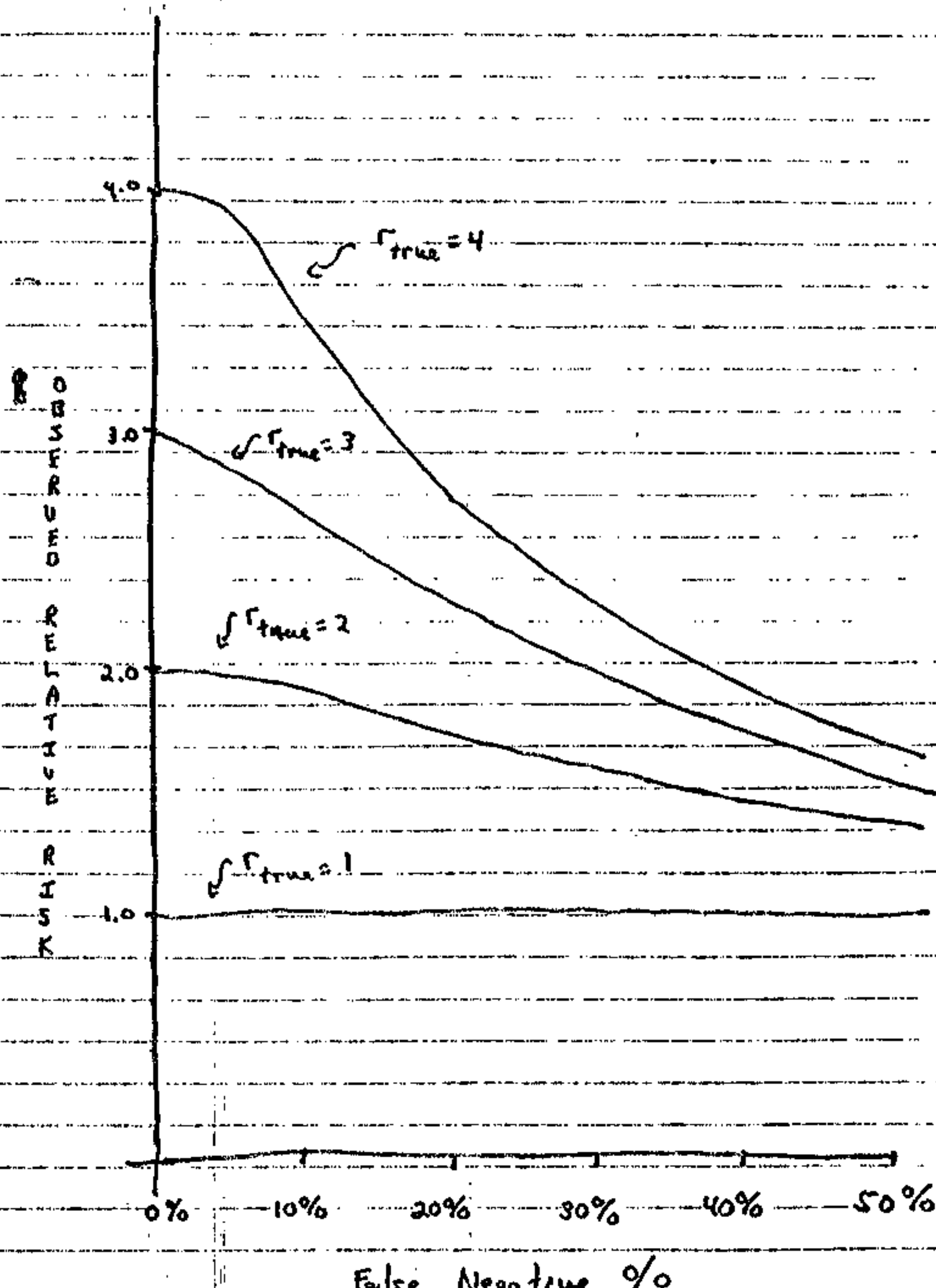
Another way to look at this example is as follows. Suppose 10% of the truly exposed become "sick." Since the true relative risk is 2.0, only 5% = 10% ÷ 2.0 of the truly non-exposed become "sick." However, because of the false negatives, we will announce that 10% ÷ 1.6 = 6.25% of the reported non-exposed become "sick."

TABLE 1 (Selected Values)

True Relative Risk	False Negative %	Observed Relative Risk
1.0	0	1.0
1.0	50	1.0
2.0	0	2.0
	10%	1.82
	25%	1.60
	50%	1.33
3.0	0	3.0
	10%	2.5
	25%	2.0
	50%	1.5
4.0	0	4.0
	10%	3.08
	25%	2.29
	50%	1.60

THE EFFECT OF FALSE NEGATIVES ON RELATIVE RISK.

Figure # 1



APPENDIX #1 (8/14/82)

$p_1 = P_r(\text{Sick/Exposed})$

$p_2 = P_r(\text{Sick/Not Exposed})$

True relative risk of exposure is

$$r_{true} = p_1/p_2$$

$q_1 = P_r(\text{Sick/Reported Exposed})$

$q_2 = P_r(\text{Sick/Reported Not Exposed})$

Suppose there are no false positives, i.e.,

$$\Pr(\text{Exposed/Reported Exposed}) = 1.$$

Further, suppose a certain percentage of false negatives is possible,

$$\Pr(\text{Exposed/Reported Not Exposed}) = \delta.$$

Then, the probabilities q_1, q_2 solve

$$q_1 = p_1$$

$$q_2 = \Pr(\text{Sick/Exposed, Report Not Exposed})\delta$$

$$+ \Pr(\text{Sick/Not Exposed, Report Not Exposed})(1-\delta).$$

If we further assume that whether the person becomes sick depends only on exposure and not on reported exposure, we get

$$q_2 = \delta p_1 + (1-\delta)p_2$$

Thus, by using the misclassified table, you will be estimating an *observed* relative risk of exposure as

$$r_{obs} = \frac{q_1}{q_2} = \frac{p_1}{\delta p_1 + (1-\delta)p_2} = \frac{(p_1/p_2)}{[\delta p_1 +)1-\delta)p_2]/p_2}$$

$$= r_{true}/\{\delta r_{true} + (1-\delta)\}.$$

In this instance, the observed relative risk will underestimate the true relative risk, and this bias depends heavily on the false positive rate δ.

What about the excess number of cases? Per 1,000 individuals,

$$1{,}000\ p_1 \text{ truly exposed become sick}$$

$$1{,}000\ p_2 \text{ truly not exposed become sick}$$

True excess number/1,000 is

$$1{,}000\ (p_1-p_2) = 1{,}000\ (r_{true}-1)p_2\ .$$

Thus if 1% of the truly not exposed become sick and $r_{true} = 2.0$, then from 1,000 truly exposed individuals we can expect 10 more sick persons than we can from 1,000 truly not exposed.

However, if we have a false negative rate of .50 (50%), then our estimated excess number will not be 10 but will be

$$1{,}000\ (r_{observed}-1)p_2 = 1{,}000(1.33-1)(.01) = 3.3\ .$$

(see Table 2)

This is a rather dramatic difference.

A technical note. My definition of false negative is as on page 4 of Fleiss, not as on his page 135 (Section 11.2). His calculations in Section 11.2 are for a retrospective study (# of sicks and wells fixed in advance).

TABLE 2 (More Details than Table 1)

True Relative Risk	False Negative Rate (Not %)	Observed Relative Risk
1	0.000	1.00
1	0.025	1.00
1	0.050	1.00
1	0.075	1.00
1	0.100	1.00
1	0.125	1.00
1	0.150	1.00
1	0.175	1.00
1	0.200	1.00
1	0.225	1.00
1	0.250	1.00
1	0.275	1.00
1	0.300	1.00
1	0.325	1.00

TABLE 2 (More Details than Table 1) cont'd

True Relative Risk	False Negative Rate (Not %)	Observed Relative Risk
1	0.350	1.00
1	0.375	1.00
1	0.400	1.00
1	0.425	1.00
1	0.450	1.00
1	0.475	1.00
1	0.500	1.00
2	0.000	2.00
2	0.025	1.95
2	0.050	1.90
2	0.075	1.86
2	0.100	1.81
2	0.125	1.77
2	0.150	1.73
2	0.175	1.70
2	0.200	1.66
2	0.225	1.63
2	0.250	1.60
2	0.275	1.56
2	0.300	1.53
2	0.325	1.50
2	0.350	1.48
2	0.375	1.45
2	0.400	1.42
2	0.425	1.40
2	0.450	1.37
2	0.475	1.35
2	0.500	1.33
3	0.000	3.00
3	0.025	2.85
3	0.050	2.72
3	0.075	2.60
3	0.100	2.50
3	0.125	2.40
3	0.150	2.30
3	0.175	2.22
3	0.200	2.14
3	0.225	2.06
3	0.250	2.00
3	0.275	1.93
3	0.300	1.87
3	0.325	1.81
3	0.350	1.76
3	0.375	1.71
3	0.400	1.66
3	0.425	1.62
3	0.450	1.57
3	0.475	1.53
3	0.500	1.50
4	0.000	4.00
4	0.025	3.72
4	0.050	3.47
4	0.075	3.26
4	0.100	3.07
4	0.125	2.90
4	0.150	2.75
4	0.175	2.62
4	0.200	2.50
4	0.225	2.38
4	0.250	2.28
4	0.275	2.19
4	0.300	2.10
4	0.325	2.02
4	0.350	1.95
4	0.375	1.88
4	0.400	1.81
4	0.425	1.75
4	0.450	1.70
4	0.475	1.64
4	0.500	1.60

APPENDIX #2 (8/14/82)

Suppose we sample N_1 who are reported exposed and N_2 who are reported unexposed. What is the probability of detecting a relative risk different from 1 if we ignore the effects of misclassification? For this prospective study, the usual test says (assuming a fairly large percentage (7.5%) are diseased) says that the that the relative risk differs significantly from 1 if

$$|\log_e \hat{r}_{obs}| > 1.96\sqrt{\frac{(1-q_1)}{N_1} + \frac{(1-q_2)}{N_2}}$$

$$= 1.96\ \sigma(q_1,q_2,N_1,N_2)\ .$$

If the normal probability function is called Φ, the statistical power is

$$2 - \Phi(1.96 - (\log_e r_{obs})/\sigma(q_1,q_2,N_1,N_2))$$

$$- \Phi(1.96 + (\log_e r_{obs})/\sigma(q_1,q_2,N_1,N_2))$$

Now,

$$\log_e r_{obs} = \log_e r_{true} - \log_e\{\delta r_{true}+(1-\delta)\}$$

If $N_1 = N_2$, then

$$\sigma^2(q_1,q_2,N_1,N_2) = \frac{1}{N}[(1-p_1) + 1 - \delta p_1 - (1-\delta)p_2]$$

$$= \frac{1}{N}[2 - p_1 - \delta p_1 - (1-\delta)p_2]$$

$$= \frac{p_2}{N}[\frac{2}{p_2} - (1+\delta)r_{true}-(1-\delta)]$$

Hence the power of the usual test depends not only on the relative risk but also on the probability of disease in the truly unexposed population.

APPENDIX #3 (8/14/82)

Suppose we observed a table such as

	Reported Exposed	Reported Not Exposed
sick	N_E	N_{NE}
well	N_1-N_E	N_2-N_{NE}
	N_1	N_2

How can we estimate the true relative risk, and how much power do we have for detecting relative risks different from one? Recall, I am assuming q_2 is fairly large. From Appendix #1 (8.14.82), page 4,

$$r_{true} = \frac{(1-\delta)\ r_{obs}}{[1 - \delta\ r_{obs}]}$$

Since an estimate of q_1 is

$$\hat{q}_1 = N_E/N_1$$

and an estimate of q_2 is

$$\hat{q}_2 = N_{NE}/N_2 \ ,$$

the estimate of r_{obs} is

$$\hat{r}_{obs} = \frac{\hat{q}_1}{\hat{q}_2} = \frac{(N_E/N_1)}{(N_{NE}/N_2)}$$

and the estimate of r_{true} is

$$\hat{r}_{true} = \frac{(1-\delta)\ \hat{r}_{obs}}{[1 - \delta\ \hat{r}_{obs}]}$$

It turns out that $\log_e \hat{r}_{true}$ is almost normally distributed with mean $\log r_{true}$ and variance

$$(*) \qquad \frac{\sigma^2(q_1,q_2,N_1,N_2)}{[1 - \delta\ r_{obs}]^2} \ ,$$

$$= \sigma_*^2(q_1,q_2,N_1,N_2,\delta, r_{obs}) \ .$$

The power of the test for relative risk different from one is

$$2 - \Phi(1.96 - \log_e r_{true}/\sigma_*(q_1,q_2,N_1,N_2,\delta,r_{true}))$$

$$- \Phi(+1.96 + \log_e r_{true}/\sigma_*(q_1,q_2,N_1,N_2,\delta,r_{true}) \ .$$

(Highly technical, but to keep all the details written down) I want to show that (*) is true. We have

$$\log \hat{r}_{true} - \log r_{true}$$

$$= [\log \hat{r}_{obs} - \log r_{obs}]$$

$$- [\log\{1-\delta\hat{r}_{obs}\} - \log\{1-\delta r_{obs}\}]$$

$$\doteq \{\hat{r}_{obs} - r_{obs}\}\{\frac{1}{r_{obs}} + \frac{\delta}{1 - \delta\ r_{obs}}\}$$

$$\doteq \frac{\{\hat{r}_{obs} - r_{obs}\}}{r_{obs}} \times \{\frac{1}{1 - \delta\ r_{obs}}\}$$