7. DESIGN CONSIDERATIONS

CHAPTER 7

DESIGN CONSIDERATIONS

7.1 Introduction

In Chapter 1, we considered a range of questions concerned with the implementation of a cohort study. In this chapter, we concentrate on the more formal aspects of study design, in particular power, efficiency and study size. The design issues considered initially in this chapter are based, in large part, on the analytical methods of Chapters 2 and 3, comprising simple comparisons of a group with an external standard, internal comparisons within a cohort, and tests for trend using the approach of §3.6. Power considerations based on the modelling approach of Chapters 4 and 5 are only touched on.

The design of case-control studies is considered at some length. The motivation comes principally from the concept of risk-set sampling introduced in Chapter 5, but the results apply to general case-control studies. Topics discussed include the choice of matching criteria, the number of controls to select, and the effects that control of confounding or an interest in interaction will have on study size requirements. Attention is focused on the simple situation of one, or a small number, of dichotomous variables.

Two approaches are taken to the evaluation of different study designs; the first is based on calculation of the power function, the second is based on the expected standard errors of the relevant parameters. The power considerations are based on one-sided tests of significance unless specifically stated to the contrary, since in most studies the direction of the main effect of interest is an inherent part of the specification of the problem under study. The discussion of the design of cohort studies assumes that external rates are known, even though the analysis may be based on internal comparison and does not use external rates. The reason is evident – that evaluation of the potential performance of a study before it is carried out must be based on information exterior to the study. Since in this chapter all expected numbers are based on external rates, we have dispensed with the notation used in earlier chapters, where expected numbers based on external rates are starred.

It needs stressing strongly that power calculations are essentially approximate. The size, age composition and survival of the cohort will usually not be known with any great accuracy before the study is performed. In addition, calculations are generally based assuming a Poisson distribution for the observed events, since they derive from the statistical methods of Chapters 2 and 3. Many data may be affected by extra

Poisson variation, which will augment the imprecision in probability statements. Furthermore, the level of excess risk that one decides that it is important to detect is to some extent arbitrary.

7.2 Sample size for cohort studies – comparison with an external standard

This section considers the design of studies in which the total cohort experience is to be compared to an external standard. It is assumed that analyses are in terms of the SMR, with tests of significance and construction of confidence intervals following the methods of Chapter 2.

The number of deaths, D, of the disease of interest (or number of cases if cancer registry material is available) is to be determined in the cohort, and compared with the number expected, E, based on rates for some external population, whether national or local. The relative risk is measured by the ratio D/E, the SMR. Tests of significance for departures of the SMR from its null value of unity and the construction of confidence intervals were discussed in §2.3. The capacity of a given study design to provide satisfactory inferences on the SMR can be judged in two ways: first, in terms of the capacity of the design to demonstrate that the SMR differs significantly from unity, when in fact it does, and, second, in terms of the width of the resulting confidence intervals, and the adequacy of the expected precision of estimation.

The first approach proceeds as follows. For an observed number of deaths, D, to be significantly greater than the expected number, E, using a one-sided test at the $100\alpha\%$ level, it has to be greater than or equal to the α point of the Poisson distribution with mean E, a point that we shall denote by $C(E, \alpha)$. (For a two-sided test, α is replaced by $\alpha/2$.) Since the Poisson is a discrete distribution, the exact α point does not usually exist, and we take $C(E, \alpha)$ to be the smallest integer such that the probability of an observation greater than or equal to $C(E, \alpha)$ is less than or equal to α. Table 7.1 gives the value of $C(E, \alpha)$ for $\alpha = 0.05$ and 0.01, and a range of values of E. If, however, the true value of the SMR is equal to R, then the observed number of deaths will follow a Poisson distribution with mean RE. The probability of a significant result is then the probability that D, following a Poisson distribution with mean RE, is greater than or equal to $C(E, \alpha)$. For given values of E and α, this probability depends only on R. It is simple if somewhat laborious to calculate and is known as the power function of the study. Common practice is to choose a value of R that one feels is the minimum that should not pass undetected, and to calculate the power for this value. Table 7.2 gives the power for a range of values of E and R, for α equal to 0.05 and 0.01, respectively. The values in the column $R = 1$ are, of course, simply the probabilities of rejecting the null hypothesis when in fact it is true, and so give the real significance of the test, rather than the nominal 5% or 1%; one can see in Table 7.2a that they are all less than 5%, and in Table 7.2b all less than 1%.

Example 7.1

Suppose that with a given study cohort and the applicable mortality rates, there is an expected number of 20 deaths. Then, all observed values greater than or equal to 29 will be significant at the 5% level, and all values greater than or equal to 32 will be significant at the 1% level (Table 7.1). These are the values $C(20, 0.05)$ and $C(20, 0.01)$, respectively. If the true value of the relative risk is 1.5, then the true expected

Table 7.1 5% and 1% points of the Poisson distribution for different values of the mean. The numbers tabulated are the smallest integers for which the probability of being equalled or exceeded is less than 5% and 1% (designated $C(E, 0.05)$ and $C(E, 0.01)$), respectively.

Mean of Poisson distribution, E	$C(E, 0.05)$	$C(E, 0.01)$	Mean (E)	$C(E, 0.05)$	$C(E, 0.01)$
1	4	5	20	29	32
2	6	7	25	34	38
3	7	9	30	40	44
4	9	10	35	46	50
5	10	12	40	52	56
6	11	13	45	57	62
7	13	15	50	63	68
8	14	16	60	74	80
9	15	18	70	85	91
10	16	19	80	96	103
11	18	20	90	107	114
12	19	22	100	118	125
13	20	23			
14	21	24			
15	23	26			

Table 7.2 Comparison with an external standard

(a) Probability (%) of obtaining a result significant at the 0.05 level (one-sided) for varying values of the expected value E assuming no excess risk, and of the true relative risk R

Expected number of cases assuming no excess risk ($R = 1$)	True relative risk (R)									
	1.0	1.5	2.0	3.0	4.0	5.0	7.5	10.0	15.0	20.0
1.0	1.90	7	14	35	57	74	94	99	100	100
2.0	1.66	8	21	55	81	93	100			
3.0	3.35	17	39	79	95	99				
4.0	2.14	15	41	84	98	100				
5.0	3.18	22	54	93	100					
6.0	4.26	29	65	97	100					
7.0	2.70	26	64	98	100					
8.0	3.42	32	73	99	100					
9.0	4.15	38	79	100						
10.0	4.87	43	84	100						
11.0	3.22	39	83	100						
12.0	3.74	44	87	100						
13.0	4.27	48	90	100						
14.0	4.79	53	93	100						
15.0	3.27	49	92	100						
20.0	3.43	60	97	100						

Table 7.2 (*contd*)

Expected number of cases assuming no excess risk ($R = 1$)	True relative risk (R)									
	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
20.0	3.43	9	18	30	45	60	73	83	90	94
25.0	4.98	13	26	42	59	74	85	92	96	98
30.0	4.63	13	27	46	64	79	89	95	98	99
35.0	4.25	13	29	49	69	83	92	97	99	100
40.0	3.87	13	30	52	72	86	94	98	99	100
45.0	4.73	16	36	60	79	91	97	99	100	
50.0	4.24	16	37	61	81	93	98	99	100	
60.0	4.42	18	42	69	88	96	99	100		
70.0	4.48	19	47	75	92	98	100			
80.0	4.46	21	51	80	94	99	100			
90.0	4.39	22	55	83	96	99	100			
100.0	4.28	23	58	86	97	100				

(b) Probability (%) of obtaining a result significant at the 0.01 level (one-sided) for varying values of the expected value E assuming no excess risk, and of the true relative risk R

Expected number of cases assuming no excess risk ($R = 1$)	True relative risk (R)									
	1.0	1.5	2.0	3.0	4.0	5.0	7.5	10.0	15.0	20.0
1.0	0.37	2	5	18	37	56	87	97	100	100
2.0	0.45	3	11	39	69	87	99	100		
3.0	0.38	4	15	54	84	96	100			
4.0	0.81	8	28	76	96	99	100			
5.0	0.55	8	30	82	98	100				
6.0	0.88	12	42	91	99	100				
7.0	0.57	11	43	93	100					
8.0	0.82	16	53	97	100					
9.0	0.53	14	53	97	100					
10.0	0.72	18	62	99	100					
11.0	0.93	22	69	99	100					
12.0	0.61	20	69	100						
13.0	0.76	24	75	100						
14.0	0.93	28	80	100						
15.0	0.62	26	79	100						
20.0	0.81	38	91	100						

Table 7.2 (*contd*)

Expected number of cases assuming no excess risk ($R = 1$)	True relative risk (R)									
	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
20.0	0.81	3	7	14	25	38	52	66	77	86
25.0	0.92	3	9	19	33	49	65	78	87	93
30.0	0.97	4	11	23	40	58	74	85	93	97
35.0	0.98	4	12	27	46	65	81	91	96	98
40.0	0.97	5	14	31	52	71	86	94	98	100
45.0	0.93	5	15	34	57	76	89	96	99	100
50.0	0.89	5	17	37	61	81	92	97	99	100
60.0	0.78	5	19	43	68	87	96	99	100	
70.0	0.91	6	24	51	77	92	98	100		
80.0	0.76	6	25	55	81	95	99	100		
90.0	0.83	7	29	62	87	97	100			
100.0	0.88	9	34	68	91	98	100			

value will be $20 \times 1.5 = 30$. The probability that an observation from a Poisson distribution with mean 30 is greater than or equal to 29 is 60% (Table 7.2) and that it is greater than or equal to 32 is 38% (Table 7.2). There is thus 60% power of obtaining a result significant at the 5% level, and 38% power of obtaining a result significant at the 1% level, if the true relative risk is 1.5.

An alternative way of expressing the power of a study is to give the relative risk for which the power is equal to a certain quantity, such as 80% or 95%. Table 7.3 gives the relative risks for a range of values of E and of the power, for 0.05 and 0.01 levels of significance, respectively.

Example 7.1 (contd)

To continue the previous example, with E equal to 20, using a 5% significance test, 50% power is obtained if the relative risk is 1.43, 80% power if R is 1.67 and 95% power if R is 1.92. The corresponding figures for 1% significance are relative risks of 1.58, 1.83 and 2.09.

The values given in Tables 7.2 and 7.3 are based on exact Poisson probabilities. To calculate power values for other values of E and R, one can use one of the approximations to the Poisson distribution suggested in Chapter 2. For example, one can use expression (2.12), the square root transformation, from which the quantity

$$\chi = 2(D^{1/2} - E^{1/2})$$

is approximately a standard normal deviate. If Z_α is the α point of the normal distribution, then for D to be significant at the 5% level (one-sided as before) we must have

$$2(D^{1/2} - E^{1/2}) \geq Z_\alpha$$

or

$$D \geq \{E^{1/2} + (Z_\alpha)/2\}^2.$$

This value corresponds to the value $C(E, \alpha)$ of the discussion in the previous pages.

Table 7.3 Comparison with an external standard

(a) True value of the relative risk required to have given power of achieving a result significant at the 5% level (one-sided), for varying values of the expected value E assuming no excess risk ($R = 1$)

Expected cases ($R = 1$)	Probability of declaring significant ($p \leq 0.05$) difference				
	0.50	0.80	0.90	0.95	0.99
1.0	3.67	5.52	6.68	7.75	10.05
2.0	2.84	3.95	4.64	5.26	6.55
3.0	2.22	3.03	3.51	3.95	4.86
4.0	2.17	2.84	3.25	3.61	4.35
5.0	1.93	2.50	2.84	3.14	3.76
6.0	1.78	2.28	2.57	2.83	3.36
7.0	1.81	2.27	2.54	2.78	3.26
8.0	1.71	2.13	2.37	2.58	3.02
9.0	1.63	2.01	2.24	2.43	2.83
10.0	1.57	1.92	2.13	2.31	2.67
11.0	1.61	1.95	2.15	2.32	2.66
12.0	1.56	1.88	2.06	2.22	2.55
13.0	1.51	1.82	1.99	2.14	2.45
14.0	1.48	1.77	1.93	2.08	2.36
15.0	1.51	1.79	1.95	2.09	2.37
20.0	1.43	1.67	1.80	1.92	2.15
25.0	1.35	1.55	1.67	1.77	1.96
30.0	1.32	1.51	1.61	1.70	1.87
35.0	1.30	1.47	1.57	1.65	1.81
40.0	1.29	1.45	1.54	1.61	1.76
45.0	1.26	1.41	1.49	1.55	1.69
50.0	1.25	1.39	1.47	1.53	1.66
60.0	1.23	1.35	1.42	1.48	1.59
70.0	1.21	1.32	1.39	1.44	1.54
80.0	1.20	1.30	1.36	1.41	1.50
90.0	1.19	1.28	1.34	1.38	1.47
100.0	1.18	1.27	1.32	1.36	1.45

(b) True value of the relative risk required to have given power of achieving a result significant at the 1% level (one-sided), for varying values of the expected value E assuming no excess risk ($R = 1$)

Expected cases ($R = 1$)	Probability of declaring significant ($p \leq 0.01$) difference				
	0.50	0.80	0.90	0.95	0.99
1.0	4.67	6.72	7.99	9.15	11.60
2.0	3.33	4.54	5.27	5.92	7.29
3.0	2.89	3.79	4.33	4.81	5.80
4.0	2.42	3.13	3.55	3.93	4.70
5.0	2.33	2.96	3.32	3.64	4.30
6.0	2.11	2.65	2.96	3.24	3.80

Table 7.3 (*contd*)

Expected cases ($R=1$)	Probability of declaring significant ($p \leq 0.01$) difference				
	0.50	0.80	0.90	0.95	0.99
7.0	2.10	2.59	2.88	3.13	3.64
8.0	1.96	2.40	2.66	2.89	3.34
9.0	1.96	2.38	2.62	2.83	3.26
10.0	1.87	2.25	2.48	2.67	3.06
11.0	1.79	2.15	2.35	2.53	2.90
12.0	1.81	2.15	2.35	2.52	2.86
13.0	1.74	2.07	2.26	2.42	2.74
14.0	1.69	2.00	2.18	2.33	2.63
15.0	1.71	2.01	2.18	2.33	2.62
20.0	1.58	1.83	1.97	2.09	2.33
25.0	1.51	1.72	1.84	1.95	2.15
30.0	1.46	1.65	1.76	1.85	2.03
35.0	1.42	1.60	1.69	1.78	1.94
40.0	1.39	1.55	1.64	1.72	1.87
45.0	1.37	1.52	1.61	1.68	1.82
50.0	1.35	1.50	1.58	1.64	1.77
60.0	1.33	1.46	1.53	1.59	1.70
70.0	1.30	1.41	1.48	1.53	1.64
80.0	1.28	1.39	1.45	1.50	1.60
90.0	1.26	1.37	1.42	1.47	1.56
100.0	1.25	1.34	1.40	1.44	1.52

When rounded up to the next integer value, one obtains exactly the same result as in Table 7.1 on almost every occasion.

If the true value of the relative risk is R, then the observation D will have a distribution such that

$$2\{D^{1/2} - (RE)^{1/2}\}$$

is a standard normal distribution. To achieve significance at the α level, we must have

$$D \geq \{E^{1/2} + (Z_\alpha)/2\}^2,$$

which will occur with probability β when

$$(RE)^{1/2} - (E)^{1/2} = (Z_\alpha + Z_{1-\beta})/2,$$

where $Z_{1-\beta}$ is the $(1-\beta)$ point of the standard normal distribution. In other words, to have probability β of obtaining a result significant at the α level when the true relative risk is R, one needs a value of E equal to or greater than

$$(Z_\alpha + Z_{1-\beta})^2/4(R^{1/2} - 1)^2. \tag{7.1}$$

As can be simply verified, use of this expression gives values close to those shown in Tables 7.2 and 7.3. For example, with $\alpha = 1 - \beta = 0.05$, for which $Z_\alpha = Z_{1-\beta} = 1.645$, a value of R equal to 2.31 requires a value of E equal to 10.01 from expression (7.1), and a value of 10.0 from Table 7.3. Use of expression (2.11) based on the cube root

transformation will give slightly improved accuracy for small values of E – say, less than 10 – whereas use of expression (2.10), the usual χ^2 statistic, will give somewhat less accurate results. Only for very small studies in which large relative risks are expected would the accuracy of the simple expression (7.1) be inadequate.

The other approach to assessing the capacity of a given study design to respond to the questions for which answers are sought is in terms of the expected widths of the resulting confidence intervals. These widths are given, in proportional terms, in Table 2.11. Given an expected number E based on external rates and a postulated value R for the relative risk, one can read off, from Table 2.11, the lower and upper multipliers one would expect to apply to the observed SMR to construct a confidence interval.

Thus, for $E = 20$ and for different values of R, we have the following 95% confidence intervals for R if D takes its expected value of RE:

	Lower bound	Upper bound
$R = 1.5$	1.01	2.09
$R = 2.0$	1.43	2.67
$R = 3.0$	2.29	3.81

The investigator would have to decide whether confidence intervals of this expected width satisfy the objectives of the study, or whether attempts would be needed to augment the size of the study.

For values of E and R not covered in Table 2.11, we can use as before the square root transformation (see expression 2.15). For a given value of E and R, the square root of the observed number of deaths, $D^{1/2}$, will be approximately normally distributed, with mean $(ER)^{1/2}$ and variance 1/4. The resulting $100(1-\alpha)\%$ confidence intervals if D took its expected value would thus be given by

$$\{(ER)^{1/2} \pm Z_{\alpha/2}/2\}^2/E$$

or

$$R \pm Z_{\alpha/2}\left(\frac{R}{E}\right)^{1/2} + Z^2_{\alpha/2}/4E.$$

The upper limit is improved by incorporating the modification of (2.15), replacing R by $R(D+1)/D$.

7.3 Sample size for cohort studies – comparison with an internal control group

In this section, we outline power and sample size determination when it is envisaged that the main comparisons of interest will be among subgroups of the study cohort, using the analytical methods of Chapter 3. We start by considering the simplest situation, in which the comparison of interest is between two subgroups of the study cohort, one considered to be exposed, the other nonexposed. Rates for the disease of interest are to be compared between the two groups. The situation corresponds to that of §3.4, with two dose levels. As argued in the preceding chapters, use of an internal

control group is often important in order to reduce bias. Suppose that the two groups are of equal size and age structure, and that we observe O_1 events in one group (the exposed) and O_2 in the other. Since the age structures are the same, age is not a confounder, and no stratification is necessary. Following §3.4, inferences on the relative risk R are based on the binomial parameter of a trial in which O_1 successes have occurred from $O_1 + O_2$ observations, the binomial parameter, π say, and R being related by

$$R = \pi/(1-\pi) \quad \text{or} \quad \pi = \frac{R}{(R+1)}$$

as in expression (3.6).

Now if R is equal to unity, π is equal to 1/2, and the test of significance can be based on the tail probabilities of the exact binomial distribution given by

$$\sum_{x=O_2}^{O_+} \binom{O_+}{x} 2^{-O_+},$$

where $O_+ = O_1 + O_2$. For a fixed value of O_+, the power of the study can be evaluated for different values of R, using the binomial distribution with parameter $R/(R+1)$. O_+, however, is not fixed, but a random variable following a Poisson distribution with mean $E(1+R)$, where E is the expected number of events in the nonexposed group. The power for each possible value of O_+ needs to be calculated, and the weighted sum computed, using as weights the corresponding Poisson probabilities. This weighted sum gives the unconditional power.

When the groups are of unequal size, but have the same age structure, a similar approach can be adopted. Suppose that E_1 events are expected in the exposed group under the null hypothesis, and that E_2 events are expected in the control group. Then, under the null hypothesis, the number of events in the exposed group, given O_+ the total number of events, will follow a binomial distribution with probability parameter $E_1/(E_1+E_2)$. Under the alternative hypothesis with relative risk R, the binomial distribution will have parameter $RE_1/(RE_1+E_2)$. The power can be evaluated for each value of O_+, and the weighted sum computed using as weights the probabilities of the Poisson distribution with mean $RE_1 + E_2$. Gail (1974) has published power calculations when E_1 equals E_2, and Brown and Green (1982) the corresponding values when E_1 is not equal to E_2. Table 7.4 gives the expected number of events in the control group, E_2, for power of 80% and 90% and significance (one-sided) of 5% and 1% for various values of R and of the ratio E_2/E_1 (written as k).

On many occasions, particularly when O_1 and O_2 are large, the formal statistical test is unlikely to be based on the binomial probabilities, but on a normal approximation using either a corrected or uncorrected χ^2 test.

In the case of equal-sized exposed and control cohorts, the observed proportion $p = O_1/(O_1+O_2)$ is compared with the proportion under the null hypothesis, namely 1/2, using as variance that under the null. The uncorrected χ^2 test statistic is equivalent to comparing

$$2\sqrt{O_+}(p - \tfrac{1}{2})$$

with a standard normal distribution.

Table 7.4 Comparison with an internal control group

(a) Expected number of cases in the control group required to detect a difference with 5% significance and given power, for given relative risk, when the control group is k times the size of the exposed group (using exact Poisson distribution)

k[a]	Relative risk[b]							
	2	3	4	5	6	8	10	20
1/10	11.3	3.86	2.16	1.47	1.10	0.712	0.528	0.212
	15.0	5.00	2.75	1.84	1.36	0.881	0.639	0.262
1/5	12.3	4.23	2.37	1.60	1.18	0.770	0.566	0.236
	16.2	5.45	3.03	2.03	1.50	0.958	0.696	0.283
1/2	15.1	5.18	2.85	1.93	1.45	0.954	0.706	0.299
	20.2	6.80	3.74	2.48	1.83	1.19	0.873	0.363
1	20.0	6.70	3.71	2.52	1.89	1.25	0.923	0.392
	27.0	8.89	4.90	3.27	2.43	1.58	1.17	0.485
2	29.6	9.91	5.40	3.58	2.59	1.63	1.19	0.498
	40.3	13.5	7.26	4.82	3.54	2.22	1.59	0.642
5	58.6	19.5	10.8	7.21	5.21	3.33	2.44	1.00
	80.1	26.3	14.5	9.76	7.19	4.50	3.25	1.33
10	107	35.0	19.5	13.0	9.52	6.00	4.29	1.67
	146	48.2	26.5	17.7	13.0	8.27	5.93	2.31

(b) Expected number of cases in the control group required to detect a difference with 1% significance and given power, for given relative risk, when the control group is k times the size of the exposed group (using exact Poisson distribution)

k[a]	Relative risk[b]							
	2	3	4	5	6	8	10	20
1/10	17.9	6.06	3.38	2.26	1.69	1.10	0.805	0.336
	22.5	7.51	4.12	2.76	2.03	1.30	0.952	0.387
1/5	19.4	6.55	3.63	2.44	1.82	1.19	0.864	0.275
	24.5	8.15	4.47	2.97	2.20	1.42	1.03	0.416
1/2	23.9	8.03	4.46	2.96	2.19	1.41	1.03	0.431
	30.3	10.0	5.57	3.69	2.70	1.73	1.25	0.508
1	31.2	10.5	5.73	3.82	2.85	1.87	1.38	0.567
	39.8	13.2	7.27	4.79	3.52	2.28	1.68	0.689
2	46.1	15.1	8.33	5.42	3.91	2.49	1.82	0.775
	59.2	19.4	10.6	7.02	5.08	3.17	2.29	0.946
5	90.5	29.2	15.9	10.6	7.76	4.80	3.41	1.38
	116	37.9	20.5	13.6	10.0	6.32	4.47	1.75
10	164	52.8	28.5	18.6	13.5	8.50	6.07	2.41
	213	69.0	37.3	24.3	17.7	11.2	7.98	3.15

[a] Ratio of E_2/E_1, where E_2 is the number of events expected in the control group and E_1 the number expected in the exposed group under the null hypothesis

[b] The top number corresponds to a power of 80% and the bottom to a power of 90%

Under the alternative of a relative increase in risk of R, p has mean $R/(R+1)$ and variance $R/\{O_+(R+1)^2\}$. The required sample size is then given by

$$O_+ = \frac{\left(\frac{1}{2}Z_\alpha + Z_{1-\beta}\sqrt{\frac{R}{(R+1)^2}}\right)^2}{\left(\frac{R}{R+1} - \frac{1}{2}\right)^2} = \frac{\{(R+1)Z_\alpha + 2Z_{1-\beta}\sqrt{R}\}^2}{(R-1)^2}. \tag{7.2}$$

When R is close to unity, approximate solutions are given by approximating $R/(R+1)^2$ by 1/4 and rewriting the equation

$$O_+ = E_2(1+R) = (Z_\alpha + Z_{1-\beta})^2\left(\frac{R+1}{R-1}\right)^2.$$

When the two groups are of unequal size, n_1 and n_2, say, but the same age distribution, then we have

$$O_+ = \frac{\left(\sqrt{\frac{n_1 n_2}{(n_1+n_2)^2}}Z_\alpha + \sqrt{\frac{Rn_1 n_2}{(Rn_1+n_2)^2}}Z_{1-\beta}\right)^2}{\left(\frac{Rn_1}{Rn_1+n_2} - \frac{n_1}{n_1+n_2}\right)^2}. \tag{7.3}$$

Following Casagrande *et al.* (1978b) and Ury and Fleiss (1980), more accurate values are given by incorporating Yates' correction in the χ^2 significance test, which for groups of equal size results in multiplying the right-hand side of (7.3) by the term

$$\tfrac{1}{4}[1 + \sqrt{1 + 4(p_1 - p_2)/A)}]^2,$$

where

$$A = \left(\tfrac{1}{2}Z_\alpha + \frac{\sqrt{R}}{(R+1)}Z_{1-\beta}\right)^2, \qquad p_1 = \frac{R}{R+1}, \qquad p_2 = \tfrac{1}{2}.$$

When the groups are of unequal size, n_1 and n_2, respectively, the corresponding correction factor is given by

$$\tfrac{1}{4}[1 + \sqrt{1 + A'}]^2,$$

where

$$A' = 2\left(\frac{Rn_1}{Rn_1+n_2} - \frac{n_1}{n_1+n_2}\right)\Big/\left(Z_\alpha\sqrt{\frac{n_1 n_2}{(n_1+n_2)^2}} + Z_{1-\beta}\sqrt{\frac{Rn_1 n_2}{(Rn_1+n_2)^2}}\right)^2. \tag{7.4}$$

Table 7.5 gives the number of cases that would need to be expected in the nonexposed group for a range of values of the relative risk R, of the relative sizes of the exposed and unexposed group, and of α and β. The numbers are based on expression (7.3), modified by incorporating Yates' correction. The values in Table 7.5 are very close to the corresponding values based on exact binomial probabilities given in Table 1 of Brown and Green (1982). They are slightly smaller than the values in Table 7.4 for the more extreme values of R and of the ratio of the sizes of the two groups; the values in Table 7.4 took account of the Poisson variability of O_+.

Table 7.5 Sample size requirements in cohort studies when the exposed group is to be compared with a control group of k times the size. The numbers in the table are those expected in the control group (using χ^2 approximation)

	k	Relative risk				
		1.5	2.0	2.5	5.0	10.0
Significance, 5%	1.00	30.9	10.0	5.5	1.5	0.6
Power, 50%	2.00	43.7	13.7	7.3	1.9	0.7
	4.00	69.2	20.9	10.9	2.6	0.9
	10.00	145.6	42.8	21.7	4.9	1.7
	100.00	1292.6	370.1	184.4	38.9	12.3
	0.50	24.6	8.2	4.6	1.3	0.5
	0.25	21.4	7.3	4.1	1.2	0.5
	0.10	19.5	6.7	3.8	1.1	0.5
Significance, 5%	1.00	64.9	19.8	10.3	2.4	0.8
Power, 80%	2.00	95.4	28.7	14.8	3.3	1.1
	4.00	156.5	46.6	23.8	5.2	1.7
	10.00	340.0	100.5	51.1	11.2	3.5
	100.00	3094.3	911.3	463.0	102.0	33.0
	0.50	49.7	15.4	8.1	2.0	0.7
	0.25	42.2	13.2	7.0	1.7	0.6
	0.10	37.6	11.9	6.3	1.6	0.6
Significance, 5%	1.00	88.1	26.4	13.5	3.0	1.0
Power, 90%	2.00	131.1	39.0	19.9	4.3	1.4
	4.00	217.5	64.6	32.9	7.1	2.2
	10.00	477.1	142.0	72.4	15.8	4.9
	100.00	4374.8	1305.8	669.8	151.3	49.7
	0.50	66.7	20.1	10.4	2.4	0.8
	0.25	56.0	17.0	8.9	2.1	0.7
	0.10	49.7	15.2	8.0	1.9	0.7
Significance, 5%	1.00	110.0	32.6	16.6	3.6	1.1
Power, 95%	2.00	165.1	48.9	24.8	5.3	1.6
	4.00	275.8	82.0	41.7	8.9	2.6
	10.00	608.7	182.0	93.2	20.3	6.2
	100.00	5607.8	1689.6	872.6	200.7	66.7
	0.50	82.6	24.6	12.5	2.8	0.9
	0.25	69.0	20.6	10.6	2.4	0.8
	0.10	60.9	18.3	9.4	2.2	0.7
Significance, 1%	1.00	58.0	18.2	9.7	2.5	0.9
Power, 50%	2.00	81.7	24.6	12.7	3.1	1.1
	4.00	128.9	37.3	18.8	4.2	1.4
	10.00	270.5	75.5	37.0	7.6	2.4
	100.00	2394.9	649.1	310.2	58.0	16.7
	0.50	46.2	15.0	8.2	2.2	0.9
	0.25	40.3	13.4	7.4	2.1	0.8
	0.10	36.8	12.4	7.0	2.0	0.8

Table 7.5 (*contd*)

	k	Relative risk				
		1.5	2.0	2.5	5.0	10.0
Significance, 1%	1.00	103.2	31.2	16.1	3.8	1.3
Power, 80%	2.00	150.2	44.4	22.6	5.0	1.6
	4.00	244.2	71.0	35.7	7.6	2.4
	10.00	526.4	150.9	75.1	15.5	4.7
	100.00	4761.5	1352.1	688.4	136.4	41.4
	0.50	79.8	24.6	12.9	3.1	1.1
	0.25	68.1	21.3	11.3	2.8	1.0
	0.10	61.1	19.4	10.4	2.6	1.0
Significance, 1%	1.00	132.3	39.5	20.2	4.5	1.5
Power, 90%	2.00	194.7	57.3	29.0	6.3	2.0
	4.00	319.8	93.3	46.9	9.8	3.0
	10.00	695.8	201.7	101.0	21.0	6.3
	100.00	6338.3	1831.8	917.5	194.0	60.4
	0.50	101.2	30.6	15.9	3.7	1.3
	0.25	85.7	26.2	13.7	3.3	1.1
	0.10	76.4	23.6	12.4	3.0	1.1
Significance, 1%	1.00	159.1	47.1	23.9	5.0	1.7
Power, 95%	2.00	235.9	69.3	34.9	7.4	2.3
	4.00	390.2	114.1	57.4	12.0	3.6
	10.00	854.0	249.4	125.6	26.3	7.8
	100.00	7816.1	2286.2	1155.3	250.3	79.3
	0.50	120.8	36.1	18.5	4.2	1.4
	0.25	101.7	30.7	15.9	3.7	1.2
	0.10	90.3	27.5	14.3	3.4	1.2

Comparison of Table 7.5 with Table 7.2 indicates that, for given α, β and R, roughly twice as many cases must be expected in the nonexposed control group when an internal comparison group of equal size is used. Since there are two groups, this implies that roughly four times as many individuals must be followed. This increase represents the price to be paid for using internal rather than external comparisons.

Since power calculations are essentially approximate, an alternative and simple approach is obtained by using the variance stabilizing arcsin transformation, given by

$$\arcsin[\{O_1/(O_1+O_2)\}^{1/2}].$$

This transformed variable is approximately normally distributed with variance equal to $1/\{4(O_1+O_2)\}$. The mean if the two groups are of equal size is given by $\arcsin\{R/(R+1)\}^{1/2}$.

Under the null hypothesis, R equals unity, so that a result significant at the α level is obtained if

$$\arcsin\{O_1/(O_1+O_2)\}^{1/2} \geqslant \arcsin(\tfrac{1}{2})^{1/2} + 0.5Z_\alpha(O_1+O_2)^{-1/2}.$$

If the relative risk among the exposed is equal to R, then this inequality will hold with

probability at least β if

$$(O_1 + O_2) \geq (Z_\alpha + Z_{1-\beta})^2/4\left\{\arcsin\left(\frac{1}{R+1}\right)^{1/2} - \arcsin(\tfrac{1}{2})^{1/2}\right\}^2, \qquad (7.5)$$

whre $Z_{1-\beta}$ is the $(1-\beta)$ point of the normal distribution.

This expression gives the total number of events expected in the two groups combined that are required to have probability β of achieving a result significant at the α level if the true relative risk is R. An approximation closer to the equivalent χ^2 test with the continuity correction is given if one adds a correction term to the arcsin transformation, replacing, for a binomial with proportion p and denominator n, $\arcsin(p)^{1/2}$ by $\arcsin(p - \frac{1}{2}n)^{1/2}$. In the present context n is given by $O_1 + O_2$, so that (7.5) would no longer give an explicit expression for E, but would require an iterative solution. Usually one iteration would suffice.

If the exposed and nonexposed groups are not of equal size, but the age distributions are the same, then a minor modification can be made to the above inequality. The binomial parameter, previously $R/(R+1)$, now becomes $Rn_1/(Rn_1 + n_2)$, where n_1 and n_2 are the numbers of individuals in the two groups. Expression (7.2) then becomes

$$(O_1 + O_2) \geq (Z_\alpha + Z_{1-\beta})^2/4\left\{\arcsin\left(\frac{Rn_1}{Rn_1 + n_2}\right)^{1/2} - \arcsin\left(\frac{n_1}{n_1 + n_2}\right)^{1/2}\right\}^2.$$

When the age structures of the two groups are dissimilar, one could use the approach of §3.4 or §3.5, and replace n_1 and n_2 in expressions (7.3), (7.4) and (7.5) by E_1 and E_2, the expected number of cases in the two groups based on an external standard or on the pooled rates for the two groups. If the confounding due to age is at all severe, however, this procedure will suffer from appreciable bias, and one should use the preferred methods of §3.6, basing power considerations on the variance of the Mantel–Haenszel estimate of relative risk (expression 3.17) (Muñoz, 1985). The effect of confounding on sample size requirements is discussed in more detail in §7.7.

If more emphasis is to be put on the precision of estimates of relative risk, rather than on detection of an effect, then the width of expected confidence intervals is of more relevance. The equations given by (3.19) can be solved to give upper and lower limits, or alternatively one can use the simpler expression (3.18).

7.4 Tests for trend

The results of a cohort study will be more persuasive of a genuine effect of exposure on risk if one can demonstrate, in addition to a difference between an exposed and an unexposed group, a smoothly changing risk with changing exposure. It is thus important that the study be designed with this aim in view. Under favourable circumstances, one will have not just two groups – one exposed and one nonexposed – but a number of groups, each with different exposures. In the analysis of the results of such a study, the single most powerful test for an effect of exposure on risk will normally be a trend test. It will therefore be useful, when assessing the value of a given

study design, to examine the power of a trend test. For the sake of simplicity, we consider the situation in which we have K exposure groups but no further stratification by age or other confounding variables. Using the notation of Chapter 3, we shall investigate the power of the test statistic (3.12), given by

$$\chi^2 = \{\sum x_k(O_k - \tilde{E}_k)\}^2 / \{\sum x_k^2 \tilde{E}_k - (\sum x_k \tilde{E}_k)^2 / \sum \tilde{E}_k\},$$

where the $\tilde{E}_k$ are expectations based on external rates, but normalized so that

$$\sum \tilde{E}_k = \sum O_k; \quad \text{i.e.,} \quad \tilde{E}_k = E_k \frac{\sum O_j}{\sum E_j}.$$

For a one-sided test of size α for positive slope, and writing the denominator in the above expression as V, we need

$$\sum x_k(O_k - \tilde{E}_k) \geq \sqrt{V} \cdot Z_\alpha \tag{7.6}$$

to achieve significance.

V is given by

$$\{\sum x_k^2 E_k - (\sum x_k E_k)^2 / \sum E_k\} \frac{\sum O_k}{\sum E_k}$$

and so, being a multiple of $\sum O_k$, will have a Poisson distribution, multiplied by a scale factor involving the x_k and E_k. $V^{1/2}$ will then be approximately normal, with standard deviation given by 1/2 times the scale factor

If E_k are the expectations based on external rates, then the left-hand side of expression (7.6) can be written as

$$\sum_k O_k \left\{ x_k - \left(\sum_j x_j E_j \Big/ \sum_j E_j \right) \right\}.$$

In order to assess the probability that the inequality (7.6) will hold, we have to specify a range of distributions for the O_k alternative to the null distribution that $E(O_k) = E_k$ for all k.

A simple family of alternatives representing a linear trend in risk is given by

$$E(O_k) = (1 + \delta x_k) E_k,$$

from which we have

$$\text{Expectation } (\tilde{E}_k) = E_k \left(1 + \frac{\delta \sum x_j E_j}{\sum E_j} \right).$$

The power is then given by the probability that the following inequality holds:

$$\sum O_k \{x_k - (\sum x_j E_j / \sum E_j)\} - Z_\alpha . \sqrt{V} \geq 0.$$

Writing

$$V = W \sum O_k,$$

where W is a function of the x_k and E_k, then under the family of alternative distributions given above, the left-hand side will have mean m approximated by

$$m = \sum \delta x_k E_k \{x_k - (\sum x_j E_j / \sum E_j)\} - Z_\alpha W^{1/2} \{\sum (E_k + \delta x_k E_k)\}^{1/2}$$

and variance s^2 by

$$s^2 = \sum (1 + \delta x_k) E_k \{x_k - (\sum x_j E_j / \sum E_j)\}^2$$

$$- Z_\alpha W^{1/2} \sum \delta x_k E_k \{x_k - (\sum x_j E_j / \sum E_j)\} (\sum E_k)^{-1/2} + z_\alpha^2 W/4.$$

The power is then approximately the probability corresponding to the normal deviate $Z_{1-\beta}$ given by $m = s \cdot Z_{1-\beta}$.

An alternative approach to the power of tests for linear trend was given by Chapman and Nam (1968) based on the noncentral χ^2 distribution.

Example 7.2

We consider a hypothetical example, comparing power considerations based on a trend test with those based on two alternative dichotomizations of the data. Let us suppose that we have four exposure levels, 0, 1, 2, 3, and that the groups at each level are of the same size and age structure. Under the null hypothesis, they therefore have the same expected numbers of events, E, say, in each group.

We consider a family of alternative hypotheses in which the relative risk is given as above by

$$1 + \delta x_k,$$

where x_k takes the values 0, 1, 2, 3. Substituting into the expression for m and s^2 gives

$$5\delta\sqrt{E} = Z_\alpha \sqrt{(5 + 7.5\delta)} + Z_{1-\beta}\{(5 + 7.5\delta) - 5\delta E^{-1/2} Z_\alpha (5/16)^{1/2} + 5Z_\alpha^2/16E\}^{1/2},$$

an equation that can be solved for β given δ and E or, conversely, solved for E given δ and β.

It is interesting to compare the results of power calculations for the trend test to the results one would obtain by dichotomizing the data, grouping, for example, the two highest and the two lowest exposed groups. We would then have a relative risk between the two groups of

$$(2 + 5\delta)/(2 + \delta),$$

and each of the two groups would be twice the size of the original four groups.

Substituting these values in expression (7.5) gives

$$2E\left(1 + \frac{2 + 5\delta}{2 + \delta}\right) = (Z_\alpha + Z_{1-\beta})^2/4 \left\{\arcsin\left(\frac{2 + 5\delta}{4 + 6\delta}\right)^{1/2} - \arcsin(\tfrac{1}{2})^{1/2}\right\}^2$$

(the 2 at the start of the left-hand side arises since we have the sum of two groups each of size E), again an equation that can be solved for either E or for β.

Alternatively, one could base power calculations on a comparison between the two groups with highest and lowest exposure, respectively, the risk of the former relative to the latter being $1 + 3\delta$.

The three approaches give the following result for the expected number E required in each group, using a test with $\alpha = 0.05$ and $\beta = 0.95$:

δ	Trend test	Dichotomy into two equal groups	Highest against lowest
0.25	46.8	66.2	51.6
0.5	14.6	22.9	16.4
1.0	5.0	9.5	6.0
2.0	1.8	4.9	2.4

The trend test is considerably more powerful in this example than the test obtained by dichotomizing the study cohort, and marginally more powerful than the simple test of highest against lowest.

7.5 Restriction of power considerations to the follow-up period of interest

The discussion so far has treated observed and expected deaths as if all periods of follow-up were of equal interest. Usually, however, one would expect any excess risk to be concentrated in particular periods of follow-up, as outlined in Chapter 6. The carcinogenic effect of many exposures is not seen for ten years or more since the start of exposure. One is clearly going to overestimate the power of a study if one groups together all person-years of follow-up. An example comes from a study of the later cancer experience among women diagnosed with cancer of the cervix (Day & Boice, 1983). The purpose of the study was to investigate the occurrence of second cancers induced by radiotherapy given for the cervical cancer. For this purpose, three cohorts were assembled: women with invasive cancer of the cervix treated by radiotherapy, women with invasive cancer of the cervix not treated by radiotherapy, and women with in-situ carcinoma of the cervix not treated by radiotherapy. Table 7.6 gives the woman-years in different follow-up periods for the three groups, and the expected numbers of cancers in the first group, excluding the first year, and excluding the first ten years of follow-up. One can see that in the in-situ group 90% of the person-years of follow-up occurred in the first ten years, with a corresponding figure of over 70% for the women with invasive cancer. This example is extreme in the sense that cohort membership for the invasive cases is defined in terms of a life-shortening condition,

Table 7.6a Woman–years at risk by time since entry into the cohort (i.e., diagnosis of cervical cancer)

Time since diagnosis (years)	Invasive cancer		In-situ cancer
	Treated by radiotherapy	Not treated by radiotherapy	
0–9	445 990 (71%)	89 719 (74%)	485 026 (90%)
10–19	149 772	27 945	53 621
20+	29 676	3 961	2 265
Total	625 438	121 625	540 912

Table 7.6b Expected number of second cancers at selected sites among the radiation-treated group

	Excluding the first year of follow-up	Excluding the first ten years of follow-up
Stomach	210.4	86.1
Rectum	157.4	68.6
Breast	804.4	304.6
Multiple myeloma	33.9	14.8

and large-scale identification of in-situ cases by mass screening did not occur until the mid-1960s or later in many of the participating areas. For most of the cancers of interest, excesses were not seen until at least ten years after entry, so that power considerations based on the full follow-up period would seriously overestimate the potential of the study, especially in assessing the value of the in-situ cohort as a comparison group.

7.6 Case-control sampling within a cohort

(*a*) *Basic considerations of case-control design: dichotomous exposure – unmatched design*

Before discussing the specific issues of concern when sampling from a risk set in the context of §5.4, we review more generally design aspects of case-control studies. We begin with the simplest situation, of a single dichotomous exposure variable. The problem is that of comparing two independent binomial distributions, one corresponding to the cases, one to the control population, with binomial probabilities, respectively, of p_1 and p_2, say.

The approach to the comparison of two proportions that we have taken in these two volumes has been based on the exact conditional distribution of a 2×2 table, expressed in terms of the odds ratio. Tests of the null hypothesis were derived either from this exact distribution, or from the approximation to it given by the χ^2 test with continuity correction. Since sample size and power calculations should refer to the statistical test that is going to be used, most of the subsequent discussions of power refer to the exact test, or approximations to it.

When the samples of cases and controls are of the same size, n, say, then for a χ^2 test without the continuity correction the power and sample sizes are related by the equation

$$n = (Z_\alpha \sqrt{2\bar{p}\bar{q}} + Z_{1-\beta}\sqrt{p_1 q_1 + p_2 q_2})^2/(p_1 - p_2)^2, \tag{7.7}$$

where α is the size of the test, β the power, p_1 the proportion exposed among the cases and p_2 the proportion exposed among the controls (and with $q_i = 1 - p_i$, $i = 1, 2$ and $\bar{p} = 1 - \bar{q} = (p_1 + p_2)/2$.)

Incorporating the continuity correction into the χ^2 test, to make it approach the exact test more closely, results in multiplying the right-hand side of (7.7) by the factor

(Casagrande *et al.*, 1978b)

$$\tfrac{1}{4}\{1+\sqrt{1+4(p_1-p_2)/A}\}^2,$$

where

$$A=(Z_\alpha\sqrt{2\bar{p}\bar{q}}+Z_{1-\beta}\sqrt{p_1q_1+p_2q_2})^2.$$

From this expression, one can either calculate the power β from a given sample size, or the sample size n required to achieve a given power.

This result has been extended by Fleiss *et al.* (1980) to the situation of unequal sample sizes. If we have a sample of size n from the population of cases (with parameter p_1) and size nk from the controls ($0<k<\infty$), then to have probability β of achieving significance at the α level, we need

$$n=A\cdot W/k(p_1-p_2)^2,$$

where

$$A=\left(Z_\alpha\sqrt{(k+1)\bar{p}\bar{q}}+Z_{1-\beta}\sqrt{kp_1q_1+p_2q_2}\right)^2,$$

$$W=\frac{1}{4}\left(1+\sqrt{1+\frac{2(k+1)(p_1-p_2)}{A}}\right)^2,$$

and

$$\bar{p}=1-\bar{q}=(p_1+kp_2)/(1+k).$$

In any particular study, sample size considerations would normally be based on an estimate of p_2, the prevalence of the exposure in the general population, and a value R for the relative risk that the investigator feels it would be important not to miss. In terms of the previous discussion, we would then have

$$\frac{p_1(1-p_2)}{p_2(1-p_1)}=R$$

or $p_1=Rp_2/(1-p_2+Rp_2)$.

Table 7.7 gives the required number of cases for a range of values of R, p_2, α, β and k, the ratio of the number of controls to the number of cases, for the χ^2 test with continuity correction. The values are close to those obtained using the exact conditional test (Casagrande *et al.*, 1978a).

An alternative, simple approximation is obtained using the variance stabilizing arcsin transformation, with which the sample size needed from each of the two populations to achieve one-sided significance at the α level with probability β is given by

$$n=(Z_\alpha+Z_{1-\beta})^2/2(\arcsin p_1^{1/2}-\arcsin p_2^{1/2})^2.$$

If there are nk controls and n cases, this expression becomes

$$n=(k+1)(Z_\alpha+Z_{1-\beta})^2/4k(\arcsin p_1^{1/2}-\arcsin p_2^{1/2})^2. \tag{7.8}$$

Consideration has recently been given to exact unconditional tests for equality of two proportions (Suissa & Shuster, 1985), approximations to which would be given by the

Table 7.7 Unmatched case-control studies. Number of cases required in an unmatched case-control study for different values of the relative risk, proportion of exposed among controls, significance level, power and number of controls per case. The three numbers in each cell refer to case-control ratios of 1:1, 1:2 and 1:4.

(*a*) *Significance = 0.05; power = 0.80*

Relative risk	Proportion exposed in control group											
	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80
1.5	6672	1415	763	550	448	390	356	325	325	352	419	571
	4901	1041	563	407	332	290	265	243	244	265	317	434
	4009	854	462	335	274	240	219	202	203	222	266	365
2.0	2087	449	246	181	150	133	123	115	119	133	162	226
	1512	327	180	133	110	98	91	86	89	100	123	173
	1220	264	146	108	90	81	75	72	75	84	104	146
2.5	1114	243	135	101	84	76	71	69	72	82	102	146
	799	175	98	74	62	56	53	51	54	62	78	112
	638	140	79	60	51	46	43	42	45	52	66	95
3.0	732	161	91	69	58	53	50	49	53	61	78	112
	521	116	66	50	43	39	37	37	40	47	59	86
	412	92	53	40	35	32	31	31	33	39	50	73
4.0	420	94	55	42	37	34	33	33	37	43	56	82
	296	67	39	31	27	25	24	25	28	33	43	64
	231	53	31	25	22	20	20	21	23	28	36	54
5.0	290	66	39	31	27	26	25	26	29	35	46	69
	203	47	28	22	20	19	19	20	22	27	36	54
	157	37	22	18	16	15	15	16	19	23	30	46
7.5	161	39	24	20	18	17	17	19	22	27	36	55
	112	27	17	14	13	13	13	14	17	21	28	43
	85	21	13	11	10	10	10	12	14	17	24	36
10.0	111	28	18	15	14	14	14	16	19	24	32	50
	77	20	13	11	10	10	11	12	14	18	25	39
	58	15	10					10	12	15	21	33
15.0	69	19	13	11	11	11	12	13	16	21	29	45
	48	13						10	12	16	22	35
	36	10							10	13	19	29
20.0	51	15	10	10		10	10	12	15	19	27	42
	35	10							11	15	21	33
	26									12	17	28

(*b*) *Significance = 0.05; power = 0.95*

Relative risk	Proportion exposed in control group											
	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80
1.5	11381	2413	1301	938	763	665	606	553	553	600	713	974
	8527	1808	975	703	572	498	454	414	414	449	534	730
	7089	1503	811	585	476	415	378	345	345	374	445	607
2.0	3505	753	413	302	250	221	205	193	199	221	271	379
	2622	564	309	226	187	166	154	144	149	166	203	284
	2171	467	256	188	155	138	128	120	123	138	168	235

Table 7.7 (*contd*)

(*b*) *Significance = 0.05; power = 0.95*

Relative risk	Proportion exposed in control group											
	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80
2.5	1852	403	224	166	139	125	117	113	119	135	169	241
	1383	301	167	124	104	93	88	84	89	101	126	180
	1140	248	138	103	86	77	73	70	74	84	104	149
3.0	1208	265	149	112	95	86	82	80	86	100	127	184
	900	198	111	84	71	64	61	60	64	75	94	137
	739	163	92	69	59	53	50	50	53	62	78	113
4.0	685	153	88	68	58	54	52	53	58	70	90	133
	509	114	66	50	44	40	39	39	44	52	67	99
	416	93	54	41	36	33	32	32	36	42	55	81
5.0	469	107	63	49	43	40	39	41	46	56	74	111
	348	79	47	36	32	30	29	31	34	42	55	82
	283	65	38	30	26	25	24	25	28	34	45	67
7.5	258	61	37	30	27	26	27	29	34	42	57	88
	191	45	28	22	20	20	20	21	25	31	42	65
	154	37	22	18	17	16	16	17	20	25	34	52
10.0	177	43	27	23	21	21	21	24	29	37	50	78
	131	32	20	17	16	15	16	18	21	27	37	57
	105	26	16	14	13	12	13	14	17	22	30	46
15.0	109	28	19	17	16	16	17	20	24	32	44	70
	81	21	14	12	12	12	12	14	18	23	32	51
	64	17	11	10		10	10	11	14	18	26	40
20.0	80	22	15	14	14	14	15	18	22	30	42	66
	59	16	11	10	10	10	11	13	16	21	30	48
	47	13						10	13	17	24	38

(*c*) *Significance = 0.01; power = 0.80*

Risk ratio	Proportion exposed in control group											
	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80
1.5	10583	2245	1211	873	711	620	565	515	515	559	664	906
	7698	1638	887	642	524	458	419	385	387	422	505	693
	6247	1332	724	525	430	377	346	319	323	354	425	585
2.0	3266	703	386	283	234	207	192	181	186	207	253	354
	2328	504	278	206	171	153	142	135	140	158	194	274
	1851	403	224	166	139	125	117	112	117	133	165	234
2.5	1728	377	210	156	131	118	110	106	112	128	159	226
	1214	267	150	113	95	86	82	80	85	98	123	177
	950	210	119	90	77	70	67	66	71	82	104	151
3.0	1128	249	140	106	90	82	78	76	82	95	119	173
	784	175	100	76	65	60	57	57	62	73	93	136
	606	136	79	61	52	48	47	47	52	61	79	116
4.0	641	144	84	64	56	52	50	51	56	66	85	126
	439	100	59	46	40	38	37	38	43	51	67	100
	333	77	46	36	32	31	30	31	36	43	57	86

Table 7.7 (*contd*)

(*c*) *Significance = 0.01; power = 0.80*

Risk ratio	Proportion exposed in control group											
	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80
5.0	440	101	60	47	41	39	38	40	45	54	70	105
	298	70	42	33	30	28	28	30	34	42	56	84
	223	53	32	26	24	23	23	25	29	35	47	72
7.5	243	58	36	29	27	26	26	28	33	41	55	83
	162	40	25	21	19	19	19	21	25	32	44	67
	119	30	19	16	15	15	16	18	21	27	37	58
10.0	167	42	27	23	21	21	21	24	28	36	48	74
	111	28	19	16	15	15	16	18	22	28	39	60
	80	21	14	13	12	12	13	15	18	24	33	52
15.0	104	28	19	17	16	16	17	20	24	31	43	67
	68	19	13	12	12	12	13	15	19	25	34	54
	49	14	10			10	10	12	16	21	29	47
20.0	76	22	16	14	14	14	15	18	22	29	40	63
	50	15	11	10	10	11	12	14	17	23	32	51
	35	11						11	14	19	28	44

(*d*) *Significance = 0.01; power = 0.95*

Risk ratio	Proportion exposed in control group											
	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80
1.5	16402	3478	1875	1352	1100	959	874	797	797	865	1028	1404
	12155	2580	1393	1006	820	715	653	597	598	650	775	1060
	10016	2128	1151	832	679	593	542	496	498	542	647	886
2.0	5018	1078	591	433	359	317	294	276	285	317	388	543
	3686	794	437	321	266	236	219	207	214	239	293	412
	3007	649	358	264	219	195	181	172	178	199	245	345
2.5	2639	574	319	237	199	178	167	161	170	193	241	344
	1926	420	235	175	147	132	125	121	128	146	183	261
	1557	341	191	143	121	109	103	100	106	122	152	219
3.0	1715	377	212	160	135	123	116	114	123	142	180	261
	1245	275	156	118	100	91	87	86	92	107	137	199
	1000	222	126	96	82	75	71	71	77	89	114	166
4.0	968	217	125	96	83	77	74	75	83	99	128	189
	698	157	91	70	61	57	55	56	62	74	97	144
	554	126	73	57	50	46	45	46	52	62	81	120
5.0	662	151	88	69	61	57	56	58	66	80	105	157
	474	109	64	51	45	42	41	43	49	60	79	120
	373	86	51	41	36	34	34	36	41	50	66	99
7.5	362	86	52	43	39	37	38	41	48	60	81	123
	258	62	38	31	28	28	28	30	36	45	61	94
	200	48	30	25	23	22	23	25	29	37	50	78
10.0	248	61	38	32	30	30	30	34	41	52	71	110
	176	44	28	24	22	22	22	25	30	39	54	84
	136	34	22	19	18	18	18	20	25	32	44	69
15.0	153	40	27	23	22	23	24	28	34	45	62	98
	108	29	19	17	17	17	18	21	26	34	47	74
	82	22	15	13	13	13	14	17	21	27	38	61
20.0	111	31	21	19	19	20	21	25	32	41	58	92
	79	22	16	14	14	15	16	19	23	31	44	70
	60	17	12	11	11	12	12	15	19	25	36	57

χ^2 test without continuity correction. Sample sizes for the latter can be calculated directly from expression (7.7).

A comparison of the sample size requirements, for 80% power and a test at the 0.05 level, is given in Table 7.8, for the exact conditional test, the exact unconditional test, the χ^2 test with and without correction, and for the arcsin approximation. It is noteworthy that in each case the exact unconditional test is more powerful than the exact conditional test. At present, however, the advantages of working within a unified structure of inference based on Cox regression methods and conditional likelihood, of which the conditional exact test is an example, more than outweigh this slight loss of power.

(*b*) *Basic considerations of case-control design: dichotomous exposure – matched design*

In matched designs, two problems have to be faced: how many controls to choose per case, and how many case-control sets to include, given the number of controls per case. We consider the second question first.

For the sake of simplicity, we shall assume that each case is matched to the same number of controls, k, say. The method of analysis is described in Chapter 5 of Volume 1. When $k = 1$, a matched-pairs design, the analysis concentrates on the discordant pairs. Suppose we have T discordant pairs, among O_1 of which the case is exposed. If risk for disease is unaffected by exposure, then O_1 is binomially distributed with proportion 1/2. If exposure increases the relative risk by R, then O_1 is binomially distributed with proportion $R/(R+1)$. The situation is discussed in §7.3, and similar power considerations apply.

Expression (7.2), with the continuity correction factor and with $n_1 = n_2$, gives the number of discordant case-control pairs that will be required to detect a relative risk of R with probability β at significance level α. Table 7.5, based on expression (7.2) and in the context of a cohort study, gives the expected number of cases required in the nonexposed group. To obtain the expected number of discordant case-control pairs required in a 1:1 matched case-control study, which corresponds to the total number of cases in the exposed and nonexposed groups combined in the context of Table 7.5, the quantities in the part of Table 7.5 referring to equal numbers in the exposed and nonexposed groups must be multiplied by $(1+R)$.

The total number of case-control pairs that is required must be evaluated. If, as in the previous section, the probability of exposure is p_1 among the cases and p_2 among the controls, then the probability of a pair being discordant is simply

$$p_1(1-p_2)+p_2(1-p_1).$$

In a situation in which a matched design is thought appropriate, the probability of exposure would vary among pairs. The above expression then, strictly speaking, requires integration over the distribution of exposure probabilities. For the approximate purposes of sample size determination, however, it would usually be sufficient to use the average exposure probabilities, $\bar{p}_1$ and $\bar{p}_2$. The number of matched pairs, M,

Table 7.8 Comparison of minimum sample sizes to have 80% power of achieving 5% significance for comparing two independent binomial proportions[a], for five different test procedures[b]

p_2	p_1	n_e	n_r	n_p	n_{as}	n^*
0.05	0.15	126	130	111	105	107
	0.20	67	72	59	55	56
	0.25	45	48	39	35	38
	0.30	34	36	28	25	28
	0.35	25	28	21	19	22
	0.40	20	22	17	15	18
	0.45	17	19	14	12	13
0.10	0.25	89	92	79	76	79
	0.30	56	58	49	47	49
	0.35	39	42	34	32	35
	0.40	30	31	25	24	26
	0.45	24	25	20	19	21
	0.50	19	20	16	15	17
	0.55	16	17	13	12	13
	0.60	13	14	11	10	10
0.15	0.30	106	108	95	94	96
	0.35	65	67	57	56	59
	0.40	46	46	39	38	40
	0.45	34	35	28	28	29
	0.50	26	27	22	21	23
	0.55	22	22	17	17	18
	0.60	17	18	14	13	14
	0.65	15	15	11	11	13
0.20	0.35	121	122	109	108	111
	0.40	73	74	64	64	68
	0.45	49	50	43	42	45
	0.50	36	37	31	30	32
	0.55	27	28	23	23	26
	0.60	23	23	18	18	20
	0.65	17	18	14	14	15
	0.70	15	15	12	12	13
0.25	0.40	132	133	120	119	123
	0.45	78	79	70	69	71
	0.50	54	53	46	46	48
	0.55	37	39	32	32	33
	0.60	30	30	24	24	26
	0.65	23	23	19	19	20
	0.70	18	19	15	15	17
	0.75	15	15	12	12	13

[a] From Suisa and Shuster (1985)

[b] n_e = Fisher's exact test; n_r = corrected chi-squared approximation; n_p = uncorrected chi-squared approximation; n_{as} = arcsin formula; n^* = unconditional exact test; p_2 = proportion exposed in control group; p_1 = proportion exposed among cases

Table 7.8 (*contd*)

p_2	p_1	n_e	n_r	n_p	n_{as}	n^*
0.30	0.45	142	141	128	128	132
	0.50	84	83	74	73	77
	0.55	55	55	48	48	50
	0.60	41	40	33	33	37
	0.65	31	30	25	25	27
	0.70	23	23	19	19	20
0.35	0.50	143	147	134	134	136
	0.55	85	86	76	76	79
	0.60	56	56	49	49	51
	0.65	41	40	34	34	37
0.40	0.55	144	149	136	136	144
	0.60	85	86	77	77	79

required is then given by

$$T/\{\bar{p}_1(1-\bar{p}_2)+\bar{p}_2(1-\bar{p}_1)\},$$

where T is the number of discordant pairs. Table 7.9 with $M=1$ indicates the number of matched pairs required for different values of R, p_2, α and β.

For studies involving $1:M$ matching, the approach is similar, if more complicated. We use the data layout and notation of §5.14, Volume 1, as below:

		Number of controls positive			
		0	1	...	M
Cases	Positive	$n_{1,0}$	$n_{1,1}$	$n_{1,2}$	$n_{1,M}$
	Negative	$n_{0,0}$	$n_{0,1}$	$n_{0,2}$	$n_{0,M}$

and we write $T_i = n_{1,i-1} + n_{0,i}$.

The usual test of the null hypothesis without the continuity correction is

$$\chi = \frac{\left\{\sum_{m=1}^{M}\left(n_{1,m-1} - \frac{mT_m}{(M+1)}\right)\right\}}{\left\{\frac{1}{(M+1)^2}\sum_{m=1}^{M} T_m m(M-m+1)\right\}^{1/2}}, \tag{7.9}$$

which, for significance at level α, we can write in the form

$$\sum_{m=1}^{M} n_{1,m-1} - E_{R=1}\left(\sum_{m=1}^{M} n_{1,m-1}\right) \geq Z_\alpha\left\{\mathrm{Var}_{R=1}\left(\sum_{m=1}^{M} n_{1,m-1}\right)\right\}^{1/2}.$$

Under the alternative hypothesis of a non-null relative risk R, we have (see §5.3, Volume 1)

$$E_R(n_{1,m-1}) = \frac{T_m mR}{mR + M - m + 1}$$

Table 7.9 Matched case-control studies. Number of case-control sets in a matched case-control study required to achieve given power at the given level of significance, for different values of the relative risk and different matching ratios

M[a]	Relative risk								
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	10.0
Proportion exposed = 0.1; significance = 5%; power = 80%									
1	757	241	131	87	65	52	43	37	17
2	559	176	95	63	46	37	30	26	11
4	460	144	77	51	37	29	24	21	9
10	400	124	66	43	32	25	21	17	7
20	380	118	63	41	30	24	19	16	7
Proportion exposed = 0.1; significance = 5%; power = 95%									
1	1283	398	211	138	101	79	65	55	23
2	963	299	158	103	76	59	48	41	17
4	804	250	133	87	63	49	40	34	14
10	708	221	118	77	56	44	36	31	12
20	677	211	113	74	54	42	35	29	12
Proportion exposed = 0.1; significance = 1%; power = 80%									
1	1204	380	206	137	102	81	67	58	27
2	881	274	146	96	71	56	46	40	18
4	720	221	117	76	56	44	36	31	13
10	623	189	99	64	47	36	30	25	10
20	591	178	93	60	44	34	28	23	10
Proportion exposed = 0.1; significance = 1%; power = 95%									
1	1855	575	305	200	147	115	95	81	35
2	1380	425	224	147	107	84	69	58	24
4	1142	351	185	120	88	68	56	47	19
10	1000	306	161	105	76	59	49	41	16
20	953	292	153	100	72	56	46	39	16
Proportion exposed = 0.3; significance = 5%; power = 80%									
1	355	122	71	50	39	32	28	25	14
2	264	90	52	37	29	24	20	18	10
4	219	74	43	30	23	19	17	15	8
10	191	65	37	26	20	17	14	13	7
20	182	62	35	24	19	16	13	12	6
Proportion exposed = 0.3; significance = 5%; power = 95%									
1	602	201	114	79	60	49	42	37	19
2	452	152	86	59	46	37	32	28	14
4	377	126	72	49	38	31	26	23	12
10	331	111	63	43	33	27	23	20	10
20	316	106	60	41	32	26	22	19	10

[a] M = number of controls per case

Table 7.9 (*contd*)

M	Relative risk								
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	10.0
Proportion exposed = 0.3; significance = 1%; power = 80%									
1	565	192	111	78	61	51	44	39	22
2	419	142	81	57	44	37	32	28	16
4	346	116	66	46	36	30	25	22	12
10	301	101	57	40	31	25	22	19	10
20	287	95	54	37	29	24	20	18	10
Proportion exposed = 0.3; significance = 1%; power = 95%									
1	870	291	165	114	88	72	62	54	29
2	651	217	123	85	65	54	46	40	21
4	540	180	102	70	54	44	38	33	17
10	474	157	89	61	47	38	32	28	15
20	452	150	84	58	44	36	31	27	14
Proportion exposed = 0.5; significance = 5%; power = 80%									
1	324	118	72	52	42	36	31	28	18
2	243	89	54	39	32	27	24	21	13
4	203	74	45	33	26	22	20	18	11
10	178	65	39	29	23	20	17	16	10
20	170	62	37	27	22	19	16	15	9
Proportion exposed = 0.5; significance = 5%; power = 95%									
1	550	195	115	83	65	55	48	42	24
2	413	147	87	63	50	42	36	32	18
4	344	123	73	52	41	35	30	27	15
10	302	108	64	46	36	30	26	23	13
20	289	102	61	43	34	29	25	22	13
Proportion exposed = 0.5; significance = 1%; power = 80%									
1	516	187	112	82	66	56	49	45	28
2	387	140	85	62	50	42	37	34	21
4	323	117	70	51	41	35	31	28	17
10	284	103	62	45	36	31	27	24	15
20	271	98	59	43	34	29	26	23	14
Proportion exposed = 0.5; significance = 1%; power = 95%									
1	795	282	167	120	95	80	69	62	36
2	597	212	126	91	72	61	53	47	28
4	498	177	105	76	60	50	44	39	23
10	437	155	92	66	52	44	38	34	20
20	417	148	88	63	50	42	36	33	19

Table 7.9 (*contd*)

M	Relative risk								
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	10.0
Proportion exposed = 0.7; significance = 5%; power = 80%									
1	417	160	100	75	61	53	47	43	28
2	316	122	77	58	47	41	37	34	22
4	265	102	65	49	40	35	31	29	19
10	234	91	57	43	36	31	28	26	17
20	224	87	55	42	34	30	27	25	16
Proportion exposed = 0.7; significance = 5%; power = 95%									
1	707	263	161	118	95	81	71	64	39
2	531	199	122	90	73	62	55	49	30
4	443	166	102	75	61	52	46	41	25
10	390	146	90	66	54	46	40	37	23
20	372	139	86	63	51	44	39	35	22
Proportion exposed = 0.7; significance = 1%; power = 80%									
1	663	252	157	117	96	83	74	67	44
2	504	193	121	91	75	65	58	53	35
4	424	163	103	78	64	56	50	46	31
10	376	146	92	70	57	50	45	41	28
20	360	140	88	67	55	48	43	40	27
Proportion exposed = 0.7; significance = 1%; power = 95%									
1	1022	381	233	172	138	118	104	94	58
2	771	289	178	132	107	91	81	73	46
4	645	242	150	111	90	77	68	62	39
10	569	214	132	98	80	69	61	55	35
20	544	205	127	94	76	66	58	53	34
Proportion exposed = 0.9; significance = 5%; power = 80%									
1	1045	417	268	205	170	148	134	123	83
2	798	322	209	161	134	118	107	98	68
4	674	275	179	139	116	103	93	86	60
10	599	246	162	125	106	93	85	79	56
20	575	237	156	121	102	90	82	76	54
Proportion exposed = 0.9; significance = 5%; power = 95%									
1	1772	688	432	323	264	226	201	182	113
2	1331	519	327	246	202	174	155	141	90
4	1111	434	275	207	170	147	132	120	78
10	979	384	244	184	152	132	118	108	71
20	936	367	233	177	146	127	113	104	68

Table 7.9 (*contd*)

M	Relative risk								
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	10.0
Proportion exposed = 0.9; significance = 1%; power = 80%									
1	1663	657	421	320	266	232	208	191	129
2	1278	514	333	256	214	188	170	157	109
4	1086	442	289	224	188	166	150	139	99
10	970	399	262	204	172	152	139	129	92
20	932	384	254	198	167	148	135	125	90
Proportion exposed = 0.9; significance = 1%; power = 95%									
1	2562	993	624	468	383	329	293	266	168
2	1942	760	481	363	299	259	231	211	137
4	1631	643	409	311	257	223	200	183	122
10	1445	573	367	279	232	202	181	167	112
20	1384	549	352	269	233	195	175	161	109

and

$$\mathrm{Var}_R(n_{1,m-1}) = T_m \frac{mR(M-m+1)}{(mR+M-m+1)^2}.$$

Sample size requirements are therefore determined from the equation

$$\sum_{m=1}^{M} \{E_R(n_{1,m-1}) - E_{R=1}(n_{1,m-1})\}$$
$$= Z_\alpha \left\{ \sum_{m=1}^{M} \mathrm{Var}_{R=1}(n_{1,m-1}) \right\}^{1/2} + Z_{1-\beta} \left\{ \sum_{m=1}^{M} \mathrm{Var}_R(n_{1,m-1} \right\}^{1/2}. \quad (7.10)$$

This equation involves the quantities $T_1, \ldots, T_M$. The probability P_m that an individual matched set contributes to a specific T_m is given in terms of p_1 and p_2 by

$$P_m = Pr(\text{matched set contributes to } T_m)$$
$$= \binom{M}{m}(1-p_1)p_2^m(1-p_2)^{M-m} + \binom{M}{m-1}p_1p_2^{m-1}(1-p_2)^{M-m+1}. \quad (7.11)$$

As in the case of matched pairs, for approximate sample size calculations we can use the mean values of p_1 and p_2 over all matched sets in this expression, rather than integrating it over the distribution of the p's over the matched sets. The quantities T_m in expression (7.10) are then replaced by NP_m, where N is the total number of matched sets and P_m is evaluated for the mean values of p_1 and p_2. Expression (7.10) can then be solved for N given α, β, p_1, p_2 and M.

More complex situations in which the number of controls per case varies can clearly be handled in the same way (Walter, 1980), with the numerator and denominator of (7.9) summed over all relevant sets. There is usually little point, however, in introducing fine detail into what are essentially rather crude calculations.

A continuity correction can be incorporated into the test given by expression (7.9) by subtracting one half from the absolute value of the numerator. The resulting sample sizes differ from those obtained by omitting the continuity correction by a factor A, given by

$$A = \tfrac{1}{4}\{1 + \sqrt{1 + 2(E_R - E_{R=1})/(Z_\alpha V_{R=1}^{1/2} + Z_{1-\beta} V_R^{1/2})^2}\}^2, \tag{7.12}$$

where

$$E_R = \sum_{m=1}^{M} E_R(n_{1,m-1})$$

and

$$V_R = \sum_{m=1}^{M} \mathrm{Var}_R(n_{1,m-1}).$$

Sample size calculations incorporating the continuity correction into the statistical test are comparable to the sample sizes given in Table 7.7 for unmatched studies.

Table 7.9 gives the number of matched sets required for a range of values of M, R, p_2, α and β using the continuity correction. The values can be compared with those in Table 7.7 for the number of cases required in unmatched analyses, to indicate the effect of matching on the sample size. As a case of special interest, we have included in Table 7.9 a large value of M. This corresponds to the situation in which one uses all available individuals as controls, of interest in the context of §5.4, where the entire risk set is potentially available.

We now turn to the question of how many controls should be selected for each case. There are several contexts in which this issue can be discussed, as outlined in Chapter 1. We may be in a situation, as in §5.4, in which all data are available and sampling from the risk sets is done solely for convenience and ease of computing. We should then want the information in the case-control series to correspond closely to the information in the full cohort, and we should select sufficient controls per case for the information loss to be acceptably small. Thus, in Table 7.9, we compare the power achieved by a given value of M with the value obtained when M is infinite, or, more generally, use expression (7.11) to evaluate the power (i.e., $Z_{1-\beta}$) for a range of values of M and R.

In other situations, the cohort may be well defined and the cases identified but information on the exposures of interest not readily available and the cost of obtaining it a serious consideration. One should then assess the marginal gain in power associated with choosing more controls.

On other occasions, as would arise in many conventional case-control studies, the investigator may be able to decide on both the number of case-control sets and the number of controls per case. The question would then be to decide on the optimal combination of controls per case and number of cases.

Several authors have considered optimal designs in terms of the costs of inclusion in the study of cases and controls (Schlesselman, 1982). On occasion, the separate costs of cases and controls may be available, and a formal economic calculation can then be

made. The more usual situation, however, is one in which one wants to know the cost in terms of the number of individuals required in the study, for different case-control ratios. For example, the rate at which cases are registered may be a limiting factor, and one would like to assess the cost, in terms of the number of extra controls required, of reducing the duration of the study by half, i.e., halving the number of cases, keeping the power constant.

The values in Table 7.9 can be used to provide answers to all three of these questions.

7.7 Efficiency calculations for matched designs

As an alternative to the criterion of power to compare different designs, one can use the efficiency of estimation of the parameter of interest, given by the expectation of the inverse of the variance of the estimate. The parameter of interest is often taken as the logarithm of the relative risk. As a comparative measure, the efficiency has attractions, since interest is usually centred more on parameter estimation than on hypothesis testing. For parameter values close to the null, power and efficiency considerations give, of course, very similar results. For parameter values distant from the null, however, the two approaches may diverge considerably. Efficiency considerations have the additional advantage that, at least in large samples, they can be derived directly from the second derivative of the likelihood function evaluated at just one point in the parameter space (see §7.11).

(*a*) *Relative size of the case and control series in unmatched studies*

In the simplest situation, of a single dichotomous variable, the results of a case-control study can be expressed as

	Exposure		Total
	+	−	
Case	a	b	n_1
Control	c	d	n_2

If p_1 is the probability of exposure for a case, and p_2 the corresponding probability for a control, then

$$E(a) = n_1 p_1 \qquad E(c) = n_2 p_2,$$

and in large samples the variance of the estimate of $\log R$ is given by

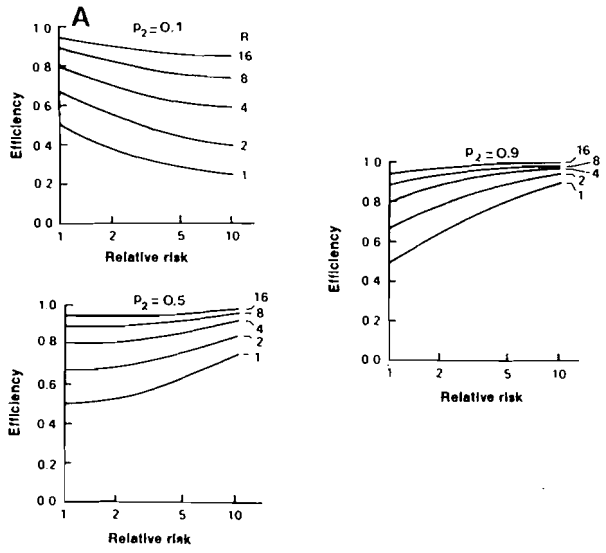
$$\frac{1}{n_1 p_1} + \frac{1}{n_1(1-p_1)} + \frac{1}{n_2 p_2} + \frac{1}{n_2(1-p_2)}. \tag{7.13}$$

When n_2 is large compared to n_1, as it typically would be in a cohort study, the variance is dominated by the first two terms. If we write $n_2 = kn_1$, so that k is the number of controls per case, then we can clearly evaluate (7.13) for different values of

p_2, R and k. When the relative risk is close to unity, then the efficiency relative to using the entire cohort for different values of k is well approximated by $(1+1/k)^{-1}$. The relative efficiency with $k=1$ is thus 50%, and with $k=4$ is 80%. Clearly, the marginal increase in relative efficiency as k increases beyond 4 becomes slight, hence, the conventional dictum that it is not worth choosing more than four controls per case. This is true, however, only when the expected relative risk is close to unity. As the relative risk diverges from one, considerably more than four controls per case may be necessary to achieve results close to those given by the entire cohort. Figure 7.1A

Fig. 7.1 Efficiency of case-control designs for differing values of the relative risk for a single dichotomous exposure E

The efficiency of a design, defined as v_k/v_∞, where v_k represents the asymptotic variance of the estimated log relative risk when using k controls per case, depends on both the relative risk and the control exposure probability p_2. Efficiencies for unmatched designs were computed from the unconditional likelihood (A). From Whittemore and McMillan (1982). Efficiencies for matched designs were computed from the conditional likelihood, assuming control exposure probabilities p_2 are constant across matching strata (B). From Breslow *et al.* (1983)



shows the change in efficiency for changing k, relative to using the entire cohort, for a number of values of p_2 and R.

(*b*) *Number of controls per case in a matched study*

With M controls per case and the layout of §7.6(*b*), the maximum likelihood equation for R is given by

$$\sum_{m=1}^{M} n_{1,m-1} = \sum_{m=1}^{M} \frac{T_m mR}{mR + M - m + 1}$$

(see §5.17 in Volume 1), from which the expectation of the inverse of the variance of $\log R$ is given by

$$[\text{Var} \log R]^{-1} = \sum_{m=1}^{M} \frac{T_m mR(M - m + 1)}{(mR + M - m + 1)^2}. \tag{7.14}$$

Using approximate values for T_m given by (7.11), we can evaluate this expression for given values of R, M and p_2. As in the previous paragraph, large values of M correspond to the inclusion of the entire risk set (see §5.4), and the relative values one obtains for small M give the relative efficiency of choosing a small number of controls per risk set. Results are given in Figure 7.1B, taken from Breslow *et al.* (1983), which can be compared with Figure 7.1A. From both figures it is clear that as the relative risk increases, for small values of p_2, a substantial loss is sustained by selecting only a small number of controls. When $R = 1$, one has the same result as in the previous section, that the efficiency relative to a large number of controls is given by $M/(M + 1)$. This result is a convenient rule of thumb when R is close to 1; but, as R increases, for many values of p_2 it becomes increasingly misleading.

7.8 Effect of confounding on sample size requirements

We now consider the effect on the required sample size if account must be taken of a confounding factor. We consider the situation in which we have a single polytomous confounding variable, C, which can take K different values. We assume that the situation is given by the following layout for each stratum, and for simplicity treat the case of equal numbers of cases and controls. We assume further that there is no interaction.

Exposure	Total control population	Stratum i (C takes value i)	
		Number of controls	Relative risk of disease
$E+$	nP	nPp_{1i}	$R_E R_{C_i}$
$E-$	$n(1-P)$	$n(1-P)p_{2i}$	R_{C_i}

where n is the total number of controls. Thus, R_E is the exposure-related relative risk for disease given C, R_{C_i} is the relative risk of the ith level of the confounder given E,

p_{1i} is the proportion of those exposed to E also exposed to C_i, p_{2i} is the proportion of those not exposed to E who are exposed to C_i, and P is the proportion exposed to E in the control population. We have taken $R_{C_1} = 1$.

When C is not a confounder, inferences on R_E can be based on the pooled table given by

	Case	Control
Exposed	nPR_E/Σ	nP
Not exposed	$n(1-P)/\Sigma$	$n(1-P)$

where $\Sigma = (PR_E + 1 - P)$.

For a given value of R_E, power β and significance α, the required number of cases is obtained by solving the equation

$$\log R_E = Z_\alpha \sqrt{V_N} + Z_{1-\beta}\sqrt{V_A}, \tag{7.15}$$

where V_N is the variance of the estimate of $\log R_E$ under the null hypothesis that $R_E = 1$, and V_A the equivalent variance with the given value of R_E. They are given when inferences are based on the pooled table by

$$nV_N = 4\Sigma[1/\{P(R_E + \Sigma)\} + 1/(1-P)(1+\Sigma)]$$

and

$$nV_A = \left(\frac{1}{P} + \frac{1}{1-P} + \frac{\Sigma}{PR_E} + \frac{\Sigma}{1-P}\right).$$

When C is a confounder, then stratification is required to give unbiased estimates of R_E. The variances in equation (7.15) now have to be replaced by the variances of the stratified estimate of R_E. An approximation to the variance of the Wolff estimate of the logarithm of R_E (see expression 3.16) which has often been used in the past (Gail, 1973; Thompson, W.D. *et al.*, 1982, Smith & Day, 1984) is given by

$$V_W = \left(\Sigma \frac{1}{V_i}\right)^{-1},$$

where V_i is the variance of the logarithm of the odds ratio derived from stratum i (given by the expression from stratum i corresponding to V_N and V_A of the previous paragraph). V_W can be calculated for the null case ($R_E = 1$), $V_{W,N}$, say, and for values of R_E of interest $V_{W,A}$, say. We then solve for

$$\log R_E = Z_\alpha \sqrt{V_{W,N}} + Z_{1-\beta}\sqrt{V_{W,A}}.$$

Writing

$$\Sigma' = PR_E \sum_{i=1}^{K} p_{1i} R_{C_{2i}} + (1-P) \sum_{i=1}^{K} p_{2i} R_{C_{1i}}$$

we have

$$nV_{i,A} = \frac{1}{Pp_{1i}} + \frac{1}{(1-P)p_{2i}} + \frac{\Sigma'}{Pp_{1i}R_{C_i}R_E} + \frac{\Sigma'}{(1-P)p_{2i}R_{C_i}}$$

and

$$nV_{i,N} = T_i^3/W_{1i}W_{2i}W_{3i}W_{4i},$$

where

$T_i = W_{1i} + W_{2i} = W_{3i} + W_{4i}$

$W_{1i} = Pp_{1i} + (1 - P)p_{2i}$ = proportion of controls in stratum i

$W_{2i} = (Pp_{1i}R_{C_i}R_E + (1 - P)p_{2i}R_{C_i})/\Sigma'$ = proportion of cases in stratum i

$W_{3i} = Pp_{1i}(1 + R_{C_i}R_E/\Sigma')$ = proportion exposed in stratum i

$W_{4i} = (1 - P)p_{2i}(1 + R_{C_i}/\Sigma')$ = proportion nonexposed in stratum i.

In the situation with only two strata, extensive tabulations have been published (Smith & Day, 1984) for a range of values of P, p_{1i}, p_{2i}, R_E and R_C. Some of the results are given in Table 7.10. The main conclusion to be drawn is that, unless C and E are strongly related, or C strongly related to disease (meaning by 'strongly related' an odds ratio of 10 or more), an increase of more than 10% in the sample size is unlikely to be needed. An alternative approach is through approximations to the variance of estimates obtained through the use of logistic regression, which has been used to investigate the joint effect of several confounding variables (Day *et al.*, 1980). Results using this approach restricted to the case of two dichotomous variables are also given in Table 7.10; for values of R_C near to one, the approximation is close to the approach given above. For several confounding variables that are jointly independent, conditional on E, as a rough guide one could add the extra sample size requirements for each variable separately.

7.9 Change in sample size requirements effected by matching

If a matched design is adopted, then equal numbers of cases and controls are included in each stratum. Usually, the numbers in each stratum would be determined by the distribution of cases rather than of controls (i.e., one chooses controls to match the available cases), so that they would be given by n times the W_{2i} of the preceding section. The computation then proceeds along similar lines to that of the previous section, and the sample size is given by

$$\log R_E = Z_\alpha\sqrt{V^1_{\mathrm{W},N}} + Z_{1-\beta}\sqrt{V^1_{\mathrm{W},A}},$$

where $V^1_{\mathrm{W},N}$ and $V^1_{\mathrm{W},A}$ correspond to $V_{\mathrm{W},A}$ and $V_{\mathrm{W},N}$ but with the constraint of matching. Alternatively, one can compare the relative efficiencies of matched and unmatched designs, in terms of the variance of the estimates. Table 7.11, from Smith and Day (1981), compares the efficiency of the matched and unmatched designs. The main conclusion is that unless C is strongly related to disease (odds ratio greater than 5) there is little benefit from matching. A similar derivation is given by Gail (1973).

Table 7.10 Increase in sample size required to test for a main effect if the analysis must incorporate a confounding variable. The ratio (×100) of the sample sizes, n_C and n, required to have 95% power to detect an odds ratio associated with exposure, R_E, at the 5% level of significance (one-sided) where n_C = sample size required allowing for stratification on confounding variable C and n = sample size required if stratification on C is ignored

P	p_1	p_2	R_{CE} [a]	$R_E = 2.0$				$R_E = 5.0$				$R_E = 10.0$			
				$R_C =$ 1.0	2.0	5.0	10.0	$R_C =$ 1.0	2.0	5.0	10.0	$R_C =$ 1.0	2.0	5.0	10.0
0.1	0.5	0.5	1.0 100	100	102	113	124	100	102	113	124	100	102	114	125
	0.6	0.4	2.3 102	101	100	105	112	102	103	111	119	103	106	116	126
	0.7	0.3	5.4 107	109	103	103	107	112	111	117	123	116	119	130	139
	0.8	0.2	16.0 120	126	115	112	113	138	135	140	146	148	153	167	177
	0.9	0.1	81.0 164	185	164	154	154	223	218	225	234	256	269	297	315
0.5	0.5	0.5	1.0 100	100	102	114	125	100	102	114	125	100	102	113	123
	0.6	0.4	2.3 104	104	109	124	138	103	112	131	146	103	113	133	150
	0.7	0.3	5.4 119	118	127	149	166	117	133	163	187	115	135	170	197
	0.8	0.2	16.0 156	155	171	204	230	151	180	232	270	145	182	245	290
	0.9	0.1	81.0 278	275	310	378	431	264	331	443	523	248	330	465	561
0.9	0.5	0.5	1.0 100	100	102	113	124	100	102	111	121	100	101	109	117
	0.6	0.4	2.3 102	101	110	132	151	100	111	131	149	100	111	131	148
	0.7	0.3	5.4 107	105	123	159	192	103	123	161	193	102	124	162	194
	0.8	0.2	16.0 120	115	145	207	265	110	143	211	271	107	143	214	275
	0.9	0.1	81.0 164	148	203	327	456	134	193	328	466	125	187	328	470
P	p_1	p_2	R_{CE} [a]	$R_E = 0.5$				$R_E = 0.2$				$R_E = 0.1$			
				$R_C =$ 1.0	2.0	5.0	10.0	$R_C =$ 1.0	2.0	5.0	10.0	$R_C =$ 1.0	2.0	5.0	10.0
0.1	0.5	0.5	1.0 100	100	102	113	124	100	102	111	121	100	101	109	117
	0.6	0.4	2.3 102	101	97	100	106	100	96	97	101	100	94	93	96
	0.7	0.3	5.4 107	105	95	92	94	103	92	86	87	102	89	81	80
	0.8	0.2	16.0 120	115	98	89	87	110	91	80	77	107	86	73	69
	0.9	0.1	81.0 164	148	118	100	94	134	101	81	75	125	91	70	64
0.5	0.5	0.5	1.0 100	100	102	114	125	99	102	114	125	100	102	113	123
	0.6	0.4	2.3 104	104	104	113	123	103	101	107	114	103	99	102	108
	0.7	0.3	5.4 119	118	116	123	132	117	109	109	114	115	104	100	102
	0.8	0.2	16.0 156	155	149	153	162	151	134	127	128	145	122	109	107
	0.9	0.1	81.0 278	275	258	259	270	264	221	196	192	248	194	159	149
0.9	0.5	0.5	1.0 100	100	102	114	125	100	102	113	124	100	102	113	124
	0.6	0.4	2.3 102	101	109	128	145	102	107	123	138	103	106	119	133
	0.7	0.3	5.4 107	109	122	152	180	112	120	142	165	116	119	135	154
	0.8	0.2	16.0 120	126	148	197	245	138	150	184	221	148	151	173	201
	0.9	0.1	81.0 164	185	224	319	421	223	244	307	381	256	259	296	349

[a] Approximation to $(n_C/n) \times 100$ based on the normal approximation to logistic reqression $= 1/(1 - q^2)$, where q = correlation coefficient between E and C, $q^2 = P(1-P)(p_1 - p_2)^2/\{(Pp_1 + (1-P)p_2)(1 - Pp_1 - (1-P)p_2)\}$. See Smith and Day (1984).

P = proportion of controls exposed to E;

p_1 = proportion exposed to E who were also exposed to C;

p_2 = proportion not exposed to E who were exposed to C;

R_{CE} = odds ratio measure of association between E and C

Table 7.11 Relative efficiency of an unmatched to a matched design, in both cases with a stratified analysis, when the extra variable is a positive confounder. The body of table shows the values of $100 \times V_{MS}/V_S$[a] (where MS = 'matched stratified'; S = 'stratified')

P	p_1	p_2	R_{CE}	$R_E = 2.0$				$R_E = 5.0$				$R_E = 10.0$			
				$R_C =$ 1.0	2.0	5.0	10.0	$R_C =$ 1.0	2.0	5.0	10.0	$R_C =$ 1.0	2.0	5.0	10.0
0.1	0.5	0.5	1.0	100	97	87	79	100	98	88	79	100	98	89	80
	0.6	0.4	2.3	99	91	78	69	98	90	78	70	98	91	79	71
	0.4	0.2	2.7	99	90	73	60	98	90	74	62	98	90	76	64
	0.8	0.6	2.7	99	92	84	80	98	91	84	79	97	91	84	79
	0.7	0.3	5.4	98	84	69	60	93	82	68	61	92	82	71	63
	0.9	0.1	81.0	89	69	51	43	79	66	53	47	80	71	62	57
0.3	0.5	0.5	1.0	100	97	87	79	100	97	87	80	100	97	88	81
	0.6	0.4	2.3	100	95	84	77	101	97	88	80	102	100	91	84
	0.4	0.2	2.7	100	96	83	71	101	99	88	76	102	102	92	82
	0.8	0.6	2.7	100	96	89	85	100	97	91	87	102	100	94	90
	0.7	0.3	5.4	99	93	82	74	102	99	89	81	107	106	96	88
	0.9	0.1	81.0	96	88	76	69	108	105	94	85	130	129	115	102
0.5	0.5	0.5	1.0	100	97	88	80	100	97	89	83	100	98	91	87
	0.6	0.4	2.3	100	99	90	82	101	100	93	87	102	101	96	91
	0.4	0.2	2.7	101	100	90	79	101	102	94	84	102	102	97	89
	0.8	0.6	2.7	100	99	93	89	101	100	96	93	102	101	98	95
	0.7	0.3	5.4	101	100	92	84	105	106	99	92	107	108	102	96
	0.9	0.1	81.0	106	107	98	90	129	133	121	108	152	157	139	122
0.7	0.5	0.5	1.0	100	97	88	82	100	98	91	87	100	99	94	91
	0.6	0.4	2.3	101	101	93	87	101	101	96	91	101	101	97	94
	0.4	0.2	2.7	101	102	95	84	101	102	97	89	101	102	98	93
	0.8	0.6	2.7	101	101	96	93	101	101	98	96	101	101	99	97
	0.7	0.3	5.4	102	105	99	92	104	106	102	96	103	105	102	98
	0.9	0.1	81.0	112	122	117	107	131	141	133	120	138	146	137	124
0.9	0.5	0.5	1.0	100	97	89	83	100	98	93	89	100	99	96	94
	0.6	0.4	2.3	100	102	96	90	100	101	97	94	100	100	98	96
	0.4	0.2	2.7	100	103	98	88	100	101	97	92	100	101	98	95
	0.8	0.6	2.7	100	102	98	95	100	101	99	97	100	101	99	98
	0.7	0.3	5.4	101	106	103	97	101	104	101	98	101	102	101	99
	0.9	0.1	81.0	109	125	128	120	112	122	122	116	109	115	115	111

[a] From Smith and Day (1981)

7.10 Interaction and matching

Occasionally, the major aim of a study is not to investigate the main effect of some factor, but to examine the interaction between factors. One might, for example, want to test whether obesity is equally related to pre- and post-menopausal breast cancer, or whether the relative risk of lung cancer associated with asbestos exposure is the same among smokers and nonsmokers. The basic question of interest is whether two relative risks are equal, rather than if a single relative risk is equal to unity. For illustrative purposes, we consider the simplest situation of two 2×2 tables, with a layout as before but restricted to two strata and with an interaction term, R_I, added.

Exposure	Proportion of population	Confounder	Proportion of population	Relative risk of disease
$E+$	P	$C+$	Pp_1	$R_E R_C R_I$
		$C-$	$P(1-p_1)$	R_E
$E-$	$1-P$	$C+$	$(1-P)p_2$	R_C
		$C-$	$(1-P)(1-p_2)$	1

If $\hat{\psi}_1$ is the odds ratio associating E with disease in the stratum with $C+$, and $\hat{\psi}_2$ the corresponding estimate in the stratum with $C-$, then

$$\text{Var}(\log R_I) = \text{Var}\{\log(\hat{\psi}_1/\hat{\psi}_2)\} = \text{Var}(\log \hat{\psi}_1) + \text{Var}(\log \hat{\psi}_2),$$

and the required sample size is given by the solution of

$$(\log R_I)^2 = (Z_\alpha \sqrt{V_N} + Z_{1-\beta}\sqrt{V_A}),$$

where V_N is the expected value of $\text{Var}(\log R_I)$ in the absence of interaction, and V_A is the expected value of $\text{Var}(\log R_I)$ at the value R_I. Some results are shown in Figures 7.2, 7.3 and 7.4. The most striking results are perhaps those of Figure 7.4, in which the

Fig. 7.2 Sample size for interaction effects between dichotomous variables. Size of study required to have 95% power to detect, using a one-sided test at the 5% level, the difference between a two-fold increased risk among those exposed to E and C and no increased risk among those exposed to E but not to C ($R_E = 1$; $R_I = 2$). The variable C is taken to be not associated with exposure ($p_1 = p_2 = p$) and not associated with disease among those not exposed to E ($R_C = 1$). From Smith and Day (1984)

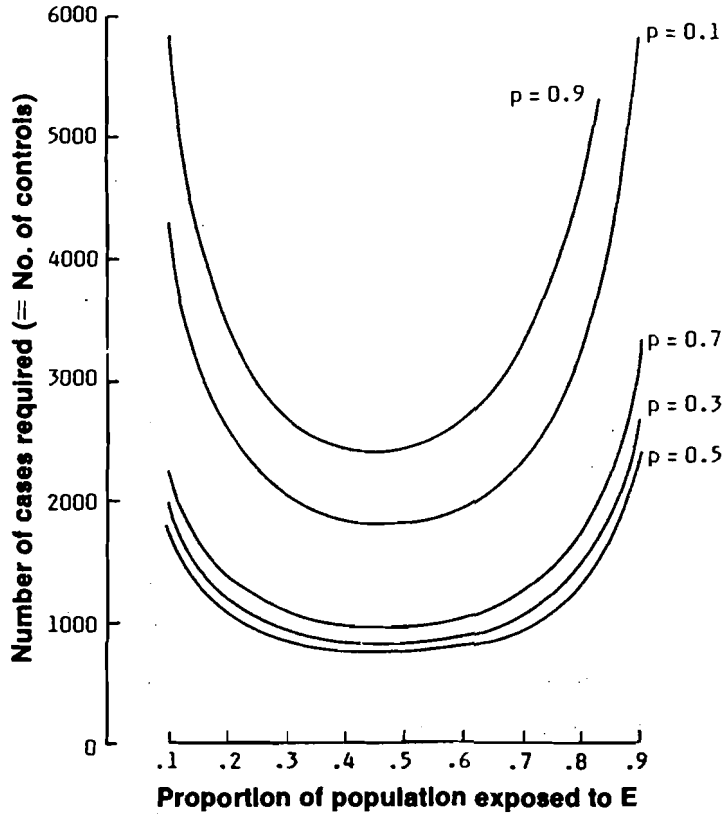
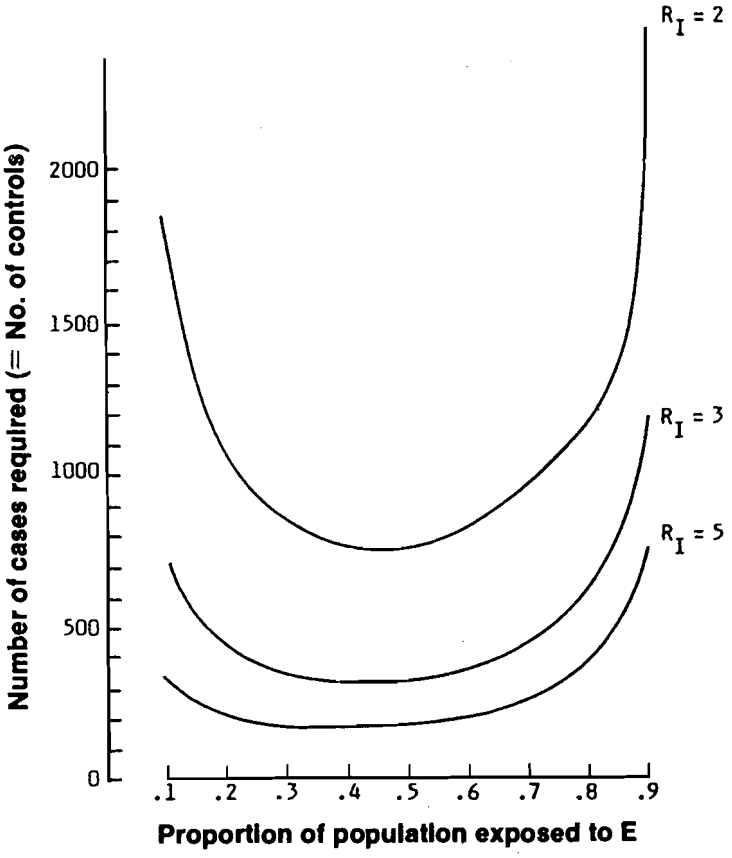


Fig. 7.3 Sample size for interaction effects between dichotomous variables. Size of study required to have 95% power to detect, using a one-sided test at the 5% level, the difference between no increased risk among those exposed to E but not to C ($R_E = 1$) and an R_I-fold increased risk among those exposed to both E and C. It has been assumed that 50% of the population are exposed to C ($p_1 = p_2 = 0.5$) and C is not associated with disease among those not exposed to E ($R_C = 1$). From Smith and Day (1984)



sample size required to detect an interaction of size R_I is compared to the sample size required to detect a main effect of the same size. The former is always at least four times the latter, and often the ratio is considerably larger. This difference can be seen intuitively, for, whereas

$$\text{Var}(\log R_I) = v_1 + v_2,$$

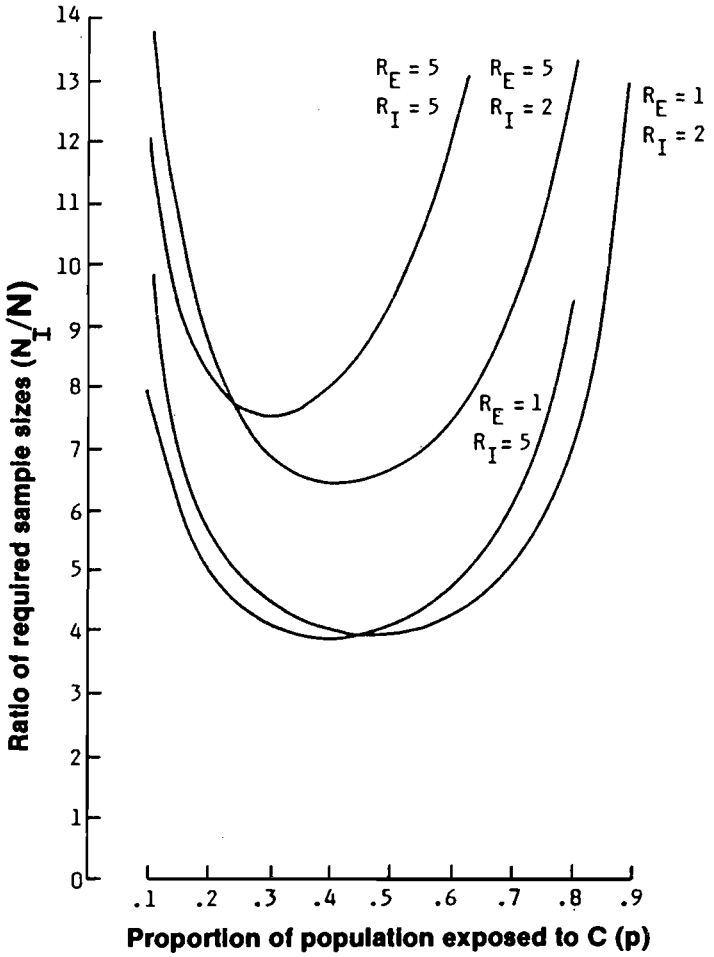
we have

$$\text{Var}(\log R_E) = v_1 v_2/(v_1 + v_2), \quad \text{approximately},$$

and the ratio $(v_1 + v_2)^2/v_1 v_2$ is always greater than or equal to 4, increasing the greater the disparity between v_2 and v_1.

One might imagine that matching, by tending to balance the strata, would improve tests for interaction, but in general the effect is slight (Table 7.12). Matching can, on occasion, have an adverse effect.

Fig. 7.4 Ratio of sample sizes required to have 95% power to detect, using a one-sided test at the 5% level, (i) an interaction of strength R_I and (ii) a main effect of strength R_I (relative risk of R_E for exposure to E for both levels, assuming 50% of the population exposed to E, $p_1 = p_2 = p$ and C not associated with disease among those not exposed to E ($R_C = 1$)). From Smith and Day (1984)



7.11 More general considerations

The previous sections have considered the simple case of dichotomous variables and power requirements for essentially univariate parameters. A more comprehensive approach can be taken in terms of generalized linear models. If interest centres on a p-dimensional parameter $\boldsymbol{\theta}$, then asymptotically the maximum likelihood estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, say, is normally distributed with mean $\boldsymbol{\theta}_0$, the true value, and variance covariance matrix given by the inverse of $I(\boldsymbol{\theta})$, the expected information matrix, the i,jth term of which is given by

$$-E\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\theta_i\,\delta\theta_j}\right], \qquad j = 1, \ldots, p,$$

Table 7.12 Effect of matching on testing for a non-null interaction. The ratio (×100) of the sample sizes, $n_I(MS)$ and $n_I(S)$ required to have 95% power to detect a difference at the 5% level of significance between an odds ratio associated with exposure E of R_E among those not exposed to C and an odds ratio for E of $R_E R_I$ among those exposed to C, where $n_I(MS)$ = sample size required in a matched stratified study and $n_I(S)$ = sample size required in an unmatched study[a]

P		$R_I = 2.0$						$R_I = 0.5$					
		$R_E = 1.0$			$R_E = 2.0$			$R_E = 1.0$			$R_E = 2.0$		
	$p_1(=p_2)$	$R_C = 1.0$	2.0	5.0	1.0	2.0	5.0	1.0	2.0	5.0	1.0	2.0	5.0
0.1	0.1	96	72	57	92	67	52	87	67	60	78	62	59
	0.3	98	93	107	96	93	112	96	102	135	96	109	158
	0.5	100	112	147	101	118	165	106	134	197	115	155	246
	0.7	102	129	176	106	141	207	116	164	244	133	200	315
	0.9	105	145	197	110	163	240	126	191	279	151	243	368
0.5	0.1	84	64	53	82	65	55	64	53	53	65	57	61
	0.3	93	92	107	93	93	109	92	101	126	95	106	125
	0.5	102	114	140	102	114	135	113	131	149	114	126	134
	0.7	109	130	158	110	128	148	128	144	154	124	130	131
	0.9	116	143	168	117	137	154	137	149	155	125	128	128
0.9	0.1	76	59	50	82	68	60	54	47	52	67	61	70
	0.3	90	91	107	94	95	107	91	103	120	99	110	121
	0.5	102	113	132	102	110	120	112	120	121	112	116	114
	0.7	111	126	144	108	116	123	117	118	115	112	111	107
	0.9	118	132	144	111	118	123	113	112	110	105	104	102

P		$R_I = 0.5$						$R_I = 0.2$					
		$R_E = 2.0$			$R_E = 4.0$			$R_E = 5.0$			$R_E = 10.0$		
	$p_1(=p_2)$	$R_C = 1.0$	2.0	5.0	1.0	2.0	5.0	1.0	2.0	5.0	1.0	2.0	5.0
0.1	0.1	105	73	48	110	72	46	126	77	43	151	89	46
	0.3	102	88	89	106	87	86	116	87	76	133	93	74
	0.5	100	102	123	101	101	125	106	97	111	115	98	106
	0.7	98	114	145	96	115	156	96	107	137	96	102	134
	0.9	96	124	158	92	127	178	87	115	152	78	106	156
0.5	0.1	116	88	62	117	93	67	137	115	80	125	115	90
	0.3	109	95	92	110	97	92	128	109	92	124	111	95
	0.5	102	101	116	102	100	112	113	101	103	114	103	100
	0.7	93	106	134	93	103	126	92	93	112	95	93	105
	0.9	84	111	147	82	106	137	64	84	119	65	81	109
0.9	0.1	118	98	74	111	99	81	113	111	98	105	106	99
	0.3	111	99	94	108	99	95	117	111	99	112	108	99
	0.5	102	100	112	102	100	106	112	103	100	112	104	100
	0.7	90	101	126	94	100	115	91	89	101	99	94	100
	0.9	76	102	138	82	100	123	54	71	103	67	79	100

[a] From Smith and Day (1984)

where $\ell(\boldsymbol{\theta})$ is the logarithm of the likelihood function. An overall test that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is given by comparing

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' I(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{7.16}$$

with a χ^2 distribution on p degrees of freedom.

Power and sample size considerations are then approached through the distribution of the quadratic form (7.16) under alternative values for the true value of $\boldsymbol{\theta}$. In the general case, for an alternative $\boldsymbol{\theta}_1$, $\hat{\boldsymbol{\theta}}$ will have mean $\boldsymbol{\theta}_1$ and variance–covariance matrix $I^{-1}(\boldsymbol{\theta}_1)$, which will differ from $I^{-1}(\boldsymbol{\theta}_0)$. Power calculations will then require evaluation of the probability that a general quadratic form exceeds a certain value, necessitating direct numerical integration. Some special situations, however, give more tractable results. Whittemore (1981), for example, has given a sample size formula for the case of multiple logistic regression with rare outcomes. In the univariate case, expression (7.16) leads directly to the following relationship between sample size N and power β:

$$N = \{Z_\alpha I^{-1/2}(\theta_0) + Z_{1-\beta} I^{-1/2}(\theta_1)\}/(\theta_1 - \theta_0)^2,$$

where now I refers to the expected information in a single observation.

Table 7.13 Degree of approximation in sample size calculation assuming that the test statistic has the same variance under the alternative as under the null hypothesis – example of an unmatched case-control study with no continuity correction in the test statistic; equal number of cases and controls. Significance = 0.05; power = 0.80

(a) Sample sizes calculated using expression (7.7), without the continuity correction

Proportion exposed in control population	Relative risk				
	1.5	2.0	2.5	5.0	10.0
0.1	717	223	119	32	13.5
0.3	334	111	62	20	10.6
0.5	305	107	63	24	14.3
0.7	393	146	90	38	25
0.9	992	387	247	114	81

(b) Sample sizes calculated using expression (7.17)

Proportion exposed in control population	Relative risk				
	1.5	2.0	2.5	5.0	10.0
0.1	764	247	136	40	19.0
0.3	357	124	72	26	15.4
0.5	325	120	73	30	20.0
0.7	420	163	103	74	33
0.9	1056	430	282	140	103

More generally, in the multivariate situation, asymptotically only alternatives close to $\boldsymbol{\theta}_0$ are of interest, since power for distant alternatives will approach 100%. One can then take $I(\boldsymbol{\theta}_1)$ to be approximately the same as $I(\boldsymbol{\theta}_0)$. Under the alternative hypothesis, the statistic

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' I(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

will then follow a noncentral χ^2 distribution on p degrees of freedom, with noncentrality parameter

$$(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)' I(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0),$$

and the power will be given by the probability that this noncentral χ^2 distribution exceeds the α point of the central χ^2 distribution on p degrees of freedom. Greenland (1985) discusses this approach in a number of situations.

An example of the degree of approximation used in this approach is given in Table 7.13, for unmatched case-control studies without the continuity correction. The relationship between power and sample size provided by this approach is, using the notation of expression (7.7),

$$n = (Z_\alpha \sqrt{2\bar{p}\bar{q}} + Z_{1-\beta}\sqrt{2\bar{p}\bar{q}})/(p_1 - p_2)^2. \tag{7.17}$$

In Table 7.13, the results of using this expression in place of (7.7) are compared, no continuity correction being used in the latter. For moderate values of the relative risk, the difference is some 5% to 10%; for values of the relative risk of 5 or greater, the approximation can overestimate the required sample size by as much as 50%.

Since, on many occasions, the likelihood function and its derivatives take relatively simple values under the null hypothesis, this approach clearly has considerable utility when interest centres mainly on detecting weak or moderate excess risks.