Biometrics 62, 929–933
September 2006

READER REACTION

Approaches for Calculating Power for Case–Cohort Studies

Mimi Y. Kim,* Xiaonan Xue, and Yunling Du

Division of Biostatistics, Department of Epidemiology and Population Health,
Albert Einstein College of Medicine, Bronx, New York 10461, U.S.A.
**email:* mikim@aecom.yu.edu

Summary. An approach for determining the power of a case–cohort study for a single binary exposure variable and a low failure rate was recently proposed by Cai and Zeng (2004, *Biometrics* **60,** 1015–1024). In this article, we show that computing power for a case–cohort study using a standard case–control method yields nearly identical levels of power. An advantage of the case–control approach is that existing sample size software can be used for the calculations. We also propose an additional formula for computing the power of a case–cohort study for the situation when the event is not rare.

Key words: Case–cohort design; Power; Sample size.

1. Introduction

Cai and Zeng (2004) recently addressed the important problem of how to determine the sample size and power of a case–cohort study with a single binary exposure variable. They showed that when the censoring distributions are the same in the two exposure groups, the number of failures is very small in the full cohort, and no failure times are tied, the power of their proposed log-rank-type tests for case–cohort studies can be approximated by

$$P_{SP} = \Phi\left\{ Z_\alpha + \tilde{n}^{1/2}\theta\sqrt{\frac{p_1 p_2 p_D}{q+(1-q)p_D}}\right\}, \qquad (1)$$

where Z_α denotes the lower α tail point of the standard normal distribution, θ is the log-hazard ratio for the two exposure groups, p_j $(j = 1, 2)$ is the proportion of the population in the jth group, $\tilde{n}$ is the total number of subjects in the subcohort, p_D is the proportion of failures in the full cohort, and q is the sampling fraction of the subcohort.

Previously, case–cohort studies have often been designed using conventional methods for case–control studies because of the absence of more appropriate approaches. We were therefore interested in comparing the power calculated using this "naive" case–control approach and the proposed methods of Cai and Zeng. In this article, we show that the two approaches for computing power for a case–cohort study yield nearly identical levels of power. We also propose an additional formula for computing the power of a case–cohort study for the situation when the event is not rare.

2. Power Calculation Using the Case–Control Approach

With the case–control approach, the power can be computed using the standard method for comparing two independent binomial proportions (Breslow and Day, 1993):

$$P_{CC} = \Phi\left\{ \frac{Z_\alpha\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_D}+\frac{1}{n_C}\right)} + (p_{E|D} - p_{E|\bar{C}})}{\sqrt{\frac{p_{E|D}(1-p_{E|D})}{n_D} + \frac{p_{E|C}(1-p_{E|C})}{n_C}}}\right\}, \qquad (2)$$

where $p_{E|D}$ = Pr(exposed | case), $p_{E|C}$ = Pr(exposed | control), n_D = number of cases, n_C = number of controls, and $\bar{p} = (n_D p_{E|D} + n_C p_{E|C})/(n_D + n_C)$. In order to apply this to evaluate the power of a case–cohort study, the following relationships are used: $n_D = p_D\tilde{n}/q$; $n_C = \tilde{n}(1-p_D)$; $p_{E|C} = p_1$; $p_{E|D} = (e^\theta p_1)/\{1 + p_1(e^\theta - 1)\}$. The last result is based on the well-known approximation of the relative risk by the odds ratio when the outcome is rare.

To illustrate this approach, suppose that the full cohort size is 1000 subjects; $q = 20\%$ of the cohort randomly selected for the subcohort; $p_D = 0.10$; and $p_1 = 0.30$. The goal is to assess the power to detect a log relative risk (θ) of 0.5. Under these assumptions, we expect that $n_D = 0.10 \times 1000 = 100$ cases; $n_C = 0.90 \times 0.20 \times 1000 = 180$ controls; $p_{E|C} = p_1 = 0.3$; and $p_{E|D} = (e^{0.5})(0.3)/\{1 + 0.3(e^{0.5} - 1)\} = 0.41$. Then using these values with (2), the power of the study is estimated to

Table 1
Comparison of power computed using the method of Cai and Zeng (2004) *and the case–control approach*

Full cohort size (n)	Failure proposition p_D (%)	p_1	θ	q (%)	$P_{SP} = P_V$	P_{CC}	$P_{SP} - P_{CC}$
1000	10	0.3	0.5	10	0.507	0.496	0.011
				20	0.615	0.610	0.004
			1.0	10	0.953	0.958	−0.005
				20	0.987	0.989	−0.002
		0.5	0.5	10	0.567	0.522	0.045
				20	0.680	0.630	0.050
			1.0	10	0.976	0.953	0.024
				20	0.995	0.988	0.008
	5	0.3	0.5	10	0.382	0.397	−0.014
				20	0.434	0.457	−0.023
			1.0	10	0.852	0.877	−0.025
				20	0.905	0.927	−0.022
		0.5	0.5	10	0.430	0.404	0.025
				20	0.488	0.458	0.029
			1.0	10	0.902	0.865	0.037
				20	0.943	0.918	0.025
5000	5	0.3	0.5	1	0.437	0.428	0.008
				2	0.620	0.623	−0.003
			1.0	1	0.907	0.929	−0.021
				2	0.988	0.993	−0.005
		0.5	0.5	1	0.490	0.475	0.015
				2	0.686	0.662	0.024
			1.0	1	0.945	0.923	0.021
				2	0.995	0.991	0.005
	1	0.3	0.5	1	0.310	0.322	−0.012
				2	0.375	0.401	−0.025
			1.0	1	0.743	0.784	−0.041
				2	0.844	0.881	−0.038
		0.5	0.5	1	0.348	0.338	0.009
				2	0.422	0.408	0.014
			1.0	1	0.805	0.772	0.033
				2	0.895	0.869	0.026

P_{SP} = power based on SP_n; P_V = power based on V_n in Cai and Zeng (2004); P_{CC} = power based on case–control approach.

be 61%. Using the approach of Cai and Zeng, the power is very similar: 62%.

We reproduced the results from Table 1 in Cai and Zeng (2004) and added columns for the power using the case–control approach and the difference in power between the two methods. As shown in Table 1, there is a high degree of agreement between the power results for the two approaches. The median absolute difference in power is 0.022. In 90% of the cases, the difference in power was 0.038 or less; the maximum difference was 0.050. In the Appendix we show that mathematically the power formulas using (1) and (2) will give similar results under the rare disease assumption.

We also explored how the two methods perform when the event rate is higher: 25% or 40%, and compared the results with the empirical power of SP_n from a simulation study. Following Cai and Zeng (2004), failure times in the simulation study were generated from exponential distributions with hazard rates $\lambda_1 = 0.10$ in group 1 and $\lambda_2 = (0.15, 0.20)$ in group 2 and censoring times from a uniform distribution between $[0, \tau]$. The parameter τ was chosen to yield overall event rates (across both groups) of 25% and 40%. Results, which are shown in Table 2, indicate that while P_{CC} tends to yield conservative estimates of power, P_{SP} produces higher estimates of power compared to the empirical power. We note that our empirical power estimates for an event rate of 25% were higher than those reported in Table 4 of Cai and Zeng (2004). The discrepancy in results could be explained by the choice of τ in Cai and Zeng's paper, which may have been based on achieving a 25% event rate in group 1 only and not across both groups.

3. Power Calculation When the Event Is Not Rare

An approach for estimating the power of a case–cohort study that does not depend on a low event rate can be derived with a few additional assumptions. Let T_j denote the potential failure time for a subject in group j and $S_j(t) = P(T_j \geq t)$. Cai and Zeng showed that under the null hypothesis where $S_1(t) = S_2(t) = S(t)$ and when the weight function in the log-rank test is equal to $w(t) = 1$ then

$$\sigma^2_{SP} = \sigma^2 + \frac{2p_1p_2(1-q)}{q} \int \frac{\pi_1(t)\pi_2(t)}{p_1\pi_1(t) + p_2\pi_2(t)} S(t)\Lambda(t)\, d\Lambda(t), \tag{3}$$

Table 2
Comparison of power using different approaches with simulated power of SP_n when the event is not rare

λ_2	N	P_1	Censoring proportion $1 - p_D$	Sampling proportion q (%)	Simulated power	P_{SP}	P_{CC}	P_{SP2}
0.15	200	0.3	0.75	30	0.238	0.276	0.238	0.248
				40	0.273	0.302	0.263	0.278
		0.5	0.75	30	0.282	0.308	0.253	0.277
				40	0.317	0.338	0.276	0.311
	400	0.3	0.75	30	0.372	0.435	0.373	0.390
				40	0.416	0.478	0.414	0.440
		0.5	0.75	30	0.444	0.488	0.399	0.438
				40	0.485	0.535	0.437	0.494
0.20	200	0.3	0.75	30	0.457	0.558	0.490	0.502
				40	0.515	0.609	0.542	0.564
		0.5	0.75	30	0.559	0.621	0.503	0.562
				40	0.624	0.674	0.549	0.627
	400	0.3	0.75	30	0.706	0.812	0.744	0.755
				40	0.778	0.857	0.797	0.818
		0.5	0.75	30	0.805	0.867	0.757	0.816
				40	0.860	0.906	0.807	0.873
0.15	200	0.3	0.60	30	0.270	0.328	0.238	0.280
				40	0.319	0.372	0.271	0.326
		0.5	0.60	30	0.331	0.368	0.264	0.313
				40	0.381	0.418	0.296	0.367
	400	0.3	0.60	30	0.411	0.520	0.378	0.442
				40	0.497	0.586	0.434	0.518
		0.5	0.60	30	0.503	0.581	0.414	0.497
				40	0.582	0.651	0.469	0.579
0.20	200	0.3	0.60	30	0.525	0.657	0.498	0.568
				40	0.607	0.728	0.568	0.655
		0.5	0.60	30	0.641	0.722	0.526	0.631
				40	0.723	0.791	0.590	0.721
	400	0.3	0.60	30	0.781	0.894	0.759	0.822
				40	0.861	0.938	0.828	0.894
		0.5	0.60	30	0.874	0.935	0.778	0.876
				40	0.902	0.966	0.841	0.884

P_{SP2} = power based on SP_n without the rare disease assumption.

where σ^2 is the asymptotic variance of the log-rank test from the full cohort, $\pi_j(t) = P(C_j \geq t)$, C_j is the potential censoring time for a subject in group j, and $\Lambda(t)$ is the cumulative hazard function.

Let $A = \int \frac{\pi_1(t)\pi_2(t)}{p_1\pi_1(t)+p_2\pi_2(t)} S(t)\Lambda(t)d\Lambda(t)$. We make the following distributional assumptions:

(1) The study period is between 0 and 1, without loss of generality.
(2) The censoring variable C_{ij} follows a uniform distribution between [0, 1] in both groups such that $\pi_1(t) = \pi_2(t) = 1 - t$ for $t \in (0, 1)$.
(3) The failure times under H_0 follow a standard exponential distribution with common parameter λ for both groups.

Under these assumptions we have

$$A = \exp(-\lambda) + 2 \times p_D - 1, \tag{4}$$

Table 3
Values of λ and A for computing power when the event is not rare

p_D	λ	A
0.15	0.3343	0.0316
0.20	0.4642	0.0572
0.25	0.6058	0.0912
0.30	0.7614	0.1340
0.35	0.9336	0.1862
0.40	1.1262	0.2486
0.45	1.3439	0.3216
0.50	1.5936	0.4064

where λ can be found by solving

$$1 - p_D = \frac{1}{\lambda} \times \{1 - \exp(-\lambda)\}.$$

Table 3 provides values of λ and A for several different event rates.

The variance of the usual log-rank statistic, σ^2, in (3) is assumed to be estimated by $\frac{1}{n}(p_2^2 D_1 + p_1^2 D_2)$ and $D_j \approx p_j p_D n$ as in Cai and Zeng. Using these assumptions and the result in (4) to approximate σ_{SP}^2, the power of SP_n when the event rate is not rare can in turn be approximated by

$$P_{SP2} = \Phi\left\{ Z_\alpha + \tilde{n}^{1/2}\theta \sqrt{\frac{p_1 p_2 p_D}{q + (1-q)2A/p_D}} \right\}. \qquad (5)$$

By comparing (1) and (5), P_{SP2} will always yield more conservative estimates of power than P_{SP} because it can be shown that $\frac{2A}{p_D} > p_D$ or $A > \frac{p_D^2}{2}$. Table 2 indicates that the power estimated by P_{SP2}, which does not depend on a rare events assumption, is generally closer to the empirical power compared to the power using the other approaches, and is intermediate between P_{CC}, which tends to be conservative, and P_{SP}, which is anticonservative.

We assumed above that the censoring distribution is uniform over the entire study period. The power formula in (5) can be generalized to other censoring distributions simply by specifying the appropriate form for A. For example, in study designs where subjects are accrued at a uniform rate from time 0 to T_0 and then followed until the time of the event or to time T, where $T > T_0$, it can be shown that the same power formula specified in (5) can be utilized with

$$A = 1 + e^{-\lambda T} + \left(T - T_0 + \frac{2}{\lambda}\right)\left(\frac{1}{T_0}\right)\left\{e^{-\lambda T} - e^{-\lambda(T-T_0)}\right\}.$$

The parameter λ can be determined from

$$1 - p_D = \frac{1}{\lambda T_0}\left\{e^{-\lambda(T-T_0)} - e^{-\lambda T}\right\}.$$

For other survival and censoring distributions, numerical integration may be needed to evaluate A.

In summary, Cai and Zeng presented a much needed practical method for determining the power and sample size of a case–cohort study. However, we found that for rare events, computing power using a standard method for case–control studies yielded nearly identical levels of power. This suggests that the case–control approach may also be a reasonable way for approximating the power of a case–cohort study when the outcome rate is low. An advantage of the case–control approach is that existing sample size software can be used for the calculations. For the situation when the event is not rare, we extended the approach of Cai and Zeng and provided a sample size formula based on the common assumption of exponentially distributed failure times that was shown to perform well when compared to the empirical power from simulation studies.

References

Breslow, N. and Day, N. (1993). *Statistical Methods in Cancer Research.* Oxford: Oxford University Press.

Cai, J. and Zeng, D. (2004). Sample size/power calculation for case–cohort studies. *Biometrics* **60,** 1015–1024.

Received January 2005. *Revised September* 2005. *Accepted October* 2005.

Appendix

Similarity in Power between Case–Cohort and Case–Control Approaches under Rare Events Assumption

The power using the case–cohort approach of Cai and Zeng under a rare events assumption is equal to

$$P_{SP} = \Phi\left\{ Z_\alpha + \tilde{n}^{1/2}\theta \sqrt{\frac{p_1 p_2 p_D}{q + (1-q)p_D}} \right\} \qquad (A.1)$$

and the power using the case–control approach is equal to

$$P_{CC} = \Phi\left\{ \frac{Z_\alpha \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_D} + \frac{1}{n_C}\right)} + (p_{E|D} - p_{E|C})}{\sqrt{\frac{p_{E|D}(1-p_{E|D})}{n_D} + \frac{p_{E|C}(1-p_{E|C})}{n_C}}} \right\}. \qquad (A.2)$$

We assume that $\bar{p}(1-\bar{p})(\frac{1}{n_D} + \frac{1}{n_C}) \approx \frac{p_{E|D}(1-p_{E|D})}{n_D} + \frac{p_{E|C}(1-p_{E|C})}{n_C}$. Then P_{CC} reduces to

$$P_{CC} = \Phi\left\{ Z_\alpha + \frac{(p_{E|D} - p_{E|C})}{\sqrt{\frac{p_{E|D}(1-p_{E|D})}{n_D} + \frac{p_{E|C}(1-p_{E|C})}{n_C}}} \right\}.$$

Under the rare events assumption, $(p_{E|D}/1-p_{E|D})(p_{E|C}/1-p_{E|C})^{-1} \approx e^\theta$, such that $p_{E|D} = (e^\theta p_{E|C})/(1 - p_{E|C} + e^\theta p_{E|C})$. Also, $n_D = (\tilde{n}/q)p_D, n_C \approx \tilde{n}$, and $p_{E|C} \approx p_1$. Then after some algebra, we have

$$\frac{(p_{E|D} - p_{E|C})}{\sqrt{\frac{p_{E|D}(1-p_{E|D})}{n_D} + \frac{p_{E|C}(1-p_{E|C})}{n_C}}} = \tilde{n}^{1/2}(e^\theta - 1)\sqrt{\frac{p_1 p_2 p_D}{e^\theta q + p_D\left\{1 + p_1(e^\theta - 1)\right\}^2}}$$

such that P_{CC} simplifies further to

$$P_{CC} = \Phi\left[Z_\alpha + \tilde{n}^{1/2}(e^\theta - 1)\sqrt{\frac{p_1 p_2 p_D}{e^\theta q + p_D\left\{1 + p_1(e^\theta - 1)\right\}^2}} \right]. \qquad (A.3)$$

Similarity between P_{SP} and P_{CC} is demonstrated if the term $\theta/(q + (1-q)p_D)^{1/2}$ in (A.1) can be shown to be similar to $(e^\theta - 1)/(e^\theta q + p_D[1 + p_1(e^\theta - 1)]^2)^{1/2}$ in (A.3). Let R be the ratio of these two terms such that

$$R = \frac{\frac{\theta}{\sqrt{q + (1-q)p_D}}}{\frac{(e^\theta - 1)}{\sqrt{e^\theta q + p_D\left\{1 + p_1(e^\theta - 1)\right\}^2}}}$$

or

$$R^2 = \left\{\frac{\theta^2 e^\theta}{(e^\theta - 1)^2}\right\}\left[\frac{q + p_D\left\{1 + p_1(e^\theta - 1)\right\}^2 e^{-\theta}}{q + p_D(1-q)}\right].$$

A second-order Taylor series expansion of e^θ around 0 yields $e^\theta \approx 1 + \theta + \theta^2/2$. Then R can be reexpressed as

$$R = \sqrt{\frac{A_1(\theta)q + A_2(\theta, p_1)p_D}{q + (1-q)p_D}},$$

where $A_1(\theta) = \frac{1}{(1+\theta/2)^2} + \frac{\theta}{1+\theta/2}$ and $A_2(\theta, p_1) = (\frac{1}{1+\theta/2} + p_1\theta)^2$.

The goal is to determine R_L and R_U, the lower and upper bounds for R, respectively, over realistic values for the parameters. We assume that the ranges of interest for θ and p_1 are (0, 1.1] and (0, 0.5], respectively, where the upper bound for θ of 1.1 corresponds to a relative risk of 3.0. Then because $A_1(\theta)$ is an increasing function of θ, $A_1(\theta)$ will be maximized at $A_1(1.1) = 1.13$ and $A_2(\theta, p_1)$ will be a maximum at $A_2(1.1, 0.5) = 1.43$. Therefore, an upper limit for R can be expressed as

$$R_U = \sqrt{\frac{1.13q + 1.43p_D}{q + (1-q)p_D}}.$$

For $p_D \in [0.01 - 0.10]$ and $q \in [0.10 - 0.50]$, the maximum value of R_U is equal to 1.16. When the upper bound of θ is increased to 1.4 (corresponding to a relative risk of 4.0) $R_U = 1.22$. Using similar arguments, it can be shown that the minimum value of R_L is equal to 0.90.

In order to make this result more interpretable, if P_{SP} is expressed as $P_{SP} = \Phi\{Z_\alpha + K(\tilde{n}^{1/2}, \theta, p_1, p_D, q)\}$ then the upper bound for the difference in power, $(P_{SP} - P_{CC})$, will be

$$\Phi\big\{Z_\alpha + K\big(\tilde{n}^{1/2}, \theta, p_1, p_D, q\big)\big\} - \Phi\big\{Z_\alpha + K\big(\tilde{n}^{1/2}, \theta, p_1, p_D, q\big)\big/1.16\big\}.$$

Similarly, the lower bound for $(P_{SP} - P_{CC})$ will be

$$\Phi\big\{Z_\alpha + K\big(\tilde{n}^{1/2}, \theta, p_1, p_D, q\big)\big\} - \Phi\big\{Z_\alpha + K\big(\tilde{n}^{1/2}, \theta, p_1, p_D, q\big)\big/0.90\big\}.$$

For example, for $P_{SP} = 90\%$ and $\alpha = 0.05, K(\tilde{n}^{1/2}, \theta, p_1, p_D, q) = 2.92$. It follows that $-4.6\% < (P_{SP} - P_{CC}) < 9.0\%$. Our results show that in most cases, the difference in power will be much smaller than the limits indicated.

The authors replied as follows:

We thank the authors for their interest in our article. The new way of interpreting our power formula (1) in this article is welcome. It is also interesting to see both the theoretical derivation and the numerical examples in determining the proximity between their formula and ours. It is not surprising that the case–control approach should give a similar expression to ours: as the event is rare, the follow-up times would be much less informative in comparing two groups than the number of cases; thus, the power formula for a case–cohort study could well be approximated by that of a case–control study.

It is also appreciated that the authors provide a formula extending to a nonrare event. However, assuming uniform distribution for censoring time and exponential distribution for failure time is restrictive and may not be appropriate in practice. For other censoring and failure time distributions, separate formulae need to be derived. In addition, their nonrare event formula is developed under the situation where all the cases are included in the case–cohort sample. However, when the event is not rare, investigators cannot usually afford to include all the cases. It is more practical and less expensive to select a subset of failures instead of all the failures, and the proposed formula is less attractive in this case. In our follow-up work to Cai and Zeng (2004), we are able to develop an approximate formula for the general case–cohort design including rare and nonrare events. Moreover, our formula allows arbitrary censoring and failure time distributions. The details of our formula will be discussed in a separate communication.

Reference

Cai, J. and Zeng, D. (2004). Sample size/power calculation for case–cohort studies. *Biometrics* **60,** 1015–1024.

Jianwen Cai and Donglin Zeng
Department of Biostatistics
CB 7420
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-7420
U.S.A.