SAMPLE SIZE/POWER CALCULATION FOR STRATIFIED CASE-COHORT DESIGN

Wenrong Hu,
Applied Statistics, Department of Mathematical Sciences, The University of Memphis;
Department of Biostatistics, CSL Behring

Jianwen Cai, and
Department of Biostatistics, School of Public Health, University of North Carolina Chapel Hill

Donglin Zeng
Department of Biostatistics, School of Public Health, University of North Carolina Chapel Hill

Abstract

The Case-cohort (CC) study design usually has been used for risk factor assessment in epidemiologic studies or disease prevention trials for rare diseases. The sample size/power calculation for the CC design is given in Cai and Zeng [1]. However, the sample size/power calculation for a stratified case-cohort (SCC) design has not been addressed before. This article extends the results of Cai and Zeng [1] to the SCC design. Simulation studies show that the proposed test for the SCC design utilizing small sub-cohort sampling fractions is valid and efficient for situations where the disease rate is low. Furthermore, optimization of sampling in the SCC design is discussed and compared with proportional and balanced sampling techniques. An epidemiological study is provided to illustrate the sample size calculation under the SCC design.

Keywords

Case-cohort design; Power calculation; Sample size; Sampling technique; Stratified case-cohort design

1. Introduction

Time-to-event is a commonly used endpoint for the risk factor assessment in epidemiologic studies or disease prevention trials [2–7]. The case-cohort (CC) design, originally proposed by Prentice [8], has often been used in studying the time to event when the disease is rare and the cost of collecting the risk factor information is high. A CC sample consists of a sub-cohort, which is a random sample of the full cohort, and all the subjects with the event (cases). Statistical analysis methods for analyzing data from the CC study design have been described in many publications [8–20]. For rare diseases, Cai and Zeng [1] proposed a log-rank type of test statistic, which is equivalent to the score test based on a pseudo-partial likelihood function, similar to that was described in Self and Prentice [9]. Furthermore, Cai and Zeng [1] provided an explicit procedure for calculating the sample size and power based on their proposed test.

In studies where the study populations are not homogenous or the original cohort is assembled through a stratified design, a stratified case-cohort (SCC) design may be more appropriate [21–22]. The SCC sample consists of the stratified sub-cohorts selected by a stratified random sampling from the full cohort, and all the cases. For example, the MONICA, Risk, Genetics, Archiving, and Monograph (MORGAM) study [23] is a multinational collaborative cohort study that prospectively followed the development of coronary heart disease (CHD) and stroke events. One goal of this study was to identify risk genotypes for predicting a CHD event. Since the CHD incidence rate differs by gender, and genotyping is expensive, a possible cost-effective design can be a stratified case-cohort design, where the gender is considered as a stratification factor so that the different proportion of sub-cohort samples is selected for each gender group.

Although stratified methods for analyzing data from the SCC design have been studied extensively [18, 24], the sample size and power calculations of the SCC design have not been previously addressed. This paper aims to fill this gap. Specifically, we propose a stratified log-rank statistic and derive expressions for sample size and power calculations. In addition, we compare different sampling strategies including proportional sampling, balanced sampling, and optimal sampling designs. Several simulation studies are presented to evaluate the proposed method using the MORGAM study. We further compare the stratified design/test with the unstratified design/test in the conclusion and discussion section.

2. Stratified case-cohort design and stratified log-rank test

2.1 Notation

Assume that there are n subjects and L strata in a stratified full cohort, and n_l subjects in stratum l ($l = 1, …, L$). Assume two groups indicating the expensive and dichotomous exposure status (for example, the standard versus the wild type single nucleotide polymorphism) and assume n_{lj} subjects in exposure group j ($j = 1, 2$) of stratum l. Assuming that T_{lij} represents the event time and C_{lij} the censoring time for subject i in exposure group j and stratum l ($i = 1, …, n_{lj}$), it is reasonable to assume the T_{lij}s are independent of each other. Let $X_{lij} = T_{lij} \wedge C_{lij}$ be the observed time, where $a \wedge b$ denotes the minimum of a and b, and $\delta_{lij} = I(T_{lij} \leq C_{lij})$ the failure indicator, in which $\delta_{lij} = 1$ denotes observed failure and $\delta_{lij} = 0$ denotes censoring.

In the SCC design, the exposure status is obtained for all the cases and a stratified sub-cohort sample. Specifically, we assume that $\tilde{n}_l$ subjects are randomly sampled into a sub-cohort from n_l subjects in stratum l, and the sub-cohort size is $\tilde{n} = \sum_{l=1}^{L} \tilde{n}_l$. Let $\xi_{lij} = 1$ denote that subject i in group j and stratum l is selected into the sub-cohort and $\xi_{lij} = 0$ otherwise. Let γ_l be the proportion of subjects in group 1 and $(1-\gamma_l)$ the proportion of subjects in group 2 in stratum l. All subjects in the sub-cohort and all events in the L strata make up the stratified case-cohort sample.

2.2 Test statistic

A log-rank type of test is used to compare the hazard rates between the two groups in SCC. The null hypothesis is $H_0: \Lambda_{l1}(t) = \Lambda_{l2}(t)$, $l = 1, \ldots, L$, $t \in [0, \Gamma]$, where Γ is the length of study period and $\Lambda_{lj}(t)$ the cumulative hazard function of the event time t in group j in stratum l. To construct a log-rank type test for the stratified case-cohort sample, we first notice that a weighted stratified log-rank test statistic for the full cohort [9] may be expressed as $W_n^* = \sum_{l=1}^{L} \int_0^\Gamma \frac{\omega(t)\overline{Y}_{l1}(t)\overline{Y}_{l2}(t)}{\overline{Y}_{l1}(t)+\overline{Y}_{l2}(t)} \left\{ \frac{d\overline{N}_{l1}(t)}{\overline{Y}_{l1}(t)} - \frac{d\overline{N}_{l2}(t)}{\overline{Y}_{l2}(t)} \right\}$, where $\overline{Y}_{lj}(t)$ is the number of subjects at risk and $N_{lj}(t)$ is a counting process representing the number of events at time t in group j and stratum l, and $\omega(t)$ is a weight function. The formula above can also be expressed as

$$W_n^* = \sum_{l=1}^{L}\sum_{i=1}^{n_{l1}} \frac{\Delta_{li1}\omega(X_{li1})\overline{Y}_{l2}(X_{li1})}{\overline{Y}_{l1}(X_{li1})+\overline{Y}_{l2}(X_{li1})} - \sum_{l=1}^{L}\sum_{i=1}^{n_{12}} \frac{\Delta_{li2}\omega(X_{li2})\overline{Y}_{l1}(X_{li2})}{\overline{Y}_{l1}(X_{li2})+\overline{Y}_{l2}(X_{li2})}. \quad (1)$$

For the full cohort, the log-rank test statistic is known to be the same as the score function of the Cox partial likelihood function [1, 9].

The test statistic W_n^* requires the covariate information of the full cohort; in a SCC sample the covariate information is only available for the subjects in the sub-cohort and the cases. We propose to use the sub-cohort data to approximate $\overline{Y}_{lj}(t)$ by $\tilde{Y}_{lj}(t)/p_l$, where $\tilde{Y}_{lj}(t)$ is the number of subjects at risk for group j and stratum l in the sub-cohort, and p_l is the sampling fraction of the sub-cohort in stratum l. Hence, we obtain the following stratified case-cohort test statistic:

$$W_n \equiv \sum_{l=1}^{L} W_{nl} \equiv \sum_{l=1}^{L}\sum_{i=1}^{n_{l1}} \frac{\Delta_{li1}\omega(X_{li1})\tilde{Y}_{l2}(X_{li1})}{\tilde{Y}_{l1}(X_{li1})+\tilde{Y}_{l2}(X_{li1})} - \sum_{l=1}^{L}\sum_{i=1}^{n_{12}} \frac{\Delta_{li2}\omega(X_{li2})\tilde{Y}_{l1}(X_{li2})}{\tilde{Y}_{l1}(X_{li2})+\tilde{Y}_{l2}(X_{li2})}, \quad (2)$$

where $\tilde{Y}_{lj}(t) = \sum_{i=1}^{\tilde{n}_{lj}} I(X_{lij} \geq t)$, and $\tilde{n}_{lj}$ is the number of subjects in group j and stratum l in the sub-cohort. Since all the quantities in the summation contribute to W_n only if $\Delta_{li1} = 1$ or $\Delta_{li2} = 1$, W_n can be obtained based on the observed data. It is also easy to verify that this test statistic is the score function of the stratified version of the pseudo partial likelihood function, and, following the results in [9], W_n has an asymptotic normal distribution.

2.3 Asymptotic variance

The asymptotic variance of W_n is the summation of the asymptotic variance of W_{nl} from all the strata. The traditional case-cohort design is considered as a special case of SCC with the number of strata $L = 1$ [9, 18]. Assume the proportion of subjects in group 1 is $\gamma_l = n_{l1}/n_l$, $\gamma_l \in (0, 1)$, and $\tilde{n}_l/n_l$ converges to p_l in stratum l *as n goes to* ∞ (i.e., $p_l = \lim \tilde{n}_l/n_l$). According to Self and Prentice [9], under H_0, $n^{-1/2}W_n$ has an asymptotic normal distribution:

$n^{-1/2}W_n \to_D N(0, \sigma^2 + \psi)$, where $\sigma^2 = \sum_{l=1}^{L} v_l \sigma_l^2$ and $\psi = \sum_{l=1}^{L} v_l \psi_l$ with $v_l = n_l/n$, where σ_l^2 and ψ_l correspond to the asymptotic variance of the log-rank test based on stratum l in the full

cohort, and the variation resulting from sampling from stratum l for the sub-cohort, respectively. Under the null hypothesis H_0: $\Lambda_{l1}(t) = \Lambda_{l2}(t) = \Lambda_l(t)$, $t \in [0, \Gamma]$, let $S_l(t) = S_{lj}(t) = P(T_{lj} \geq t)$, and $\pi_{lj}(t) = P(C_{lj} \geq t)$, then the results in Self and Prentice [9] give

$$\psi_l = \iint \frac{\omega(t)}{\gamma_l \pi_{l1}(t) + (1-\gamma_l)\pi_{l2}(t)} \frac{\omega(w)}{\gamma_l \pi_{l1}(w) + (1-\gamma_l)\pi_{l2}(w)} Q_l(t,w) \frac{dS_l(w)}{S_l(w)} \frac{dS_l(t)}{S_l(t)}, \text{ where}$$
$$Q_l(t,w) = \frac{1-p_l}{p_l} \gamma_l(1-\gamma_l) S_l(t \vee w) \left[(\gamma_l \pi_{l1}(t)\pi_{l1}(w)\pi_{l2}(t \vee w) + (1-\gamma_l)\pi_{l2}(t)\pi_{l2}(w)\pi_{l1}(t \vee w) \right]$$

with event time $w \in [0, \Gamma]$ and $a \vee b$ denoting the maximum of a and b.

The estimator for the asymptotic variance for W_n, $\hat{\sigma}^2_{W_n}$, can be derived based on the

arguments similar to those in Cai and Zeng [1]. Specifically, $\hat{\sigma}^2_{W_n}$ is given by $\hat{\sigma}^2_{W_n} = \hat{\sigma}^2 + \hat{\psi}$, where

$$
\begin{aligned}
\hat{\psi} = &\frac{1}{n} \sum_{l=1}^{L} 2(1-\hat{p}_l) \sum_{j=1}^{2} \sum_{i=1}^{n_{lj}} \left\{ \frac{\Delta_{lij}\omega(X_{lij})\tilde{Y}_{l1}(X_{lij})\tilde{Y}_{l2}(X_{lij})}{(\tilde{Y}_{l1}(X_{lij})+\tilde{Y}_{l2}(X_{lij}))^2} \times \sum_{j'=1}^{2}\sum_{i'=1}^{n_{lj'}} \frac{\Delta_{li'j'}\omega(X_{li'j'})I(X_{li'j'} \leq X_{lij})}{\tilde{Y}_{l1}(X_{li'j'})+\tilde{Y}_{l2}(X_{li'j'})} \right\} \\
&-\frac{1}{n} \sum_{l=1}^{L} (1-\hat{p}_l) \sum_{j=1}^{2} \sum_{i=1}^{n_{ij}} \frac{\Delta_{lij}\omega(X_{lij})^2 \tilde{Y}_{l1}(X_{lij})\tilde{Y}_{l2}(X_{lij})}{(\tilde{Y}_{l1}(X_{lij})+\tilde{Y}_{l2}(X_{lij}))^3},
\end{aligned}
\tag{3}
$$

with $\hat{p_l} = \tilde{n}_l/n_l$ being the estimate of p_l, and $\hat{\sigma^2}$ being the estimate of σ^2 given by

$$\hat{\sigma}^2 = \frac{1}{n} \left\{ \sum_{l=1}^{L}\sum_{i=1}^{n_{l1}} \frac{\Delta_{li1}\omega(X_{li1})\tilde{Y}_{l2}(X_{li1})^2}{(\tilde{Y}_{l1}(X_{li1})+\tilde{Y}_{l2}(X_{li1}))^2} + \sum_{l=1}^{L}\sum_{i=1}^{n_{i2}} \frac{\Delta_{li2}\omega(X_{li2})\tilde{Y}_{l1}(X_{li2})^2}{(\tilde{Y}_{l1}(X_{li2})+\tilde{Y}_{l2}(X_{li2}))^2} \right\}.$$

Since all the quantities expressed above contribute to $\hat{\sigma^2}$ and $\hat{\psi}$ only when $_{li1} = 1$ or $_{li2} =$

1, $\hat{\sigma}^2_{W_n}$ can be obtained from the observed data. The derivations are given in the Web Appendix.

Therefore, to test the equality of the cumulative hazard function of the event time between the two groups in SCC, i.e., to test the null hypothesis H_0: $\Lambda_{l1}(t) = \Lambda_{l2}(t)$, $l = 1, \ldots, L$, $t \in [0, \Gamma]$ vs. the alternative hypothesis H_A: $\Lambda_{l1}(t) \neq \Lambda_{l2}(t)$ (two-sided) at the significance level α,

we reject H_0 if $\left| n^{-1/2}W_n / \sqrt{\hat{\sigma}^2_{W_n}} \right| > Z_{1-\alpha/2}$, where Z_α is the $(100\alpha)^{th}$ percentile of the standard normal distribution.

3. Sample size and power calculation

The sample size and power estimation formula is derived and simplified based on the alternative hypothesis H_A: $\Lambda_{l1}(t) = e^\theta \Lambda_{l2}(t)$, $t \in [0, \Gamma]$ where $\theta = O(1/\sqrt{n})$, where the log-hazards ratios between the two exposure groups are assumed to be constant across the strata. We further assume the following conditions: (i) the censoring distributions are the same in the two groups; (ii) the number of failures is very small (i.e., failure proportion 0 italic> p_D

$\ll 1$) in the full cohort; and (iii) there are no ties of failures. For the sample size and power calculation, we consider the test statistic with $\omega(t) = 1$.

Under the alternative hypothesis H_A, the asymptotic expectation of $n^{-1/2}W_n$ is the same as the asymptotic expectation of the usual log-rank test statistic for the full cohort under H_A and can be approximated by

$$n^{-1/2}\sum_{l=1}^{L}\int_0^{\Gamma}\frac{\overline{Y}_{l1}(t)\overline{Y}_{l2}(t)}{\overline{Y}_{l1}(t)+\overline{Y}_{l2}(t)}[d\Lambda_{l1}(t)-d\Lambda_{l2}(t)] \approx n^{-1/2}\sum_{l=1}^{L}\theta(1-\gamma_l)D_{l1},$$ where D_{lj} is the total number of failures in group $j(j = 1,2)$ in stratum l. Additionally, $\hat{\sigma}^2$ can be approximated by

$1/n\sum_{l=1}^{L}((1-\gamma_l)^2 D_{l1}+\gamma_l^2 D_{l2})$ following the exact approximation and algebra as Cai and Zeng [1] for each stratum. To simplify $\hat{\psi}$, since the failures are much fewer than the stratum

sizes, we approximate $\sum_{j=1}^{2}\overline{Y}_{lj}(t)$ by $(n_l - D_l/2)$, where $D_l = D_{l1} + D_{l2}$. Since the size of the risk set in stratum l of the sub-cohort is about p_l times the size of the risk set in stratum l

of the full cohort, $\hat{\psi}$ can be approximated by $\frac{1}{n}\sum_{l=1}^{L}\frac{(1-p_l)}{(n_l-D_l/2)p_l}\gamma_l(1-\gamma_l)(D_{l1}+D_{l2})^2$. Hence, the non-centrality parameter for $n^{-1/2}W_n/\sqrt{\hat{\sigma}^2_{W_n}}$ under the alternative is approximately

$$\frac{n^{-1/2}\sum_{l=1}^{L}\theta(1-\gamma_l)D_{l1}}{\sqrt{1/n\sum_{l=1}^{L}((1-\gamma_l)^2 D_{l1}+\gamma_l^2 D_{l2})+1/n\sum_{l=1}^{L}\frac{(1-p_l)}{(n_l-D_l/2)p_l}\gamma_l(1-\gamma_l)(D_{l1}+D_{l2})^2}},$$ which can be

simplified as $\dfrac{n^{1/2}\theta\sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l)}{\sqrt{\sum_{l=1}^{L}\left\{\gamma_l(1-\gamma_l)p_{Dl}v_l(1+\frac{(1-p_l)}{(1-p_{Dl}/2)p_l}p_{Dl})\right\}}}$, where p_{Dl} is the failure

proportion in stratum l and v_l is the proportion of stratum l in the full cohort ($v_l = n_l/n$). Consequently, the power function is

$$\Phi\left(Z_{\alpha/2}+n^{1/2}|\theta|\frac{\sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l)}{\sqrt{\sum_{l=1}^{L}\left\{(\gamma_l(1-\gamma_l)p_{Dl}v_l)(1+\frac{1-p_l}{(1-p_{Dl}/2)p_l}p_{Dl})\right\}}}\right),\quad (4)$$

where n is the total number of subjects in the full cohort, θ is the log hazard ratio, α is the significance level, p_{Dl} is the failure proportion in stratum l, v_l is the proportion of stratum l, γ_l is the proportion of subjects in group 1 and $(1-\gamma_l)$ is the proportion of subjects in group 2 in stratum l, and p_l is the sub-cohort sampling fraction in stratum l. For rare diseases, p_{Dl} is very small. By dropping $p_{Dl}/2$, the formula (4) can be further simplified as

$$\Phi\left(Z_{\alpha/2}+n^{1/2}|\theta|\frac{\sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l)}{\sqrt{\sum_{l=1}^{L}\left\{(\gamma_l(1-\gamma_l)p_{Dl}v_l)(1+(1/p_l-1)p_{Dl})\right\}}}\right).$$

When $L = 1$, the above function can be further simplified as

$$\Phi\left(Z_{\alpha/2}+\tilde{n}^{1/2}|\theta|\sqrt{\frac{\gamma(1-\gamma)p_D}{p+(1-p)P_D}}\right),$$ in which p_D is the failure proportion and $\tilde{n} = np$. This is

the same power function of the CC design as reported in Cai and Zeng [1]. When $p_l = 1$, we obtain the power function of the stratified log-rank test for the full cohort, which is given by:

$$\Phi(Z_{\alpha/2} + n^{1/2}|\theta| \sqrt{\sum_{l=1}^{L} (\gamma_l(1-\gamma_l)p_{Dl}v_l)}. \quad (5)$$

4. Proportional, balanced, and optimal designs

This section describes the power issues for two commonly used stratified sampling methods, namely the proportional and balanced designs. Also described is an allocation strategy that maximizes the power.

4.1 Proportional design

The proportional design is commonly used in stratified studies. Under the proportional design, the number of subjects in the sub-cohort at each stratum is proportional to the size of the stratum in the population. For example, consider the full cohort size $n = 2,000$, and there are 4 strata with the strata proportion of 0.1, 0.2, 0.3 and 0.4; i.e., there are 200, 400, 600, and 800 subjects in the 4 strata, respectively. The sub-cohort consists of 200 subjects. With the proportional design, the numbers of samples in each stratum are 20, 40, 60, and 80, respectively. Under such a design, the sub-cohort sampling proportions are the same for all strata, i.e., $p_l = p$ for l.

To detect a log hazard ratio of θ with power β and significance level α, the required total sub-cohort size is at least:

$$\tilde{n} = \left[\frac{n\sum_{l=1}^{L} \left(\gamma_l(1-\gamma_l)p_{Dl}^2 v_l / (1-p_{Dl}/2) \right)}{B_2^2 - \sum_{l=1}^{L} \left(\gamma_l(1-\gamma_l)p_{Dl}v_l \left(1-p_{Dl}/(1-p_{Dl}/2)\right) \right)} \right], \quad (6)$$

where [x] denotes the smallest integer that is bigger than x, and

$B_2 = \dfrac{n^{1/2}\theta\sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l)}{Z_{1-\alpha/2}+Z_{\beta}}$. The sampling proportion $p = \tilde{n}/n$, and the required number of subjects in stratum l is $\tilde{n}_l = \tilde{n}v_l$, $l = 1, 2, \ldots, L$.

4.2 Balanced design

Another popular stratified sampling approach is the balanced design. Under such a design, the number of subjects in a sub-cohort is the same across the strata. For example, consider the full cohort size $n = 2,000$ with 4 strata, and a total of 200 subjects is required for the sub-cohort, each stratum would contain 50 sampled subjects. To detect a log hazard ratio θ with a power β and a significance level α, the required total sub-cohort size $\tilde{n}$ is at least

$$\left[\frac{Ln\sum_{l=1}^{L} \left(\gamma_l(1-\gamma_l)p_{Dl}^2 v_l^2 / (1-p_{Dl}/2) \right)}{B_2^2 - \sum_{l=1}^{L} \left(\gamma_l(1-\gamma_l)p_{Dl}v_l \left(1-p_{Dl}/(1-p_{Dl}/2)\right) \right)} \right] \quad (7)$$

The sub-cohort size in stratum l is $\tilde{n}_l = \tilde{n}/L$ and the sub-cohort sampling proportion $p_l = \tilde{n}_l/n_l = \tilde{n}/(Lnv_l)$.

4.3 Optimal design

In many studies, the number of subjects that can be included in sub-studies is limited because of financial and resource constraints. In these studies, we are given the total number of subjects in the sub-cohort. The distribution of the number of subjects to each of the stratum in the sub-cohort needs to be determined. We consider an optimal design strategy which provides the highest power under such situation. Specifically, we propose an optimal design with a set of p_l which provides the highest power for a given $\tilde{n}$. This optimization problem is solved by using the Lagrange multipliers method following the steps below.

Maximizing the power function for a given $\tilde{n}$ is equivalent to minimizing the denominator

$\sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l)(1+\frac{1-p_l}{(1-p_{Dl}/2)\,p_l}p_{Dl})$ in the formula (4), a function of p_l, subject to

$\sum_{l=1}^{L}p_lv_l=\tilde{n}/n$, a constraint function of p_l. We obtain the Lagrange function

$$\Xi(p_l,\lambda)=\sum_{l=1}^{L}\left((\gamma_l(1-\gamma_l)p_{Dl}v_l)(1+\frac{1-p_l}{(1-p_{Dl}/2)\,p_l}p_{Dl})\right)+\lambda*\left(\sum_{l=1}^{L}(p_lv_l)-\frac{\tilde{n}}{n}\right).$$

Furthermore, we have

$$\frac{\partial\Xi(p_l,\lambda)}{\partial p_l}=\frac{\partial\sum_{l=1}^{L}\left((\gamma_l(1-\gamma_l)p_{Dl}v_l)(1+\frac{1-p_l}{(1-p_{Dl}/2)p_l}p_{Dl})\right)}{\partial p_l}+\lambda*\frac{\partial\left(\sum_{l=1}^{L}(p_lv_l)-\frac{\tilde{n}}{n}\right)}{\partial p_l}$$

$$=-\frac{\gamma_l(1-\gamma_l)p_{Dl}v_lp_{Dl}/(1-p_{Dl}/2)}{p_l^2}+\lambda*v_l=0, \text{ and } \frac{\partial\Xi(p_l,\lambda)}{\partial\lambda}=\frac{\partial\lambda\left(\sum_{l=1}^{L}p_lv_l-\frac{\tilde{n}}{n}\right)}{\partial\lambda}=\sum_{l=1}^{L}p_lv_l-\frac{\tilde{n}}{n}=0.$$

After solving these two sets of equations, we obtain the optimal sub-cohort sampling proportion

$$p_l=\frac{\tilde{n}\sqrt{\gamma_l(1-\gamma_l)/(1-p_{Dl}/2)}p_{Dl}}{n\sum_{l=1}^{L}(\sqrt{\gamma_l(1-\gamma_l)/(1-p_{Dl}/2)}p_{Dl}v_l)}. \quad (8)$$

Hence, the optimal power for a given $\tilde{n}$ is calculated as

$$\Phi\left(Z_{\alpha/2}+\frac{\tilde{n}^{1/2}|\theta|\sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l)}{\sqrt{\tilde{n}/n\sum_{l=1}^{L}((\gamma_l(1-\gamma_l)p_{Dl}v_l\,(1-p_{Dl}/\,(1-p_{Dl}/2)))+\left(\sum_{l=1}^{L}\left(\sqrt{\gamma_l(1-\gamma_l)/(1-p_{Dl}/2)}p_{Dl}v_l\right)\right)^2}}\right).$$

To achieve a power β with a significance level α based on the optimal design, the required total sub-cohort size is given by

$$\tilde{n} = \left[\frac{n\left(\sum_{l=1}^{L} \sqrt{\gamma_l(1-\gamma_l)/(1-p_{Dl}/2)}p_{Dl}v_l\right)^2}{B_2^2 - \sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l(1-p_{Dl}/(1-p_{Dl}/2)))} \right],$$

where $B_2 = \dfrac{n^{1/2}\theta\sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l)}{Z_{1-\alpha/2}+Z_\beta}$. Therefore,

$$p_l = \frac{\left(\sqrt{\gamma_l(1-\gamma_l)/(1-p_{Dl}/2)}p_{Dl}\right)\left(\sum_{l=1}^{L}\sqrt{\gamma_l(1-\gamma_l)/(1-p_{Dl}/2)}p_{Dl}v_l\right)}{B_2^2 - \sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l(1-p_{Dl}/(1-p_{Dl}/2)))}. \quad (9)$$

From the formula (8), we observe that under the situation when γ_l is similar across the strata and p_{Dl} is very small (disease is rare), the optimal $\tilde{n}_l = \dfrac{p_{Dl}}{p_D}\tilde{n}$ or $\tilde{n}_l = \dfrac{D_l}{D}\tilde{n}$. Furthermore, under the homogeneous situation where p_{Dl} is similar across the strata, the optimal p_l is close to $\tilde{n}/n$ the estimate p from the proportional design. It means that the proportional design is nearly optimal when the event rate is homogeneous across the strata.

We obtain the number of subjects in stratum l using $\tilde{n}_l = p_l n v_l$, and the SCC total sample size $n_{scc} = n\sum_{l=1}^{L}(p_l v_l + (1-p_l)p_{Dl}v_l)$, in which p_l is obtained from the formulae in the sections for the proportional, balanced, or optimal design above, depending on the desired design.

4.4 Practical note: minimal detectable log-hazard ratio

The denominator of the total sub-cohort size $\tilde{n}$ formula in the previous section needs to be positive. This condition is written as

$$|\theta| > \theta_0 \equiv (Z_{1-\alpha/2}+Z_\beta)\frac{\sqrt{\sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l) - \sum_{l=1}^{L}\left(\gamma_l(1-\gamma_l)p_{Dl}^2 v_l/(1-p_{Dl}/2)\right)}}{n^{1/2}\sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l)}.$$

Since the failure rate p_{Dl} is usually fairly small for the case-cohort studies, $p_{Dl} - p_{Dl}^2/\left(1-\dfrac{p_{Dl}}{2}\right) \approx p_{Dl}$. Hence, $\theta_0 \approx \dfrac{(Z_{1-\alpha/2}+Z_\beta)}{n^{1/2}\sqrt{\sum_{l=1}^{L}(\gamma_l(1-\gamma_l)p_{Dl}v_l)}}$, which is the log-hazard ratio that can be detected with the entire cohort. This condition implies that the stratified case-cohort design will not be able to detect a hazard ratio smaller than the one that can be detected by using the entire cohort, which is a reasonable restriction.

5. Numeric results

5.1 Theoretical power

Table 1 shows the theoretical power of the SCC design, as well as the power of the full cohort and the sub-cohort. The power function (4) is used to calculate P_{SCC}, the power of the SCC design, while formula (5) is used to calculate P_{Full}, the power of the full cohort.

The sub-cohort power P_{Sub} is obtained by substituting n with $\tilde{n}$ in the full cohort power function, where $\tilde{n}$ is the sub-cohort size $\tilde{n} = n \sum_{l=1}^{L} v_l p_l$. The power P_{Full}, P_{SCC}, and P_{Sub} are calculated for the different combinations of the full cohort size n, the event proportion p_{Dl}, the group 1 proportion γ_l, the log-hazard ratio θ, and the sub-cohort sampling fraction p_l in stratum l. The significant level is set at $\alpha = 0.05$ and the number of strata is $L = 4$. The event proportion p_D in the table is a mean value over all strata. For instance, at the mean value of 10%, p_{Dl} are set to 9%, 8%, 11%, and 10% for each of the 4 strata, respectively. Similarly, at the mean value of 5% (1%), p_{Dl} are set to 4%, 5%, 4.5%, and 6% (0.8%, 1%, 1.2%, and 0.9%) for each of the 4 strata, respectively. In the example where the full cohort size $n =$ 2,000, the event proportion $p_D = 10\%$, the group 1 proportion $\gamma_l = 0.3$, and the log-hazard ratio $\theta = 0.5$, the SCC sample with the 10% sub-cohort sampling proportion yields a power of 0.634, while the power for the full cohort and for the stratified random sample are 0.894 and 0.172, respectively. In another example where the full cohort size $n = 10,000$, the event proportion $p_D = 1\%$, the group 1 proportion $\gamma_l = 0.3$, and the log-hazard ratio $\theta = 1.0$, the SCC sample with the 1% sub-cohort sampling yields a power of 0.898 while the powers for the full cohort and for the stratified random sample are 0.996 and 0.067, respectively. The results in Table 1 suggest that the SCC design is an efficient and attractive solution in situations with low event proportions and small sub-cohort sampling fractions.

5.2 Type I error and power for the stratified log-rank test

Simulation studies are conducted to evaluate the empirical type I error and the empirical power for the stratified log-rank test using the SCC, the full cohort, and the sub-cohort data. The simulation procedures and their results are presented in the Web Appendix (Tables A and B).

Appendix Table A shows the empirical type I error for the stratified log-rank test using the SCC (*SCC*), the full cohort (*Full*), and the sub-cohort (*Sub*) samples. The significance level α is set at 0.05 and the number of strata $L = 4$. Various values are considered for the full cohort size n, the stratum proportion v_l, the event proportion p_{Dl}, the group 1 proportion γ_l, and the sub-cohort sampling fraction p_l in stratum l. Overall, the empirical type I error rates in the SCC samples are fairly close to the nominal 0.05 level.

Appendix Table B presents the empirical power for the log-rank tests in the SCC the full cohort and the sub-cohort samples. In addition, the theoretical power is compared with the empirical power. It is observed that the test based on the SCC design is more powerful than using the sub-cohort, and the power based on the full cohort provides the upper bound. Note that in real studies, it is usually impossible to collect all the full cohort information required to conduct the log-rank test. As illustrated in Appendix Table B, using only a small fraction of the subjects, the power of the SCC design is over 50% of the power with the full cohort. As expected, when the sampling rate increases, the power of the SCC increases. Overall, the empirical power is very close to the theoretical powers. In the additional simulations, we consider the different group 1 proportions across strata and the results are similar.

5.3 Proportional, balanced, and optimal designs comparison

Power comparison under homogeneous and heterogeneous event rates—We compare the proportional, balanced, and optimal sampling methods in order to investigate which one is more efficient in the SCC design. Two situations where the event rates are relatively homogeneous or heterogeneous over the strata are considered for comparison. In the situation where the event rates are homogeneous, the event proportion p_{Dl} at each stratum is relatively similar to each other. In the situation where the event rates are heterogeneous, the event proportions p_{Dl} over the strata have a wide range. The corresponding analysis results in both homogeneous and heterogeneous situations are presented in Table 2.

Results for the SCC with homogeneous event rates are presented for a theoretical power based on proportional, balanced, and optimal sampling for SCC with various combinations of the full cohort size n, the event proportion p_{Dl}, the group 1 proportion γ_l, the log hazard ratio θ, and the sub-cohort size $\tilde{n}$. The number of strata is $L = 4$ with the stratum proportions (v_l) of 0.1, 0.2, 0.3, and 0.4, respectively. The event proportion p_D in the table is a mean value over all strata. Specifically, at the level of 10%, p_{Dl} s are set to 9%, 8%, 11%, and 10% for each stratum. Similarly, at $p_D = 5\%$, 4 strata have 4%, 5%, 4.5%, and 6%, respectively. The sub-cohort sampling fractions p_l in stratum l for the proportional, balanced, and optimal designs are calculated by $\tilde{n}/n$, $\tilde{n}/Lnv_l$, and the formula (8), respectively. The total SCC sizes n_{scc} *(prop)*, n_{scc} *(bal)*, and n_{scc} *(opt)* are then calculated using the formula $n\sum_{l=1}^{L}(p_l v_l + (1-p_l)p_{Dl}v_l)$. The theoretical powers P_{prop}, P_{Bal}, and P_{opt} are calculated using the power formula (4). The power ratio (P_{Bal} vs. P_{prop}) is presented in percent (%).

Table 2 indicates that the total SCC sample sizes from the three methods are generally similar under homogeneous circumstances. For instance, where the full cohort size $n = 2,000$, the event proportion $p_D = 10\%$, the group 1 proportion $\gamma_l = 0.3$, the log hazard ratio $\theta = 0.5$, and the stratified sub-cohort size $= 200$, the total SCC sample sizes are 376, 377, and 376 for proportional, balanced, and optimal samplings, respectively. The results show that the power from proportional method P_{prop} is at least equal to or larger than P_{Bal} in all the situations and the power ratio (P_{Bal} vs. P_{prop}) has a range from 83% to 100%. These results suggest that, when the event rates are homogeneous over the strata, the proportional sampling is more efficient than the balanced sampling. Furthermore, we observe that the powers from the proportional method and the optimal design remain close, which indicates that, when the event rates are homogeneous and the exposure group 1 proportion γ_l is the same over strata, the proportional method is close to the optimal sampling strategy.

Table 2 also provides results for situations with heterogeneous event rates over strata. The set-up is similar to the homogeneous situation, except that the event rates are set to a wide range over strata. Two sets of combination of p_{Dl} ($l = 1, 2, 3, 4$) are examined. Set1 gives the values of p_{Dl} to 9%, 30%, 5%, and 20% for the 4 strata and Set2 gives the values of p_{Dl} to 4%, 25%, 10%, and 6% for the 4 strata, respectively. Results in Table 2 indicate that for the given set up and given $\tilde{n}$, in a heterogeneous situation, the total SCC sample sizes from the proportional and balanced methods are similar. The power for these two methods is also

similar with slightly more power for the proportional method in most cases in Set1 and Set2. As expected, among all three methods, the optimal design yields the highest theoretical power (P_{opt}) with the smallest total SCC sample size. For instance, where the full cohort size $n = 2,000$, the event proportion p_{Dl} is as in Set1, the group 1 proportion $\gamma_l = 0.3$, the log hazard ratio $\theta = 0.5$, and the stratified sub-cohort size = 200, the powers (n_{scc}) are 0.637 (495) for the proportional, 0.590 (496) for the balanced, and 0.731 (485) for the optimal design. Thus, under the heterogeneous event rate situation, the optimal design indeed provides more powerful test over the other two designs.

Additional simulation studies are conducted to examine whether the sample size formulae for each design produce sufficient power, specifically, for a full cohort size $n = 2,000$ with 4 strata and overall disease rate p_D of 5% (4%, 5%, 4.5%, or 6% over the strata) or 10% (9%, 8%, 11%, or 10% over the strata). The group 1 proportion γ_l is set to 0.3 for all strata and the log hazard ratio is set to 0.55 or 0.693.

To target a power of 80% at the significance level of 0.05, we first calculate the sub-cohort size at each stratum $\tilde{n}_l$, the stratum sampling proportion p_l, the total sub-cohort size $\tilde{n}$, and the total number of subjects in SCC n_{scc} by using the formulae given in Section 4 for each of the proportional, balanced and optimal sampling designs. We then carry out simulations using the derived sample sizes to examine whether the empirical powers achieve the target 80%. The simulation procedure is similar to that for Appendix Table B. The results are summarized in Table 3. From Table 3, we observe that the sample sizes calculated from the formulae do provide close to sufficient power empirically in each design. The results in Table 3 also indicate that to achieve the same power, the optimal design gives the smallest sample size among the three designs, the proportional is the second smallest, and the balanced has the largest for all 3 samples. The average sub-cohort size saving of the optimal vs. the balanced approach is approximately 20%.

6. The MORGAM Study

This section presents the MORGAM study [23] as an example to illustrate the efficiency of a SCC design. The MORGAM study is a multinational collaborative cohort study prospectively followed the development of CHD and stroke events. A total of 4,559 subjects including 2,282 males and 2,277 females were assessed at the baseline visit in 1997; by 2003, ninety-six CHD events were observed in males (CHD incidence $p_{Dl} = 0.042$) and 24 in females (CHD incidence $p_{Dl} = 0.011$). The CHD incidence rates differ by gender, and the testing for genotyping is expensive, so a cost-effective SCC design may be needed. The SCC design examines the relationship between the genetic risk factor and the CHD incidence where gender is considered as a stratification factor. The study is designed with 80% power and a 0.05 significance level, and assumes the genetic risk factor frequency is about 0.4 for both the male and the female strata. The full cohort and strata information for this design are displayed in Table 4.

Assume that a hazard ratio of 2 is to be detected. Note that the minimal detectable hazard ratio based on the entire MORGAM study is 1.9. Table 4 presents the sample size calculation using the proportional, balanced, and optimal sampling methods. Under the

optimal (proportional) design, a total of 154 (210) subjects is required for the sub-cohort, 123 (105) of which are from the male stratum and 31 (105) from the female stratum; the total SCC sample size is 269 (325). The balanced design requires a sample size similar to that of the proportional method because of the similar strata proportion v_l for the male and the female (i.e., 2,282 subjects are in the male stratum and 2,277 in the female stratum). However, both the proportional and balanced methods require approximate 20% more sub-cohort subjects than the optimal design.

Interestingly, under the optimal design, the sub-cohort size at stratum l is proportional to the ratio of the number of the events at stratum l vs. all events, that is, $\tilde{n}_l = \dfrac{D_l}{D} \tilde{n}$. For instance, D ninety-six events were observed in the male stratum, which is 80% of the total number of events observed in the full cohort (120). The required sub-cohort size at the male stratum is 123, 80% of the overall sub-cohort size (154).

The non-event vs. event ratio has been examined for all three sampling methods. All methods yield a ratio greater than 1 to ensure the good precision of testing. The optimal method has the smallest overall non-event vs. event ratio of 1.2 among all methods, supporting the conclusion that the optimal method is the most efficient among others.

7. Conclusion and discussion

We have proposed a stratified log-rank type test statistic for the SCC design and provided the power calculation formula. We have investigated the proportional, balanced, and optimal sampling methods, and derived the corresponding sample size calculation formulae. The simulation studies show that the proposed stratified log-rank type test statistic is valid for the finite SCC samples. The simulations also indicate that the power of the SCC design can be fairly high compared with the full cohort when the event rate is low. The empirical power is similar to the theoretical power.

Additional simulation studies have also been conducted to compare the proportional, balanced, and optimal samplings methods. The results show that when the event rates are relatively homogeneous across the strata, the proportional method is superior to the balanced method and is close to the optimal method. However, when the event rates are heterogeneous over the strata, the power for the proportional method is slightly higher than that for the balanced in most of the finite samples. Overall, the optimal method yields the highest power along with the smallest required sample size among all three methods.

Stratified sampling is commonly used in the survey sampling to improve the estimation precision for the population quantity of interest. In some situations, the stratified sampling may be unnecessary but it often leads to the more efficient estimators as compared with the unstratified design, e.g., a more precise estimation of the exposure risk effects, especially when subjects from the same stratum are homogeneous (due to the strong association between the exposure group and the stratum). Furthermore, the stratified design ensures the representation of the small subgroups in the population. When the sampling is stratified, it is natural to consider a stratified test, although an unstratified test statistic can be used when

the association between the stratum and the outcome is proportional. Our proposed stratified and nonparametric test statistic naturally accounts for the non-proportionality if it exists.

Our paper only considers the combination of stratified sampling and stratified test when strata in both the design stage and the test stage are the same. In practice, the stratified sampling and the stratified test may be used very differently: when there is a strong association between stratified variable and exposure, the stratified sampling may be used to improve the design efficiency; however, if one believes a strong non-proportional association between failure time and exposure variable across strata, the stratified test needs to be adopted to ensure the test validity. In the Web Appendix III, we use the power formula (4) to compare the stratified design with the unstratified design analytically. The results show that in general, the stratified design tends to have a higher power than the unstratified design with stratified or unstratified test. Therefore, when both associations are present, it is necessary to take the current approach with both stratified sampling and stratified test. In the situation when the disease proportions or the strata distribution are not available, we suggest to conduct a pilot study to obtain this information before planning a stratified case-cohort study.

The situation becomes more complex when the stratified variable in the design stage is not the same as the stratified variable in the test stage. Generalizing our sample size/power calculation to address this complex situation will be an interesting future study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Cai J, Zeng D. Sample Size/Power Calculation for Case-Cohort Studies. Biometrics. 2004; 60:1015–1024. [PubMed: 15606422]

2. Kalbfleisch J, Lawless J. Likelihood analysis of multi-state models for disease incidence and mortality. Statistics in Medicine. 1988; 7:149–60. [PubMed: 3353602]

3. ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: Design and Objectives. American Journal of Epidemiology. 1989; 129:687–702. [PubMed: 2646917]

4. Schouten E, Dekker J, Kok F, Le Cessie S, Van Houwelingen H, Pool J, Vandenbroucke J. Risk ratio and rate ratio estimation in case-cohort designs: Hypertension and cardiovascular mortality. Statistics in Medicine. 1993; 12:1733–1745. [PubMed: 8248665]

5. Liao D, Cai J, Rosamond WD, Barnes RW, Hutchinson RG, Whitsel EA, Rautaharju P, Heiss G. Cardiac Autonomic Function and Incident Coronary Heart Disease: A Population Based Case-Cohort Study - The ARIC Study. American Journal of Epidemiology. 1997; 145:696–706. [PubMed: 9125996]

6. Savitz DA, Cai J, van Wijngaarden E, Loomis D, Mihlan G, Dufort V, Kleckner RC, Nylander-French LA, Kromhout H, Zhou H. Case-Cohort Analysis of Brain Cancer and Leukemia in Electric Utility Workers Using a Refined Magnetic Field Job-Exposure Matrix. American Journal of Industrial Medicine. 2000; 38:417–425. [PubMed: 10982982]

7. Ballantyne CM, Hoogeveen RC, Bang HJ, et al. Lipoprotein associated phospholipase A(2), high-sensitivity C-reactive protein, and risk for ischemic stroke in middle-aged men and women in the Atherosclerosis Risk in Communities (ARIC) study. Circulation. 2004; 109(7):837–842. [PubMed: 14757686]

8. Prentice R. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika. 1986; 73:1–11.

9. Self SG, Prentice R. Asymptotic distribution theory and efficiency results for case-cohort studies. The Annals of Statistics. 1988; 16:64–81.

10. Barlow WE, Prentice RL. Residuals for relative risk regression. Biometrika. 1988; 75:65–74.

11. Lin DY, Ying Z. Cox regression with incomplete covariate measurements. Journal of American Statistical Association. 1993; 88:1341–1349.

12. Barlow WE. Robust variance estimation for the case-cohort design. Biometrics. 1994; 50:1064–1072. [PubMed: 7786988]

13. Barlow WE, Ichikawa L, Rosner D, et al. Analysis of case cohort designs. Journal of Clinical Epidemiology. 1999; 52(12):1165–1172. [PubMed: 10580779]

14. Chen K, Lo SH. Case-cohort and Case-control analysis with Cox's model. Biometrika. 1999; 86:755–764.

15. Chen KN. Generalized case-cohort sampling. Journal of the Royal Statistical Society, Series B. 2001; 63:791–809.

16. Chen H. Weighted semiparametric likelihood method for fitting a proportional odds regression model to data from the case-cohort design. Journal of American Statistical Association. 2001; 96:1446–1457.

17. Chen H. Fitting semiparametric transformation regression models to data from a modified case-cohort design. Biometrika. 2001; 88:255–268.

18. Borgan Ø, Langholz B, Samuelsen S, Goldstein L, Pogoda J. Exposure stratified case-cohort designs. Lifetime Data Analysis. 2000; 6:39–58. [PubMed: 10763560]

19. Kulich M, Lin DY. Improving the efficiency of relative-risk estimation in Case-cohort studies. Journal of American Statistical Association. 2004; 99:832–844.

20. Kang S, Cai J. Marginal hazards regression for retrospective studies within cohort with possibly correlated failure time data. Biometrics. 2009; 65:405–14. [PubMed: 18565164]

21. Boice J, Monson R. Breast cancer in women after repeated fluoroscopic examinations of the chest. Journal of National Cancer Institute. 1977; 59(3):823–32.

22. Hrubec Z, Boice J, Monson R, Rosenstein M. Breast cancer after multiple chest fluoroscopes: Second follow-up of Massachusetts women with tuberculosis. Cancer Research. 1989; 49:229–34. [PubMed: 2908849]

23. Kulathinal S, Karvanen J, Saarela O, Kuulasmaa K. Case-cohort design in practice –experiences from the MORGAM. Project Epidemiologic Perspectives & Innovations. 2007; 4:15.10.1186/1742-5573-4-15

24. Breslow N, Lumley T, Ballantyne C, Chambless L, Kulich M. Using the Whole Cohort in the Analysis of Case-Cohort Data. American Journal of Epidemiology. 2009; 169(11):1398–1405. [PubMed: 19357328]

Table 1

Theoretical Power of Stratified Case-Cohort Design

n	p_D	γ_l	θ	p_l	P_{Full}	P_{SCC}	P_{Sub}
2,000	10%	0.3	0.5	10%	0.894	0.634	0.172
				20%	0.894	0.769	0.300
			1.0	10%	1.000	0.996	0.527
				20%	1.000	1.000	0.818
		0.5	0.5	10%	0.938	0.710	0.197
				20%	0.938	0.836	0.347
			1.0	10%	1.000	0.999	0.600
				20%	1.000	1.000	0.879
	5%	0.3	0.5	10%	0.643	0.479	0.110
				20%	0.643	0.559	0.179
			1.0	10%	0.996	0.968	0.312
				20%	0.996	0.988	0.548
		0.5	0.5	10%	0.718	0.548	0.124
				20%	0.718	0.633	0.205
			1.0	10%	0.999	0.986	0.361
				20%	0.999	0.996	0.621
4,000	1%	0.3	0.5	1%	0.305	0.174	0.035
				2%	0.305	0.218	0.040
			1.0	1%	0.826	0.533	0.047
				2%	0.826	0.657	0.061
		0.5	0.5	1%	0.352	0.199	0.036
				2%	0.352	0.251	0.041
			1.0	1%	0.885	0.606	0.050
				2%	0.885	0.732	0.065
10,000	1%	0.3	0.5	1%	0.630	0.365	0.042
				2%	0.630	0.464	0.051
			1.0	1%	0.996	0.898	0.067
				2%	0.996	0.962	0.095

n	p_D	γ_1	θ	p_1	P_{Full}	P_{SCC}	P_{Sub}
	0.5	0.5	0.5	1%	0.705	0.421	0.044
				2%	0.705	0.532	0.054
			1.0	1%	0.999	0.941	0.072
				2%	0.999	0.983	0.105

n =full cohort size, pD =mean event proportion, γ1 =group 1 proportion, θ =log-hazard ratio, p1 =sub-cohort sampling fraction in stratum l. PSCC =theoretical power of SCC, PFull =theoretical power of full cohort, and PSub =theoretical power of sub-cohort. Significant level α = 0.05.

Table 2

Theoretical Power of Proportional, Balanced, and Optimal Sampling in SCC

p_D	γ_1	θ	$\tilde{n}$	p_1	Proportional		Balanced			Optimal		
					n_{scc}	P_{prop}	n_{scc}	P_{Bal}	$\frac{P_{Bal}}{P_{prop}}$	n_{scc}	P_{Opt}	$\frac{P_{prop}}{P_{Opt}}$
with Homogeneous Event Rate												
10%	0.3	0.5	200	10%	376	0.634	377	0.581	92%	376	0.637	100%
			400	20%	557	0.769	558	0.732	96%	556	0.770	100%
		1	200	10%	376	0.996	377	0.991	100%	376	0.996	100%
			400	20%	557	1.000	558	0.999	100%	556	1.000	100%
	0.5	0.5	200	10%	376	0.710	377	0.656	93%	376	0.713	100%
			400	20%	557	0.836	558	0.804	96%	556	0.838	100%
		1	200	10%	376	0.999	377	0.997	100%	376	0.999	100%
			400	20%	557	1.000	558	1.000	100%	556	1.000	100%
5%	0.3	0.5	200	10%	293	0.479	293	0.442	92%	292	0.482	99%
			400	20%	482	0.559	484	0.533	95%	482	0.561	100%
		1	200	10%	293	0.968	293	0.952	98%	292	0.969	100%
			400	20%	482	0.988	484	0.983	100%	482	0.988	100%
	0.5	0.5	200	10%	293	0.548	293	0.507	93%	292	0.551	99%
			400	20%	482	0.633	484	0.606	96%	482	0.635	100%
		1	200	10%	293	0.986	293	0.977	99%	292	0.987	100%
			400	20%	482	0.996	484	0.994	100%	482	0.996	100%
with Heterogeneous Event Rate												
10% (Set1)	0.3	0.5	200	10%	380	0.610	381	0.532	87%	377	0.643	95%
			400	20%	560	0.757	562	0.700	92%	554	0.779	97%
		1	200	10%	380	0.994	381	0.983	99%	377	0.996	100%
			400	20%	560	1.000	562	0.999	100%	554	1.000	100%
	0.5	0.5	200	10%	380	0.686	381	0.605	88%	377	0.719	95%
			400	20%	560	0.826	562	0.774	94%	554	0.846	98%
		1	200	10%	380	0.998	381	0.994	100%	377	0.999	100%
			400	20%	560	1.000	562	1.000	100%	554	1.000	100%

p_D	γ_l	θ	$\tilde{n}$	Proportional			Balanced			Optimal		
				p_l	n_{scc}	P_{prop}	n_{scc}	P_{Bal}	$\dfrac{P_{Bal}}{P_{prop}}$	n_{scc}	P_{Opt}	$\dfrac{P_{prop}}{P_{Opt}}$
5% (Set2)	0.3	0.5	200	10%	290	0.425	290	0.398	94%	286	0.479	89%
			400	20%	480	0.520	481	0.499	96%	471	0.558	93%
		1	200	10%	290	0.943	290	0.925	98%	286	0.968	97%
			400	20%	480	0.980	481	0.975	99%	471	0.988	99%
	0.5	0.5	200	10%	290	0.489	290	0.459	94%	286	0.548	89%
			400	20%	480	0.592	481	0.570	96%	471	0.632	94%
		1	200	10%	290	0.971	290	0.960	99%	286	0.986	98%
			400	20%	480	0.992	481	0.990	100%	471	0.996	100%

pD =mean event proportion, γ_l =group 1 proportion, θ =log-hazard ratio, $\tilde{n}$ =sub-cohort size, nSCC =SCC sample size, p_l =sub-cohort sampling fraction in stratum l. Pprop=proportional power, PBal =balanced power, and POpt =optimal power. Set1 (Set2): event proportion=9%, 30%, 5%, and 20% (4%, 25%, 10%, and 6%) for strata 1–4.

Table 3

Sample Size Comparison in Optimal, Proportional, and Balanced Designs

Design	v_l	n_l	p_{Dl}	p_l	$ñ_l$	$ñ$	n_{SCC}	NE:E	T
Optimal (θ=0.55)	0.1	200	9%	13.7%	28	300	466	270:196	80%
	0.2	400	8%	12.1%	49				
	0.3	600	11%	16.8%	101				
	0.4	800	10%	15.2%	122				
Proportional (θ=0.55)	0.1	200	9%	15.1%	31	305	469	273:196	80%
	0.2	400	8%	15.1%	61				
	0.3	600	11%	15.1%	91				
	0.4	800	10%	15.1%	122				
Balanced (θ=0.55)	0.1	200	9%	47.1%	95	380	538	342:196	79%
	0.2	400	8%	23.6%	95				
	0.3	600	11%	15.7%	95				
	0.4	800	10%	11.8%	95				
Optimal (θ=0.693)	0.1	200	4%	10.8%	22	282	369	272:103	81%
	0.2	400	5%	13.6%	55				
	0.3	600	4.5%	12.2%	74				
	0.4	800	6%	16.4%	131				
Proportional (θ=0.693)	0.1	200	4%	14.3%	29	288	375	266:103	80%
	0.2	400	5%	14.3%	58				
	0.3	600	4.5%	14.3%	86				
	0.4	800	6%	14.3%	115				
Balanced (θ=0.693)	0.1	200	4%	46.3%	93	372	455	352:103	79%
	0.2	400	5%	23.1%	93				
	0.3	600	4.5%	15.4%	93				
	0.4	800	6%	11.6%	93				

All samples have a full cohort size =2000, group 1 proportion =0.3, significant level α =0.05, and power =80%. n_l =size of stratum l in full cohort, v_l =proportion of stratum l in full cohort, p_{Dl} =event proportion in stratum l, θ =log-hazard ratio, p_l =sub-cohort sampling fraction in stratum l, $ñ$ ($ñ_l$)=sub-cohort size in stratum l, $ñ$ ($ñ_l$)=sub-cohort size (at stratum l), n_{SCC} =SCC sample size, NE:E= Non-event : Event in SCC. T =empirical power. The sample size is rounded up to the nearest integer as appropriate.

Table 4

MORGAM Study Sample Size Calculation

	Full Cohort and Strata Information					
Stratum No.	**Strata Description**	**n_l**	**Event**	**pD_l**	**v_l**	**γ_l**
1	Male	2,282	96	0.042	0.5	0.4
2	Female	2,277	24	0.011	0.5	0.4
Overall		4,559	120	0.026	1.0	0.4

	Sample Size Calculation ($\theta = 0.693$)					
	Stratum No.	**n_l**	**Sub-cohort**	**p_l**	**Non-event**	**NE:E ratio**	**n_{scc}**
Proportional	1	2,282	105	0.046	101	1.0	197
	2	2,277	105	0.046	104	4.4	128
	Overall	4,559	210	0.046	205	1.7	325
Balanced	1	2,282	105	0.046	101	1.0	197
	2	2,277	105	0.046	104	4.4	128
	Overall	4,559	210	0.046	205	1.7	325
Optimal	1	2,282	123	0.054	118	1.2	214
	2	2,277	31	0.014	31	1.3	55
	Overall	4,559	154	0.034	149	1.2	269

n_l =size of stratum l in full cohort, v_l =proportion of stratum l, pD_l =event proportion in stratum l, p_l =sub-cohort sampling fraction in stratum l, θ =log-hazard ratio, p_l =group 1 proportion in stratum l, γ_l =group 1 proportion in stratum l, n_{SCC} =SCC sample size, Sub-cohort=sub-cohort size in stratum l, Non-event=number of subjects with non-event in stratum l in SCC, Event=number of subjects with event in stratum l in SCC, NE:E=Ratio of the number of non-events to the number of events. Significant level $\alpha = 0.05$. Power =80%. The sample size is rounded up to the nearest integer as appropriate.