A Simulation Study of Control Sampling Methods for Nested Case-Control Studies of Genetic and Molecular Biomarkers and Prostate Cancer Progression

Ming-Hsi Wang,[1,2] Yin Yao Shugart,[1] Stephen R. Cole,[1,3] and Elizabeth A. Platz[1,4,5]

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; [2]Department of Medicine, Saint Agnes Hospital, Baltimore, Maryland; [3]Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill; [4]The James Buchanan Brady Urological Institute and [5]The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins Medical Institutions, Baltimore, Maryland

Abstract

Background: **Incidence density sampling is typically the least biased efficient method for control sampling in nested case-control studies. However, in studies of genetic variants and prostate cancer progression, some argue that controls should be sampled from men who did not progress by end of follow-up. Thus, we examined the validity of relative risk (RR) estimates of prostate cancer progression using three methods for control sampling from cohorts of men with prostate cancer generated by Monte Carlo simulation.**

Methods: **Data were simulated for nine scenarios for combinations of genotype frequency (10%, 30%, and 50%) and association (RR, 1.0, 1.5, and 2.0) using prostate progression rates from Johns Hopkins Hospital. RRs estimated from conditional logistic regression for the genetic association from case-control studies nested in the nine cohort scenarios using three control sampling methods, (*a*) incidence density sampling, (*b*) incidence density sampling without replacement of selected controls, and (*c*) ''pure'' control sampling (i.e., men who did not progress by end of long-term follow-up), were compared with the true RRs.**

Results: **Use of controls selected by incidence density sampling produced unbiased RR estimates of progression. In our setting, only a slight bias was produced by use of incidence density sampling without replacement. In contrast, use of controls selected by pure control sampling produced biased RR estimates, except when there was no association; extent of bias increased with increasing size of the association and duration of follow-up.**

Conclusions: **Nested case-control studies designed to estimate the association of genetic variants with risk of prostate cancer progression should use incidence density sampling to provide a valid RR estimate.**
(Cancer Epidemiol Biomarkers Prev 2009;18(3):706–11)

Introduction

Prostate cancer has been a major public health problem in the United States and Western Europe for decades. In U.S. men, prostate cancer is the most commonly diagnosed cancer and the second most common cause of cancer death (1). In recent years, men diagnosed with early-stage prostate cancer have a greater than a 99% 5-year relative survival.[6] This high survival is likely due to earlier detection via screening for elevated prostate-specific antigen (PSA) coupled with more effective subsequent therapy. However, ~30% of men with clinically organ-confined prostate cancer who are treated by removal of their prostate may experience biochemical progression (i.e., PSA reelevation from a postsurgical nondetectable level by 10 years after surgery) or overt recurrence of their prostate cancer at local, regional, or distant sites (2, 3).

Men who experience biochemical progression after prostatectomy have a higher risk of developing metastasis and dying from their prostate cancer subsequently (4). Identification of genetic or other molecular biomarkers prognostic of prostate cancer progression would contribute to understanding mechanisms of disease progression and provide complementary information to that obtained from traditional histopathologic features or clinical parameters for predicting progression after prostatectomy (3, 5).

As for genetic and molecular biomarker studies of disease incidence, two types of observational epidemiologic study designs can be used to evaluate markers of progression, namely, the cohort study and the case-control study. A prospective cohort study has the advantage of the correct temporal sequence between observation of the marker and the outcome; that is, marker status is assessed before the progression occurred. However, the time and cost involved in collecting marker and covariate information on all members of the cohort may be substantial, and further testing all samples may deplete the biospecimen repository of the cohort rapidly. In genetic and molecular biomarker studies, an efficient design is needed to reduce cost while preserving

Received 9/9/08; revised 12/12/08; accepted 12/31/08; published OnlineFirst 3/3/09.

Grant support: Department of Defense grant DAMD 17-03-0273.

Note: The content of this article is solely the responsibility of the authors and does not necessarily represent the official view of the Department of Defense.

Requests for reprints: Elizabeth A. Platz, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Room E6132, 615 North Wolfe Street, Baltimore, MD 21205. Phone: 410-614-9674; Fax: 410-614-2632. E-mail: eplatz@jhsph.edu

doi:10.1158/1055-9965.EPI-08-0839

[6] http://seer.cancer.gov/csr/1975_2005/

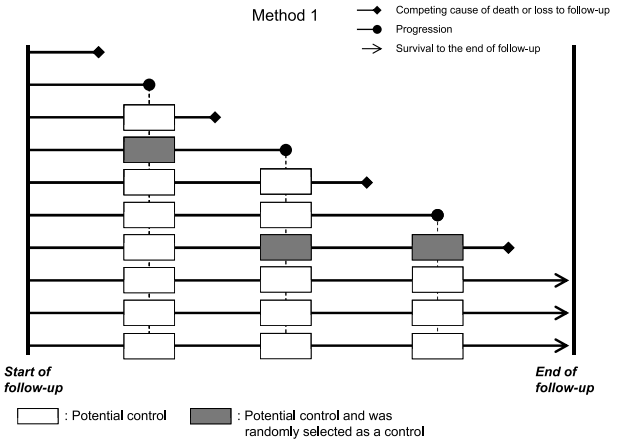

Figure 1. Control sampling method 1: ''incidence density sampling with replacement,'' an unbiased method.

temporality: a commonly used approach is to conduct a case-control study nested within a preexisting cohort (6-8). Under an appropriate control sampling method, valid estimates of the relative risk (RR) can be obtained in a more efficient way, that is, by offering impressive reductions in the cost and effort of sample testing and with relatively minor loss in statistical efficiency (9). Alternative efficient designs nested in prospective cohort studies, such as case-cohort (10), will not be considered here.

In the setting of nested case-control studies, one such control sampling method is incidence density or risk set sampling (Fig. 1; ref. 11), which is defined as follows: a control is randomly selected from all persons at risk, excluding the index case, at the time of the index case occurrence. This is repeated for each risk set. A selected control (*a*) is still eligible to be selected again as a control for another case occurring at later time, if that person still has not had the outcome of interest and is still alive and under follow-up, and (*b*) may become a case at a later time in follow-up (12, 13). The efficiency of this approach is gained by estimating the at-risk experience of all members of the risk set, as done in Cox proportional hazards regression, by the at-risk experience of the one control sampled from the risk set. Costs associated with biospecimen testing may also be reduced with this approach because resampled individuals may be tested only once. The concept of incidence density sampling of controls in genetic and molecular biomarker studies of disease etiology, including incident prostate cancer, is well established and has been shown to provide unbiased results (11, 12), irrespective of left truncation (i.e., late or staggered entries) or right censoring associated with exposure (14). However, perhaps because of the complexity of the sampling method or a reluctance of investigators to count the same subject in an analysis more than once because they perceive this to be incorrect, relatively few groups (6) have used this method of control sampling.

A widely used alternative (7, 15, 16) to incidence density sampling in the study of genetic and molecular biomarkers of disease incidence is ''incidence density sampling without replacement'' (Fig. 2). In this sampling method, once a control is sampled, he/she may not be sampled again as a control for any future case, although a control may later become a case (13). This approach preserves individual participants' biospecimens and avoids the perceived ''problem'' of counting some controls in the analysis more than once.

Another control sampling method used for nested case-control studies of disease incidence is akin to cumulative incidence sampling in a closed cohort (i.e., a cohort in which all members enter follow-up together and exit only for the outcome of interest) but can be applied to an open cohort in which all members enter together, but individuals exit for reasons other than the outcome of interest, including death from other causes (Fig. 3). This approach has been called ''pure'' control sampling; that is, controls are sampled only from the group of individuals who did not experience the outcome of interest at the end of study follow-up (12). In our planned studies on genetic and molecular biomarkers of prostate cancer progression, some colleagues advised using pure control sampling given that 30% of men treated by prostatectomy do experience biochemical progression by 10 years after surgery (2, 3). To ensure the nonprogressor phenotype, controls would be sampled after only long-term follow-up, which would be on the order of 10 or more years for progression after prostatectomy (3). However, we contend that restricting controls to those individuals who do not progress during a long follow-up period would lead to biased estimates of the RR, as it does in studies of disease incidence (12).

Here, we present the results of a Monte Carlo simulation study conducted to assess the influence of these three control sampling methods, incidence density sampling, incidence density sampling without replacement, and pure control sampling, on estimates of the RR relative to its true value for the association between a genetic variant and prostate cancer progression in the setting of long-term follow-up of a hypothetical group of men diagnosed with clinically organ-confined prostate cancer and who were surgically treated.

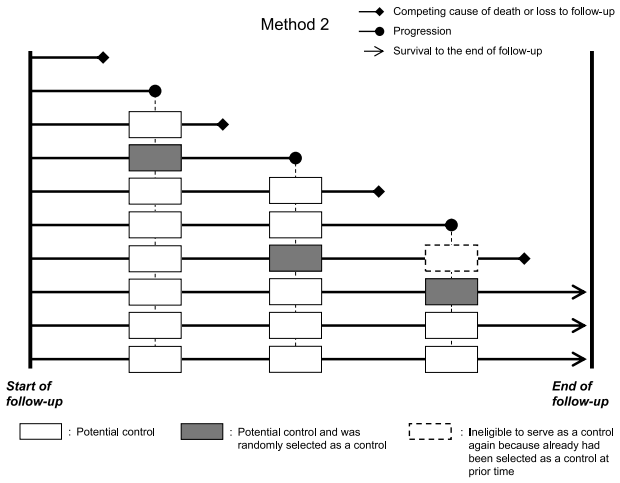

Figure 2. Control sampling method 2: incidence density sampling without replacement, a biased method.

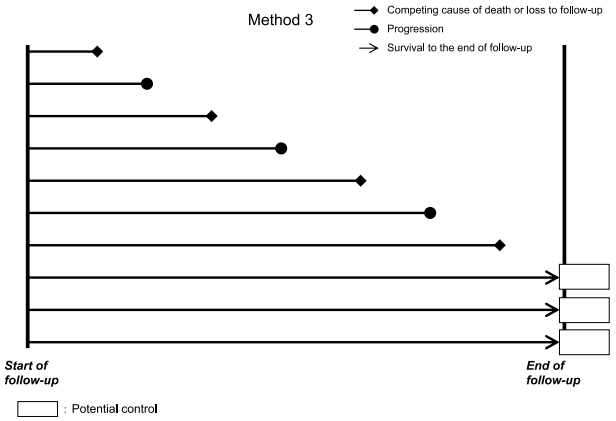

Figure 3. Control sampling method 3: "pure control sampling," a biased method.

Materials and Methods

Generation of Hypothetical Cohort. Using published long-term progression-free survival data from Johns Hopkins Hospital (3), which we combined into three risk strata, we simulated cohorts of men who underwent radical prostatectomy and were followed for 15 y for disease progression. We assumed that the likelihood of recurrence after 15 y was negligible (17). Progression was defined as PSA reelevation after being undetectable postoperatively, local recurrence, or distant recurrence with or without local recurrence as in Han et al. (3).

Pathologic stage and Gleason sum are two important predictors of disease progression; among men with clinically organ-confined prostate cancer, men with pathologically extraprostatic extension or higher Gleason sum are more likely to progress than men with pathologically organ-confined and lower Gleason sum disease (18). To reflect these patterns, we simulated three strata of patients each with different 5-, 10-, and 15-y progression-free survival proportions: stratum 1, Gleason sum <7 with or without extraprostatic extension, with 96%, 90%, and 80% for 5-, 10-, and 15-y progression-free survival proportions, respectively; stratum 2, Gleason sum ≥7 with extraprostatic extension, with 82%, 69%, and 59% for 5-, 10-, and 15-y progression-free survival proportions, respectively; and stratum 3, seminal vesicle or lymph node involvement irrespective of Gleason sum, with 68%, 51%, and 43% for 5-, 10-, and 15-y progression-free survival proportions, respectively (3).

In these simulations, X is a genetic variant that causes disease progression independent of all other known and unknown factors. X was modeled as a binary exposure variable derived from dichotomizing an independent uniform (0, 1) random variable. We generated the survival time t_{ij} for the i^{th} person in the j^{th} stratum using a Weibull distribution, $(\gamma_j/\alpha_j)^*[\exp(X_{ij}\beta)^*(t_{ij}/\alpha_j)^{(\gamma j\ -\ 1)}]$, where X_{ij} is the genetic exposure variable for the i^{th} person in the j^{th} stratum, and β is the coefficient for the natural logarithm of the RR (i.e., the relative hazard) for X; γ_j is the shape parameter and α_j is the scale parameter, these two parameters for each j^{th} stratum (Table 1).

When men diagnosed with clinically organ-confined prostate cancer die, the cause is frequently other common diseases, such as heart attack or stroke. Therefore, we simulated a competitive death probability using vital statistics data for U.S. white men in 1988 to 1992, which we assumed followed an approximate Weibull distribution (Table 1; ref. 19).

A total of nine scenarios were generated: combinations of three different magnitudes for the true association between genetic variant X and progression (RR, 1.0, 1.5, and 2.0) that are of the size typically expected in genetic and molecular biomarker epidemiologic studies, and three different prevalences of genetic variant X (10%, 30%, and 50%). For each scenario, we did 2,000 replicates, which provides a simulation SE of ~0.5% for the 95% confidence interval coverage or the type I error. In each replicate, the total sample size was fixed at 2,000, with 1,400 men (70%) in stratum 1, 400 men (20%) in stratum 2, and 200 men (10%) in stratum 3, where strata are defined by stage and grade as above. In addition, we assumed that each simulated cohort was observed for a maximum of 15 y. Progressors (events) were defined as men who post-prostatectomy experienced PSA reelevation, local recurrence, regional or distant metastasis, or death from prostate cancer before the end of follow-up and also before death due to competing causes. The remaining men were censored at death or 15 y since prostatectomy, whichever came first.

Selection of Cases and Controls. All observed events were selected as cases from the cohort. The three different control sampling methods as described in Introduction were used. When investigating genetic and molecular biomarkers of progression, which may also be associated with pathologic tumor characteristics, it is essential to adjust for possible confounding by pathologic stage and Gleason sum. To improve the efficiency of adjusting for potential confounders by

Table 1. Parameters used to simulate cohorts of men diagnosed with clinically organ-confined prostate cancer and who underwent radical prostatectomy, each cohort is composed of three different strata for risk of progression and followed up for 15 y

Stratum	Sample size	Progression-free probability			Progression hazard rate function					Competitive death rate function				
		5 y	10 y	15 y	5 y	10 y	15 y	γ*	α*	5-y cumulative deaths	10-y cumulative deaths	15-y cumulative deaths	γ*	α*
1	1,400	96%	90%	80%	0.007	0.010	0.013	1.5	60	55	89	156	8.74	95.59
2	400	82%	69%	59%	0.023	0.021	0.019	0.85	52	9	24	44	8.74	95.59
3	200	68%	51%	43%	0.032	0.027	0.023	0.75	40	5	8	13	8.74	95.59

*Under Weibull distribution assumption, $(\gamma_j/\alpha_j)^*[\exp(X_{ij}\beta)^*(t_{ij}/\alpha_j)^{(\gamma_j-1)}]$, where γ is the shape parameter and α is the scale parameter.

Table 2. Bias in the average estimated RR using three control sampling methods

True RR	Proportion exposed	Average RR			
		Full cohort design	Case-control designs		
			Method 1: incidence density sampling without replacement	Method 2: incidence density sampling with replacement	Method 3: pure control sampling
1.00	10%	0.99	1.00	1.00	0.99
	30%	1.00	1.00	1.00	1.00
	50%	1.00	1.00	1.00	1.00
1.50	10%	1.50	1.51	1.51	1.63
	30%	1.50	1.50	1.49	1.62
	50%	1.50	1.50	1.49	1.62
2.00	10%	2.00	2.02	2.01	2.34
	30%	2.00	2.00	1.98	2.33
	50%	2.00	2.00	1.98	2.34

NOTE: For a variant prevalence of 10%, the average simulated number of cases when the true RR was 1.0, 1.5, or 2.0 was 339.0 ± 16.5, 353.3 ± 16.4, and 365.5 ± 15.9, respectively. Case-control ratios for the simulated full cohorts when the true RRs were 1.0, 1.5, and 2.0 were 1:4.9, 1:4.7, and 1:4.5. For a variant prevalence of 30%, the average simulated number of cases when the true RRs were 1.0, 1.5, and 2.0 was 339.5 ± 16.4, 381.6 ± 16.4, and 417.8 ± 17.2, respectively. Case-control ratios for the simulated full cohorts when the true RRs were 1.0, 1.5, and 2.0 were 1:4.9, 1:4.2, and 1:3.8. For a variant prevalence of 50%, the average simulated number of cases when the true RRs were 1.0, 1.5, and 2.0 was 339.8 ± 16.3, 408.7 ± 17.1, and 469.2 ± 18.2, respectively. Case-control ratios for the simulated full cohorts when the true RRs were 1.0, 1.5, and 2.0 were 1:4.9, 1:3.9, and 1:3.3.

ensuring equal numbers of cases and controls within each stratum, controls may be sampled so that they are matched to cases on key characteristics. Thus, in each scenario, one control per case was selected and matched on pathologic stage and Gleason sum.

Analysis of Simulated Data. The relative hazard was estimated for each of the three control sampling methods for each of the 2,000 replicates under each of the nine scenarios using conditional logistic regression. Convergence was achieved for all replicates. We report an estimate of the bias, which was defined as the average difference between $\beta_{estimate}$ from the nested case-control studies using control sampling methods 1, 2, or 3 and the natural logarithm of true RR. The estimated rejection probability was defined as the fraction of replicates for which [$\beta_{estimate}$ / SE of $\beta_{estimate}$] exceeded 1.96 under $\alpha = 0.05$, for a two-sided test of $\beta = 0$. The significance level of this estimated rejection probability is the observed type I error in the scenario of RR = 1.0 and is the observed power in the scenarios of RR = 1.5 or 2.0. Statistical Analysis System version 9.1 was used for all analyses (SAS Institute).

Results

There was no bias evident in $\beta_{estimate}$ under control sampling method 1 for any of the nine scenarios (Table 2). For control sampling method 2, $\beta_{estimate}$ was not appreciably different from the true value and not obviously different from $\beta_{estimate}$ from method 1 for any of the nine scenarios (Table 2). On the other hand, persistent bias that overestimated the true RR was observed for method 3 when the true RR was 1.5 or 2.0 and the prevalence of the genetic variant was 10%, 30%, or 50%; no bias was observed when the true RR was 1.0 (Table 2). When the true RR was above the null, on the absolute scale, the bias in $\beta_{estimate}$ increased with increasing size of the true RR, and on the relative scale, the bias in $\beta_{estimate}$ increased at a faster rate than the increase in true RR; the extent of the bias did not vary with the prevalence of the genetic variant.

When the true RR was 1.0, the estimated rejection probabilities for testing the null hypothesis of $\beta = 0$ (i.e., type I error) were within the sampling error of their nominal values for each control sampling method, with the exception of method 3 when the variant prevalence was 30%, which exceeded the 5% significance level. Method 2 provided a more conservative type I error than did method 1 (Table 3). When the true RR was 1.5 or 2.0, power to reject the null hypothesis of $\beta = 0$ was similar across the three variant prevalences and the magnitude of the true RR for both control sampling methods 1 and 2, but was overstated for method 3.

In the 2,000 simulated replicates for each of the nine scenarios, the true RR coverage probability of the 95% confidence intervals was consistently ~95% in all scenarios of simulation for the full cohort and for each of the control sampling methods (Supplementary Table S1).

To evaluate length of follow-up since prostatectomy on the extent of the bias for control sampling method 3, we conducted an additional simulation using the same parameters for the true RR and the prevalence of the genetic variant and then varied the minimum follow-up time to be eligible to be a control from 5 to 10 to 15 years. The extent of the bias increased with the length of follow-up (data not shown); for instance, when the variant prevalence was 50% and the true RR was 2.0, the extent of bias in $\beta_{estimate}$ was +0.05, +0.134, and +0.164 for minimum follow-up times of 5, 10, and 15 years, respectively.

Discussion

As is well known in the epidemiologic literature, it is obvious that different case-control sampling methods could lead to unbiased (11) or biased (12, 13) estimates of the true RR. From this Monte Carlo simulation study on prostate cancer progression, we showed that unbiased

estimates can be ensured by selecting controls randomly from all those at risk at each case event time t (i.e., method 1, incidence density sampling). However, substantial bias can be induced by selecting as controls only men who have not progressed by the end of long-term follow-up (i.e., method 3, pure control sampling).

Understanding how incidence rates of progression are calculated in a prospective cohort study can provide insight into why control sampling method 1 should be unbiased, whereas sampling method 3 is biased. In a prospective cohort study of prostate cancer progression, the progression incidence rate is defined as follows: (no. progressed) / (Σ person*time at risk). Each man who progresses contributes not only to the numerator but also to the person-time denominator of this rate, which is tallied up to the time when he is no longer at risk. Therefore, each man who progresses contributes to the rate person-time at risk before he progresses, and thus, he would be eligible to be sampled as a control before he progresses. Viewed another way, a man sampled as a control at time t is also eligible to become a case at any subsequent time. Furthermore, a man selected as a control remains eligible to be selected again as a control later as long as he continues to contribute person-time at risk to the rate denominator. Thus, in a nested case-control study, the same man may appear in the control group one or more times and in both the case and control groups (12, 13).

Although we expected the control sampling theory for disease incidence to apply to disease progression, we nevertheless undertook this simulation in the context of prostate cancer progression because there is a tendency to view pure controls as preferable. A researcher might use control sampling method 3 to conduct a nested case-control study for practical reasons, such as not knowing event times. Another reason for adopting this approach could be based on (faulty) intuition; that is, a progressor is a progressor irrespective of when he progressed and, thus, never eligible to be a control. However, our simulation study provides evidence that control sampling method 3 yields biased estimates of the RR for progression. This bias results because the exposure distribution among pure controls is distorted relative to that in the source population from which the cases are derived. In method 3, the controls cannot have progressed by the end of 15 years of follow-up. If genetic variant X is a risk factor for progression, then the frequency of the genetic variant in the control group sampled using method 3 must differ from that in the source population that gave rise to the cases. The probability of carrying the genetic variant among controls would be much lower than the source population, and therefore, the estimated RR would be overestimated. Furthermore, other potential sources of bias in the estimate of RR that can be induced when using the pure control sampling method are when men who are lost to follow-up differ from men who remain in the cohort with respect to the probability of carrying genetic variant and of the outcome of interest or another outcome; we did not explore these sources of bias in this simulation. Wacholder et al. (20) have described the principle of unbiased control sampling from risk sets in a cohort.

We expected that control sampling method 2 (incidence density sampling without replacement) would lead to bias in estimating the RR of progression when the true RR was not equal to 1.0. However, we noted differences in the β_{estimate} between method 2 and the truth only when the true RR was large, and we also noted that when the truth was no association, the type 1 error was too conservative. The lack of appreciable bias for method 2 when the true RR was small likely resulted because when the baseline rate of disease and the RR are small, there is a greater pool of person-time at risk from which to sample controls, and thus, the likelihood of resampling the same man as a control more than once would be quite limited. The only difference between methods 1 and 2 is that the former allows a control to be repeatedly selected as a control if the man remains at risk, but the latter does not allow for that possibility. To document this conceptual difference, we conducted another simulation analysis by increasing the hazard of progression, which resulted in a higher chance of a control being selected more than once because the number of incident cases was larger. Results of this analysis showed biased β_{estimate} when using method 2, whereas method 1 continued to be unbiased (Supplementary Table S2).

We observed that the extent of bias varied with the minimum length of follow-up since prostatectomy required to be eligible to be a control given the same

Table 3. Type 1 error and power for the estimated RR using three control sampling methods

True RR	Proportion exposed	Type 1 error or power*			
		Full cohort design	Case-control designs		
			Method 1: incidence density sampling without replacement	Method 2: incidence density sampling with replacement	Method 3: pure control sampling
1.00	10%	5%	5%	3%	5%
	30%	5%	5%	3%	6%
	50%	5%	5%	3%	5%
1.50	10%	72%	40%	38%	50%
	30%	97%	74%	75%	87%
	50%	98%	83%	84%	92%
2.00	10%	>99%	88%	89%	95%
	30%	>99%	>99%	>99%	>99%
	50%	>99%	>99%	>99%	>99%

*Under the null of a true RR = 1, the type 1 error is given; under the alternatives of a true RR = 1.5 or 2, the power is given.

prostate cancer progression rate, genetic variant prevalence, and sample size. We did another simulation for method 3 using the same parameters except for minimum length of follow-up required to be eligible to be a control, and observed the extent of bias increased with length of follow-up. As we mentioned earlier, the most important reason for bias in method 3 is that the variant prevalence among sampled controls does not represent the prevalence in the source population for the cases. More specifically, the prevalence of a high-risk genetic variant in the controls sampled using method 3 will be lower than the variant prevalence in the source population. The longer the time since prostatectomy, the greater the difference in the distribution of the variant between the controls sampled using method 3 and underlying source population, and this difference creates more biased estimate of RR of progression.

In this simulation study, we chose shape and scale parameters such that we allowed the strata defined by pathologic stage and grade to have an effect on the rate of progression, but we did not allow the hazard ratio to be nonproportional over time. Future work may wish to explore the implications on bias, power, or confidence interval coverage for when the hazard ratio of progression is not constant over follow-up time.

The evidence of biased estimates of the RRs using the pure control sampling method provided in this study becomes more important especially when a study is designed to replicate and confirm genetic association findings from previous studies, and even more important for subsequent meta-analyses. Investigators may be tempted to take advantage of the characteristics of method 3 pure control sampling, specifically its higher power in studies designed to search for unknown genes associated with disease outcome, but researchers should be alert to two possible limitations of this method: (*a*) survival bias (not evaluated in this simulation) and (*b*) overestimation of the true RR when the genetic variant increases the risk of disease outcome (evaluated in this simulation). When the study aim is to accurately estimate the true RR of genes discovered earlier as done in confirmation studies (e.g., the subsequent steps in Cancer Genetic Markers of Susceptibility[7]), the valid incidence density sampling method should be applied.

In conclusion, using a nested case-control study to search for genetic variants contributing to prostate cancer progression within a defined cohort provides an unbiased estimate of the RR when a valid control sampling method, such as method 1 in this simulation study, is used. This method of incidence density sampling is well known for use in studies of genetic and molecular biomarkers of disease incidence. This rule can be applied to nested case-control studies of any outcome of interest, whether disease incidence or progression, with the certainty of achieving an unbiased estimate of the RR.

[7] http://cgems.cancer.gov/

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Katherine Zarfas for programming assistance.

References

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2008. CA Cancer J Clin 2008;58:71–96.
2. Roehl KA, Han M, Ramos CG, Antenor JA, Catalona WJ. Cancer progression and survival rates following anatomical radical retropubic prostatectomy in 3,478 consecutive patients: long-term results. J Urol 2004;172:910–4.
3. Han M, Partin A, Pound C, Epstein J, Walsh P. Long-term biochemical disease-free and cancer-specific survival following anatomic radical retropubic prostatectomy: the 15-year Johns Hopkins experience. Urol Clin North Am 2001;28:555–65.
4. Pound CR, Partin AW, Eisenberger MA, Chan EW, Pearson JD, Walsh PC. Natural history of progression after PSA elevation following radical prostatectomy. JAMA 1999;281:1591–7.
5. Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. J Natl Cancer Inst 1998;90:766–71.
6. Michaud DS, Daugherty SE, Berndt SI, et al. Genetic polymorphisms of interleukin-1B (IL-1B), IL-6, IL-8, and IL-10 and risk of prostate cancer. Cancer Res 2006;66:4525–30.
7. Platz EA, Leitzmann MF, Rifai N, et al. Sex steroid hormones and the androgen receptor gene CAG repeat and subsequent risk of prostate cancer in the prostate-specific antigen era. Cancer Epidemiol Biomarkers Prev 2005;14:1262–9.
8. Hayes VM, Severi G, Padilla EJ, et al. 5α-Reductase type 2 gene variant associations with prostate cancer risk, circulating hormone levels and androgenetic alopecia. Int J Cancer 2007;120:776–80.
9. Ernster VL. Nested case-control studies. Prev Med 1994;23:587–90.
10. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 1986;73:1–11.
11. Prentice RL, Breslow NE. Retrospective studies and failure time models. Biometrika 1978;65:153–8.
12. Lubin JH, Gail MH. Biased selection of controls for case-control analyses of cohort studies. Biometrics 1984;40:63–75.
13. Robins JM, Gail MH, Lubin JH. More on ''Biased selection of controls for case-control analyses of cohort studies.'' Biometrics 1986;42:293–9.
14. Flanders WD, Louv WC. The exposure odds ratio in nested case-control studies with competing risks. Am J Epidemiol 1986;124:684–92.
15. Visvanathan K, Helzlsouer KJ, Boorman DW, et al. Association among an ornithine decarboxylase polymorphism, androgen receptor gene (CAG) repeat length and prostate cancer risk. J Urol 2004;171:652–5.
16. Li H, Kantoff PW, Giovannucci E, et al. Manganese superoxide dismutase polymorphism, prediagnostic antioxidant status, and risk of clinical significant prostate cancer. Cancer Res 2005;65:2498–504.
17. Allaf ME, Partin AW, Walsh PC. Who is cured? How long after radical prostatectomy can PSA rise? American Urological Association Annual Meeting; 2006.
18. Epstein JI, Partin AW, Sauvageot J, Walsh PC. Prediction of progression following radical prostatectomy. A multivariate analysis of 721 men with long-term follow-up. Am J Surg Pathol 1996;20:286–92.
19. Pickle LW, Mungiole M, Jones GK, White AA. Atlas of United States mortality. Hyattsville (MD): National Center for Health Statistics; 1996.
20. Wacholder S, Gail M, Pee D. Selecting an efficient design for assessing exposure-disease relationships in an assembled cohort. Biometrics 1991;47:63–76.

Correction

Correction: Article on Sampling in Studies of Prostate Cancer Progression

In the article on sampling in studies of prostate cancer progression in the March 2009 issue of *Cancer Epidemiology, Biomarkers & Prevention*, there were errors in Tables 2 and 3 (1). The correct headings are as follows.

Method 1: incidence density sampling with replacement.

Method 2: incidence density sampling without replacement.

Reference

1. Wang MH, Shugart YY, Cole SR, Platz EA. A simulation study of control sampling methods for nested case-control studies of genetic and molecular biomarkers and prostate cancer progression. Cancer Epidemiol Biomarkers Prev 2009;18:706–11.

doi:10.1158/1055-9965.EPI-18-7-COR2

A Simulation Study of Control Sampling Methods for Nested Case-Control Studies of Genetic and Molecular Biomarkers and Prostate Cancer Progression

Ming-Hsi Wang, Yin Yao Shugart, Stephen R. Cole, et al.

Cancer Epidemiol Biomarkers Prev 2009;18:706-711.