



HHS Public Access

Author manuscript

Eur J Epidemiol. Author manuscript; available in PMC 2016 March 01.

Published in final edited form as:

Eur J Epidemiol. 2015 March ; 30(3): 197–207. doi:10.1007/s10654-014-9974-4.

A New Comparison of Nested Case-Control and Case-Cohort Designs and Methods

Ryung S. Kim*

Department of Epidemiology and Population Health, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, New York, 10461 USA

Abstract

Existing literature comparing statistical properties of nested case-control and case-cohort methods have become insufficient for present day epidemiologists. The literature has not reconciled conflicting conclusions about the standard methods. Moreover, a comparison including newly developed methods, such as inverse probability weighting methods, is needed. Two analytical methods for nested case-control studies and six methods for case-cohort studies using proportional hazards regression model were summarized and their statistical properties were compared. The answer to which design and method is more powerful was more nuanced than what was previously reported. For both nested case-control and case-cohort designs, inverse probability weighting methods were more powerful than the standard methods. However, the difference became negligible when the proportion of failure events was very low (<1%) in the full cohort. The comparison between two designs depended on the censoring types and incidence proportion: with random censoring, nested case-control designs coupled with the inverse probability weighting method yielded the highest statistical power among all methods for both designs. With fixed censoring times, there was little difference in efficiency between two designs when inverse probability weighting methods were used; however, the standard case-cohort methods were more powerful than the conditional logistic method for nested case-control designs. As the proportion of failure events in the full cohort became smaller (<10%), nested case-control methods outperformed all case-cohort methods and the choice of analytic methods within each design became less important. When the predictor of interest was binary, the standard case-cohort methods were often more powerful than the conditional logistic method for nested case-control designs.

Keywords

Nested Case-Control; Case-Cohort; Simulation Study; Inverse Probability Weighting

*Corresponding author. Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Block 316, 1300 Morris Park Ave, Bronx, New York, 10461 USA. Tel: 1-347-886-0027. Fax: 1-718-839-7977. ryung.kim@einstein.yu.edu.

Conflict of Interest
None declared.

Introduction

Case-cohort and nested case-control designs are the most common approaches for reducing the costs of exposure assessment in prospective epidemiologic studies. Exposure data in these designs are obtained on a subset of the full cohort. Nested case-control designs (or equivalently, incidence density sampling designs) include all cases and a pre-specified number of controls randomly chosen from the risk set at each failure time [1]. Case-cohort designs include all cases and one randomly selected sub-cohort from the risk set at baseline [2].

A few studies, all published before the year 2000, compared the statistical properties of the two designs. Prentice [2] and Self & Prentice [3] reported that case-cohort designs coupled with their respective analysis methods yielded higher statistical efficiency than nested case-control designs coupled with conditional logistic approach proposed by Thomas [1]. Langholz & Thomas [4] later pointed out that these conclusions had not accounted for repeated sampling of same persons in nested case-control designs. In their own simulation studies, nested case-control designs coupled with conditional logistic method was more efficient than the case-cohort designs coupled with Self & Prentice method [3] when there was moderate random censoring or staggered entries. Barlow et al. [5] reported that Prentice [2]'s method was more efficient than Barlow [6] which in turn was more efficient than Self & Prentice [3]. They reported that all these methods were more efficient than nested case-control designs coupled with conditional logistic approach when estimating the relative risk with respect to a binary predictor. They, however, did not find meaningful difference for a continuous predictor.

These reports have become insufficient for present day practitioners. First, more efficient analytical methods including inverse probability weighting methods have been developed, and a comparison of the new methods is needed. Second, the literature has not reconciled the seemingly conflicting conclusions even about the traditional methods. Self & Prentice [3] concluded case-cohort designs coupled with their method is more efficient than the nested case-control designs coupled with conditional logistic method. On the contrary, Langholz & Thomas [4] concluded that conditional logistic method out-performed Self & Prentice [3] when there was moderate random censoring. By Barlow et al [5], the order is reversed yet again and, in fact, both were outperformed by Prentice [2]. All these studies implied there should be a single answer to what is the best design and analysis method, considered one non-zero value of log relative risk, one cohort size, and a binary predictor. The real answer may be more nuanced. A more comprehensive investigation with varying magnitude of relative risks, cohort sizes, and incidence proportion may sort out what affects the relative performance. In addition, the relative performance is unknown for continuous predictors for which conditional logistic approach suffers less from sparse samples. Third, the literature compared efficiency of the methods but not power. However, a seemingly large difference in efficiency may not yield practically meaningful difference in power. For example, when the true relative risk is large, a large difference in efficiency may lead to only a moderate difference in power because the power will reach near the upper limit of one. In the author's opinion, what has happened in practice is that researchers would choose one of the designs without a clear understanding about the ramification of the choice on statistical power.

This article first summarizes various analytical methods for nested case-control and case-cohort designs using Cox proportional hazards model. We perform simulation studies to compare bias, efficiencies, type 1 errors, and powers of the methods by varying the following design factors: the type of predictors (continuous or binary), the magnitude of hazard ratio, cohort size, the number of controls, censoring type, and the proportion of failure events. Two analytical methods for nested case-control studies are considered: conditional logistic approach of Thomas [1] and the inverse inclusion probability method of Samuelsen [7] coupled with an approximate jackknife standard error [8, 9]. Six methods for case-cohort studies are considered: Prentice [2], Self & Prentice [3], Lin & Ying [10], Barlow [6], and both Prentice [2] and Binder [11] coupled with approximate jackknife standard errors.

Methods

Notations and Methods

Consider a cohort study with N subjects. Subject i enters the study at a fixed time a_i . The subject has the failure time T_i and the censoring time W_i . The investigator observes only $Y_i = \min(T_i, W_i)$. The failure time and censoring time are assumed to be independent. $X_i(t)$ is a time dependent covariate vector of the subject. Assume that the hazard function $\lambda_i(t)$ of the failure time follows the model $\lambda_i(t) = \lambda_0(t) \exp(X_i(t)\beta)$ where $\lambda_0(t)$ is an unspecified baseline hazard function and β is the parameter vector of interest. Then, inferences are typically made by maximizing Cox partial likelihood:

$$L_c(\beta) = \prod_{i=1}^N \left[\frac{e^{X_i(Y_i)\beta}}{\sum_{j \in R_i} e^{X_j(Y_i)\beta}} \right]^{\delta_i} \quad (1)$$

δ_i is one if subject i failed during the study, otherwise it is zero; $R_i = \{j: Y_j > a_j\}$ is a risk set, an index set of subjects at risk at the failure time of subject i .

In nested case-control studies, for each case, m controls are sampled without replacement at each Y_i where $\delta_i=1$ from $R_i \cap \{i\}^c$, i.e. from the subjects still at risk at the time of the failure of the case [1]. Notice that the controls may include both failures and non-failures. Case-cohort studies include all cases and a randomly selected sub-cohort from the risk set at baseline [2], i.e., controls are randomly selected once from R_0 , the risk set at baseline. Let C_i denote the set of controls selected from R_i and C denote the union of all controls. For case-cohort design, $C = C_0$ is the subcohort. For nested case-control designs, $C = \cup_{i: \delta_i=1} C_i$. Then $S = \{i: \delta_i=1\} \cup C$ is the index set of subjects that were ever included in the sample. $R_i \cap S$ is a risk set in the sample, an index set of subjects included in the sample who are also at risk at failure time of subject i . By changing the weighting strategies, the following pseudo/partial-likelihood was maximized by Thomas [1] and Samuelsen [7] for nested case-control designs, by Prentice [2], Self & Prentice [3], and Barlow [6] for case-cohort designs, by Binder [11] for complex survey data, and by Lin & Ying [10] for incomplete data:

$$L_w(\beta) = \prod_{i \in S} \left[\frac{e^{X_i(Y_i)\beta}}{\sum_{j \in R_i \cap S} w_j e^{X_j(Y_i)\beta}} \right]^{\delta_i} \quad (2)$$

The top section of Table 1 shows how the methods use different weights in the pseudo/partial-likelihood (2). Barlow [5] explained the weights in the first three methods (from the top). Unlike these methods, the inverse probability weighting methods, Binder [11] and Samuelsen [7], include the non-subcohort case in earlier risk sets and use knowledge about the future case status of that individual. Barlow [5] speculated that such use might bias the estimates in case-cohort designs but Binder [11] and Lin [12] proved the consistency of the estimator for complex sampling designs. Samuelsen [7] proved the consistency for nested case-control designs. The same argument of the proofs can be applied to case-cohort designs. In short, this is because the pseudo-likelihood (2) is a design-consistent estimator of Cox's partial likelihood (1) conditional on the full cohort and considering only the randomness of the indicators of which subjects are to be included in the particular case-cohort study. This property ensures that inverse probability weighted estimator to be a design-consistent estimator of the Cox's estimator, which in turn is a model-consistent estimator of β under the proportional hazards assumption. We note that (2) for the case of Thomas [1] is a partial likelihood, and not a pseudo-likelihood, since its contributions are score unbiased and the variance of the score is the expected information.

In case-cohort studies, the proportion of sub-cohort is fixed at, say, π . For nested case-control studies, Samuelsen [7] calculated inclusion probabilities and Kim [8] extended them to account for ties and additional matching. The inclusion probability of subject i is the following:

$$p_i = \begin{cases} 1 & \text{if } \delta_i = 1 \\ 1 - \prod_{j: a_i < Y_j < Y_i} \left(1 - \min\left(1, \frac{mb_{ji}}{k_{ji} - b_{ji}}\right) \right) & \text{in nested case-control study if } \delta_i = 0 \\ \pi & \text{in case cohort study if } \delta_i = 0 \end{cases} \quad (3)$$

Let H_i be the index set of the subjects with the same values of matching variables as subject i . k_{ji} is the size of $R_j \cap H_i$ or the number of subjects at risk at t_j with the same values of matching variables as subject i ; b_{ji} is the number of subjects in H_i that failed exactly at t_j ; m is the number of controls per each failure.

The weights in the top section of Table 1 give insight into Barlow [5]'s finding of the higher efficiency of Prentice [2] over Self & Prentice [3]'s method which uses less data to estimate the covariate contribution, and Barlow [6]'s method which always uses equal or greater weights. It is reasonable to hypothesize that inverse probability weighting methods would have higher efficiencies than others in both designs because they use the non-subcohort cases in earlier risk sets. It has been shown that such was the case for the nested case-control designs [7, 9] but not yet for the case-cohort designs.

The stated assumption that failure and censoring being independent was in fact more than necessary. The appropriate conditions are that the form of the intensity (or the rate ratio) does not change due to censoring or sampling. See conditions for full cohort in Anderson and Gill [13], for case-cohort in Self and Prentice [3], and for nested case-control in Borgan et al [14].

Standard Error Estimation

We studied if various standard error estimators meaningfully affect the performance of the aforementioned methods. Prentice [2] and Self & Prentice [3] each proposed asymptotic variance estimators for their respective methods. Lin and Ying [10]'s estimating equation for incomplete data reduces to the pseudo-likelihood score function of the Self & Prentice [3] for case-cohort designs. Therefore, Lin and Ying [10]'s proposed approximate jackknife variance estimator is an alternative variance estimator for Self & Prentice's method [3]. Barlow [6] also proposed an approximate jackknife variance estimator for his method. Binder [11] originally provided a variance estimator that accounts only for the sampling variation from a finite cohort but Lin [12] later provided the variance estimator for Binder's method accounting for the sampling of the cohort from an infinite super-population.

For nested case-control designs, Goldstein & Langholz [4] proved that the typical standard errors from standard conditional logistic software are valid for the conditional logistic approach of Thomas [1]. Samuelsen [7]'s proposed an asymptotic variance estimator for the inverse probability weighting method that is similar to the variance estimator of Lin [12].

The second section of Table 1 summarizes the various standard errors. Therneau [15] pointed out that Self & Prentice [3] variance converges to the usual Cox model variance as the sample size increases to the size of the full cohort. This is also true for Lin [12], Thomas [1], and Samuelsen [7]. On the other hand, the Lin & Ying [10] estimate converges to the approximate jackknife estimate of Lin & Wei [16]. These approximate jackknife type standard errors have an advantage that they are available in standard software such as SAS and R without any extra programming. We did not calculate Lin [12] and Samuelsen [7]'s variance estimators for inverse probability weighting methods. They both require computational memory in the order of $O(n^2)$ to compute all pair-wise co-inclusion probabilities. Instead, we used approximate jackknife estimators. Kim [8, 9] showed that they accurately estimated the empirical variance and the estimates were remarkably close to that of Samuelsen [7]. We also did not calculate the variance proposed by Prentice [2] and used the form of variance estimator proposed by the Self & Prentice [3] as well as an approximate jackknife estimator.

Programming

All analyses were programmed in R environment [17]. We followed Therneau and Li [15]'s guidance in programming Self & Prentice [3] and Lin & Ying [10]. And we followed Barlow [5]'s guidance in programming Prentice [2] and Barlow [6]. Inverse probability weighting methods and approximate jackknife type variance estimators are straightforward to program in R using weights and cluster options in coxph function. See reference [18] for the R code.

Results

Exponential failure times T with rate $\exp(\beta_1 X_1 + \beta_2 X_2)$ were generated for full cohorts with size $N = 500, 1,000, \text{ or } 1,500$. In addition, X_1 was assumed to be distributed as a standard normal variable, and X_2 was specified as independent Bernoulli variables with success probability of $(1 + \exp(-X_1))^{-1}$. The distribution of covariates was set up so that a mild multicollinearity existed. The true log hazard ratio β_1 assumed the values of 0, 0.1, 0.2... 0.8. The hazard ratios, therefore, ranged between 1 and 24.5 by an increase of the variable equivalent to four standard deviations. The log hazard ratio for X_2 was set as $\beta_2 = 0.5$. Censoring times were uniformly distributed between 0 and c , the upper limit of censoring, which was chosen so that the proportion of failure events was, on average, 15% in the full cohort. For each subject, either the failure or censoring time was observed, whichever occurred earlier. The log hazard ratios and their standard errors in the full cohort were estimated under the Cox proportional hazards model.

Nested case-control samples were then selected with varying numbers of controls, $m = 1, 2, \text{ or } 5$, at each failure time. When each nested case-control sample was selected, a case-cohort sample was also selected. To make the average sample size of the two designs the same, the sampling proportion for the subcohort of the case-cohort sample was set as the number of non-failures in the nested case-control sample divided by the number of non-failures in the full cohort. For simplicity, additional matching factors were not used.

For each simulated nested case-control and case-cohort data set, log hazard ratios were estimated according to the pseudo-likelihood (2) using weights defined in Table 1 along with the standard errors methods in the table. This overall process including the generation of the full cohort, nested case-control sample, and case-cohort sample was repeated 5,000 times. For the estimation of empirical type 1 error (i.e. when $\beta_1 = 0$), the overall process was repeated 20,000 times.

Figure 1 shows the empirical biases in estimating β_1 from the full cohort analysis, two nested case-control methods, and four case-cohort methods. When sample sizes were small to moderate ($N=500, m=1$, or $N=500, m=2$, or $N=1,000, m=1$), Self & Prentice [3] and Barlow [6]'s methods over-estimated non-zero β_1 about 5–20%. The biases of other methods were less than 5% of β_1 except for the smallest sample sizes considered ($m=1, N=500$) when the average sample size (n^*) was 131.9. The biases became negligible as the sample size increased.

Figure 2 shows the empirical standard errors by the methods. Consistent with the report by Barlow et al [5], we observed higher efficiency of Prentice [2] over the methods of Self & Prentice [3] and Barlow [6]. Notably, inverse probability weighting methods yielded higher efficiency compared to other methods in both case-cohort and nested case-control designs. This is because they use more data, namely the non-subcohort cases in earlier risk sets, to estimate the covariate contribution in (2). As we expected, nested case-control designs showed higher efficiency over case-cohort when inverse probability weighting methods are used in both designs. Interestingly, contrary to Barlow et al [5]'s report, the conditional logistic approach of Thomas [1] outperformed Prentice [2], Self & Prentice [3], and Barlow

[6]. The discrepancy between the findings comes from X_1 being a continuous predictor for which conditional logistic approach suffers less from the sparse samples with small numbers of discordant pairs. Barlow et al. [5] had reported that these methods were more efficient than the conditional logistic approach when estimating a relative risk with respect to a binary predictor. The figure also affirmed the theoretical finding by Zhang and Goldstein [19] that for Self-Prentice [3] to obtain greater than 80% optimal efficiency at $\beta=0$, the sampling fraction needs to be at least as three to five times the proportion of failure events.

Controlling the nominal type 1 error rate at 0.05 in testing $H_0: \beta_1=0$, the empirical power and type 1 error rates were measured. All methods yielded empirical type 1 error rates close to the nominal rate (Table 2) with the exception of the inverse probability weighting methods and Prentice [2] that were mildly inflated when the sample sizes were small (<0.06 ; $N=500, m=1$ or $N=1,000, m=1$). Figure 3 shows the empirical power of the methods. The conclusions are similar to that from the empirical standard errors. Nested case-control designs coupled with the inverse probability weighting method by Samuelsen [7] was most powerful. Also among the case-cohort methods, inverse probability method by Binder [11] was most powerful, and then it was Prentice [2] over the methods of Barlow [6] and Self & Prentice [3]. The first column of the Table 3 demonstrates the power of the methods when $N=500, m=2$, and $\beta_1=0.5$. For example, the empirical powers by Samuelsen [7], Binder [11], Thomas [1], Prentice [2], Barlow [6], Self-Prentice [3] were 0.90, 0.85, 0.85, 0.83, 0.83, and 0.80. Notice that the empirical variances of Binder, Thomas, Prentice, Barlow, and Self-Prentice were 22%, 23%, 47%, 80%, and 91% greater than that of Samuelsen in the same setting. The relative difference of power was moderate even when the difference of variance seemed dramatic.

Next, to study the relative performance for binary predictors, we repeated the simulation study but this time for β_2 , the log hazard ratio with respect to the binary predictor X_2 . The β_2 assumed the values of 0, 0.2, 0.4... 2.0. The log hazard ratio for X_1 was set as $\beta_1=0.5$. Again, empirical biases (Web Figure 1) were less than 5% of β_2 for all methods except for the smallest sample sizes considered ($m=1, N=500$). Inverse probability weighting methods yielded the highest efficiency in both case-cohort and nested case-control designs (Web Figure 2). We observed higher efficiency of Prentice [2] over the methods of Self & Prentice [3] and Barlow [6]. And nested case-control designs yielded higher efficiency over case-cohort when inverse probability weighting methods were used in both designs. In short, some conclusions from the analysis of β_1 were confirmed in the analysis β_2 .

However, the conditional logistic approach of Thomas [1] was less stable for the binary predictor when the sample sizes were small ($N=500, m=1$ or $N=1,000, m=1$; Web Figure 2). There were a number of sparse samples in which hazard ratios were inestimable by the conditional logistic approach and, with less frequency, Self & Prentice [3]. Such sparse samples were more frequent with large β_2 , which makes in unlikely for a subject with $X_2=1$ to fail. For example, when $N=500, m=1, \beta_2=2$, 4% of the samples did not have any risk set that had both at least one case with $X_2=0$ and at least one control with $X_2=1$. The relative risk estimates by the conditional logistic approach in these samples were infinity. For fair comparison, these sparse samples were excluded from the analysis by all considered

methods. Notice that the differences in power between Thomas [1], Prentice [2], Barlow [6], and Self-Prentice [3] were still moderate (Web Figure 3).

To study how incidence proportion affects the relative performance, we then repeated the simulation study this time with varying proportion of failure events in the full cohort. Here the upper limit of random censoring was varied so that the proportion of failure events in the full cohorts was, on average, 0.1%, 1%, 5%, or 10%. The sizes of the full cohort were adjusted to contain the same average number of cases as the above simulation studies. For example, when the proportion of failure events was set at 1%, the cohort size was $N = 7,500$, 15,000, or 22,500. The β_1 assumed the values of 0, 0.1, 0.2... 0.8. The log hazard ratio for X_2 was set as $\beta_2=0.5$. In testing $H_0: \beta_1=0$, both nested case-control methods were more powerful than any case-cohort method when the proportion of failure events was less than or equal to 10% and the difference in power grew greater as the proportion became smaller. Interestingly, in both designs, the difference in power between inverse probability weighting method and the standard methods (Prentice [2] for case-cohort studies and the conditional logistic method for the nested case-control studies) became ignorable as the proportion of failure events became very low (< 1%; Table 3 shows when $m=2$, $\beta_1=0.5$). This is because the inverse probability weights for the controls become too large compared to the weights for the cases when the proportion of failure events is low. There seems to be a balancing effect on efficiency: inverse probability weighting methods gain efficiency by using more controls in the risk sets but lose it when variation among the weights is large.

We also suspected that the random censoring might favor nested case-control design as case-cohort design was intended for the situation where most of the censoring was administrative at the end so that controls would serve in many risk sets. Intuitively, nested case-control methods gain power by over-sampling from controls with longer follow-ups but such gain would be lost when censoring and entry times are the same across controls. In order to explore the effect of censoring type, we repeated the simulation study this time with fixed censoring times, which were chosen so that the proportion of failure events was 15% in the full cohort. The β_1 assumed the values of 0, 0.1, 0.2... 0.8. The log hazard ratio for X_2 was set as $\beta_2=0.5$. While the standard case-cohort methods, Prentice [2], Self-Prentice [3], and Barlow [6], were all more powerful than the conditional logistic method for nested case-control designs across different values of β_1 , the inverse probability weighting methods were still more efficient than other methods within each design (Table 3; Only when $N=500$, $m=2$, $\beta_1=0.5$ is shown). Interestingly, there was little difference in efficiency between two designs when inverse probability weighting methods were used.

Discussion

For both designs, inverse probability weighting methods were more powerful than the standard methods. This is because they use more data, namely the non-subcohort cases in earlier risk sets, to estimate the covariate contribution in the pseudo-likelihood. However, the difference became negligible when the proportion of failure events was very low (<1%).

The comparison between two designs depended on the censoring types and incidence proportion. With random censoring, nested case-control designs coupled with the inverse

probability weighting method proposed by Samuelsen [7] showed the highest statistical power among all methods for both designs. With fixed censoring times, there was little difference in efficiency between two designs when inverse probability weighting methods were used, however, the standard case-cohort methods, Prentice [2], Self-Prentice [3], and Barlow [6], were often more powerful than the conditional logistic method for nested case-control designs. As the proportion of failure events became smaller (<10%), both nested case-control methods outperformed all case-cohort methods and the choice of analytic methods within each type of design became less important. When the predictor was binary, again, the standard case-cohort were often more powerful than the conditional logistic method for nested case-control designs.

This explains the discrepancy between the reports by Langholz & Thomas [4] and Barlow et al [5]: the former report was based on random censoring and the latter report was based on type 2 censoring. It shows that the answer to which design and method is more efficient is more nuanced than what was previously reported. In addition to what we investigated, the relative performance may also depend on the type of left truncation, and the degree of stratification.

Furthermore, statistical power is not the sole determinant of choosing study designs or analytical methods. For example, the inverse probability methods require the retrospective access to the outcome and the matching variables of the full cohort in order to compute inclusion probabilities. Moreover, there are model spaces under which certain designs and analyses are invalid. In particular, the inverse probability weighting methods are invalid when the cohort is “finely stratified” (i.e., the number of strata increases with sample size) since the methods require the consistency of the weights. Goldstein and Zhang [20] proved that when highly stratified cohorts are followed over a short period, conditional logistic approach achieves optimal efficiency under the stratified proportional hazards model. In other cases, case-cohort designs has been preferred for the ease in designing and analyzing a follow-up study with respect to new secondary outcomes [21].

As the new analytical methods are being developed, the conclusion about the relative performance may still change. For example, Chen [22, 23] improved inverse probability weighting methods for case-cohort and nested case-control designs by refining weights by averaging the observed covariates from subjects with similar failure times to estimate contribution from unselected controls. The relative performance would depend on the correlation between failure times and covariates for local averaging. In addition, there have been developments to non-parametrically model predictors of interest conditioned on other available covariates. In some important situations, these maximum semiparametric likelihood has shown to increase the efficiency [24, 25].

Left truncation, or delayed entries, must be accounted in the pseudo/partial likelihoods to avoid bias in estimating hazard ratios under proportional hazards model. When truncation and failure times are independent, the full cohort likelihoods can be adjusted by excluding subjects from the risk sets corresponding to the times before their enrollment. In nested case-control studies, controls are selected only from those at risk that are already enrolled in the

studies. Small sample properties of the case-cohort and the nested case-control methods under varying degrees of left-truncation have not been reported yet.

Staggered entries, like random right censorings, may favor nested case-control methods over case-cohort methods because the former gain power by over-sampling from controls with longer follow-ups especially when the incidence proportion is not very small.

For recurrent events, proportional intensity models have been proposed based on total time of follow-up [13, 26] or gap time [26] while accounting for dependency among the repeated measurements with a marginal ‘working independence’ variance. An alternate approach is to model dependency by random effects. A modified nested case-control sampling strategy for recurrent events were proposed by Lubin [27], along with the usual conditional logistic likelihood, as an extension of recurrent events method by Prentice et al [26]. In case-cohort studies, Zhang et al [28] extended Andersen and Gill’s model [13] using the inverse probability weighting method and Chen & Chen [29] modeled the impact of earlier events on the subsequent events using Prentice’s approach for case-cohort studies [2].

The proportional hazards assumption can have substantial importance [30]. For example, the effects on breast cancer metastases of both higher tumor grades and negative hormone receptor diminished over time [30, 31]. Recently, methods have been developed to assess proportional hazards assumption in case-cohort and nested case-control studies. For example, correlation tests between Schoenfeld residuals and event time were extended to case-cohort studies [30] and a goodness-of-fit test based on inverse probability weighting method was developed for nested case-control studies [32]. Sometimes, the violation of proportional hazards assumption can be remedied through time-invariance predictors. While it is out of scope of this paper, we suspect the reported relative performances may differ with time-varying predictors.

Informative censoring can cause bias when, for example, those who do well drop out from treatment group and those who do worse drop out from control group [33]. In full cohort studies, informative censoring has been modeled explicitly, for example, via the relationship of hazard functions [34] or survival functions [35] between censored and uncensored subjects in proportional hazards regression models. Some used data collected after censoring to address the issue [36]. These methods have not been extended to case-cohort or nested case-control study designs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health Grants 1UL1RR025750-01, P30 CA01330-35; and the National Research Foundation of Korea Grant NRF-2012-. The author is deeply thankful for the constructive comments from the anonymous referees, which led to significant improvement of this work.

References

1. Thomas D. Addendum to 'Methods of cohort analysis: Appraisal by application to asbestos mining' by Liddell FDK, McDonald JC, Thomas DC. *Journal of the Royal Statistical Society*. 1977; A 140:469–91.
2. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986; 73:1–11.
3. Self SG, Prentice RL. Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics*. 1988; 16:64–81.
4. Langholz B, Thomas D. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. *American Journal of Epidemiology*. 1990; 131:169–76. [PubMed: 2403467]
5. Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of Case-Cohort Designs. *Journal of Clinical Epidemiology*. 1999; 52:1165–72. [PubMed: 10580779]
6. Barlow WE. Robust variance estimation for the case-cohort design. *Biometrics*. 1994; 50:1064–72. [PubMed: 7786988]
7. Samuelsen S. A pseudo-likelihood approach to analysis of nested case-control studies. *Biometrika*. 1997; 84:379–94.
8. Kim RS, Kaplan R. Analysis of Secondary Outcomes in Nested Case-Control Study Designs. *Statistics in Medicine*. 2014 In press. 10.1002/sim.6231
9. Kim RS. Analysis of Nested Case-Control Study Designs: Revisiting the Inverse Probability Weighting Method. *Communications for Statistical Applications and Methods*. 2013; 20:455–66.
10. Lin DY, Ying Z. Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association*. 1993; 88:1341–9.
11. Binder DA. Fitting Cox's proportional hazards models from survey data. *Biometrika*. 1992; 79:139–47.
12. Lin DY. On fitting Cox's proportional hazards models to survey data. *Biometrika*. 2000; 87:37–47.
13. Anderson PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. *Annals of Statistics*. 1982; 10:1100–20.
14. Borgan O, Goldstein L, Langholz B. Methods for the Analysis of Sampled Cohort Data in the Cox Proportional Hazards Model. *Annals of Statistics*. 1995; 23:1749–78.
15. Therneau TM, Li H. Computing the Cox model for case cohort designs. *Lifetime Data Analysis*. 1999; 5:99–112. [PubMed: 10408179]
16. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*. 1989; 84:1074–8.
17. R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: 2010.
18. R code: six case-cohort and two nested case-control methods.
19. Zhang H, Goldstein L. Information and asymptotic efficiency of the case-cohort sampling design in Cox's regression model. *Journal of Multivariate Analysis*. 2003; 85:292–317.
20. Goldstein L, Zhang H. Efficiency of the maximum partial likelihood estimator for nested case control sampling. *Bernoulli*. 2009; 15:569–97.
21. Wacholder J. Practical Considerations in Choosing between the Case-Cohort and NCC Designs. *Epidemiology*. 1991; 2:155–8. [PubMed: 1932316]
22. Chen KN. Generalized case-cohort sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2001; 63:791–809.
23. Chen KN. Statistical estimation in the proportional hazards model with risk set sampling. *Annals of Statistics*. 2004; 32:1513–32.
24. Chen HY. Double-Semiparametric Method for Missing Covariates in Cox Regression Models. *Journal of American Statistical Association*. 2002; 97:565–76.
25. Scheike TH, Juul A. Maximum likelihood estimation for Cox's regression model under nested case-control sampling. *Biostatistics*. 2004; 5:193–206. [PubMed: 15054025]
26. Prentice RL, Williams BJ, Peterson AV. On the Regression Analysis of Multivariate Failure Time Data. *Biometrika*. 1981; 68:373–9.

27. Lubin JH. Case-control methods in the presence of multiple failure times and competing risks. *Biometrics*. 1985; 41:49–54. [PubMed: 4005386]
28. Zhang H, Schaubel DE, DKJ. Proportional Hazards Regression for the Analysis of Clustered Survival Data from Case–Cohort Studies. *Biometrics*. 2011; 67:18–28. [PubMed: 20560939]
29. Chen F, Chen K. Case-cohort analysis of clusters of recurrent events. *Lifetime Data Analysis*. 2014; 20:1–15. [PubMed: 23832308]
30. Xue X, Xie X, Gunter M, Rohan TE, Wassertheil-Smoller S, Ho GY, et al. Testing the proportional hazards assumption in case-cohort analysis. *BMC Medical Research Methodology*. 2013; 13:1–10. [PubMed: 23297754]
31. Bellera C, MacGrogan G, Debled M, de Lara C, Brouste V, Mathoulin-Pelissier S. Variables with time-varying effects and the cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Medical Research Methodology*. 2010; 10:1–12. [PubMed: 20053272]
32. Lu W, Liu M, Chen Y-H. Testing Goodness-of-Fit for the Proportional Hazards Model based on Nested Case–Control Data. *Biometrics*. 2014; 10.1111/biom.12239
33. Ranganathan P, Pramesh CS. Censoring in survival analysis: Potential for bias. *Perspectives in Clinical Research*. 2012; 3:40. [PubMed: 22347702]
34. Meier, EN. Master’s thesis. University of Washington; 2012. A Sensitivity Analysis for Clinical Trials with Informatively Censored Survival Endpoints.
35. Braekers R, Veraverbeke N. Cox’s Regression Model Under Partially Informative Censoring. *Communications in Statistics - Theory and Methods*. 2005; 34:1793–811.
36. Lin DY, Robins JM, Wei LJ. Comparing two failure time distributions on the presence of dependent censoring. *Biometrika*. 1996; 83:381–93.

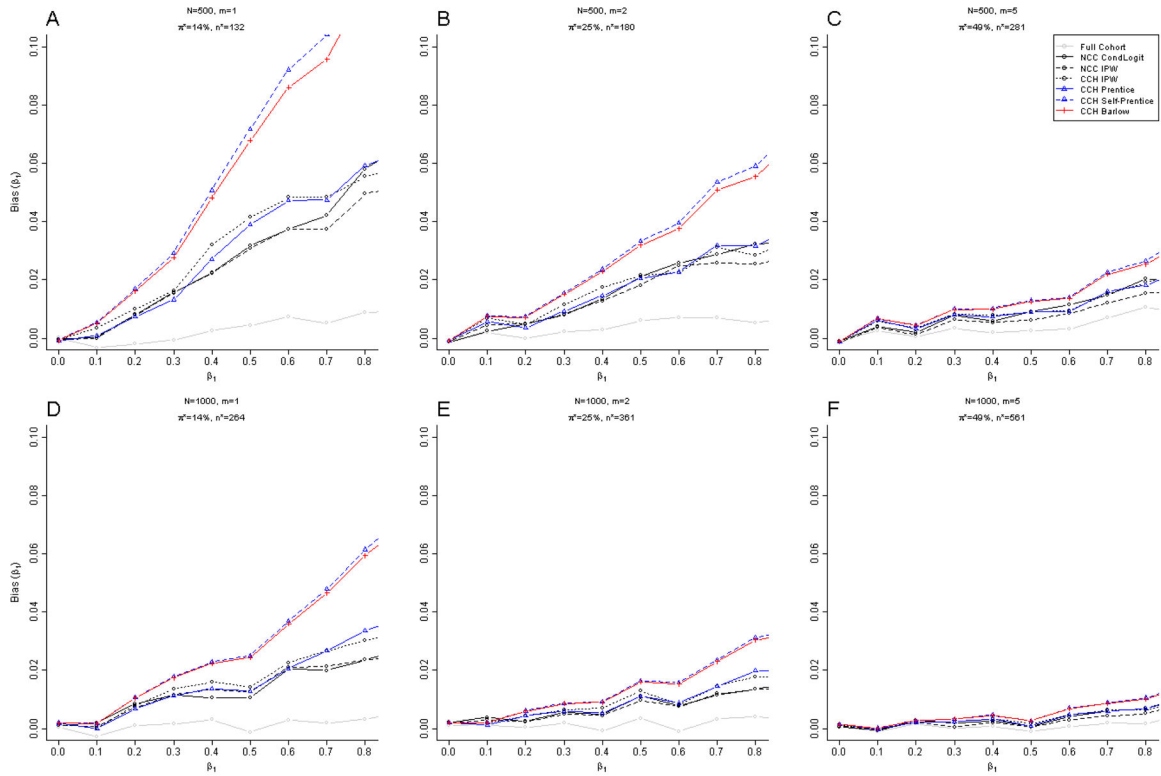


Fig. 1. The Empirical Biases of the Estimators of β_1

The considered methods are the full cohort analysis, two nested case-control (NCC) methods, which are the conditional logistic approach by Thomas (1977), and the inverse probability weighting method by Samuelsen (1997), and four case-cohort (CCH) methods, which are the inverse probability weighting method by Binder (1992), and the methods by Prentice (1986), Self & Prentice (1988), and Barlow (1994). The average sample size n^* and the average subcohort proportion π^* are shown in the titles. Only the results for $N=500$, 1,000 are shown. See Web Figure 4 for the result when $N=1,500$.

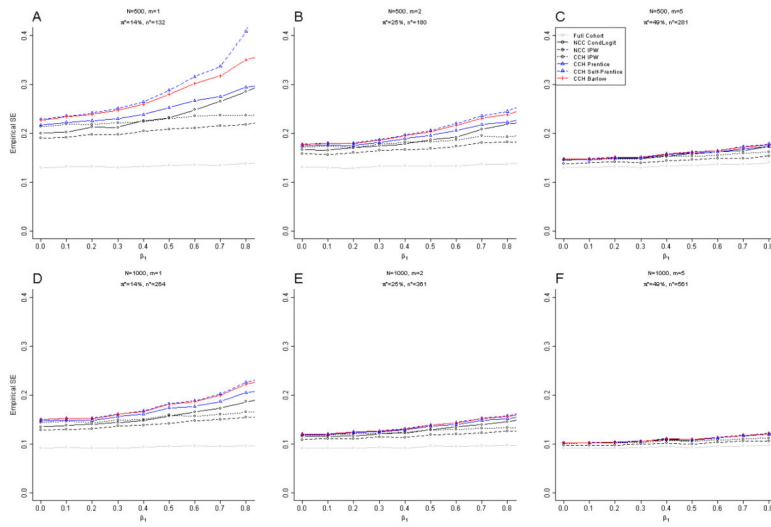


Fig. 2. The Empirical Standard Errors of the Estimators of β_1
 The empirical standard errors of β_1 estimators are shown for the full cohort analysis, two nested case-control (NCC) methods, which are the conditional logistic approach by Thomas (1977) and the inverse probability weighting method by Samuelsen (1997), and four case-cohort (CCH) methods, which are the inverse probability weighting method by Binder (1992), and the methods by Prentice (1986), Self & Prentice (1988), and Barlow (1994). The average sample size n^* and the average subcohort proportion π^* are shown in the titles. Only the results for $N=500, 1,000$ are shown. See Web Figure 4 for the result when $N=1,500$.

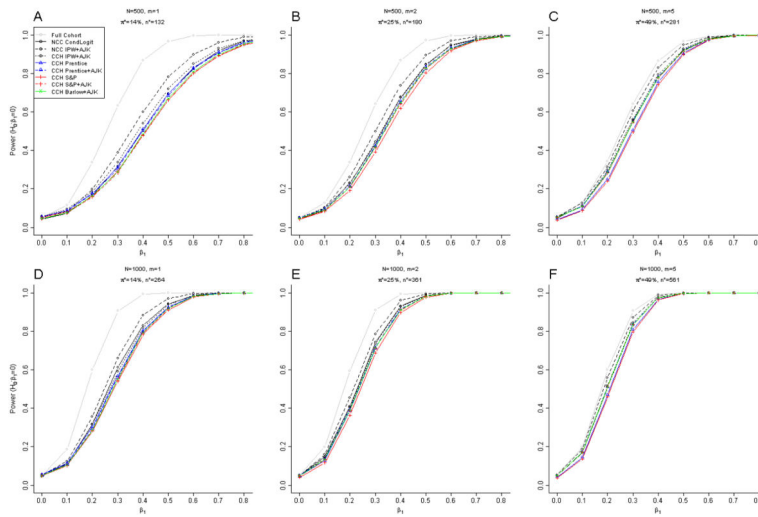


Fig. 3. Empirical Power Testing $H_0:\beta_1=0$ by Methods

The nominal type 1 error rate was 0.05. The empirical power of nine methods is measured: full cohort analysis, the conditional logistic approach by Thomas (1997), inverse probability weighting methods by Samuelsen (1997) coupled with approximate jackknife (AJK) variance estimator (Kim 2013a), the inverse probability weighting methods by Binder (1992) coupled with AJK variance estimator, Prentice (1986), Prentice (1986) coupled with AJK variance estimator, Self & Prentice (1988), Self & Prentice coupled with AJK variance estimator (i.e. Lin & Ying 1993), and Barlow (1993). The average sample size n^* and the average subcohort proportion π^* are shown in the titles. Only the results for $N=500, 1,000$ are shown. CCH and NCC are abbreviations for case-cohort and nested case-control designs, respectively. See Web Figure 4 for the result when $N=1,500$.

Table 1

Denominator Weights in the Pseudo/Partial-likelihood and Variance Estimators

I. Denominator Weights in the Pseudo/Partial-likelihood						
Outcome type and timing	Design	Failure after t_i and outside C	Failure at t_i and outside C	Failure after t_i and in C	Failure at t_i and in C	Non-failure in C
Prentice (1986)	CCH	0	1	1	1	1
Self, Prentice (1988)	CCH	0	0	1	1	1
Barlow (1994)*	CCH	0	1	$1/\pi$	1	$1/\pi$
Binder (1992)	CCH	1	1	1	1	$1/\pi$
Thomas (1977)	NCC	0	1	$1(j \text{ in } C_j)$	1	$1(j \text{ in } C_j)$
Samuelsen (1997)	NCC	1	1	1	1	$1/p_j$

II. Variance Estimators of the Methods

Methods for β estimation	Design	Type A Variance Estimators	Type B Variance Estimators
Prentice (1986)	CCH	Prentice (1986)*, Self, Prentice (1988)	Included in the simulation
Self, Prentice (1988)	CCH	Self, Prentice (1988)	Lin, Ying (1993)
Barlow (1994)	CCH	-	Barlow (1994)
Binder (1992)	CCH	Lin (2000)*	Included in the simulation
Thomas (1977)	NCC	Thomas (1977)	Not included
Samuelsen (1997)	NCC	Samuelsen (1997)*	Kim (2013)

In the top section, the denominator weights in the pseudo-likelihood for Prentice (1986), Self & Prentice (1988), and Barlow (1994) are reported by Barlow et al (1999). Binder (1992) and Samuelsen (1997) are inverse probability weighting (IPW) methods. C is the union of all controls. p_j is the probability of j^{th} subject to be included in a nested case-control sample and π is the subcohort sampling proportion for case-cohort designs.

* Barlow (1994) originally proposed using $\pi(t_i) = r_{sc}(t_i)/rch(t_i)$ where $r_{sc}(t_i)$ is the number of subjects at risk among the subcohort at t_i and $r_{ch}(t_i)$ is the number of subjects at risk in the cohort at t_i but also suggested approximating it by π . Later Barlow et al. (1999) used only π . $1(j \text{ in } C_j)=1$ if j^{th} subject is among C_j and 0 otherwise. In the bottom section, Type A variance estimators converge to usual Cox model variance as sample size increases to the size of full cohorts. Type B variance estimators converge to approximate jackknife estimators (or 'robust' estimators by Lin & Wei, 1989).

* We did not calculate the variance estimators in the original form of Prentice (1986), Samuelsen (1997), or Lin (2000) in the simulation studies. CCH and NCC are abbreviations for case-cohort and nested case-control designs, respectively.

Table 2

Empirical Type 1 Error Testing $H_0: \beta_1=0$

	N=500			N=1,000			N=1,500		
	m=1	m=2	m=5	m=1	m=2	m=5	m=1	m=2	m=5
Full Cohort	.048	.052	.052	.050	.050	.054	.049	.051	.051
NCC CondLogit	.044	.045	.049	.045	.049	.051	.046	.051	.048
NCC IPW+AJK	.054	.052	.053	.051	.049	.053	.049	.049	.051
CCH IPW+AJK	.059	.051	.055	.056	.051	.051	.050	.050	.049
CCH Prentice	.056	.048	.041	.057	.047	.040	.052	.048	.037
CCH Prentice+AJK	.053	.049	.054	.052	.050	.051	.048	.051	.050
CCH Self-Prentice	.051	.042	.039	.050	.041	.038	.046	.041	.035
CCH LinYing	.046	.046	.052	.049	.049	.050	.047	.050	.049
CCH Barlow	.048	.047	.053	.050	.049	.051	.047	.050	.049

The rows from the top to bottom are type 1 errors by full cohort analysis, conditional logistics approach of Thomas (1977) for nested case-control (NCC) designs, inverse probability weighting methods (IPW) by Samuelsen (1997) coupled with approximate jackknife (AJK) variance estimator, IPW by Binder (1992) coupled with AJK variance estimator for case-cohort (CCH) designs, Prentice (1986), Prentice (1986) coupled with AJK variance estimator, Self & Prentice (1988), Self & Prentice (1988) coupled with AJK variance estimator (i.e. Lin & Ying 1993), and Barlow (1994). CCH and NCC are abbreviations for case-cohort and nested case-control designs, respectively.

Table 3

Empirical Power for $m=2$

Proportion of failure	15%		15%		10%		5%		1%		0.1%	
	Random	Fixed	Random	Fixed	Random	Fixed	Random	Fixed	Random	Fixed	Random	Fixed
Censoring type	500	500	500	500	750	750	1500	1500	7500	7500	75000	75000
N	500	500	500	500	750	750	1500	1500	7500	7500	75000	75000
True β ($H_0: \beta=0$) [†]	$\beta_1=0.5$		$\beta_2=1.0$ §		$\beta_1=0.5$		$\beta_1=0.5$		$\beta_1=0.5$		$\beta_1=0.5$	
Full Cohort	.971(1.08)	.971(1.11)	.951(1.11)	.970(1.09)	.968(1.10)	.972(1.12)	.972(1.12)	.988(1.09)	.991(1.08)			
NCC CondLogit	.845(.94)	.797(.93)	.856(.96)	.842(.96)	.844(.98)	.844(.98)	.844(.98)	.906(1.00)	.918(1.00)			
NCC IPW+AJK	.895(1) [‡]	.858(1)	.891(1)	.891(1)	.879(1)	.865(1)	.865(1)	.906(1)	.917(1)			
CCH IPW+AJK	.849(.95)	.804(.94)	.887(1)	.887(1)	.824(.94)	.820(.95)	.820(.95)	.869(.96)	.873(.95)			
CCH Prentice	.826(.92)	.782(.91)	.873(.98)	.873(.98)	.811(.92)	.813(.94)	.813(.94)	.870(.96)	.877(.96)			
CCH Prentice+AJK	.836(.93)	.796(.93)	.880(.99)	.880(.99)	.816(.93)	.822(.95)	.822(.95)	.873(.96)	.879(.96)			
CCH Self-Prentice	.804(.90)	.764(.89)	.858(.96)	.858(.96)	.793(.90)	.800(.92)	.800(.92)	.867(.96)	.878(.96)			
CCH LinYing	.826(.92)	.785(.91)	.878(.99)	.878(.99)	.805(.92)	.809(.94)	.809(.94)	.866(.96)	.872(.95)			
CCH Barlow	.829(.93)	.790(.92)	.878(.99)	.878(.99)	.807(.92)	.811(.94)	.811(.94)	.866(.96)	.872(.95)			

[†] For each setting, the table shows the power only for the smallest β where at least one method had the power greater than .8.

[‡] Also in the parentheses are the ratios of the power compared to that of the NCC IPW+AJK method for each setting.

[§] See the main text about the conditional logistic approach being the least stable method for the binary predictor when the relative risk is large ($\beta > 1$) due to sparse samples.

The methods from the top to bottom are: full cohort analysis, conditional logistcs approach of Thomas (1977) for nested case-control (NCC) designs, inverse probability weighting methods (IPW) by Samuelsen (1997) coupled with approximate jackknife (AJK) variance estimator, IPW by Binder (1992) coupled with AJK variance estimator for case-cohort (CCH) designs, Prentice (1986), Prentice (1986) coupled with AJK variance estimator, Self & Prentice (1988), Self & Prentice (1988) coupled with AJK variance estimator (i.e. Lin & Ying 1993), and Barlow (1994). CCH and NCC are abbreviations for case-cohort and nested case-control designs, respectively.