Published in final edited form as:

Biometrics. 2014 December ; 70(4): 845–851. doi:10.1111/biom.12239.

Testing Goodness-of-Fit for the Proportional Hazards Model based on Nested Case-Control Data

Wenbin Lu^{1,*}, Mengling Liu², and Yi-Hau Chen³

¹Department of Statistics, North Carolina State University, Raleigh, NC, U.S.A.

²Departments of Population Health and Environmental Medicine, New York University School of Medicine, New York, NY, U.S.A.

³Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan, R.O.C.

Summary

Nested case-control sampling is a popular design for large epidemiological cohort studies due to its cost effectiveness. A number of methods have been developed for the estimation of the proportional hazards model with nested case-control data; however, the evaluation of modeling assumption is less attended. In this paper, we propose a class of goodness-of-fit test statistics for testing the proportional hazards assumption based on nested case-control data. The test statistics are constructed based on asymptotically mean-zero processes derived from Samuelsen's maximum pseudo-likelihood estimation method. In addition, we develop an innovative resampling scheme to approximate the asymptotic distribution of the test statistics while accounting for the dependent sampling scheme of nested case-control design. Numerical studies are conducted to evaluate the performance of our proposed approach, and an application to the Wilms' Tumor Study is given to illustrate the methodology.

Keywords

Goodness-of-fit test; Nested case-control sampling; Proportional hazards model; Pseudo-likelihood estimation; Resampling method

1. Introduction

The nested case-control design (Thomas, 1977) is a cost-effective sampling method for large epidemiological cohort studies with survival outcomes. In a nested case-control study, covariate information is ascertained only for cases, i.e., subjects with the event of interest, and for a fixed number of controls who are randomly selected from the at-risk set at each case failure time excluding the case itself. A number of methods have been proposed for analyzing nested case-control data under the proportional hazards model (Cox, 1972). For example, Thomas (1977) considered a maximum partial likelihood estimation method,

*lu@stat.ncsu.edu.

Supplementary Materials

The Web Appendix referenced in Section 3 and the R codes for analysis of the Wilms' Tumor Study data are available with this paper at the Biometrics website on Wiley Online Library.

which was further studied by Oakes (1981). The large-sample properties of Thomas' estimator were investigated by Goldstein and Langholz (1992) and Borgan et al. (1995) using the martingale theory. Chen (2004) improved the efficiency of Thomas's estimator by using a kernel-weighted local averaging method. In addition, Samuelsen (1997) proposed a maximum pseudo-likelihood estimation method by weighting the subjects in the nested case-control sample with the inverse of their selection probabilities. In general, Samuelsen's estimator is more efficient than Thomas' estimator because the former utilizes cases and controls in the pseudo-likelihood whenever they are at risk, while the latter has a case and its associated controls present only at the event-time of the case. Other likelihood-based approaches, including Chen and Little (1999), Scheike and Juul (2004), Zeng et al. (2006) and Saarela et al. (2008), usually require nonparametric estimation of the distribution for covariates measured only in nested case-control samples, which can be challenging for multivariate continuous covariates. Recently, Lu and Liu (2012) proposed a weighted estimating equation approach for a general class of linear transformation models (Clayton and Cuzick, 1985) with nested case-control data.

For cohort studies, many diagnostic tools have been developed to evaluate the proportional hazards model assumption. One popular method, proposed by Lin et al. (1993), is to use martingale-based residual plots and processes, which can provide both graphical and statistical diagnostics for the functional form of the covariates and the proportional hazards assumption. A computationally convenient resampling method was also developed for computing the critical values of the test statistics. Such techniques were further extended to test the marginal proportional hazards model for correlated survival data (Spiekerman and Lin, 1996). However, goodness-of-fit tests for the proportional hazards model based on nested case-control data are less attended. Borgan and Langholz (2007) showed how to define martingale residual processes for nested case-control data and proposed a goodness-of-fit test based on grouped martingale residuals. A numerically easy procedure for computing the critical value of the corresponding supremum type test statistic has not been put forward. One difficulty is that covariate information is only available for the subjects in nested case-control samples and thus the martingale-based residuals are not directly constructible due to the incompleteness of data. Another challenge is that the nested case-control sampling scheme is a stochastic process that introduces dependence between the selection status of the cohort subjects, even in the asymptotic sense (Samuelsen, 1997). Therefore, the resampling methods for full-cohort data considered by Lin et al. (1993) and Spiekerman and Lin (1996) cannot be used to obtain the critical values of the tests since they do not accommodate the dependence induced by the nested case-control sampling. As noted in Cai and Zheng (2012), the standard bootstrap variance estimation method generally lacks a theoretical justification under the nested case-control sampling.

In this paper, based on Samuelsen's maximum pseudo-likelihood estimation method, we propose a class of goodness-of-fit statistics for testing the proportional hazards model with nested case-control data. To tackle the challenges in estimating the asymptotic null distributions for the test statistics, a new resampling scheme that properly accounts for the dependence in the asymptotic representation of the test statistics is developed to compute the critical values.

2. Data, Model and Estimation

Consider a full cohort of n study subjects. Let T_i and C_i denote the failure time and censoring time of subject i , respectively. In addition, let Z_i denote the p -dimensional vector of covariates. The survival information of the full cohort consists of

$\mathcal{F} \equiv \{(\tilde{T}_i, \delta_i), i=1, \dots, n\}$, where $\tilde{T}_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. We assume a proportional hazards model (Cox, 1972) for the failure times,

$$\lambda(t|Z_i) = \lambda_0(t) \exp(\beta_0' Z_i), \quad (1)$$

where $\lambda(t|Z_i)$ is the conditional hazard function for subject i , $\lambda_0(t)$ is the baseline hazard function and β_0 is a p -dimensional vector of regression coefficients. It is assumed that T_i and C_i are independent given Z_i . In a nested case-control design, suppose that m controls are randomly sampled from the at-risk set at each case failure time, excluding the failed subject itself. Then, the covariates are ascertained only for the cases and sampled controls. Let ξ_i be the indicator denoting whether subject i is ever selected as a control. As shown by Samuelsen (1997), we have

$$p_{0i} \equiv \text{pr}(\xi_i=1|\mathcal{F}) = 1 - \prod_{\tilde{T}_j < \tilde{T}_i} \left\{ 1 - \frac{m\delta_j}{\sum_{k=1}^n Y_k(\tilde{T}_j) - 1} \right\},$$

where $Y_k(t) = I(\tilde{T}_k \geq t)$. Define $N_i(t) = \delta_i I(\tilde{T}_i \leq t)$ and $\pi_i = \delta_i + (1 - \delta_i)\xi_i/p_{0i}$. Then, Samuelsen's maximum pseudo-likelihood estimator for β_0 is the solution to

$$\sum_{i=1}^n \int_0^\tau \left\{ \pi_i Z_i - \frac{\sum_{j=1}^n \pi_j Z_j e^{\beta' Z_j} Y_j(t)}{\sum_{j=1}^n \pi_j e^{\beta' Z_j} Y_j(t)} \right\} dN_i(t) = 0,$$

where τ is the maximum follow-up length. Moreover, given β , the estimator of the cumulative baseline hazard function, $\Lambda_0(t) \equiv \int_0^t \lambda_0(s) ds$, is

$$\hat{\Lambda}(t, \beta) = \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\sum_{j=1}^n \pi_j e^{\beta' Z_j} Y_j(s)}, \quad 0 \leq t \leq \tau.$$

Let $\hat{\beta}$ and $\hat{\Lambda}(\cdot) = \hat{\Lambda}(\cdot, \hat{\beta})$ denote the resulting estimators for β_0 and Λ_0 , respectively.

3. Goodness-of-fit Tests

Define the conventional martingale process $M_i(t) = N_i(t) - \int_0^t Y_i(s) e^{\beta_0' Z_i} d\Lambda_0(s)$. The estimated martingale residual process is given by $\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s) e^{\hat{\beta}' Z_i} d\hat{\Lambda}(s)$. With nested case-control data, $\hat{M}_i(\cdot)$ is constructible only for the cases and sampled

controls. To check the proportional hazards model assumption, we consider the following weighted summation of martingale residual processes:

$$W_n(t, z) = n^{-1/2} \sum_{i=1}^n \pi_i f(Z_i) I(Z_i \leq z) \hat{M}_i(t), \quad (2)$$

where $f(z)$ is a pre-specified bounded function of covariates, and $I(Z_i \leq z) = I(Z_{i1} \leq z_1, \dots, Z_{ip} \leq z_p)$ with $Z_i = (Z_{i1}, \dots, Z_{ip})'$ and $z = (z_1, \dots, z_p)'$. Given that $\hat{\beta}$ and $\hat{\Lambda}$ are consistent (Samuelsen, 1997) and $E(\pi_i | \mathcal{F}) = 1$, the process $W_n(t, z)$ is an asymptotically mean-zero process.

The proposed test process is different from the martingale residual processes considered by Lin et al. (1993) for checking the proportional hazards model assumption based on the full cohort data in two aspects. First, because the weights, π_i , depend on future full-cohort data information, process $W_n(t, z)$ is no longer a martingale process even when replacing $\hat{M}_i(t)$ by $M_i(t)$. Second, the individual summation terms in $W_n(t, z)$ are not independent, even asymptotically, since the π_i 's are correlated due to the nested case-control sample scheme. Therefore, it is challenging to derive the inference procedure for test process $W_n(t, z)$.

Define, for $i = 1, \dots, n$,

$$\begin{aligned} W_i^*(t, z) = & f(Z_i) I(Z_i \leq z) M_i(t) \\ & - \int_0^t \frac{H_2(u, z)}{s^{(0)}(u, \beta_0)} dM_i(u) \\ & - \left\{ H_1(t, z) + \int_0^t H_2(s, z) dA(s) \right\}' I^{-1} \int_0^\tau \{Z_i - \mu_Z(t, \beta_0)\} dM_i(t), \end{aligned}$$

where $s^{(k)}(t, \beta) = E\{Z^k e^{\beta' Z} Y(t)\}$, $k = 0, 1, 2$ with $a^0 = 1$, $a^1 = a$ and $a^2 = aa'$ for a vector a ,

$$\mu_Z(t, \beta) = s^{(1)}(t, \beta) / s^{(0)}(t, \beta), \quad H_1(t, z) = E \left\{ f(Z) I(Z \leq z) Z e^{\beta_0' Z} \int_0^t Y(s) d\Lambda_0(s) \right\},$$

$$H_2(t, z) = E \left\{ f(Z) I(Z \leq z) e^{\beta_0' Z} Y(t) \right\}, \quad n(t) = E\{N(t)\}, \text{ and}$$

$$A(t) = \int_0^t \frac{s^{(1)}(u, \beta_0)}{\{s^{(0)}(u, \beta_0)\}^2} dn(u), \quad I = E \left[\int_0^\tau \left\{ \frac{s^{(2)}(t, \beta_0)}{s^{(0)}(u, \beta_0)} - \mu_Z^2(t, \beta_0) \right\} dN(t) \right].$$

It is easy to show that given z , $W_i^*(t, z)$ is a mean-zero martingale process. To establish the asymptotic properties of $W_n(t, z)$, we assume the following regularity conditions:

(C1) The parameters β_0 belong to an interior of a known compact set.

(C2) The covariates z are bounded and if there exists a vector α such that $\alpha' z = 0$ with probability one, then $\alpha = 0$ almost surely.

(C3) The baseline cumulative hazard function satisfies $\Lambda_0(\tau) < \infty$ and $pr(\tilde{T} \geq \tau | z) > 0$ for all possible values of z .

(C4) The information matrix I is positive definite.

Theorem 1: Assume model (1) and that conditions (C1)–(C4) hold. We have

$$W_n(t, z) = n^{-1/2} \sum_{i=1}^n \pi_i W_i^*(t, z) + o_p(1),$$

and given z , $W_n(t, z)$ converges weakly to a mean-zero Gaussian process as $n \rightarrow \infty$.

The proof of Theorem 1 is given in the Supplementary Materials. Similar to Lin et al. (1993), test statistics for the proportional hazards model can be constructed as $\sup_{t,z} |W_n(t, z)|$. For example, for testing the functional form of the j th component of covariates Z , we

consider the test statistic $T_{1,n}^j = \sup_z |W_1^j(z)|$, where

$W_1^j(z) = n^{-1/2} \sum_{i=1}^n \pi_i I(Z_{ij} \leq z) \hat{M}_i(\tau)$ and z is a scalar. For testing the proportional

hazards assumption, we consider $T_{2,n}^j = \sup_t |W_2^j(t)|$, where

$W_2^j(t) = n^{-1/2} \sum_{i=1}^n \pi_i Z_{ij} \hat{M}_i(t)$. In addition, we can also use $T_{2,n}^j = \sup_t |W_2^j(t)|$ to test the proportional hazards assumption for the j th component of the covariates. Here both $W_1^j(z)$ and $W_2^j(t)$ are special cases of $W_n(t, z)$.

To obtain the critical values for the proposed test statistics, we must consistently estimate the asymptotic distribution of $\sup_{t,z} |W_n(t, z)|$. A popular method is to use Monte-Carlo simulations to approximate the asymptotic distribution, e.g., the resampling method considered by Lin et al. (1993) and Spiekerman and Lin (1996). In their methods, the corresponding test statistics can be asymptotically written as the summation of independently and identically distributed mean-zero processes, and each estimated term in the asymptotic representation is perturbed by multiplying an independent standard normal random variable to achieve the same limiting distribution conditional on data. However, such resampling methods cannot be used because they cannot accommodate the correlation among the π_i 's. Here, we propose a new resampling scheme to take into account the dependence induced by the nest case-control sampling. Specifically, we write

$$W_n(t, z) = n^{-1/2} \sum_{i=1}^n W_i^*(t, z) + n^{-1/2} \sum_{i=1}^n (\pi_i - 1) W_i^*(t, z) + o_p(1) \equiv W_{n,1}(t, z) + W_{n,2}(t, z) + o_p(1).$$

Note that $W_{n,1}(t, z)$ and $W_{n,2}(t, z)$ are uncorrelated (Samuelsen, 1997). Thus, we perturb the two terms separately. Let $\hat{W}_i^*(t, z)$ denote the estimate of $W_i^*(t, z)$, where

$$\begin{aligned}
\hat{W}_i^*(t, z) &= f(Z_i) I(Z_i \leq z) \hat{M}_i(t) \\
&\quad - \int_0^t \frac{\hat{H}_2(u, z)}{S^{(0)}(u, \hat{\beta})} d\hat{M}_i(u) \\
&\quad - \left\{ \hat{H}_1(t, z) + \int_0^t \hat{H}_2(s, z) d\hat{A}(s) \right\} \hat{I}^{-1} \int_0^t \left\{ Z_i - \hat{\mu}_Z(t, \hat{\beta}) \right\} d\hat{M}_i(t),
\end{aligned}$$

where $S^{(k)}(t, \beta) = n^{-1} \sum_{i=1}^n \pi_i Z_i^k e^{\beta' Z_i} Y_i(t)$, $k = 0, 1, 2$, $\hat{\mu}_Z(t, \beta) = S^{(1)}(t, \beta) / S^{(0)}(t, \beta)$,

$$\hat{H}_1(t, z) = n^{-1} \sum_{i=1}^n \pi_i f(Z_i) I(Z_i \leq z) Z_i e^{\hat{\beta}' Z_i} \int_0^t Y_i(s) d\hat{\Lambda}(s),$$

$$\hat{H}_2(t, z) = n^{-1} \sum_{i=1}^n \pi_i f(Z_i) I(Z_i \leq z) e^{\hat{\beta}' Z_i} Y_i(t),$$

$$\hat{A}(t) = n^{-1} \sum_{i=1}^n \int_0^t \frac{S^{(1)}(u, \hat{\beta})}{\{S^{(0)}(u, \hat{\beta})\}^2} dN_i(u) \quad \text{and}$$

$$\hat{I} = n^{-1} \sum_{i=1}^n \int_0^t \left\{ \frac{S^{(2)}(t, \hat{\beta})}{S^{(0)}(u, \hat{\beta})} - \hat{\mu}_Z^2(t, \hat{\beta}) \right\} dN_i(t).$$

The first term $W_{n,1}(t, z)$ is a summation of i.i.d. terms that can be perturbed by multiplying independent standard normal random variables. The challenge, however, is that $\hat{W}_i^*(t, z)$ is computable only for those subjects in the nested case-control sample. Therefore, we multiply each perturbed term by $\pi_i^{1/2}$ to ensure its computability and to achieve the same limiting variance. Specifically, we perturb $W_{n,1}(t, z)$ by

$$\widetilde{W}_{n,1}(t, z) = n^{-1/2} \sum_{i=1}^n G_i \pi_i^{1/2} \hat{W}_i^*(t, z),$$

where G_i , $i = 1, \dots, n$ are independent standard normal random variables. Due to the fact that $E(\pi_i | \mathcal{F}) = 1$, it can be shown that the conditional variance of $\widetilde{W}_{n,1}(t, z)$ given the observed data is $n^{-1} \sum_{i=1}^n \left\{ \hat{W}_i^*(t, z) \right\}^2$, which converges to the limiting variance of $W_{n,1}(t, z)$.

The second term $W_{n,2}(t, z)$ is a summation of correlated terms, since π_i and π_j are correlated. To perturb $W_{n,2}(t, z)$, we need to maintain not only the variance of $(\pi_i - 1) W_i^*(t, z)$ but also the covariance structure between $(\pi_i - 1) W_i^*(t, z)$ and $(\pi_j - 1) W_j^*(t, z)$. This cannot be achieved by multiplying independent standard normal random variables. This motivates us to develop a new way to perturb $W_{n,2}(t, z)$ based on the limiting distribution of $W_{n,2}(t, z)$. Specifically, we perturb $W_{n,2}(t, z)$ by

$$\widetilde{W}_{n,2}(t, z) = n^{-1/2} \sum_{i=1}^n G_i^* \frac{(1 - p_{0i})^{1/2}}{p_{0i}} (1 - \delta_i) \xi_i \hat{W}_i^*(t, z),$$

where $(G_1^*, \dots, G_n^*)'$ is a multivariate normal random vector with mean 0 and variance-covariance matrix, $\hat{\Sigma}$ and the G_i^* 's are independent of the G_i 's. Here, the diagonals of $\hat{\Sigma}$ are all 1 and the off-diagonal elements are given by

$$\hat{\sigma}_{ij} = -\frac{m}{n} \frac{\{(1-p_{0i})(1-p_{0j})\}^{1/2}}{p_{0i}p_{0j}} \int_0^\tau Y_i(t) Y_j(t) \frac{n^{-1} \sum_{i=1}^n \pi_i e^{\beta' Z_i} Y_i(t)}{\{n^{-1} \sum_{i=1}^n Y_i(t)\}^2} d\hat{\Lambda}(t), 1 \leq i \neq j \leq n.$$

The choice of $\hat{\Sigma}$ ensures the limiting variance of $\tilde{W}_{n,2}(t, z)$ to be the same as that of $W_{n,2}(t, z)$. The multiplication by the term $(1-\delta_i)\xi_i$ ensures that the perturbed term is computable based on the nested case-control sample. Define $\tilde{W}_n(t, z) = \tilde{W}_{n,1}(t, z) + \tilde{W}_{n,2}(t, z)$. We have

Theorem 2: Assume model (1) and that conditions (C1)–(C4) hold. Given the observed data and ξ_i 's, $\tilde{W}_n(t, z)$ converges weakly to the same mean-zero Gaussian process as that of $W_n(t, z)$ for a given z , as $n \rightarrow \infty$.

The proof of Theorem 2 is given in the Supplementary Materials. Based on Theorem 2, we can compute the critical values of the test statistics based on the empirical distribution of $\sup_{t,z} |\tilde{W}_n(t, z)|$ by generating a large number of the perturbed test processes $\tilde{W}_n(t, z)$. The computation of $\sup_{t,z} |W_n(t, z)|$ and $\sup_{t,z} |\tilde{W}_n(t, z)|$ are done on a common set of gridding points of t or z on its support.

4. Simulations

We conducted simulations to study the performance of our proposed methods for testing the functional form of covariates and the proportional hazards assumption under various scenarios. Specifically, under the null hypothesis, the failure times were generated from the proportional hazards model with $\lambda_0(t) \exp(\beta_{01}Z_1 + \beta_{02}Z_2)$ where Z_1 followed a uniform distribution on $[0, 1]$ and Z_2 followed a Bernoulli distribution with a success probability of 0.5. The true regression coefficients $\beta_0 = (1, -1)'$ and the baseline hazard function $\lambda(t) \equiv \lambda_0$, a positive constant. The censoring times were generated from a uniform distribution on $[0, \tau]$, where λ_0 and τ were chosen to give 10% and 20% failure rates. We considered a full cohort of 1000 subjects and conducted a nested case-control sampling with $m = 1$. Given the proportional hazards model, we tested the functional form of the continuous covariate, i.e., the first covariate in the model, using the following five alternatives: (i) indicator function: the effect of the first covariate is given by $1.5I(Z_1 > 0.5)$; (ii) quadratic function: the effect is given by $-1.5Z_1 + 4Z_1^2$; (iii) log function: the effect is given by $-0.8 \log(Z_1)$; (iv) exponential function: the effect is given by $1.5e^{Z_1}$; (v) sine function: the effect is given by $\sin(2\pi Z_1)$. All other settings were chosen the same as in the null hypothesis. We performed the tests based on proposed statistics $T_{1,n}^j$ for the first covariate at significance levels of 10% and 5%. Under each scenario, we conducted 1000 simulation runs, and the critical values of the test statistics were computed based on 1000 resampled statistics proposed in Section 3, which works well for all the considered scenarios. We also compared with the test of Lin et

al. (1993) for cohort data, which ignores the correlation introduced by the nested case-control sampling, and refer to it as the naive method. The simulation results are given in Table 1, which lists the proportions (in percentage) of rejecting the null hypothesis among 1000 runs at significance levels of 10% and 5%. Based on the results, the estimated size of the proposed test is close to the significance level when the null hypothesis is true. The proposed test also has reasonable power to detect most alternatives except for the exponential function given in (iv), and the power increases as the failure rate increases. For the exponential function alternative, we plotted the function against Z_1 and observed that it was very close to a linear line with the fitted R^2 value being 0.984. Therefore, it is expected that our test would have low powers under such an alternative. On the other hand, the naive test gives conservative type I error rates and thus has smaller power for detecting the alternatives compared with the proposed test. In summary, the power of the proposed test depends on the magnitude of deviation from the linear form and sample size. In practice, if the test indicates significant violation of the null hypothesis of linear form, one could further visualize the functional effect of the covariate using the spline methods.

To test the proportional hazards assumption, we also considered five alternatives. The first two alternatives are for the proportional hazards model with time-dependent covariates. Specifically, for the first alternative, the conditional hazard function is given by $\lambda(t/Z_i) = \lambda_0 \exp\{Z_{i1} - Z_{i2}I(t > 0.7)\}$, whereas for the second alternative, the conditional hazard function is given by $\lambda(t/Z_i) = \lambda_0 \exp[Z_{i1}\{1 + \log(t)\} - Z_{i2}]$. For the remaining three alternatives, the failure times were generated from a class of log-linear models

$$\log(T_i) = \alpha - Z_{i1} + Z_{i2} + \sigma \varepsilon_i,$$

where ε_i was chosen to follow the standard normal distribution and the distributions with the hazard function $e^{u/(1 + re^u)}$ with $r = 1$ or 2. Here, $r = 1$ corresponds to a proportional odds model. Constants α and σ were chosen to give 10% and 20% failure rates. All other settings were the same as in the previous simulations for testing the functional form of the covariate. We used proposed statistic $T_{2,n}$ to test the overall proportional hazards assumption based on 1000 runs. The critical values of the test statistics were computed based on 1000 resamplings. The naive test was also included for comparison. The simulation results are given in Table 2. The findings for the proposed test are very similar to those for testing the functional form of the covariate. However, the naive test has slightly inflated type I error rates, especially for the case with 20% failure rate.

5. An Application to Wilms' Tumor Study

To further illustrate the performance of the proposed goodness-of-fit tests, we applied our method to the data from the Wilms' Tumor Study (D'Angio et al., 1989; Green et al., 1998). The complete data contain 3915 subjects that participated in the third and fourth Wilms' tumor studies. Among them, 669 patients had tumor recurrence with the recurrence rate of 17%. Six covariates were considered: study indicator (denoted by NWTSG-4), 1 for the fourth Wilms' tumor study and 0 for the third one; tumor stage indicators (denoted by stages II–IV), which are three dummy variables corresponding to stages II–IV; unfavorable central histology status (denoted by UH); and initial tumor diameter (denoted by diameter), scaled

between 0 and 1. We conducted a nested case-control sample with $m = 1$. We first computed Samuelsen's maximum pseudo-likelihood estimators based on the proportional hazards model with the above six covariates. The corresponding estimation results are given in Table 3. We found that tumor stage indicators and unfavorable central histology status are significant. Next, we performed the proposed goodness-of-fit tests based on $T_{1,n}^j$ and $T_{2,n}$. For testing the functional form of covariates given the proportional hazards model assumption, we considered the tumor diameter since it is the only continuous covariate. The p -values for tests $T_{1,n}^j$ and $T_{2,n}$ are 0.172 and < 0.001 , respectively, based on 1000 resamplings, which supports the linear functional form for tumor diameter but rejects the overall proportional hazards assumption. As a graphical diagnostic tool, we plotted the corresponding test process, $W_1^j(z)$ for $T_{1,n}^j$, and 25 randomly sampled resampled test processes as a function of z in Figure 1. Furthermore, we tested the proportional hazards assumption for each of the six covariates using the test statistics $T_{2,n}^j$. The corresponding diagnostic plots along with the p -values are given in Figure 2. Based on the p -values of the tests, the covariates study indicator, stage IV indicator, unfavorable central histology status and initial tumor diameter showed significant deviation from the proportional hazards assumption. In addition, for those covariates with significant p -values, the deviance of the observed test processes from the perturbed processes mainly occurred at early time before 3, indicating that the proportional hazards assumption was likely to be violated for at early stage. This may suggest that a Cox model with time-varying coefficients may be more suitable for this data-set.

6. Discussion

The proposed goodness-of-fit tests can be extended to accommodate nested case-control data with left truncation by modifying the definitions of the at-risk and counting processes as in Samuelsen (1997). The tests can also be adapted to test the fit for a class of linear transformation models under nested case-control sampling by using the weighted estimating equation approach studied in Lu and Liu (2012). The proposed resampling approach can numerically approximate the distribution of the proposed test statistics without much computation cost. The number of resampling runs needed for real applications would depend on the variance of the test process and its supremum statistic. Therefore, graphical examinations of the test process and its resampled counterparts as shown in Figures 1 and 2 are helpful in practice to determine the number of resampling iterations. If the resampled test processes have relatively smaller variation compared to the original test process as for tumor diameter shown in Figure 2, it may imply that a small number of resamplings may be sufficient. In addition, the required number of resampling statistics depends on the significance level of the test, the smaller the significance level, the larger the number required. From our numerical experience, for significance level of 5%, 1000 resamplings usually suffice.

In practice, the selection of controls may require matching additional factors. When matching is within categories of reasonably sample sizes such as gender and ethnicity, the selection probability p_{0i} can still be calculated in a similar way as Samuelsen (1997) and the

proposed test and inference approach can readily accommodate the matching. When fine matching is used to select controls such as caliper matching within certain interval of continuous matching variables, Støer and Samuelsen (2012) showed that the inverse selection probability method may lose its efficiency gain over the traditional Thomas' partial likelihood estimator, partially because the small number of available controls yields large sampling probability and small weight, and may fail to reconstruct the full cohort. It is of great practical interest to further study the assessment of proportional hazards assumption with fine-matched nested case-control studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the editor, the associate editor and two referees for their comments that substantially improved the article. W. Lu and M. Liu's work was partially supported by NIH/NCI grants R01 CA140632 and R21 CA169739.

References

- Borgan Ø, Goldstein L, Langholz B. Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Annals of Statistics*. 1995; 23:1749–1778.
- Borgan, Ø.; Langholz, B. Using martingale residuals to assess goodness-of-fit for sampled risk set data. In *Festschrift for Kjell Doksum*. World Scientific Publishing; Singapore: 2007.
- Cai T, Zheng Y. Evaluating prognostic accuracy of biomarkers in nested case-control studies. *Biostatistics*. 2012; 13:89–100. [PubMed: 21856652]
- Chen HY, Little RJA. Proportional hazards regression with missing covariates. *Journal of The American Statistical Association*. 1999; 94:896–908.
- Chen K. Statistical estimation in the proportional hazards model with risk set sampling. *Annals of Statistics*. 2004; 32:1513–1532.
- Clayton DG, Cuzick J. Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society A*. 1985; 48:82–117.
- Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society Series B-Statistical Methodology*. 1972; 34:187–220.
- D'Angio GJ, Breslow N, Beckwith JB, Evans A, Baum H, de Lorimier A, Ferbach D, Hrabovsky E, Jones G, Kelalis P. Treatment of Wilms tumor. Results of the third national Wilms tumor study. *Cancer*. 1989; 64:349–360. [PubMed: 2544249]
- Goldstein L, Langholz B. Asymptotic theory for nested case-control sampling in the Cox regression-model. *Annals of Statistics*. 1992; 20:1903–1928.
- Green DM, Breslow NE, Beckwith JB, Finklestein JZ, Grundy PG, Thomas PRM, Kim T, Shochat S, Haase GM, Ritchey ML, Kelalis PP, D'Angio GJ. Comparison between single-dose and divided-dose administration of dactinomycin and doxorubicin for patients with wilms tumor: A report from the national wilms tumor study group. *Journal of Clinical Oncology*. 1998; 16:237–245. [PubMed: 9440748]
- Lin DY, Wei LJ, Ying ZL. Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*. 1993; 80:557–572.
- Lu W, Liu M. On estimation of linear transformation models with nested case-control sampling. *Lifetime Data Analysis*. 2012; 18:80–93. [PubMed: 21912975]
- Oakes D. Survival times - aspects of partial likelihood. *International Statistical Review*. 1981; 49:235–252.

- Saarela O, Kulathinal S, Arjas E, Läärä E. Nested case-control data utilized for multiple outcomes: a likelihood approach and alternatives. *Statistics in Medicine*. 2008; 27:5991–6008. [PubMed: 18792086]
- Samuelsen SO. A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika*. 1997; 84:379–394.
- Scheike TH, Juul A. Maximum likelihood estimation for Cox's regression model under nested case-control sampling. *Biostatistics*. 2004; 5:193–206. [PubMed: 15054025]
- Spiekerman CF, Lin DY. Checking the marginal cox model for correlated failure time data. *Biometrika*. 1996; 83:143–156.
- Støer NC, Samuelsen SO. Comparison of estimators in nested case-control studies with multiple outcomes. *Lifetime Data Analysis*. 2012; 18:261–283. [PubMed: 22382602]
- Thomas DC. Addendum to “Methods of cohort analysis: Appraisal by application to asbestos mining” by Liddell, F. D. K., McDonald, J. C. and Thomas, D. C. *Journal of the Royal Statistical Society A*. 1977; 140:469–491.
- Zeng D, Lin DY, Avery CL, North KE, Bray MS. Efficient semiparametric estimation of haplotype-disease associations in case-cohort and nested case-control studies. *Biostatistics*. 2006; 7:486–502. [PubMed: 16500923]

Figure 1. Original test process (black color) and 25 randomly sampled resampled test processes (grey color) for testing the functional form of tumor diameter.

Figure 2.

Original test process (black color) and 25 randomly sampled resampled test processes (grey color) for testing the proportional hazards assumption for each covariate.

Table 1
Type I error and power in percentage for testing the function form of the covariate

failure rate	method	null hypothesis		indicator function		quadratic function	
		10%	5%	10%	5%	10%	5%
10%	proposed	11	6	39	25	43	30
	naive	2	1	15	6	30	19
20%	proposed	10	6	74	59	68	57
	naive	2	1	60	37	63	51

		log function		exponential function		sine function	
		10%	5%	10%	5%	10%	5%
10%	proposed	87	78	14	9	60	47
	naive	82	73	8	3	36	18
20%	proposed	98	95	22	13	91	83
	naive	98	93	14	7	81	68

[†] proposed, the proposed tests; naive, the tests that ignore the correlation introduced by the nested case-control sampling.

Table 2
Type I error and power in percentage for testing the proportional hazards model

failure rate	method	null hypothesis		time-dependent I		time-dependent II	
		10%	5%	10%	5%	10%	5%
10%	proposed	10	6	61	47	28	17
	naive	13	8	64	49	34	20
20%	proposed	10	6	73	63	44	29
	naive	14	7	77	70	52	40

		PO model		PO2 model		normal model	
		10%	5%	10%	5%	10%	5%
10%	proposed	24	14	36	24	52	38
	naive	29	18	40	26	50	40
20%	proposed	50	36	84	71	61	44
	naive	55	41	86	73	32	22

[†] proposed, the proposed tests; naive, the tests that ignore the correlation introduced by the nested case-control sampling; time-dependent I and II, the two alternatives corresponding to the proportional hazards model with time-dependent covariates; PO, the proportional odds model; PO2, the model with $r = 2$; normal, the model with the normal error.

Table 3
Estimation results for Wilm's tumor data

covariates	estimates	SE	p-value
NWTSG-4	-0.138	0.116	0.234
stage II	0.358	0.151	0.018
stage III	0.636	0.156	0.000
stage IV	0.991	0.178	0.000
UH	1.280	0.148	0.000
diameter	1.062	0.557	0.057

[†]SE, estimated standard error.