Relaxing the Interaction Test in the Cox Proportional Hazard Model for Epidemiological Studies

Eiji Nakatani[1], Yoko Sato[1]
[1]Research Support Center, Shizuoka General Hospital, 4-27-1, Kitaando, Aoi-ku, Shizuoka, 420-8527, Japan

Abstract
The interaction test in epidemiological studies is used to determine the subclass with a strong therapeutic effect, but one of its practical problems is the less power and the larger sample size required to confirm treatment efficacy, compared with that of a randomized trial. From this perspective, we proposed a test that had enough power to find the subclasses that had significant effects in the Cox proportional hazards model. The estimator was the difference in hazard ratios for a particular factor between the subclass and the whole. The estimator and confidence interval were estimated using the bootstrap method. For the discovery of subclass, the proposed test can be used by epidemiological researchers as an alternative to the traditional interaction test and may enable easy interpretation and powerful subclass determination. We performed simulations that applied our proposed method for the interaction test.

Key Words: Cox proportional hazards model, interaction test, prediction

1. Introduction

The interaction test in epidemiological studies is used to determine the subclass that has a substantial therapeutic effect on a factor of interest, but one of its practical problems is the less power and the larger sample size required to confirm treatment efficacy, compared with that of a randomized trial.

To identify an independent prognostic factor, the consistency of the adverse effects of prognostic factors across different patient subgroups may be assessed. For example, the effect of primary tumor location as a prognostic factor of survival times in metastatic colorectal cancer was assessed in a subgroup analysis stratified by baseline factors.(Loupakis et al., 2015) On the other hand, to explore the impact of dependent prognostic factors, subgroup analyses may be performed as tree analysis.(Hauschild et al., 2018; Kros et al., 2015) However, such attempts have not been many, and this would often make new prognostic factors that are similar but not independent not useful, compared with the effect on conventional prognostic factors. Even in this cases, the prognostic factor may exert a strong impact on a specific subgroup.

To select the optimal subset for the predictor of interest, we expect these covariates to be conditional rather than to contain the interaction term, which is the product of the predictor and another variable used to select the subset. Having a statistical model that gives a good prediction of the outcome is important.(VanderWeele et al., 2019) Intuitively, if the hazard ratio (HR) for the prognostic factor is greater in a specific subgroup than in all cases, it may underscore the adverse effects of the prognostic factor in that subgroup.

Therefore, we proposed a test that had enough power to find the subclass with a significant effect on a predictive factor in the Cox proportional hazards model. The estimator used for the proposed test was the difference in HR between the subpopulation and the whole population.

2. Motivating Example

In the assessment of a new prognostic factor in the presence of several previously reported prognostic factors, discovery of a subset that has a strong adverse effect on this prognostic factor would be needed. This work was motivated by a report on an excessive adverse effect on a subset of a specific prognostic factor in patients diagnosed as stage II colon cancer.(Ueno et al., 2020)

For stage II colon cancer, grade 3 (G3, poorly differentiated) is an essential decision factor for treatment, but no unified diagnostic criteria has been established. According to previous studies, an intratumoral poorly differentiated area with no glandular formation (POR) that encompasses the microscopic field of a 40 × 5 objective lens was an essential factor that defined G3. In a randomized controlled study on adjuvant chemotherapy for stage II colon cancer (i.e., SACURA trial), the optimal criteria for G3 were prospectively validated.

This validation analysis included 991 patients with stage II colon cancer. The intensity and predictive power of poorly differentiated clusters (POR grade) were evaluated. When G3 POR was added to the prognostic model that comprised eight common factors, it was found to be a significant factor for recurrence-free survival (P = 0.040, Wald test). Furthermore, because the prognosis differed between patients who had markedly high microsatellite instability (MSI) and those who had no MSI, the HR (95% CI) for FAS was 1.93 (1.34–2.77) between the G3 and nonG3 groups and 2.61 (1.81–3.77) between the subsets microsatellite stable (MSS) and MSI low. The adverse impact of G3 POR on recurrence-free survival was higher in the MSS/ MSI low subset than in the full analysis set. Even in this case, on the interaction analysis, the adverse effect of G3 POR was not significantly different between the MSS/ MSI low and the MSI high groups (P = 0.094).

3. Comparison of Hazard Ratios between the Subset and All Cases

3.1 Notation and Assumptions

In an observational cohort study setting, the data (T, G, H, Z) for each n $i.i.d.$, observation $i = 1, \cdots, n$, was generated. T was the time to event outcome for prognosis; G was the dichotomous prognostic factor of interest (1/0); and H was the dichotomous variable generated for subset selection (1/0). A stronger adverse effect of the prognostic factor was presented by the subset of h = 1 than by another subset (h = 0). Z was a vector of the other covariates.

3.2 Comparison of Hazard Ratios between the Subpopulation and the Entire Population

In a Cox proportional hazard model, the proportional hazards assumption holds. The HR is time constant, and two hazard functions are proportional. In the proportional hazards model with the interaction term, the hazard was expressed as

$$\gamma(t; G, H, Z) = \gamma_0(t) exp\left(\beta_1 G + \beta_2 H + \beta_3 GH + \sum_{k=1}^{K} \gamma_k Z_k\right),$$

where λ (*t; G, H, Z*) is the hazard function at time *t*, which depended on $G = g$,$H = h$, $Z = z$; $\beta_1, \beta_2, \beta_3, \gamma_1, \cdots, \gamma_K$ were the linear regression parameters. Conventionally, in an interaction test with the null hypothesis, $\beta_3 = 0$ is used.

According to the notation method of Li and Chambless(Li and Chambless, 2007), we denoted $HR(g, h; z)$, in which HR depended on $\mathrm{Z} = z$, and compared with $\mathrm{G} = g$. $\mathrm{H} = h$ with the reference group $\mathrm{G} = 0$, $\mathrm{H} = 0$ was defined as follows:

$$HR(g, h; z) = \frac{\gamma(t; g, h, z)}{\gamma(t; 0, 0, z)} = exp(\beta_1 g + \beta_2 h + \beta_3 gh)$$

The respective strata-specific HRs in the groups $H = 0$ and $H = 1$ were as follows:

$HR_{H=1} = \mathrm{HR}(1,1;\mathrm{z})/\mathrm{HR}(0,1;\mathrm{z}) = \exp(\beta_1 + \beta_3)$ and $HR_{H=0} = \mathrm{HR}(1,0;\mathrm{z})/\mathrm{HR}(0,0;\mathrm{z}) = \exp(\beta_1)$

The HR in all cases was expressed as $\mathrm{HR}_{whole} = \exp(\beta_1) \cdot E_h[\exp(\beta_2 \cdot h)]$.
The HR of the interaction term effect (ITE) was
$\mathrm{HR}_{ITE} = \mathrm{HR}(1,1;\mathrm{z})/\mathrm{HR}(0,0;\mathrm{z}) = \exp(\beta_1 + \beta_2 + \beta_3)$.

The difference in HRs between the subset $H = 1$ and all cases was represented as $HR_{H=1} - HR_{whole} = exp(\beta_1) \cdot \{exp(\beta_3) - E_h[exp(\beta_2 \cdot h)]\}$. In addition, the HR ratio in the subset $H = 1$ and all cases was $HR_{H=1}/HR_{whole} = \frac{exp(\beta_3)}{E_h\,[exp(\beta_2 \cdot h)]}$. Notably,

$HR_{H=1}/HR_{H=0} = exp(\beta_3)$ and $HR_{H=1} - HR_{H=0} = \exp(\beta_1)\{\exp(\beta_3) - 1\}$

The following Uno's concordance index (Uno et al. 2011) was used to evaluate their predictive performance:

$$C_U(h; g, z) = \Pr(lp(; g_1, h_1, z_1) > lp(; g_2, h_2, z_2) | T_1 < T_2, T_1 < \tau),$$

where T_1, T_2 were times to events; $lp(; g, h, z) = \beta_1 g + \beta_2 h + \beta_3 gh + \sum_{k=1}^{n} \gamma_k z_k$. $(g_1, h_1, z_1), (g_2, h_2, z_2)$ were the covariate vectors for pairs within the population; and τ was a specified time point within the support of the censoring variable. The gain of prognostic performance by the models in a subset and all cases were measured as $C_U(1; g, z) - C_U(h; g, z)$ or $C_U(h; g, z) \,/\, C_U(1; g, z)$.

4. Simulation

4.1 Overview and Setting

We performed the simulation that was generated from the motivating example (Ueno et al., 2020). Below is the data-generating mechanism with a constant ITE. We performed this simulation using a library of "coxed" in R software.(Harden and Kropko, 2019) The sample size was 1,000 patients. The maximum period was 110 months, and the probability of censoring was 0.85.

Covariates were followed the distribution as follows: $Z_1 \sim Ber(0.24)$, $Z_2 \sim Ber(0.70)$, $Z_3 \sim Ber(0.17)$, $Z_4 \sim Ber(0.58)$, $Z_5 \sim Ber(0.61)$, $Z_6 \sim Ber(0.50)$, $Z_7 \sim Mulnorm(0.38,\ 0.33)$, and H$\sim Ber(0.90)$, G$\sim Ber(0.15)$. The HRs of ITE ($HR_{ITE}$) were set as 1, 1.5, 2, 2.5, and 3. The HRs of H,G,$Z_1, Z_2, Z_3, Z_4, Z_5, Z_6,$ and Z_7 were 1.66, 1.5, 1.30, 1.35, 2.33, 0.90, 1.16, 0.27, and 0.85, respectively. The linear predictor was $ln(1.66)\,H + ln(1.5)G + ln(\mathrm{HR}_{ITE})GH + ln(1.30)Z_1 + ln(1.35)\,Z_2 + ln(2.33)\,Z_3 + ln(0.90)\,Z_4 + ln(1.16)\,Z_5 + ln(0.27)\,Z_6 + ln(0.85)\,Z_7$. These simulations were performed 1,000 times. The 95% CIs were estimated using the bootstrap method for 10,000 times.

Table 1. Setting of simulation

Variable for subset	New prognostic factor		Strata-specific HR for the prognostic factor	HR for the prognostic factor
	G = 0	G = 1		
H = 0	HR(0,0;z) = 1	HR(1,0;z) = 1.50	1.5	1.5, 2.16, 2.8, 3.42, 4.03
H = 1	HR(0,1;z) = 1.66	HR(1,1;z) = 2.49, 3.74, 4.98, 6.23,7.47 for β_3 = 1, 1.5, 2,2.5, 3	1.5, 2.25, 3, 3.75, 4.5	

4.2 Results

The powers of the interaction test and the difference in HRs or CIs between the subset and the whole population are shown in Table 2 and Figure 1. Compared with the interaction tests, the tests for the differences in HRs for a particular factor between the subclass and all cases had higher power. The tests for the difference among the C-indices had no enough power because of the number of covariates.

Table 2. Powers of the interaction test and the tests by new measures

HR of the interaction term	Power of the interaction test	Difference in HRs			Difference in CIs		
		Mean of HRs		Power for test	Mean of CIs		Power for test
		All cases	Subset		All cases	Subset	
1.0	0.029	1.49	1.50	0.069	0.702	0.699	0.001
1.5	0.021	2.10	2.23	0.124	0.707	0.705	0.002
2.0	0.066	2.66	2.97	0.185	0.712	0.711	0.005
2.5	0.158	3.16	3.69	0.262	0.716	0.716	0.008
3.0	0.242	3.61	4.39	0.342	0.720	0.720	0.013

5. Conclusion

In conclusion, the proposed test, which estimated the differences in HRs for a particular factor between the subclass and all cases, maybe have a higher power, compared with that of the conventional interaction test. Furthermore, the proposed test enabled easy interpretation of the difference between the subgroup effect estimates and the overall effect.

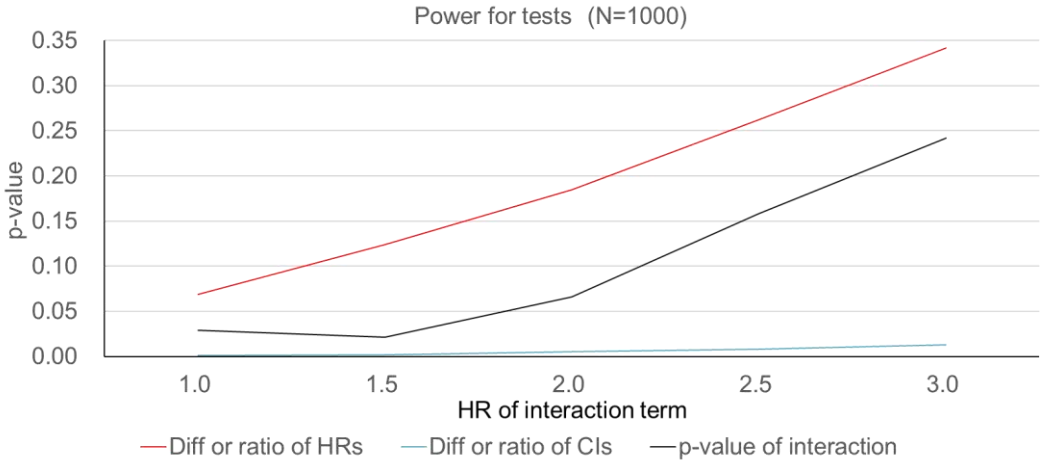

Figure 1. Powers of the interaction test and the tests by new measures

Acknowledgement

This work was supported by a Grant-in-Aid for Young Scientists (B) from the Japan Society for the Promotion of Science (17K18445). The authors would like to thank MARUZEN-YUSHODO Co., Ltd. (https://kw.maruzen.co.jp/kousei-honyaku/) for the English language editing.

References

Harden, J. J., and Kropko, J. (2019). Simulating Duration Data for the Cox Model. *Political Science Research and Methods* **7**, 921-928.

Hauschild, A., Larkin, J., Ribas, A.*, et al.* (2018). Modeled Prognostic Subgroups for Survival and Treatment Outcomes in BRAF V600-Mutated Metastatic Melanoma: Pooled Analysis of 4 Randomized Clinical Trials. *JAMA Oncol* **4**, 1382-1388.

Kros, J. M., Huizer, K., Hernández-Laín, A.*, et al.* (2015). Evidence-Based Diagnostic Algorithm for Glioma: Analysis of the Results of Pathology Panel Review and Molecular Parameters of EORTC 26951 and 26882 Trials. *J Clin Oncol* **33**, 1943-1950.

Li, R., and Chambless, L. (2007). Test for additive interaction in proportional hazards models. *Ann Epidemiol* **17**, 227-236.

Loupakis, F., Yang, D., Yau, L.*, et al.* (2015). Primary tumor location as a prognostic factor in metastatic colorectal cancer. *J Natl Cancer Inst* **107**.

Ueno, H., Ishiguro, M., Nakatani, E.*, et al.* (2020). Optimal Criteria for G3 (Poorly Differentiated) Stage II Colon Cancer: ProspectiveValidation in a Randomised Controlled Study (SACURA trial). *American Journal of Surgical Pathology*.

VanderWeele, T. J., Luedtke, A. R., van der Laan, M. J., and Kessler, R. C. (2019). Selecting Optimal Subgroups for Treatment Using Many Covariates. *Epidemiology* **30**, 334-341.