Concordance Index: Surrogacy of Progression-Free Survival for Overall Survival

Yiwei Fan and Guosheng Yin*

Department of Statistics and Actuarial Science

The University of Hong Kong, Hong Kong

*Corresponding author
Guosheng Yin, PhD
Patrick S C Poon Endowed Professor and Head of Department
Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong
Adjunct Professor
Department of Biostatistics, MD Anderson Cancer Center, Houston, TX, USA
Email: gyin@hku.hk
Tel: 852-3917-8313
Fax: 852-2858-9041

Funding: The research was supported in part by a grant No. 17308420 from the Research Grants Council of Hong Kong.

Conflict of interest: None

Abstract

In oncology clinical trials, the primary endpoint is often time to an event of clinical interest, e.g., time to disease progression or time to death. As a result, progression-free survival (PFS: the time from initiation of treatment till disease progression or death whichever occurs first) and overall survival (OS: the time from initiation of treatment till death) are the focus of statistical analysis in comparison of two treatment arms. It is often argued that PFS may serve as a surrogate endpoint for OS, while the validity of such surrogacy is still under debates in different types of cancer. In practice, one may observe a significant difference in PFS but no significant difference in OS; or vice versa. We provide a concordance index (C-index) to measure the degree of concordance between PFS and OS, and elaborate on the PFS vs OS discrepancies using the C-index using simulation studies and real trial analysis.

Keywords: Concordance index; Oncology trials; Overall survival; Progression-free survival; Surrogate; Survival analysis

1. Introduction

In cancer clinical trials, progression-free survival (PFS) is often used to evaluate the efficacy of therapies to avoid a potentially long follow-up period, which may serve as a surrogate endpoint for overall survival (OS). However, in practice, it is often seen a significant difference in PFS but no difference in OS; or vice versa. The validity of using PFS as a surrogate endpoint for OS is still under debates for different types of cancer.

Broglio and Berry[1] conducted simulation studies to examine the role of survival post progression (SPP) in understanding treatment effects, assuming no difference in SPP between treatment groups. When median SPP is short (e.g., 2 months), OS is probably statistically significant for trials with a PFS benefit. For long median SPP (e.g., 24 months), while not implying a lack of improvement in OS, clinical trials with a PFS benefit may not be able to reach statistical significance in OS. Similar results have also been found in a recent phase III randomized clinical trial of patients with operable triple-negative breast cancer[2]. PFS time was longer in one treatment arm compared with the other, while there was no statistically significant difference in OS between the two treatment arms. In contrast, two phase III clinical trials for metastatic breast cancer and KRAS wild-type metastatic colorectal cancer showed statistically significant efficacy of the treatments in OS, but not in PFS[3].

Various statistical tools have been utilized to test the validity of PFS as a surrogate endpoint of OS, including Spearman's rank correlation[4,5], and copula bivariate models[6]. We provide a concordance index (C-index), a.k.a. Kendall's tau, to measure the association between PFS and OS. We aim to identify the situations in which PFS can serve as an appropriate surrogate endpoint of OS, and the threshold value of C-index to be used in such cases.

2. Method

Let P and D denote the PFS time and OS time, respectively. For individuals i and j, the C-index is defined as the difference between the concordant probability of PFS and OS and the discordant probability[7],

$$\tau = \Pr\{(P_i - P_j)(D_i - D_j) > 0\} - \Pr\{(P_i - P_j)(D_i - D_j) < 0\}.$$

We should take the censoring information into account in survival data, and let C denote the censoring time. The censoring indicators of PFS and OS are denoted by $\Delta^{(P)} = I(P < C)$ and $\Delta^{(D)} = I(D < C)$, respectively. For individuals i and j, define

$$U_{ij}^{(P)} = \begin{cases} 1, & if\ P_i > P_j, \Delta_j^{(P)} = 1, \\ -1, & if\ P_i < P_j, \Delta_i^{(P)} = 1, \end{cases}$$

and

$$U_{ij}^{(D)} = \begin{cases} 1, & if\ D_i > D_j, \Delta_j^{(D)} = 1, \\ -1, & if\ D_i < D_j, \Delta_i^{(D)} = 1. \end{cases}$$

We call a pair (i, j) concordant, if $U_{ij}^{(P)} U_{ij}^{(D)} = 1$; and discordant if $U_{ij}^{(P)} U_{ij}^{(D)} = -1$. Let $R_{ij} = 1$ if the pair (i, j) is orderable, while due to censoring some pairs may not have a definitive order and thus $R_{ij} = 0$. Halabi et al.[8] and Ter-Minassian et al.[9] computed the C-index as[10]

$$\hat{\tau} = \sum_{i<j} R_{ij} U_{ij}^{(P)} U_{ij}^{(D)}.$$

However, this estimator, taking the sum over the orderable pairs only, is biased because of the presence of missing data. To ensure consistency, our estimator of C-index is defined as[11]

$$\hat{\tau} = \frac{\sum_{i<j} w_{ij} R_{ij} U_{ij}^{(P)} U_{ij}^{(D)}}{\sum_{i<j} w_{ij} R_{ij}},$$

where each orderable pair is weighted by $w_{ij} = 1/F\left[\max\left\{\min\left(P_i, P_j\right), \min\left(D_i, D_j\right)\right\}\right]^2$, the inverse probability of the pair being orderable with $F(\cdot)$ being the survival function of the censoring time.

3. Numerical Studies

We conducted simulations under various settings of PFS and OS to demonstrate how C-index can be used to quantify the association between PFS and OS. We also compare the performance of C-index with Spearman's rank correlation, and two copula models, namely the Clayton copula and Frank copula. All simulations are conducted by Rstudio 1.4.1103.

In a simulated trial, each subject was randomly assigned to the treatment or control group with equal probability. The accrual rate was 30 patients per month, and there was an additional 9-month follow-up period after enrollment was finished. We generated each patient's PFS time and SPP time from exponential distributions and then took the sum to obtain the OS time[1]. We explored three different configurations by first assuming a larger median value of PFS in the treatment arm (12 months) than that in the control arm (6 months), while there was no difference in SPP between the two arms. For this case, we experimented four median SPP values (3, 6, 9, and 12 months) corresponding to the first set of four scenarios ("S1", "S2", "S3" and "S4"). We next assumed that the median value of SPP differed between the treatment (12 months) and control (6 months) arms, while there was no difference in PFS between the two arms. For this case, we used four median PFS values (3, 6, 9, and 12 months) corresponding to the second set of four scenarios ("S5", "S6", "S7" and "S8"). Finally, we assumed there was no difference in PFS and SPP between two arms, and we set the median value of PFS as 6 months, and considered four median SPP values (3, 6, 9, and 12 months) corresponding to the last set of four scenarios, ("S9", "S10", "S11" and "S12"). The sample size was 600. For both treatment arms, the distribution of the survival time to death of other causes was assumed to be exponential with median 24 months. We simulated 1000 trials under each scenario. For each simulated trial, we conducted the log-rank test for both PFS and OS analysis and computed the C-index.

Table 1 summarizes the number of trials grouped by p-values from log-rank tests for the differences in PFS and OS from 12000 simulations under 12 scenarios. When the C-index is below 0.5, we cannot use PFS as a surrogate for OS because the log-rank tests lead to contradictory conclusions for PFS and OS. When the value of C-index is in $(0.6, 0.7]$, PFS can provide some information on OS, although it might still not be able to serve as an adequate surrogate endpoint. When the value of C-index is larger than 0.7, PFS can serve as an appropriate surrogate of OS. This conclusion provides some guidance about the role of C-index when evaluating the surrogacy of PFS for OS. In general, the threshold value could be adjusted according to specific cases. Table 2 presents the number of trials satisfying the value of C-index in each scenario among 12000 simulated trials.

Table 1. Summary of p-values from log-rank tests for PFS and OS in terms of C-index over 12000 simulated trials.

C-index	OS / PFS	≤ 0.01	>0.01	OS / PFS	≤0.05	>0.05	OS / PFS	≤0.1	>0.1
≤0.5	≤0.01	212	546	≤0.05	485	424	≤0.1	688	405
	>0.01	1696	2384	>0.05	1868	2061	>0.1	1895	1850
0.5-0.6	≤0.01	988	817	≤0.05	1412	471	≤0.1	1611	385
	>0.01	598	1453	>0.05	896	1077	>0.1	973	887
0.6-0.7	≤0.01	502	108	≤0.05	586	89	≤0.1	648	125
	>0.01	231	1560	>0.05	449	1277	>0.1	543	1085
>0.7	≤0.01	873	25	≤0.05	896	2	≤0.1	898	0
	>0.01	0	7	>0.05	0	7	>0.1	0	7

Table 2. Number of trials in each scenario (S1-S12) among 12000 simulated trials.

C-index	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
≤0.5	0	0	100	624	1000	948	12	0	0	173	981	1000
0.5-0.6	0	503	900	376	0	52	963	214	2	827	19	0
0.6-0.7	102	497	0	0	0	0	25	785	992	0	0	0
>0.7	898	0	0	0	0	0	0	1	6	0	0	0

We also evaluate the performance of Spearman's rank correlation, which is calculated using the R package "survSpearman" by Eden et al.[12], and two copula bivariate models (Clayton and Frank), which are computed by the R package "CopulaCenR"[13]. The detailed results are shown in Tables 3 and 4, respectively.

Table 3. Summary of p-values from log-rank tests for PFS and OS in terms of Spearman's rank correlation over 12000 simulated trials.

Spearman rank corr	OS / PFS	≤ 0.01	>0.01	OS / PFS	≤0.05	>0.05	OS / PFS	≤0.1	>0.1
≤0.6	≤0.01	311	501	≤0.05	577	355	≤0.1	756	321
	>0.01	1426	1882	>0.05	1567	1621	>0.1	1588	1455
0.6-0.7	≤0.01	786	742	≤0.05	1152	477	≤0.1	1346	425
	>0.01	767	1841	>0.05	1080	1427	>0.1	1170	1195
0.7-0.8	≤0.01	535	209	≤0.05	672	129	≤0.1	748	136
	>0.01	313	1126	>0.05	526	856	>0.1	598	701
>0.8	≤0.01	943	42	≤0.05	977	24	≤0.1	994	32
	>0.01	17	555	>0.05	38	518	>0.1	53	478

Table 4. Summary of p-values from log-rank tests for PFS and OS in terms of Clayton and Frank copula bivariate models over 12000 simulated trials.

Clayton copula	OS / PFS	≤ 0.01	>0.01	OS / PFS	≤0.05	>0.05	OS / PFS	≤0.1	>0.1
≤0.4	≤0.01	775	1147	≤0.05	1324	768	≤0.1	1622	692
	>0.01	1305	3309	>0.05	1577	2867	>0.1	1661	2561
0.4-0.5	≤0.01	750	295	≤0.05	925	158	≤0.1	1018	121
	>0.01	457	620	>0.05	668	371	>0.1	720	263
0.5-0.6	≤0.01	167	27	≤0.05	200	56	≤0.1	244	100
	>0.01	242	1459	>0.05	460	1179	>0.1	550	1001
>0.6	≤0.01	879	27	≤0.05	903	3	≤0.1	905	1
	>0.01	1	4	>0.05	1	4	>0.1	1	4
Frank copula	OS / PFS	≤ 0.01	>0.01	OS / PFS	≤0.05	>0.05	OS / PFS	≤0.1	>0.1
≤0.5	≤0.01	26	30	≤0.05	124	99	≤0.1	254	171
	>0.01	1683	2676	>0.05	1862	2330	>0.1	1896	2094

0.5-0.6	≤0.01	574	943	≤0.05	965	581	≤0.1	1133	462
	>0.01	110	693	>0.05	170	604	>0.1	184	541
0.6-0.7	≤0.01	1000	485	≤0.05	1275	274	≤0.1	1412	226
	>0.01	447	1365	>0.05	668	1080	>0.1	737	922
>0.7	≤0.01	975	38	≤0.05	1015	32	≤0.1	1046	56
	>0.01	285	670	>0.05	513	408	>0.1	594	272

It can be seen that Spearman's rank correlation and Clayton copula provide similar conclusions about the surrogacy of PFS for OS. The performance of Frank copula is relatively poor. Taking the corresponding thresholds for C-index, Spearman's correlation, Clayton and Frank copulae as 0.7, 0.8, 0.6 and 0.7, the contradictory rates, defined as the proportion of trials with contradictory conclusions, under these four measures are shown in Table 5.

Table 5. Contradictory rates using C-index, Spearman's rank correlation, Clayton copula and Frank copula models with the number of trials satisfying the threshold condition given in the parentheses.

Significant level	C-index (905)	Spearman (1557)	Clayton (911)	Frank (1968)
0.01	2.76%	3.79%	3.07%	16.41%
0.05	0.22%	3.98%	0.44%	27.69%
0.1	0%	5.46%	0.22%	33.03%

As seen from Table 5, the C-index has advantages over Clayton copula and performs much better than Spearman's rank correlation and Frank copula. The C-index is a more robust measure for the dependence of bivariate censored data because it does not rely upon any model assumption. Spearman's rank correlation is more appropriate when one or both variables are skewed or ordinal[14], which measures Pearson's linear association between ranks of the observations. The copula model depends on the parametric copula structures and may result in biased estimation if the assumed model is incorrect. There is often ambiguity in the choice of the copula family as there

is a large variety of copula structures to choose from. Therefore, we recommend C-index as the robust, model-free measurement of correlations between PFS and OS.

4. Real Trial Illustration

To illustrate the practical use of the C-index, we analyzed three real trials to explain the controversial findings between PFS and OS. The first dataset was from a colon cancer trial with 929 patients randomized to observation and two treatment groups[15]. We pooled data from both treatment arms together for simplicity. For each patient, two survival times were recorded: one for recurrence and the other for death. The second trial example was for advanced ovarian cancer with 1192 patients[16]. The third dataset was for gastric cancer with 279 patients[17]. For each patient, two endpoints, PFS and OS, were recorded. We implemented the log-rank test and estimated the C-index. Figure 1 shows the survival curves of PFS and OS by the Kaplan-Meier method. For the first two datasets, there existed statistically significant differences in both PFS and OS between the treatment and control arms. The values of C-index are 0.833 (standard error: SE=0.0114) and 0.826 (SE=0.00027) for the colon and ovarian cancer studies, respectively. For the third dataset, the difference in PFS was statistically significant while that in OS was not, and the corresponding value of C-index is 0.517 (SE= 0.00038). As a conclusion, when the value of C-index is large, PFS may serve as an appropriate surrogate endpoint of OS; otherwise, there may exist conflicting conclusions between PFS and OS.

Many factors would affect the correlation between PFS and OS, such as median SPP[1], types of investigational product[18], treatment crossover[19], and patient subgroups (e.g., sites of oncology origin)[9]. For example, the long median SPP and treatment crossover may underestimate the

treatment effect on OS, and immunotherapies might lead to smaller benefits in PFS versus OS. However, all the three trials discussed here were chemotherapy trials and there was no crossover allowed. The values of median SPP for the colon, ovarian, and gastric cancer trials were 963 days, 0.099 years, and 120 days, respectively, while the corresponding values of median PFS were 1589 days, 0.222 years and 86 days. Compared with the values of median PFS, long median SPP may be one of the reasons for the insignificance of OS for the gastric cancer trial. In addition, we examined whether associations between PFS and OS differed in patient subgroups. Taking the colon cancer trial as an example, the association between PFS and OS was somewhat weaker for patients with poor differentiation of tumor ($\hat{\tau} = 0.790$). Moreover, we did not observe significant differences in the PFS and OS associations across sex or obstruction of colon by tumor.

5. Discussion

In this paper, we focus on the association between PFS and OS at the patient level. Evaluation of surrogacy of PFS can be more comprehensively conducted with multiple trials in the same type of cancer. At the trial level, we can follow a conventional approach[20] to first estimate the treatment effects by calculating the log hazard ratios on PFS and OS within each trial. We then compute the C-index between these pairwise log hazard ratios to measure the association between PFS and OS at the trial level.

Acknowledgements

We would like to thank two anonymous referees, Associate Editor and Editor-in-Chief for their insightful suggestions that have led to great improvements in this article. The research was supported by a grant (17308420) from the Research Grants Council of Hong Kong.

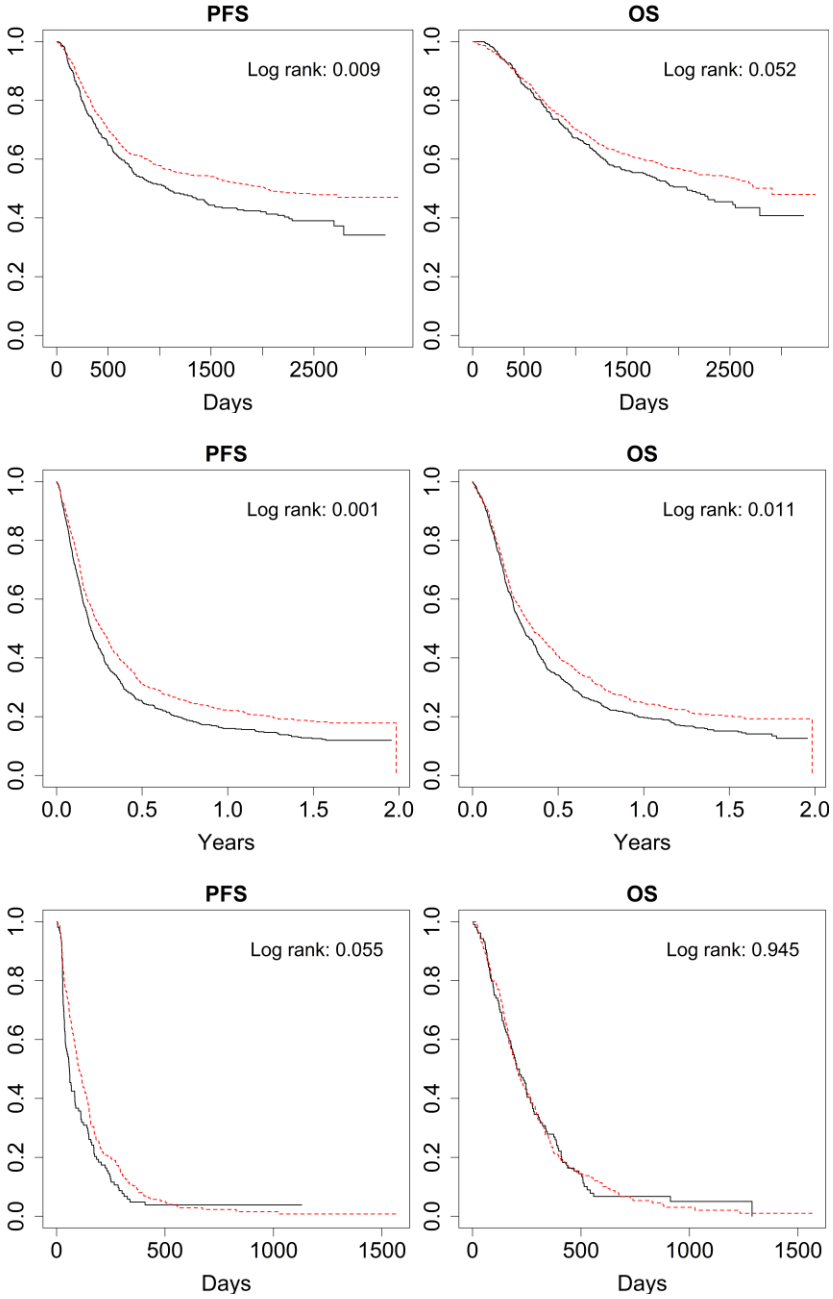

Figure 1. PFS and OS survival curves (dashed lines for treatment and solid lines for control) using the Kaplan-Meier method for the colon (top, $\hat{\tau}$ =0.833), ovarian (middle, $\hat{\tau}$ =0.826) and gastric (bottom, $\hat{\tau}$ =0.517) cancer trials with p-values from the log-rank tests.

References

1. K.R. Broglio, D.A. Berry, Detecting an overall survival benefit that is derived from progression-free survival, Journal of the National Cancer Institute 101(23) (2009) 1642-1649.
2. K.D. Yu, et al., Effect of Adjuvant Paclitaxel and Carboplatin on Survival in Women With Triple-Negative Breast Cancer: A Phase 3 Randomized Clinical Trial, JAMA Oncology 6(9) (2020) 1390-1396.
3. S. Morita, K. Sakamaki, G. Yin, Detecting overall survival benefit derived from survival postprogression rather than progression-free survival, Journal of the National Cancer Institute 107(8) (2015) djv133.
4. K. Oba, et al., Progression-free survival (PFS) as a surrogate endpoint for overall survival (OS) in advanced/recurrent gastric cancer (AGC) treatment: Individual-patient-data (IPD) based meta-analysis of randomized trials, Journal of Clinical Oncology 38(15) (2020) e16506.
5. S. Michiels, E.D. Saad, M. Buyse, Progression-free survival as a surrogate for overall survival in clinical trials of targeted therapy in advanced solid tumors, Drugs 77(7) (2017) 713-719.
6. X. Paoletti, et al., Assessment of progression-free survival as a surrogate end point of overall survival in first-line treatment of ovarian cancer: a systematic review and meta-analysis, JAMA Network Open 3(1) (2020) e1918939.
7. E.M. Weber, A.C. Titman, Quantifying the association between progression-free survival and overall survival in oncology trials using Kendall's τ, Statistics in Medicine 38(5) (2019) 703-719.
8. S. Halabi, et al., Progression-free survival as a surrogate endpoint of overall survival in patients with metastatic renal cell carcinoma, Cancer 120(1) (2014) 52-60.
9. M. Ter-Minassian, et al., Association between tumor progression endpoints and overall survival in patients with advanced neuroendocrine tumors, The Oncologist 22(2) (2017) 165-172.
10. M.G. Akritas, J. Siebert, A test for partial correlation with censored astronomical data, Monthly Notices of the Royal Astronomical Society 278(4) (1996) 919-924.
11. L. Lakhal, L.P. Rivest, D. Beaudoin, IPCW estimator for Kendall's tau under bivariate censoring, The International Journal of Biostatistics 5(1) (2009).
12. S.K. Eden, C. Li, B.E. Shepherd, Nonparametric Estimation of Spearman's Rank Correlation with Bivariate Survival Data, Biometrics (under revision).
13. T. Sun, et al., Copula-based score test for bivariate time-to-event data, with application to a genetic study of AMD progression, Lifetime Data Analysis 25(3) (2019) 546-568.
14. M.M. Mukaka, A guide to appropriate use of correlation coefficient in medical research, Malawi Medical Journal 24(3) (2012) 69-71.
15. C.G. Moertel, et al., Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage III colon carcinoma: a final report, Annals of Internal Medicine 122(5) (1995) 321-326.
16. G.A. Omura, et al., Cyclophosphamide plus cisplatin versus cyclophosphomide, doxorubicin, and cisplatin chemotherapy of ovarian carcinoma: a meta-analysis. Journal of Clinical Oncology 9(9) (1991) 1668-1674.
17. X. Paoletti, et al., Progression-free survival as a surrogate for overall survival in advanced/recurrent gastric cancer trials: a meta-analysis, Journal of the National Cancer Institute 105(21) (2013) 1667-1670.
18. B. Gyawali, S.P. Hey, A.S. Kesselheim, A comparison of response patterns for progression-free survival and overall survival following treatment for cancer with PD-1 inhibitors: a meta-analysis of correlation and differences in effect sizes, JAMA Network Open 1(2) (2018) e180416.
19. M. Hashim, et al., Do surrogate endpoints better correlate with overall survival in studies that did not allow for crossover or reported balanced postprogression treatments? An application in advanced non–small cell lung cancer, Value in Health 21(1) (2018) 9-17.
20. N.R. Foster, et al., Tumor response and progression-free survival as potential surrogate endpoints for overall survival in extensive stage small-cell lung cancer: Findings on the basis of North Central Cancer Treatment Group trials, Cancer 117(6) (2011) 1262-1271.