برآورد خطای پیش بینی برای زمان بقاء و کاربرد آن در تحلیل بقای بیماران مبتلا به سرطان روده بزرگ

احمدرضا باغستانی[۱]، محمودرضا گوهری[۲]، آرزو اروجی[۱*]، محمدامین پورحسینقلی[۳]

۱) گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی تهران

۲) گروه آمار و ریاضی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران

۳) مرکز تحقیقات گوارش، دانشگاه علوم پزشکی شهید بهشتی تهران

تاریخ دریافت: ۹۲/۴/۲۹ تاریخ پذیرش: ۹۲/۱۲/۱۸

چکیده

مقدمه: سرطان روده بزرگ جزء سرطان های شایع و کشنده است و پیش بینی این که افراد تا چه زمانی زنده می مانند از اهمیت بالایی برخوردار است. هدف از این مقاله برازش مدل کاکس و منحنی کاپلان مایر برای داده های سرطان روده بزرگ بیمارستان طالقانی و مقایسه آن دو با معیار زیان آشکار است.

مواد و روش ها: این مطالعه توصیفی-تحلیلی بر روی بیماران مبتلا به سرطان روده بزرگ مراجعه کننده به بیمارستان طالقانی شهر تهران با حداقل ۵ سال پیگیری انجام شده است. این افراد بین سال های ۱۳۸۰ تا ۱۳۸۵ به بیمارستان طالقانی مراجعه کرده اند و اطلاعاتشان در بخش سرطان مرکز تحقیقات گوارش ثبت شده و مورد پیگیری قرار گرفته اند. برای پیش بینی زمان بقاء مدل مخاطره متناسب کاکس و منحنی کاپلان مایر استفاده شده است و این دو با برآورد خطای پیش بینی به روش زیان آشکار با هم مقایسه شده اند.

یافته های پژوهش: مدل کاکس ارتباط معناداری را بین اندازه تومور و جنس با زمان بقاء نشان داد و خطای پیش بینی با روش زیان آشکار برای مدل کاکس در بعضی از زمان ها کمتر از منحنی کاپلان مایر بود.

بحث و نتیجه گیری: در این مطالعه به این نتیجه رسیدیم که برای مقایسه مدل ها می توان از برآورد خطای پیش بینی استفاده کرد.

واژه های کلیدی: مدل مخاطره متناسب کاکس، منحنی کاپلان مایر، خطای پیش بینی، زیان آشکار، سرطان روده بزرگ

*نویسنده مسئول: گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی تهران

Email: arezoo.orooji23@gmail.com

[Downloaded from sjimu.medilam.ac.ir on 2026-11-07]

مقدمه

سرطان یعنی رشد مهار گسیخته سلول ها، یکی از علت های منجر به مرگ در سراسر جهان است و پس از بیماری قلبی-عروقی دومین عامل شایع در کشورهای توسعه یافته و سومین عامل مرگ در کشورهای کمتر توسعه یافته است. سرطان کولورکتال جزء سرطان های شایع و کشنده است و چهارمین سرطان شایع در مردان و سومین سرطان شایع در زنان می باشد و زمانی رخ می دهد که گسترش سلول های سرطانی در بخشی از روده بزرگ اتفاق افتد،(۱). از نظر آناتومی روده بزرگ شامل کولون و رکتوم است که طول کولون حدود ۱۸۰ سانتی متر و طول رکتوم بین ۱۵ تا ۲۵ سانتی متر می باشد. اگر گسترش سلول‌های سرطانی در بخش رکتوم باشد به آن سرطان رکتوم و اگر در قسمت کولون باشد به آن سرطان کولون گویند. از علائم بالینی این بیماری درد شکم، تغییر عادات دفع و کم خونی فقر آهن توجیه نشده می باشد. که البته علائم آن بستگی به ناحیه درگیر دارد،(۲). از آن جایی که این بیماری از جمله سرطان‌های کشنده است پیش بینی این که بیماران تا چه زمانی زنده می مانند از اهمیت بالایی برخوردار است. اما حقیقت این است که اختلاف چشم گیری بین پیش بینی زمان بقای پزشکی و بقای حقیقی وجود دارد و به این علت است که برای بررسی این موضوع باید از مدل های آماری استفاده کنیم.(۳)

مدل های آماری اغلب برای پیش بینی متغیرهای تصادفی آینده استفاده می شوند. در مدل های آماری دو نوع پیش بینی در نظر گرفته می شود. پیش بینی نقطه ای و پیش بینی احتمالی، که پیش بینی نقطه ای یک مقدار مشخص برای متغیر پاسخ برآورد می کند. و پیش بینی احتمالی احتمال این که فرد تا زمانی زنده باشد را برآورد می‌کند. دقت پیش بینی توسط اندازه هایی که معمولاً براساس توابع زیان است مشخص می شود،(۴)، پیش بینی در بقاء می تواند با استفاده از مدل هایی مانند مدل مخاطره متناسب کاکس انجام شود،

مدل های بقاء یا مدل های زمان شکست از جمله روش های آماری می باشند که در آن ها متغیر پاسخ زمان رخداد واقعه می باشد و مثل تمام روش های آماری دیگر هدف آن ها آنالیز و پیش بینی داده هاست با این تفاوت که در این مدل ها متغیر پاسخ می تواند سانسور شود و این ویژگی بارز و چالش برانگیز داده های بقاست که آن ها را از سایر مدل های آماری مجزا می کند. پیش بینی در داده های بقاء با سایر روش های آماری به علت وجود سانسور متفاوت می باشد. وجود سانسور در متغیر زمان بقاء باعث می شود از هر تابع زیانی نتوانیم استفاده کنیم و یک چالش اصلی برای داده های بقاء، تشخیص تابع زیان مناسب می باشد.(۵)

زیان حقیقی مشاهده سانسور شده معلوم نیست به خاطر این که زمان بقای مربوطه مشاهده نشده است به این علت است که برای استفاده از توابع زیان اختیاری باید روش سانسور با وزن احتمال معکوس را بسط دهیم و هم چنین این مساله روی شیوه به دست آوردن خطای پیش بینی و فواصل اطمینان برای اندازه های عملکرد تاثیر می گذارد،(۶). مطالعات مختلفی برای تعریف مناسب تابع زیان برای داده های بقاء انجام شده است که می توان به هندرسون و کیدینگ،(۳)، جرالد لاولس و ین یوان،(۴)، شمپر،(۵)، گردز و شوماخر،(۶)، کرن و سیمون،(۷)، هندرسون،(۸)، گرف و همکاران،(۹)، گردز و شوماخر،(۱۰)، شمید و همکاران،(۱۱)، هندرسون،(۱۲)، اشاره کرد.

هدف از این مطالعه برازش مدل کاکس و شناسایی عوامل مؤثر بر پیش بینی زمان بقای بیماران مبتلا به سرطان روده بزرگ و مقایسه مدل کاکس با منحنی کاپلان مایر با استفاده از معیار زیان آشکار است.

مواد و روش ها

این مطالعه توصیفی-تحلیلی بر روی بیماران مبتلا به سرطان روده بزرگ مراجعه کننده به بیمارستان طالقانی شهر تهران با حداقل ۵ سال پیگیری انجام شده است. به منظور پیش بینی زمان بقای بیمارانی که در فاصله سال های ۸۵-۸۰ به بیمارستان طالقانی مراجعه کرده اند فایلی تهیه شده است که حاوی متغیرهای پیش آگهی دهنده ای از جمله، سن بیمار، جنسیت بیمار، اندازه تومور، محل گسترش تومور، آخرین وضعیت بیمار، زمان مرگ و... می باشد. تعداد کل افرادی که در این پژوهش مورد مطالعه قرار گرفتند ۷۲۴ نفر بودند. این افراد حداقل به مدت ۵ سال پیگیری شدند و اطلاعات لازم ثبت شد. سپس مدل کاکس و منحنی کاپلان مایر را به داده ها برازش دادیم و برای مقایسه برازش آن دو از برآورد خطای پیش بینی به روش زیان آشکار استفاده کردیم.

منحنی کاپلان مایر: منحنی کاپلان مایر یک روش ناپارامتری است که می توان با استفاده از آن برآوردی استاندارد برای تابع بقاء به دست آورد. یک محور آن زمان

[Downloaded from sjimu.medilam.ac.ir on 2026-11-07]

بقاء(t) و دیگری تابع بقاء $s(t)=p(T>t)$ را نشان می دهد. علاوه بر آن با استفاده از این منحنی و آزمون لگ رتبه و... می توان بقاء را برای گروه های مختلف به دست آورد و آن ها را با یکدیگر مقایسه کرد.

مدل مخاطره متناسب کاکس: مدل مخاطره متناسب کاکس یا رگرسیون کاکس یک روش نیمه پارامتری است که علاوه بر برآورد احتمالات بقاء در زمان های مختلف به بررسی تاثیر متغیرهای کمکی نیز می پردازد. این مدل به صورت زیر بیان می شود:

$$h(t|X) = h_0(t)\exp(\acute{B}X)$$

که در آن $h(t|X)$ نشان دهنده نرخ مخاطره در زمان t برای فردی با متغیرهای کمکی X و $h_0(t)$ نرخ خطر پایه می باشد. برای برازش مدل کاکس به داده ها ابتدا باید فرض مخاطره متناسب برقرار باشد. یعنی نسبت خطر در دو گروه در طول زمان ثابت باشد. در صورت برقرار نبودن فرض مخاطره متناسب باید از مدل کاکس طبقه بندی یا مدل کاکس تعمیم یافته استفاده شود.

خطای پیش بینی زیان آشکار: زیان مورد انتظار را خطای پیش بینی گویند که بستگی به تابع زیان و نوع متغیر دارد. یکی از روش های برآورد خطای پیش بینی استفاده از شیوه زیان آشکار است و به صورت زیر است.

$$\mathrm{AL} = \frac{1}{n}\sum L(t_i, G(x_i, \hat{\theta}))$$

که در آن t_i زمان بقاء، x_i متغیرهای کمکی و L تابع زیان مورد نظر و $G(x_i, \hat{\theta})$ بهترین پیش بینی توسط متغیرهای کمکی است و بستگی به نوع تابع زیان دارد. و انتخاب تابع زیان بستگی به متغیر مورد نظر دارد. از آن جایی که در مطالعات بقاء توزیع زمان بقاء چوله می باشد، استفاده از زیان قدرمطلق خطا توصیه می شود.

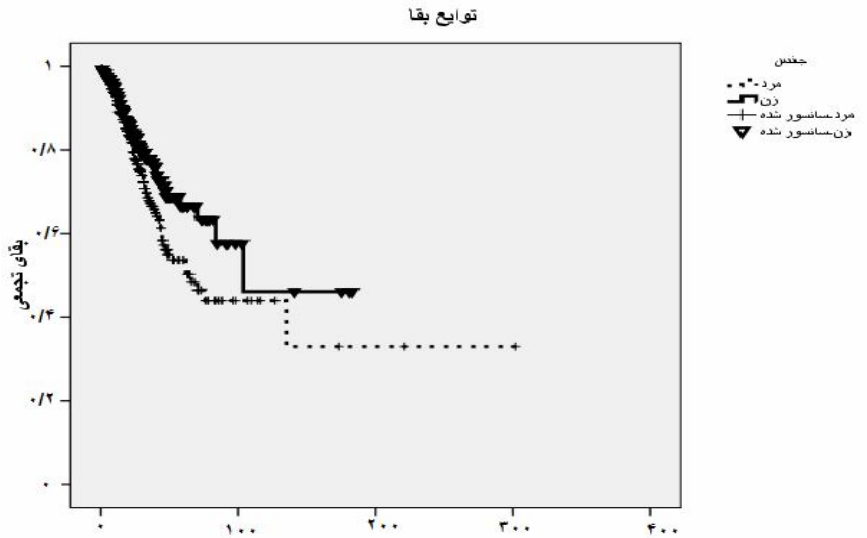
یافته های پژوهش

از ۷۲۴ بیمار، ۴۲۲(۵۸/۳ درصد) نفر مرد و ۳۰۲(۴۱/۷ درصد) نفر زن بودند. ۱۷۷(۲۴/۴ درصد) از بیماران فوت کردند که ۶۵ درصد آن ها مرد بودند. میانگین سنی این بیماران ۱۴/۸۱±۵۳/۰۳ بود. میانه بقای بیماران ۲۳/۳±۱۰۴ ماه بود. ابتدا با استفاده از منحنی کاپلان مایر و آزمون لگ رتبه معنی داری اختلاف متغیرهای جنس، محل سرطان، سن هنگام تشخیص و اندازه تومور مورد بررسی قرار گرفت. توزیع بیماران و نتایج آزمون لگ رتبه و منحنی های کاپلان مایر برای متغیرهای جنس، محل سرطان، سن هنگام تشخیص و اندازه تومور در جدول شماره ۱ و نمودارهای شماره ۱ تا شماره ۴ نشان داده شده است.

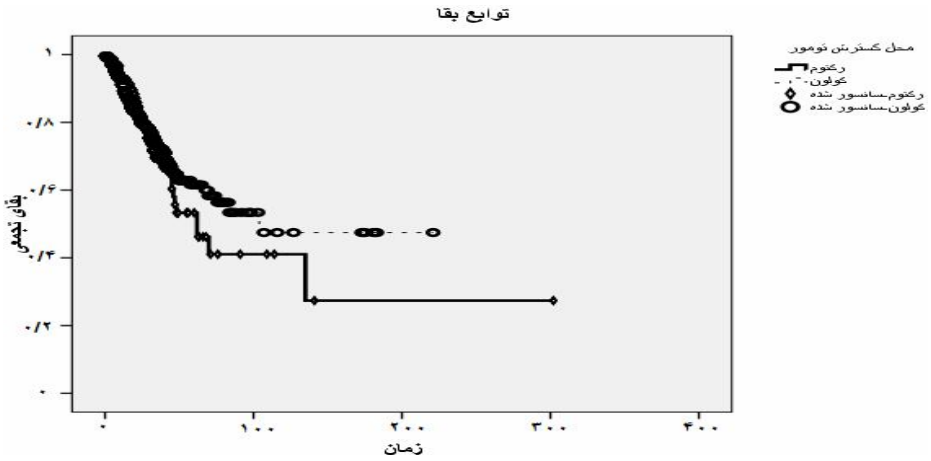
جدول شماره ۱. توزیع بیماران و نتایج آزمون لگ رتبه برای متغیرهای جنس، محل سرطان، سن هنگام تشخیص و اندازه تومور در بیماران مبتلا به سرطان روده بزرگ استخراج شده از بیمارستان طالقانی

	متغیر	تعداد بیماران	تعداد مرگ‌ها	میانگین‌زمان بقا(SE)	P
محل سرطان	کولون	۵۰۰(۶۹/۱)	۱۱۹(۲۳/۸)	۱۲۷/۲۶(۲۴/۱۰۷)	۰/۴۴۸
	رکتوم	۲۲۴(۳۰/۹)	۵۸(۲۵/۹)	۱۲۲/۸۲(۹/۹۵)	
جنس	زن	۳۰۲ (۴۱/۷)	۶۲ (۲۰/۵)	۱۱۲/۴۶(۱۱/۰۷)	۰/۰۲
	مرد	۴۲۲(۵۸/۳)	۱۱۵(۲۷/۳)	۱۳۳/۶۲۱(۱۹/۲۷)	
اندازه تومور	کمتر از ۲۰	۸۵(۱۱/۷)	۱۱(۱۲/۹)	۱۴۴/۲۰(۲۶/۲۷)	۰/۰۱
	بیشتر از ۲۰	۶۳۹(۸۸/۳)	۱۶۶(۲۶)	۱۴۵/۳۴(۱۵/۶۴)	
سن هنگام تشخیص	کمتر از ۵۰	۳۰۷(۴۲/۴)	۷۸(۲۵/۴)	۱۵۹/۵۹(۱۶/۹۳)	۰/۳۱۷
	بیشتر از ۵۰	۴۱۷(۵۷/۶)	۹۹(۲۳/۷)	۸۰/۹۱(۶/۸۷)	

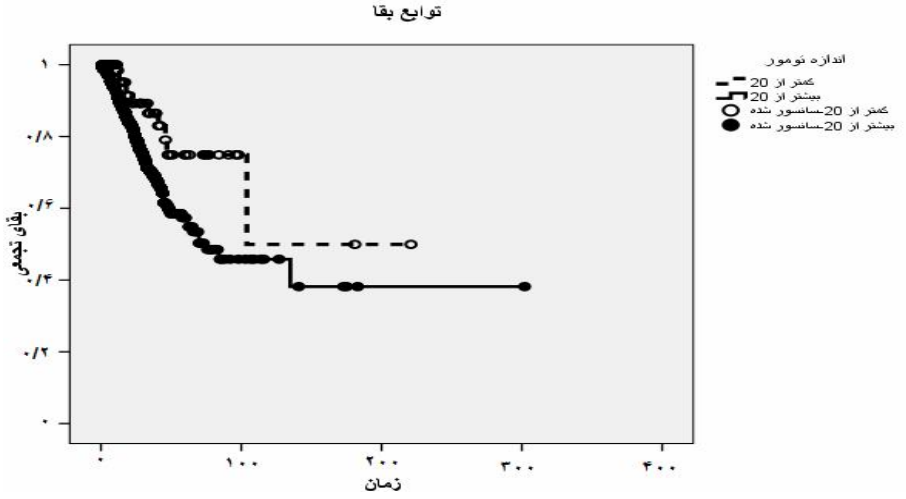
[Downloaded from sjimu.medilam.ac.ir on 2026-11-07]



نمودار شماره ۱. منحنی های کاپلان مایر برای متغیر جنس در بیماران مبتلا به سرطان روده بزرگ در بیمارستان طالقانی

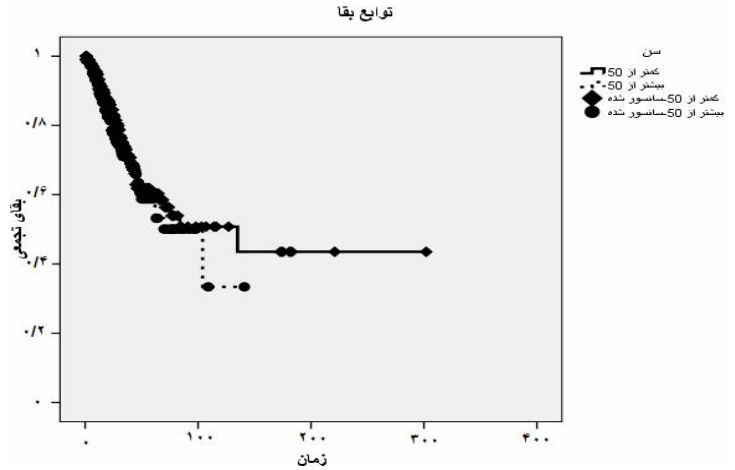


نمودار شماره ۲. منحنی های کاپلان مایر برای متغیر محل سرطان در بیماران مبتلا به سرطان روده بزرگ در بیمارستان طالقانی



نمودار شماره ۳. منحنی های کاپلان مایر برای متغیر اندازه تومور در بیماران مبتلا به سرطان روده بزرگ در بیمارستان طالقانی

[Downloaded from sjimu.medilam.ac.ir on 2026-11-07]



نمودار شماره ٤. منحنی های کاپلان مایر برای متغیر سن هنگام تشخیص در بیماران مبتلا به سرطان روده بزرگ در بیمارستان طالقانی

قبل از برازش مدل کاکس فرض خطر متناسب برای متغیرها با استفاده از باقی مانده اسکینفیلد بررسی شد و این فرض برای متغیر محل گسترش تومور برقرار نبود.($P<0.05$) در نتیجه مدل کاکس طبقه بندی به داده ها برازش داده شد. نتایج حاصل از مدل رگرسیونی کاکس طبقه بندی در جدول شماره ۲ نشان داده شده است.

جدول شماره ۲. نتایج حاصل از برازش مدل رگرسیونی کاکس طبقه بندی در بیماران مبتلا به سرطان روده بزرگ بیمارستان طالقانی

متغیر	ضریب	نسبت مخاطره	انحراف استاندارد ضریب رگرسیونی	P
جنس[۱]	-۰/۳۰۳	۰/۷۳۹	۰/۱۵۹	۰/۰۵۸
سن[۲]	۰/۱۴۸	۱/۱۵۹	۰/۱۵۳	۰/۳۴
اندازه تومور[۳]	۰/۷۴۵	۲/۱۰۷	۰/۳۱۳	۰/۰۱۷

۱: زن به مرد؛ ۲: بالای ۵۰ سال به زیر ۵۰ سال؛ ۳: بالای ۲۰ میلی متر به زیر ۲۰ میلی متر

برای مقایسه منحنی کاپلان مایر و مدل کاکس از برآورد خطای پیش بینی استفاده شد. خطای پیش بینی مدل کاکس از منحنی کاپلان مایر کمتر است و وجود متغیرهای کمکی باعث کاهش خطای پیش بینی شده است. علاوه بر آن با افزایش زمان اختلافی که بین منحنی کاپلان مایر و مدل کاکس است بیشتر می شود. نتایج آن در جدول شماره ۳ نشان داده شده است.

جدول شماره ۳. نتایج خطای پیش بینی به روش زیان آشکار در بیماران مبتلا به سرطان روده بزرگ بیمارستان طالقانی

زمان	تعداد افراد در خطر	کاپلان مایر	کاکس
۲۴/۷۵	۳۵۶	۰/۱۵۳	۰/۱۵۳
۴۹/۵۰	۸۷	۰/۲۲۸	۰/۲۲۰
۷۹/۷۵	۲۷	۰/۲۲۷	۰/۲۰۴

[Downloaded from sjimu.medilam.ac.ir on 2026-11-07]

بحث و نتیجه‌گیری

این مطالعه به منظور شناسایی عوامل موثر بر بقای بیماران سرطان کولورکتال بیمارستان طالقانی و مقایسه منحنی کاپلان مایر و مدل کاکس انجام گرفته است که این مدل ها با استفاده از برآورد خطای پیش بینی به روش زیان آشکار با یکدیگر مقایسه شده اند و مدل کاکس مدل مناسب تری برای داده های این مطالعه است.

با توجه به یافته ها تفاوت متغیرهای جنس و اندازه تومور هم با توجه به آزمون لگ رتبه و هم با توجه به نتایج مدل کاکس در سطح($P=0.01$) معنی دار شده اند. با توجه به نتایج مدل کاکس خطر برای زنان نسبت به مردان کمتر و ۰/۷۳۹ برابر است. برخی مطالعات بالا بودن زمان بقای زنان نسبت به مردان را نشان دادند،(۱۳،۱۴). مطالعه دیگری بیش تر بودن مرگ و میر مردان نسبت به زنان را در این سرطان نشان می دهد،(۱۵). برای افرادی که اندازه تومورشان بیشتر از ۲۰ بوده است نسبت به افرادی که اندازه تومورشان کمتر از ۲۰ بوده است مخاطره ۲/۱۰۷ برابر بوده است که بسیاری از مطالعات این نتیجه را تایید می کنند،(۱۶،۱۷). در این مطالعه تفاوت در محل گسترش تومور معنی دار نشد که در تضاد با مطالعه های دیگراست،(۱۸،۱۹). در سایر مطالعاتی که روی این سرطان انجام شده است متغیر سن، نیز روی زمان بقای بیماران سرطان کولورکتال تاثیر می‌گذارد،(۲۰)، اما در این مطالعه این متغیر معنی دار نشده است. برای مقایسه مدل کاکس و منحنی کاپلان مایر از برآورد خطا به روش زیان آشکار استفاده شده است که این روش برای حالتی است که مجموعه داده جدیدی در دسترس نباشد و به همین دلیل مقداری خطا را کم برآورد می کند، پیشنهاد می شود از برآورد خطای پیش بینی برای مقایسه مدل های پارامتری و نیمه پارامتری نیز استفاده شود. علاوه بر آن می توان خطای پیش بینی را با روش اعتبار متقاطع به دست آورد و مقایسه ای بین آن دو نیز انجام داد.

References

1.Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. Cancer 2005; 55:74-108.

2.Wickham R, Lassere Y, editors. The ABCs of colorectal cancer. Oncol Nurs 2007 ;7:769-84.

3.Henderson R, Keiding N. Individual survival time prediction using statistical models. J Med Ethics 2005;31:703-6.

4.Lawless JF, Yuan Y. Estimation of prediction error for survival models. Stat Med 2010;29:262-74.

5.Schemper M. Predictive accuracy and explained variation. Stat Med 2003;22:2299-308.

6.Gerds TA, Schumacher M. Consistent Estimation of the Expected Brier Score in General Survival Models with Rights Censored Event Times. Biometr J 2006; 48:1029-40.

7.Korn EL, Simon R. Measures of explained variation for survival data. Stat Med 1990; 9:487-503.

8.Henderson R, Jones M, Stare J. Accuracy of point predictions in survival analysis. Stat Med 2001; 20:3083-96.

9.Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. Stat Med 1999;18:2529-45.

10.Gerds TA, Schumacher M. Efrons type measures of prediction error for survival analysis. Biometrics 2007;63:1283-7.

11.Schmid M, Hielscher T, Augustin T, Gefeller O. A robust alternative to the Schemper–Henderson estimator of prediction error. Biometrics 2011;67:524-35.

12.Henderson R. Problems and prediction in survivals data analysis. Stat Med 1995; 14:161-84.

13.Aarts MJ, Lemmens VE, Louwman MW, Kunst AE, Coebergh JWW. Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. Eur J Cancer 2010;46:2681-95.

14.Dupont M, Pampalon R, Hamel D. Deprivation and cancer mortality among Quebec women and men, 1994-1998. Institut nationnal de santé publique du Québec; 2004.

15.Wrigley H, Roderick P, George S, Smith J, Mullee M, Goddard J. Inequalities in survival from colorectal cancer: a comparison of the impact of deprivation, treatment, and host factors on observed and cause specific

[Downloaded from sjimu.medilam.ac.ir on 2026-11-07]

survival. J Epidemiol Community Health 2003;57:301-9.

16.Bear HD, MacIntyre J, Burns HJ, Jarrett F, Wilson RE. Colon and rectal carcinoma in the west of Scotland: Symptoms, histologic characteristics, and outcome. Am J Surg 1984;147:441-6.

17.Xu F, Di M, Dong J, Wang F, Jin Y, Zhu Y, et al. [Influence of clinical and pathomorphological parameters on prognosis in colon carcinoma and rectal carcinoma]. Zhejiang da xue xue bao Yi xue ban]. J Zhejiang Uni Med Sci 2006; 35:303-10. (Chinese)

18.Wray CM, Ziogas A, Hinojosa MW, Le H, Stamos MJ, Zell JA. Tumor subsite location within the colon is prognostic for survival after colon cancer diagnosis. Dis Colon Rectum 2009;52:1359-66.

19.Meguid RA, Slidell MB, Wolfgang CL, Chang DC, Ahuja N. Is there a difference in survival between right-versus left-sided colon cancers? Ann Surg Oncol 2008;15:2388-94.

20.Møller H, Sandin F, Robinson D, Bray F, Klint □, Linklater KM, et al. Colorectal cancer survival in socioeconomic groups in England: variation is mainly in the short term after diagnosis. Eur J Cancer 2012; 48:46-53.

Estimation of Prediction Error for Survival Time with Application in Colorectal Cancer

Baghestani A[1], Gohari M[2], Orooji A[1], Pourhosseingholi A[3]*

(Recived: July 20, 2013 Accepted: March 9, 2014)

Abstract

Introduction: Colorectal cancer is one of the most widespread and killer ~~among~~ cancers and it is highly important to predict the survival time. The purpose of this paper is fitting the Cox model and Kaplan-Meier curve for the colorectal cancers data in Taleghani Hospital and comparing both with apparent loss.

Materials & Methods: This paper is related to colorectal cancer patients in Tehran Taleghani Hospital who were followed up for at least five years. These patients referred to Taleghani Hospital from 2001 to 2006 and their information was recorded in the cancer section of Digestion Research Center and was followed up for at least five years. The Cox Proportional Hazards model and Kaplan- Meier curve were used for prediction of survival times and these were compared with estimation of prediction error by apparent loss method.

Findings: Cox model showed a significant relation between size of tumor and sex with survival time. The prediction error with apparent loss method for Cox model was at times less than Kaplan-Meier curve.

Discussion & Conclusion: In this study, we concluded that estimation of prediction error could be used for comparison of models

Keywords: Cox proportional hazards model, kaplan- meier curve, prediction error, apparent loss, colorectal cancer

1.Dept of Biostatistics, Faculty of Paramedicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2.Dept of Biostatistics and Mathematic, Faculty of Health, Tehran University of Medical Sciences, Tehran, Iran
3.Gastroenterology and Liver Disease Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran
** (Corresponding author*

[Downloaded from sjimu.medilam.ac.ir on 2026-11-07]