

# برآورد خطای پیش‌بینی برای وضعیت بقا و کاربرد آن در تحلیل بقای بیماران مبتلا به سرطان روده بزرگ

نویسندگان: محمودرضا گوهری<sup>۱</sup>، احمدرضا باغستانی<sup>۲</sup>، آرزو اروجی<sup>۳\*</sup> و محمدامین پورحسینقلی<sup>۴</sup>

۱- دانشیار گروه آمار زیستی، مرکز تحقیقات مدیریت بیمارستان، دانشگاه علوم پزشکی ایران، تهران، ایران

۲- استادیار گروه آمارزیستی دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

۳- دانشجوی کارشناسی ارشد آمارزیستی، گروه آمارزیستی، دانشکده پیراپزشکی دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

۴- استادیار مرکز تحقیقات گوارش دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

E-mail: arezoo.orooji23@gmail.com

\* نویسنده مسئول: آرزو اروجی

## چکیده

مقدمه و هدف: سرطان روده بزرگ جزو سرطان‌های شایع و کشنده است و پیش‌بینی اینکه افراد در آینده چه وضعیتی دارند، بسیار اهمیت دارد و در نتیجه، ارزیابی پیش‌بینی نیز، بسیار مهم است. هدف از این مقاله، برآزش مدل کاکس و شناسایی عوامل مؤثر بر پیش‌بینی وضعیت بقای بیماران مبتلا به سرطان روده بزرگ و مقایسه مدل کاکس با منحنی کاپلان مایر با استفاده از تابع زیان برابر است.

مواد و روش‌ها: طی سال‌های ۱۳۸۰ تا ۱۳۸۵ تعداد ۷۲۴ بیمار مبتلا به سرطان روده بزرگ به بیمارستان طالقانی شهر تهران مراجعه کردند و دست‌کم پنج سال پیگیری شدند. برای تعیین عوامل مرتبط با وضعیت بقای سرطان کولورکتال مدل رگرسیون کاکس به داده‌ها برآزش داده شد و این مدل با منحنی کاپلان مایر به وسیله برآورد خطای پیش‌بینی به روش تابع زیان برابر مقایسه شد.

نتایج: 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 آنها مرد بودند؛ میانگین سنی این بیماران  $53/03 \pm 14/81$  بود. میانه بقای بیماران  $104 \pm 23/3$  ماه بود. وضعیت بقای بیماران با عوامل جنس ( $p=0/058$ ) و اندازه تومور ( $p=0/017$ ) ارتباط داشت و خطای پیش‌بینی برابر مدل کاکس در بعضی از زمان‌ها کمتر از منحنی کاپلان مایر شد.

نتیجه‌گیری: وجود متغیرهای کمکی و استفاده از مدل‌ها در پیش‌بینی وضعیت بقای افراد مؤثر است. در این مطالعه، برآزش مدل کاکس و وجود متغیرهای کمکی، باعث کاهش خطای پیش‌بینی شد.

واژگان کلیدی: مدل مخاطره متناسب کاکس، منحنی کاپلان مایر، خطای پیش‌بینی، سرطان روده بزرگ

## دانشور پزشکی

دوماهنامه علمی-پژوهشی  
دانشگاه شاهد  
سال بیست‌ویکم-شماره ۱۱۱  
تیر ۱۳۹۳

دریافت: ۱۳۹۳/۰۱/۲۴  
آخرین اصلاح‌ها: ۱۳۹۳/۰۳/۲۱  
پذیرش: ۱۳۹۳/۰۳/۲۶

## مقدمه

رخداد واقعه است و مثل تمام روش‌های آماری دیگر، هدف آنها آنالیز و پیش‌بینی داده‌هاست با این تفاوت که در این مدل‌ها متغیر پاسخ می‌تواند سانسور شود و این ویژگی بارز و چالش‌برانگیز داده‌های بقاست که آنها را از سایر مدل‌های آماری مجزای کند. پیش‌بینی در داده‌های بقا با سایر روش‌های آماری به دلیل وجود سانسور، متفاوت است. وجود سانسور در متغیر زمان بقا باعث می‌شود از هر تابع زیانی نتوانیم استفاده کنیم و چالشی اصلی برای داده‌های بقا تشخیص تابع زیان مناسب به‌شمار می‌آید (۴). زیان حقیقی مشاهده سانسور شده معلوم نیست به دلیل اینکه زمان بقای مرتبط مشاهده نشده‌است و به این دلیل است که برای استفاده از توابع زیان اختیاری باید روش سانسور با وزن احتمال معکوس<sup>۱</sup> را بسط دهیم و همچنین این مسئله روی شیوه به‌دست آوردن خطای پیش‌بینی تأثیر می‌گذارد (۵). مطالعاتی مختلف برای تعریف مناسب تابع زیان برای داده‌های بقا انجام شده‌اند که می‌توان به *جرالد لاولس* و *ین‌یوان* (۳)، *شمپر* (۴)، *گردز و شوماخر* (۵)، *کرن* و *سیمون* (۶)، *هندرسون، جونز و استایر* (۷) اشاره کرد. هدف از این مقاله، برآزش مدل کاکس و شناسایی عوامل مؤثر بر پیش‌بینی وضعیت بقای بیماران مبتلا به سرطان روده بزرگ و مقایسه مدل کاکس با منحنی کاپلان مایر با استفاده از تابع زیان برآیر<sup>۲</sup> است.

## مواد و روش‌ها

در این مطالعه پرونده ۷۲۴ بیمار مبتلا به سرطان روده بزرگ که طی سال‌های ۱۳۸۰ تا ۱۳۸۵ برای درمان به بیمارستان طالقانی شهر تهران مراجعه کرده‌اند بررسی شد؛ برای این بیماران، متغیرهای سن، جنسیت، اندازه تومور، محل سرطان، آخرین وضعیت بیمار (زنده، فوت)، زمان مرگ ثبت شد. زمان از هنگام درمان تا پایان مطالعه یا مرگ به‌عنوان متغیر زمان در نظر گرفته شد؛ این افراد دست‌کم به مدت پنج سال پیگیری شدند. برای تعیین

سرطان از تغییر ماهیت سلول‌های سالم به وجود می‌آید. در بدن تحت اصولی معین، سلول‌های بدن با نظمی مشخص تولید و تکثیر می‌شوند و رشد می‌کنند و به تدریج از بین می‌روند و سلول‌های جدید، جایگزین سلول‌های از بین رفته می‌شوند؛ اما در شرایطی سلول‌های طبیعی تغییر ماهیت می‌دهند و میزان رشد آنها از کنترل خارج می‌شود و به صورت خودمختار و بدون وقفه رشد و گسترش پیدا می‌کنند و بدین ترتیب نظم و ساختار بافت طبیعی و عضو را به هم می‌زنند و باعث ایجاد سرطان می‌شوند. سرطان کولورکتال، جزو سرطان‌های شایع و کشنده است و چهارمین سرطان شایع در مردان و سومین سرطان شایع در زنان است و زمانی رخ می‌دهد که گسترش سلول‌های سرطانی در بخشی از روده بزرگ اتفاق افتد (۱). از نظر آناتومی، روده بزرگ شامل کولون و رکتوم می‌شود که طول کولون حدود ۱۸۰ سانتی‌متر و طول رکتوم میان ۱۵ تا ۲۵ سانتی‌متر است. اگر گسترش سلول‌های سرطانی در بخش رکتوم باشد به آن سرطان رکتوم و اگر در قسمت کولون باشد به آن سرطان کولون گویند (۲)؛ از آنجاکه این بیماری از جمله سرطان‌های کشنده است پیش‌بینی اینکه بیماران در آینده چه وضعیتی دارند دارای اهمیتی بالاست.

مدل‌های آماری، اغلب برای پیش‌بینی متغیرهای تصادفی آینده استفاده می‌شوند. در مدل‌های آماری، دو نوع پیش‌بینی در نظر گرفته می‌شوند: پیش‌بینی نقطه‌ای و پیش‌بینی احتمالی که پیش‌بینی نقطه‌ای یک مقدار مشخص برای متغیر پاسخ برآورد می‌کند و پیش‌بینی احتمالی، احتمال اینکه فرد تا زمانی زنده باشد را برآورد می‌کند. دقت پیش‌بینی، توسط اندازه‌هایی که اغلب بر اساس توابع زیان است مشخص می‌شود (۳). پیش‌بینی در بقا می‌تواند با استفاده از مدل‌هایی از جمله مدل مخاطره متناسب کاکس انجام شود.

مدل‌های بقا یا مدل‌های زمان شکست از جمله روش‌های آماری هستند که در آنها متغیر پاسخ زمان

1- IPCW (inverse probability of censoring weighting)

2- Brier score

پیش‌بینی‌شده توسط مدل است. منحنی‌های پیش‌بینی برای توزیع‌های پیشگویانه از رسم زیان برابر در مقابل زمان به‌دست‌می‌آید. برای اجرای مراحل بالا از نرم‌افزار R بسته PEC استفاده کردیم.

### نتایج

از ۷۲۴ بیمار، ۴۲۲ (۵۸/۳ درصد) نفر مرد و ۳۰۲ (۴۱/۷ درصد) نفر زن بودند. ۱۷۷ (۲۴/۴ درصد) از بیماران فوت‌کردند که ۶۵ درصد آنها مرد بودند؛ میانگین سنی این بیماران  $۵۳/۰۳ \pm ۱۴/۸۱$  بود. میانه بقای بیماران  $۱۰۴ \pm ۲۳/۳$  ماه بود. ابتدا با استفاده از منحنی کاپلان مایر و آزمون لگ‌رتبه، معنی‌داری متغیرهای جنس، محل سرطان، سن هنگام تشخیص و اندازه تومور مورد بررسی قرارگرفت. توزیع بیماران و نتایج آزمون لگ‌رتبه و منحنی‌های کاپلان مایر برای متغیرهای جنس، محل سرطان، سن هنگام تشخیص و اندازه تومور در جدول ۱ و نمودار ۱ نشان‌داده شده‌است.

جدول ۱. توزیع بیماران و نتایج آزمون لگ‌رتبه برای متغیرهای جنس، محل سرطان، سن هنگام تشخیص و اندازه تومور

در بیماران مبتلا به سرطان روده بزرگ استخراج‌شده از بیمارستان طالقانی

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

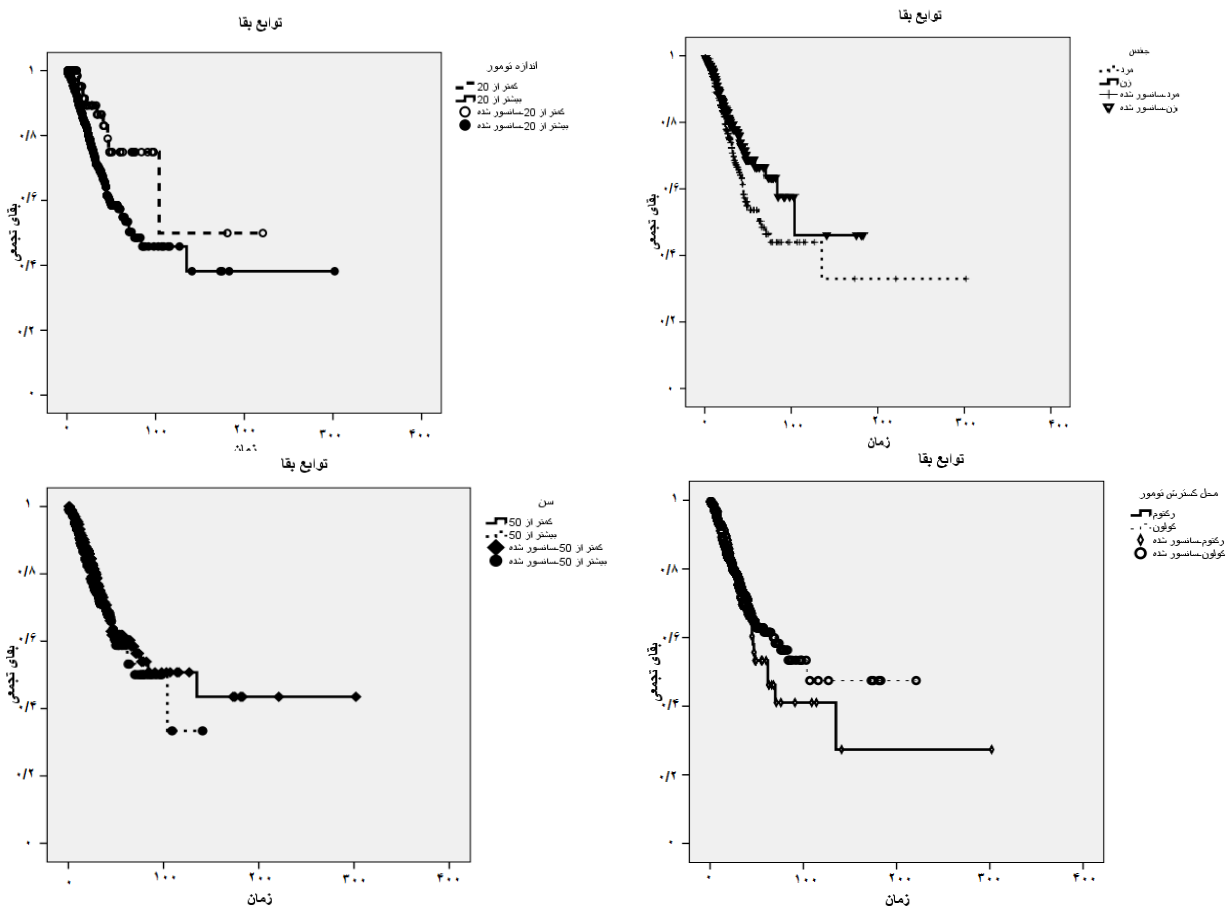
عوامل مرتبط با وضعیت بقای بیماران مدل کاکس به داده‌ها برازش‌داده‌شد و این مدل با منحنی کاپلان مایر با استفاده از تابع زیان برابر مقایسه‌شد. مدل مخاطره متناسب کاکس یا رگرسیون کاکس، روشی نیمه‌پارامتری است که علاوه بر برآورد احتمال‌های بقا در زمان‌های مختلف به بررسی تأثیر متغیرهای کمکی نیز می‌پردازد؛ این مدل به‌صورت زیر بیان‌می‌شود:

$$h(t|X) = h_0(t) \exp(BX)$$

که در آن  $h(t|X)$  نشان‌دهنده نرخ مخاطره در زمان  $t$  برای فردی با متغیرهای کمکی  $X$  و  $h_0(t)$  نرخ خطر پایه است.

در آمار برای تعیین مقدار خطای یک مدل از تابع زیان استفاده‌می‌شود که به زیان مورد انتظار، خطای پیش‌بینی گفته‌می‌شود؛ این مقدار خطا به تابع زیان و نوع متغیر بستگی دارد. یکی از توابع زیان که در مطالعات بقا بسیار کاربرد دارد، تابع زیان برای است که در واقع همان زیان توان دو وضعیت بقای مشاهده‌شده و احتمال بقای

برآورد خطای پیش‌بینی برای وضعیت بقا و کاربرد آن در تحلیل بقای بیماران مبتلا به سرطان روده بزرگ



**نمودار ۱. منحنی‌های کاپلان مایر برای متغیرهای جنس، محل سرطان، سن هنگام تشخیص و اندازه تومور در بیماران مبتلا به سرطان روده بزرگ استخراج‌شده از بیمارستان طالقانی**

با توجه به نتایج مدل کاکس، وضعیت بقای بیماران با عوامل جنس ( $p=0/058$ ) و اندازه تومور ( $p=0/017$ ) ارتباط داشت. نسبت خطر برای زنان،  $0/739$  برابر مردان شد؛ علاوه بر آن، نسبت خطر برای افرادی که اندازه تومورشان بیشتر از  $20$  میلی‌متر است  $2/107$  برابر افرادی شد که اندازه تومورشان کمتر از  $20$  میلی‌متر است. برای مقایسه منحنی کاپلان‌مایر و مدل کاکس از برآورد خطای پیش‌بینی با زیان برابر استفاده شد. خطای پیش‌بینی مدل کاکس از منحنی کاپلان‌مایر کمتر شد و وجود متغیرهای کمکی و برازش مدل کاکس، باعث کاهش خطای پیش‌بینی شد؛ علاوه بر آن با توجه به نمودار خطای پیش‌بینی تا پیش از زمان  $32$  برای دو منحنی یکسان شد اما با گذر زمان، این اختلاف نمایان‌تر شد که نتایج آن در نمودار ۲ نشان‌داده شده است.

پیش از برازش مدل کاکس، فرض خطر متناسب برای متغیرها با استفاده از باقی‌مانده شونفلد بررسی شد و این فرض برای متغیر محل گسترش تومور برقرار نبود ( $p < 0/05$ ) در نتیجه، مدل کاکس طبقه‌بندی به داده‌ها برازش داده شد. نتایج حاصل از مدل رگرسیونی کاکس طبقه‌بندی در جدول ۲ نشان‌داده شده است.

**جدول ۲. نتایج حاصل از برازش مدل رگرسیونی کاکس طبقه‌بندی در بیماران مبتلا به سرطان روده بزرگ بیمارستان طالقانی**

متغیر	ضریب	نسبت مخاطره	انحراف استاندارد ضریب رگرسیونی	p-مقدار
جنس <sup>۱</sup>	-0/303	0/739	0/159	0/058
سن <sup>۲</sup>	0/148	1/159	0/153	0/34
اندازه تومور <sup>۳</sup>	0/745	2/107	0/313	0/017

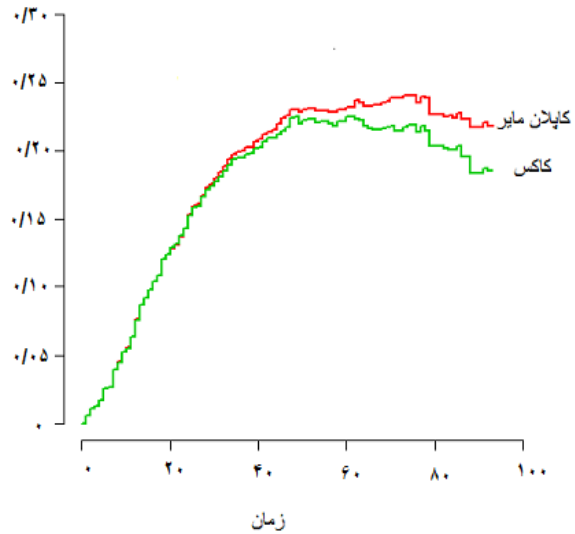
۱. زن به مرد؛ ۲. بالای ۵۰ سال به زیر ۵۰ سال و ۳. بالای ۲۰ میلی‌متر به زیر ۲۰ میلی‌متر

خطای پیش‌بینی این بیماران نشان داده شده است. در مطالعه‌ای که جلالی و همکاران در سال ۱۳۸۳ انجام دادند، به این نتیجه رسیدند که بیماران ایرانی در مقایسه با بیماران گزارش شده در جوامع غربی جوان‌ترند و سن آغاز بیماری در ایران پایین‌تر است؛ در آن مطالعه، تعداد مردانی که به سرطان کولورکتال مبتلا شدند بیشتر از زنان بود (۱۳). در این مطالعه در حالت تک‌متغیره، محل سرطان معنی‌دار نشد که در تضاد با مطالعات (۱۴ و ۱۵) است.

برای مقایسه مدل کاکس و منحنی کاپلان‌مایر از برآورد خطا با زیان برابر استفاده شد. پیشنهاد می‌شود از برآورد خطای پیش‌بینی برای مقایسه مدل‌های پارامتری و نیمه‌پارامتری نیز استفاده شود؛ علاوه بر آن می‌توان خطای پیش‌بینی را با روش اعتبارسنجی متقابل به‌دست آورد و مقایسه‌ای میان آن دو انجام داد.

### نتیجه‌گیری

در این مقاله به این نتیجه رسیده‌ایم که جنس و اندازه تومور روی پیش‌بینی وضعیت بقای بیماران مبتلا به سرطان روده بزرگ تأثیری ندارد. برای مقایسه مدل‌ها می‌توان از برآورد خطای پیش‌بینی با زیان برابر استفاده کرد؛ در این داده‌ها خطای پیش‌بینی برای مدل کاکس، کمتر از منحنی کاپلان‌مایر است و وجود متغیرهای جنس و اندازه تومور، باعث کاهش خطای پیش‌بینی شده است.



نمودار ۲. مقایسه خطای پیش‌بینی منحنی کاپلان‌مایر و مدل کاکس با زیان برابر

### بحث

این پژوهش به منظور شناسایی عوامل مؤثر بر بقای بیماران سرطان روده بزرگ انجام شد. با توجه به یافته‌ها متغیرهای جنس و اندازه تومور هم با توجه به آزمون لگ‌رتبه و هم با برازش مدل کاکس طبقه‌بندی معنی‌دار شدند. نتایج مدل کاکس طبقه‌بندی نشان دادند که مخاطره برای زنان نسبت به مردان  $0/739$  برابر و برای افرادی که اندازه تومورشان بیشتر از ۲۰ بوده است نسبت به افرادی که اندازه تومورشان کمتر از ۲۰ بوده است مخاطره  $2/107$  برابر شد. مطالعه ۸ به همین نتیجه رسید و نشان داد که زمان بقای زنان در این سرطان بالاتر از زمان بقای مردان است. مطالعه ۹ تأثیرگذار بودن اندازه تومور را تأیید کرد. سن زمان تشخیص متغیری است که در مطالعه ۱۰ به معنادار بودن تأثیر آن اشاره شده است؛ اما در این پژوهش، این متغیر معنی‌دار نشد؛ البته مطالعاتی وجود دارند که عدم تأثیر سن را نشان می‌دهند (۱۱)؛ شاید این تناقض به دلیل نقاط برش متفاوت در مطالعات مختلف و ناحیه درگیر سرطان (کولون، رکتوم) باشد. در مطالعه ۱۲ که نواحی درگیر به طور مجزا بررسی شده‌اند، سن هنگام تشخیص به‌عنوان یکی از عوامل مهم و تأثیرگذار در زمان بقای

### منابع

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA: a Cancer Journal for Clinicians. 2005;55(2):74-108.
2. Wickham R, Lassere Y, editors. The ABCs of colorectal cancer. Seminars in Oncology Nursing; 2007: Elsevier.
3. Lawless JF, Yuan Y. Estimation of prediction error for survival models. Statistics in Medicine. 2010;29(2):262-74.
4. Schemper M. Predictive accuracy and explained variation. Statistics in Medicine. 2003;22(14):2299-308.
5. Gerds TA, Schumacher M. Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. Biometrical Journal. 2006;48(6):1029-40.
6. Korn EL, Simon R. Measures of explained variation for survival data. Statistics in Medicine. 1990;9(5):487-503.
7. Henderson R, Jones M, Stare J. Accuracy of point predictions in survival analysis. Statistics in Medicine. 2001;20(20):3083-96.
8. Dupont M, Pampalon R, Hamel D. Deprivation and cancer mortality among Quebec women and men, 1994-1998: Institut National de Santé Publique du Québec; 2004.
9. Bear HD, MacIntyre J, Burns HJ, Jarrett F, Wilson RE. Colon and rectal carcinoma in the west of Scotland: Symptoms, histologic characteristics, and outcome. The American Journal of Surgery. 1984;147(4):441-6.
10. Molaei M, Mansoori BK, Ghiasi S, Khatami F, Attarian H, Zali M. Colorectal cancer in Iran: immunohistochemical profiles of four mismatch repair proteins. International Journal of Colorectal Disease. 2010;25(1):63-9.
11. Moghimi-Dehkordi B, Safaee A, Zali MR. Prognostic factors in 1,138 Iranian colorectal cancer patients. International Journal of Colorectal Disease. 2008;23(7):683-8.
12. Asghari Jafarabadi M, Hajizadeh E, Kazemnejad A, Fatemi S. Comparison the role of BMI, pathologic stage and hereditary related factors on survival between colon and rectal cancers: frailty competing risks model. Iranian Journal of Epidemiology. 2010;6(3):35-49.
13. Jalali S, Kordjazi I, Jalali S. Epidemiological characteristics of colorectal cancer in patients referred to Imam Khomeini Hospital during (1981-2001). Journal of Iran University of Medical Sciences. 2005.
14. 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. Diseases of the Colon & Rectum. 2009;52(8):1359-66.
15. Meguid RA, Slidell MB, Wolfgang CL, Chang DC, Ahuja N. Is there a difference in survival between right-versus left-sided colon cancers? Annals of Surgical Oncology. 2008;15(9):2388-94.

Daneshvar

Medicine

*Scientific-Research  
Journal of Shahed  
University  
21st Year, No.111  
June- July, 2014*

Received: 13/04/2014

Last revised: 11/06/2014

Accepted: 16/06/2014

## Estimation of prediction error for survival status with application in colorectal cancer

Mahmoudreza Gohari<sup>1</sup>, Ahmadreza Baghestani<sup>2</sup>, Arezou Orooji<sup>3\*</sup>, Mohammadamin Pourhosseingholi<sup>4</sup>

1. Department of Biostatistics, Iran University of Medical Sciences, Tehran, Iran
2. Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3. Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Tehran, Iran
4. Department of Health System Research, Gastroenterology and Liver Disease Research Center, Tehran, Iran

**E-mail: arezoo.orooji23@gmail.com**

### Abstract

**Background and Objective:** Colorectal cancer is one of the most widespread and fatal cancers. It is important that we could predict what status people have in the future. The purpose of this study was comparison of the Cox model and Kaplan-Meier curve with IBS and also identifying the factors about predicted survival time of patients with colon cancer.

**Materials & Methods:** This paper is related to colorectal cancer patients in Tehran Taleghani hospital that at least were followed up for five years. These patients referred to Taleghani hospital from 1380 to 1385 and their information was recorded in the cancer section of Digestion Research Center and at least was followed up for five years. The Cox Proportional Hazards model and Kaplan-Meier curve were used for prediction of survival status in different times and these were compared with estimation of prediction error by IBS method.

**Results:** Among 724 patients, there was 422 men (58.3%) and 302 women (41.7%). A total of 177 patients died, among them 65% were men. Survival median of patients was  $104 \pm 23.3$ . Cox model showed significant relation between size of tumor and sex with survival time. Sometimes, prediction error with IBS method for Cox model was less than Kaplan-Meier curve and Cox model is more perfect for these data.

**Conclusion:** Presence of covariates and the use of models in predicting survival of patients is effective. In this study, the presence of covariates and Cox model reduces the prediction error.

**Keywords:** Cox proportional hazards model, Kaplan-Meier curve, Prediction error, Colorectal cancer