

Evaluation of parametric models with estimating the prediction error by apparent loss method in analyzing survival of colorectal patients

Baghestani AR¹, Gohari MR², Orooji A^{1*}, Pourhosseingholi MA³

1- Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2- Department of Biostatistics, Hospital Management Research Center, Iran University of Medical Sciences, Tehran, Iran

3- Gastroenterology Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Received: 2 Jun 2014, Accepted: 17 Dec 2014

Abstract

Background: Colorectal cancer is the most common gastrointestinal cancer. Investigating the factors that predict survival time for these patients is important. The purpose of this study was comparison of parametric models by estimating the prediction error and also identifying the effective factors on predicted survival time of patients with colorectal cancer.

Materials and Methods: This cohort study was conducted with 600 patients who were suffered from colorectal cancer in Taleghani Hospital of Tehran between 2001 to 2005 and they were followed up for at least 5 years. For identifying the effective factors on survival time, of the patients we analyzed the data by some parametric models such as Weibull, Exponential and Log logistic and compared these models with the estimation of prediction error by apparent loss method.

Results: Among 600 patients there was 344 men (57.3%) and 256 women (42.7%). Of total, 151 patients were died that 62.3% of them were men. Univariate analysis showed that the effect of BMI, sex, staging of tumor, tumor site were significant but in multivariate model staging of tumor and BMI were significant. By the estimation of prediction error, the best model was Log logistic.

Conclusion: With respect to the importance of survival time prediction, we found that we can use the prediction error to compare the parametric models. In addition, because of effectiveness of tumor stages and BMI in the patients' survival time, survival time could be increased by an on-time diagnosis and an appropriate controlled diet.

Keywords: Colorectal Cancer, Prediction, Parametric Model

*Corresponding Author:

Address: Department of Biostatistics, Faculty of paramedical, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Email: arezoo.orooji23@gmail.com

ارزیابی مدل‌های پارامتری با برآورد خطای پیش‌بینی به روش زیان آشکار در بررسی بقای بیماران مبتلا به سرطان روده بزرگ

احمدرضا باغستانی^۱، محمودرضا گوهری^۲، آرزو اروچی^{۳*}، محمد امین پورحسینقلی^۴

۱- استادیار، گروه آمار زیستی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

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

۳- دانشجوی کارشناسی ارشد آمار زیستی، گروه آمار زیستی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

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

تاریخ دریافت: ۹۳/۳/۱۲ تاریخ پذیرش: ۹۳/۹/۲۶

چکیده

زمینه و هدف: سرطان کولورکتال شایع‌ترین سرطان دستگاه گوارش می‌باشد. بررسی عوامل مؤثر بر زمان بقای این بیماران حائز اهمیت است. هدف از این پژوهش مقایسه مدل‌های پارامتری با برآورد خطای پیش‌بینی و شناسایی عوامل مؤثر بر زمان بقای بیماران مبتلا به سرطان روده بزرگ است.

مواد و روش‌ها: مطالعه حاضر از نوع مطالعات هم‌گروهی بود. در این مطالعه ۶۰۰ بیمار مبتلا به سرطان روده بزرگ طی سال‌های ۱۳۸۰ تا ۱۳۸۵ به بیمارستان طالقانی شهر تهران مراجعه نمودند و حداقل ۵ سال پی‌گیری شدند. برای تعیین عوامل مؤثر بر زمان بقای بیماران مدل‌های پارامتری وایبل، نمایی و لگ-لجستیک به داده‌ها برازش داده شد و این مدل‌ها با برآورد خطای پیش‌بینی به روش زیان آشکار با یکدیگر مقایسه شدند.

یافته‌ها: از ۶۰۰ بیمار، ۳۴۴ (۵۷/۳ درصد) نفر مرد و ۲۵۶ (۴۲/۷ درصد) نفر زن بودند. از کل بیماران ۱۵۱ (۲۵/۲ درصد) نفر فوت کردند که ۶۲/۳ درصد آنها مرد بودند. در حالت تک متغیره زمان بقای بیماران با عوامل مرحله بیماری، شاخص توده بدنی، محل سرطان و جنسیت بیمار ارتباط معنی‌دار داشت اما در حالت چند متغیره تنها مرحله بیماری و شاخص توده بدنی عوامل مؤثر بر بقای این بیماران بودند. با استفاده از برآورد خطای پیش‌بینی به روش زیان آشکار، مدل لگ لجستیک بهترین مدل بود.

نتیجه‌گیری: با توجه به این که پیش‌بینی زمان بقای بیماران از اهمیت بالایی برخوردار است لذا می‌توان نتیجه گرفت برای مقایسه مدل‌ها از برآورد خطای پیش‌بینی استفاده شود. علاوه بر آن از آنجایی که مرحله بیماری و شاخص توده بدنی از عوامل مؤثر در زمان بقای بیماران شناخته شد می‌توان با تشخیص به موقع و رعایت رژیم غذایی مناسب زمان بقای بیماران را افزایش داد.

واژگان کلیدی: سرطان روده بزرگ، پیش‌بینی، مدل پارامتری

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

Email: arezoo.orooji23@gmail.com

مقدمه

در علم پزشکی پیش‌بینی از اهمیت ویژه‌ای برخوردار است و پزشکان باید بتوانند زمان بقای بیماران را پیش‌بینی کنند و متناسب با وضعیت بیمار درمان‌های مناسب را تجویز نمایند. علاوه بر آن پیش‌بینی زمان بقا برای مطلع ساختن بیمار نیز از اهمیت بالایی برخوردار است. اما حقیقت این است که اختلاف چشم‌گیری بین پیش‌بینی زمان بقای پزشکی و بقای حقیقی وجود دارد و به این علت است که برای بررسی این موضوع باید از مدل‌های آماری استفاده کنیم (۱). آماردان‌ها از مدل‌ها برای تقریب رابطه‌ای که بین یک مجموعه متغیر کمکی و متغیر پیامد است، استفاده می‌کنند و هدف اصلی آنها از ساخت مدل توضیح و پیش‌بینی متغیر پاسخ با استفاده از متغیرهای کمکی می‌باشد. پیش‌بینی در بقا می‌تواند با استفاده از مدل‌های پارامتری مانند نمایی، وایبل، لگ نرمال و لگ-لجستیک انجام شود (۲). مدل‌های بقا یا مدل‌های زمان شکست از جمله روش‌های آماری می‌باشند که در آنها متغیر پاسخ زمان تا رخداد واقعه می‌باشد 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 چالش برانگیز داده‌های بقاست که آنها را از سایر مدل‌های آماری مجزا می‌کند (۳). دقت پیش‌بینی توسط اندازه‌هایی که معمولاً براساس توابع زیان است مشخص می‌شود (۴)، اما به علت وجود سانسور باید از روش‌هایی مثل جانهای (imputation) و وزن دهی (weighting) استفاده کرد (۵).

سرطان کولورکتال یکی از شایع‌ترین بدخیمی‌ها و شایع‌ترین سرطان دستگاه گوارش می‌باشد، به طوری که سالانه ۶۰۰۰۰۰ مورد جدید آن در آمریکا تشخیص داده می‌شود (۶). در سال‌های اخیر طول عمر این بیماران بهبود یافته است (۷). علت این بهبود کاملاً مشخص نیست. تغییر در سیر طبیعی بیماری، تغییر عادات تغذیه‌ای و پیشرفت در روش‌های تشخیصی از علل احتمالی عنوان شده‌اند. کرن و سیمون در سال ۱۹۹۰، هندرسون در سال ۱۹۹۵ و گرف و همکاران در سال ۱۹۹۹ در مورد برآورد خطای پیش‌بینی با

استفاده از شیوه تابع زیان بحث کردند (۲، ۸، ۹). هندرسون و همکاران در سال ۲۰۰۱، کیدینگ و همکاران در سال ۲۰۰۴، کرن و سیمون در سال ۱۹۹۰ در مورد تابع زیان مناسب برای داده‌های بقا بحث کردند (۸، ۱۰). جرال لاولس و یون در سال ۲۰۰۹ در مورد برآورد خطای پیش‌بینی زمانی که از مدل‌های رگرسیونی برای پیش‌بینی زمان بقا استفاده می‌شود، بحث کردند. آنها پیش‌بینی زمان بقا را با استفاده از برآوردهای نقطه‌ای و فاصله‌ای به دست آوردند و هم‌چنین انتخاب مدل و متغیر را مورد بررسی قرار دادند (۴). هدف از این پژوهش استفاده از مدل‌های پارامتری برای تعیین عوامل مؤثر بر پیش‌بینی زمان بقای بیماران مبتلا به سرطان روده بزرگ بخش گوارش بیمارستان طالقانی و انتخاب مدل بهتر برای این داده‌هاست.

مواد و روش‌ها

مطالعه حاضر از نوع مطالعات هم‌گروهی بود. در این مطالعه پرونده ۶۰۰ بیمار مبتلا به سرطان روده بزرگ که طی سال‌های ۱۳۸۰ تا ۱۳۸۵ برای درمان به بیمارستان طالقانی شهر تهران مراجعه نمودند مورد بررسی قرار گرفت. برای بیماران متغیرهای سن بیمار، جنسیت بیمار، اندازه تومور، شاخص توده بدنی، بافت شناسی تومور، محل گسترش تومور، سابقه خانوادگی، مصرف دخانیات، مشاهده شدن خون، کاهش وزن، آخرین وضعیت بیمار، زمان مرگ ثبت شد. اما به علت این که برای بعضی از بیماران بعضی از متغیرهای کمکی ثبت نشده بود تنها به بررسی متغیرهای جنس، شاخص توده بدنی، بافت شناسی، مرحله سرطان، سابقه خانوادگی و محل گسترش تومور پرداختیم. این افراد حداقل به مدت ۵ سال پی‌گیری شدند. سپس مدل‌های وایبل، نمایی و لگ-لجستیک به داده‌ها برازش داده شد و برای مقایسه پیش‌بینی آنها از برآورد خطای پیش‌بینی به روش زیان آشکار استفاده شد. هر چه خطا کمتر باشد مدل، مدل مناسب‌تری است. مدل پارامتری مدلی است که در آن فرض می‌شود زمان بقا از توزیع شناخته شده‌ای پیروی کند. وقتی این مدل‌های پارامتری برازش مناسبی به داده‌ها فراهم

برای تجزیه و تحلیل داده‌ها از نرم افزار R استفاده شد. سطح معنی داری نیز ۰/۰۵ در نظر گرفته شد.

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

یافته‌ها

از ۶۰۰ بیمار، ۳۴۴ (۵۷/۳ درصد) نفر مرد و ۲۵۶ (۴۲/۷ درصد) نفر زن بودند. از کل بیماران ۱۵۱ (۲۵/۲ درصد) نفر فوت کردند که ۶۲/۳ درصد آنها مرد بودند. ابتدا با استفاده از آزمون لگرتبه معنی داری متغیرهای جنس، بافت شناسی، محل گسترش تومور، شاخص توده بدنی و سابقه خانوادگی مورد بررسی قرار گرفت. توزیع بیماران و نتایج آزمون لگرتبه برای این متغیرها در جدول ۱ نشان داده شده است.

نمایند، برآوردهای دقیق‌تری (نسبت به مدل نیمه پارامتری) از کمیت‌های مورد بررسی به دست می‌آید چرا که این برآوردها مبتنی بر تعداد پارامترهای کمتری است.

خطای پیش‌بینی زیان آشکار

زیان مورد انتظار را خطای پیش‌بینی گویند که بستگی به تابع زیان و نوع متغیر دارد. یکی از روش‌های برآورد خطای پیش‌بینی استفاده از شیوه زیان آشکار است و به صورت زیر است. t_i زمان بقا، x_i متغیرهای کمکی و L تابع زیان مورد نظر و $G(x_i, \theta)$ بهترین پیش‌بینی توسط متغیرهای کمکی است و بستگی به نوع تابع زیان دارد و انتخاب تابع زیان بستگی به متغیر مورد نظر دارد. از آنجایی که در مطالعات بقا توزیع زمان بقا چوله می‌باشد، استفاده از زیان قدرمطلق خطا توصیه می‌شود.

جدول ۱. توزیع بیماران و نتایج آزمون لگرتبه برای متغیرهای جنس، بافت شناسی، محل گسترش تومور، شاخص توده بدنی و سابقه خانوادگی در بیماران مبتلا به سرطان روده بزرگ استخراج شده از بیمارستان طالقانی

متغیر	تعداد بیماران (درصد)	تعداد شکست‌ها (درصد)	میانگین زمان بقا ± انحراف معیار	آماره آزمون	p
شاخص توده بدنی	کمتر از ۲۵ (۶۰/۳)۳۶۲	۱۱۳ (۳۱/۲)	۷۱/۵۶۵ ± ۵/۰۵۵	۲۸/۵۵	< ۰/۰۰۰۱
	بیشتر از ۲۵ (۳۹/۷)۲۳۸	۳۸ (۱۶)	۱۲۸/۰۳۲ ± ۹/۷۱۹		
محل سرطان	کولون (۶۹/۷)۴۱۸	۹۶ (۲۳)	۱۰۹/۱۶۶ ± ۶/۹۹	۴/۹۳۶	۰/۰۲۶
	رکتوم (۳۰/۳)۱۸۲	۵۵ (۳۰/۲)	۷۷/۳۲ ± ۶/۸۲		
سابقه خانوادگی	دارد (۵۷/۵)۳۴۵	۸۸ (۲۵/۵)	۷۴/۵۵۷ ± ۶/۵۲۳	۲/۰۳۴	۰/۱۵۴
	ندارد (۴۲/۵)۲۵۵	۶۳ (۲۴/۷)	۱۱۵/۱۷۸ ± ۶/۸۱۲		
بافت شناسی	موسینوسی (۱۰/۷)۶۴	۱۸ (۲۸/۱)	۱۰۵/۱۱۳ ± ۱۴/۰۰۶	۰/۷۶۶	۰/۳۸۱
	غیر موسینوسی (۸۹/۳)۵۳۶	۱۳۳ (۲۴/۸)	۹۹/۷۹۹ ± ۶/۳۶۹		
جنس	زن (۴۲/۷) ۲۵۶	۵۷ (۲۲/۳)	۱۱۱/۵۰۳ ± ۹/۳۴۸	۳/۸۷	۰/۰۴۹
	مرد (۵۷/۳)۳۴۴	۹۴ (۲۷/۳)	۸۸/۱۱۴ ± ۸/۳۳۴		

نسبت خطر (بخت) محاسبه شده برای هر سه مدل تقریباً یکسان شد. با توجه به نتایج بهترین مدل (لگ لجستیک) نسبت بخت برای افرادی که در مرحله پیشرفته بیماری قرار دارند ۱/۹۷ برابر افرادی است که در مرحله اولیه بیماری هستند. علاوه بر آن نسبت بخت برای افرادی که شاخص توده بدنی آنها کمتر از ۲۵ است، ۲/۱۹ برابر افرادی است که شاخص توده بدنی آنها بیشتر از ۲۵ است.

سپس به دلیل اهمیت بالینی این متغیرها تمام آنها در مدل‌های مختلف بقا (وایبل، نمایی و لگ-لجستیک) وارد شد، برای هر سه مدل متغیرهای شاخص توده بدنی و مرحله سرطان معنی دار شد ($p < ۰/۰۰۰۱$). پس از آن خطای پیش‌بینی به روش زیان آشکار برای این مدل‌ها محاسبه شد. مقدار خطا برای مدل وایبل ۲/۴۳، مدل نمایی ۲/۵۸ و مدل لگ-لجستیک ۲/۰۳ محاسبه شد.

بحث

در این مطالعه در حالت تک متغیره و با توجه به آزمون لگ رتبه متغیرهای جنس، شاخص توده بدنی، محل گسترش تومور و مرحله سرطان معنی دار شد اما در حالت چند متغیره تنها متغیرهای شاخص توده بدنی و مرحله سرطان معنی دار شدند. نسبت خطر (بخت) محاسبه شده برای هر سه مدل تقریباً یکسان شد. در این مطالعه که به منظور ارزیابی مدل های پارامتری با برآورد خطای پیش بینی به روش زیان آشکار انجام شد مقدار خطای پیش بینی به روش زیان آشکار برای مدل لگ لجستیک از همه کمتر و این مدل به عنوان بهترین مدل پیش بینی انتخاب شد. در مطالعه ای که توسط روشی و همکاران انجام شد مدل وایبل بهترین برآزش را داشت اما در مطالعه ای که توسط آنها انجام شد مدل های پارامتری با استفاده از معیار آکائیکه با یکدیگر مقایسه شدند در حالی که در این مطالعه به علت اهمیت پیش بینی در مطالعات بقا این مدل ها با برآورد خطای پیش بینی با یکدیگر مقایسه شدند (۱۱). با توجه به نتایج مدل رگرسیونی لگ لجستیک متغیر جنس معنی دار نشد. مطالعه ای بالا بودن زمان بقای زنان نسبت به مردان را نشان می دهد (۱۲). مطالعه دیگری بیش تر بودن مرگ و میر مردان نسبت به زنان را در این سرطان نشان می دهد (۱۳). در این مطالعه محل گسترش تومور معنی دار نشد که در تضاد با مطالعه های مگد و همکاران و ری و همکاران است (۱۴، ۱۵). در آنالیزهای تک متغیره و چند متغیره مرحله بیماری به عنوان فاکتور مهمی در پیش آگهی بیماران می باشد به طوری که بقای بیماران با بالا رفتن مرحله بیماری کاهش می یابد و هر چه مرحله بیماری پایین تر باشد، درمان ساده تر و راحت تر بوده و بقا بیماران بهتر خواهد بود (۱۶). در مطالعاتی که توسط فلود و همکاران (۱۷، ۱۸) و وی و همکاران (۱۹) صورت گرفته چنین نتیجه گیری شد که اثر متغیرهای تغذیه ای بر سرطان کولون و رکتوم متفاوت است. در این مطالعه شاخص توده بدنی معنی دار شد. مطالعه ای که توسط آرتس و همکاران انجام شد این نتیجه را تایید می کند (۱۲). اما در مطالعه ریگلی شاخص توده بدنی

معنی دار نشد (۱۳). در این مطالعه سابقه خانوادگی معنی دار نشد. فوجس و همکاران (۲۰) و وی و همکاران (۱۹) نشان دادند که تاثیر سابقه خانوادگی در ابتلا به سرطان کولون و رکتوم متفاوت است.

نتیجه گیری

از آنجایی که پیش بینی زمان بقای بیماران اهمیت زیادی دارد، برای مقایسه مدل ها در داده های بقا و پزشکی از برآورد خطای پیش بینی به روش زیان آشکار می توان استفاده کرد و علاوه بر آن پیشنهاد می شود به مقایسه روش های برآورد خطای پیش بینی پرداخته شود و همچنین از برآورد خطای پیش بینی برای مقایسه مدل های دیگر همراه با فریلتی (frailty) استفاده شود.

تشکر و قدردانی

این مقاله برگرفته از پایان نامه دوره کارشناسی ارشد آمار زیستی مصوب دانشگاه علوم پزشکی شهید بهشتی تحت عنوان برآورد خطای پیش بینی برای مدل های بقا و کاربرد آن در تحلیل بقای بیماران مبتلا به سرطان روده بزرگ می باشد. بدین وسیله بر خود لازم می دانیم از کلیه افرادی که در مرکز تحقیقات گوارش با ما همکاری داشتند تشکر و قدردانی نماییم. قابل به ذکر است که در این مطالعه از هیچ منبعی کمک مالی دریافت نشده است.

منابع

1. Henderson R, Keiding N. Individual survival time prediction using statistical models. *Journal of Medical Ethics*. 2005;31(12):703-6.
2. Henderson R. Problems and prediction in survival-data analysis. *Statistics in medicine*. 1995;14(2):161-84.
3. Rosthøj S, Keiding N. Explained variation and predictive accuracy in general parametric statistical models: the role of model misspecification. *Probability, Statistics and Modelling in Public Health*: Springer; 2006. p. 392-404.

4. Lawless JF, Yuan Y. Estimation of prediction error for survival models. *Statistics in medicine*. 2010;29(2):262-74.
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. Casciato DA, Territo MC. *Manual of clinical oncology*. 4thed. New York: Lippincott Williams & Wilkins Inc; 2009.p.182-3.
7. De Leon MP, Di Gregorio C. Pathology of colorectal cancer. *Digestive and Liver Disease*. 2001;33(4):372-88.
8. Korn EL, Simon R. Measures of explained variation for survival data. *Statistics in medicine*. 1990;9(5):487-503.
9. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*. 1999;18(17-18):2529-45.
10. Henderson R, Jones M, Stare J. Accuracy of point predictions in survival analysis. *Statistics in medicine*. 2001;20(20):3083-96.
11. Roshanaie G, Komijani A, Sadighi A, Faradmal J. Prediction of survival in patients with colorectal cancer referred to the Hamadan MRI center using Weibull parameter model and determination of its risk factor during 2005-2013. *Arak Journal of Medical Sciences*. 2014; 16(80):41- 9.[Persian]
12. 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. *European Journal of Cancer*. 2010;46(15):2681-95.
13. 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 survival. *Journal of epidemiology and community health*. 2003;57(4):301-9.
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.
16. Chapuis P, Dent O, Fisher R, Newland R, Pheils M, Smyth E, et al. A multivariate analysis of clinical and pathological variables in prognosis after resection of large bowel cancer. *British journal of surgery*. 1985;72(9):698-702.
17. Flood A, Rastogi T, Wirfalt E, Mitrou PN, Reedy J, Subar AF, et al. Dietary patterns as identified by factor analysis and colorectal cancer among middle-aged Americans. *The American journal of clinical nutrition*. 2008; 88(1):176-84.
18. Flood A, Peters U, Chatterjee N, Lacey JV, Schairer C, Schatzkin A. Calcium from diet and supplements is associated with reduced risk of colorectal cancer in a prospective cohort of women. *Cancer Epidemiology Biomarkers & Prevention*. 2005;14(1):126-32.
19. Wei EK, Giovannucci E, Wu K, Rosner B, Fuchs CS, Willett WC, et al. Comparison of risk factors for colon and rectal cancer. *International journal of cancer*. 2004; 108(3): 433-42.
20. Fuchs CS, Giovannucci EL, Colditz GA, Hunter DJ, Speizer FE, Willett WC. A prospective study of family history and the risk of colorectal cancer. *New England Journal of Medicine*. 1994;331(25):1669-74.