

## **Application of penalized spline model for determining the factors affecting the survival rate of gastric cancer patients**

Noorkojuri H(M.Sc)<sup>1</sup>, Hajizadeh E(Ph.D)<sup>1\*</sup>, Baghestani AR(Ph.D)<sup>2</sup>, Pourhoseingholi MA(Ph.D)<sup>3</sup>

1- Department of Biostatistics, Tarbiat Modares University, Tehran, Iran

2- Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

3- Department of Biostatistics, Gastroenterology and Liver Diseases Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Received: 3 Jul 2012, Accepted: 3 Oct 2012

### **Abstract**

**Background:** Gastric cancer is the second leading cause of mortality due to cancer worldwide and it is the most common type of cancer in Iran. The aim of this study was to assess the effects of prognostic factors on the survival of patients with gastric cancer using the penalized spline in Cox model and compare it with Cox proportional hazards model.

**Materials and Methods:** In this retrospective study, information obtained from a total of 216 gastric cancer patients who underwent surgery at Taleghani Hospital of Tehran between 2003 and 2008 years was included. Cox proportional hazards model and penalized spline in Cox model were used. R software was used for data analysis. The efficacy of these models was compared according to Akaike information criterion.

**Results:** The five-year survival rate was 30% and the mean follow-up time was 16.60±15.28 months. Survival mean and median were 46 and 30 months, respectively. The analysis of Cox proportional hazards and penalized spline models resulted in age at diagnosis and tumor size as prognostic factors for survival time ( $P<0.05$ ). Also, Akaike information criterion and survival curve for patients with a tumor size over 35 mm and age at diagnosis over 45 years were equal in both models.

**Conclusion:** Cox proportional hazards and penalized spline models generated similar results; thus, it is more appropriate to use Cox proportional hazards model because of its ease of interpretation and capability of modeling both continuous and discrete covariates. This study also showed if gastric cancer is diagnosed early, the relative risk of death will reduce.

**Keywords:** Cox proportional hazards model, gastric cancer, penalized spline model

\*Corresponding author:

\* Address: Tarbiat Modares University, Jalal Ale-Ahmad Highway, Tehran, Iran

Email: hajizadeh@modares.ac.ir

## استفاده از مدل اسپلین جریمه شده در تعیین عوامل موثر بر بقای بیماران مبتلا به سرطان معده

هدی نورکجوری<sup>1</sup>، ابراهیم حاجی زاده<sup>2\*</sup>، احمدرضا باغستانی<sup>3</sup>، محمد امین پورحسینقلی<sup>4</sup>

1- دانشجوی کارشناسی ارشد آمار زیستی، دانشگاه تربیت مدرس، تهران، ایران

2- دانشیار، گروه آمار زیستی، دانشگاه تربیت مدرس، تهران، ایران

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

4- استادیار، گروه آمار زیستی، مرکز تحقیقات بیماری‌های گوارش و کبد، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

تاریخ دریافت: 91/4/13 تاریخ پذیرش: 91/7/12

### چکیده

**زمینه و هدف:** سرطان معده دومین علت مرگ ناشی از سرطان‌ها در سراسر دنیا است و در ایران شایع‌ترین سرطان به شمار می‌رود. این مطالعه به منظور تحلیل عوامل مرتبط با بقای بیماران مبتلا به سرطان معده با استفاده از اسپلین جریمه شده در مدل کاکس و مقایسه آن با مدل مخاطرات متناسب کاکس انجام پذیرفت.

**مواد و روش‌ها:** در این مطالعه گذشته‌نگر، 216 بیمار مبتلا به سرطان معده که طی سال‌های 1382 تا 1387 در بیمارستان طالقانی تهران تحت عمل جراحی قرار گرفتند وارد مطالعه شدند. برای تحلیل داده‌ها از مدل‌های مخاطرات متناسب کاکس و اسپلین جریمه شده در کاکس استفاده گردید. کلیه محاسبات با نرم افزار R انجام شد. معیار مقایسه کارایی مدل‌ها، ملاک آکاییکه بود.

**یافته‌ها:** میزان بقای پنج ساله بیماران مورد بررسی 30 درصد به دست آمد. میانگین مدت زمان پی‌گیری  $16/60 \pm 15/28$  ماه بود. میانگین و میانه بقای بیماران به ترتیب 46 و 30 ماه بود. متغیرهای سن هنگام تشخیص و اندازه تومور با طول عمر بیماران ارتباط معنی‌داری داشتند ( $p < 0/05$ ). ملاک آکاییکه و نمودار بقای افراد با اندازه تومور بالای 35 میلی‌متر و سن بالای 45 سال نیز در هر دو مدل یکسان شده است.

**نتیجه‌گیری:** با توجه به این که ملاک آکاییکه در هر دو مدل یکسان شد، استفاده از مدل مخاطرات متناسب کاکس در داده‌های پزشکی به دلیل تفسیر راحت‌تر و محدود نبودن در استفاده از انواع متغیرهای کمکی، مناسب‌تر است. همچنین این مطالعه نشان داد که در صورت تشخیص این بیماری در مراحل اولیه پیشرفت بیماری، خطر نسبی مرگ کاهش می‌یابد.

**واژگان کلیدی:** مدل مخاطرات متناسب کاکس، سرطان معده، مدل اسپلین جریمه شده

\* نویسنده مسئول: تهران، جلال آل احمد، پل نصر، دانشگاه تربیت مدرس

## مقدمه

معده بافتی غده‌ای دارد که از آن انواع آنزیم‌ها و مواد هضم کننده‌ی غذا ترشح می‌شود. غذا پس از ورود به معده برای چند ساعت در آن توقف کرده و در این مدت دچار هضم شیمیایی و فیزیکی (با حرکات معده) می‌شود. سپس برای هضم بیشتر و جذب، وارد روده کوچک می‌گردد. ترشح موکوس محافظتی غلیظ از برخی سلول‌های معده، مانع آسیب دیدن آنها در برابر اسید می‌شود. سرطان معده عبارت است از رشد بدون کنترل سلول‌های بدخیم در معده که در آن بیشتر افراد تا مراحل پیشرفته بیماری، علامتی ندارند. سرطان معده سرطانی شایعی در سرتاسر جهان است که میزان آن در برخی کشورها چون ایران به مراتب بالاتر از متوسط جهانی است. ژاپن، کره جنوبی، چین، شیلی و کاستاریکا میزان بالایی از مبتلایان را به خود اختصاص داده‌اند. محققان علت این امر را عادات و روش‌های غلط زندگی عنوان می‌کنند. سرطان معده معمولاً بدون علامت است و تنها با بررسی‌های دقیق و مکرر در افراد در معرض خطر قابل تشخیص است. متأسفانه در بیش از نیمی از حالات پیشرفته سرطان معده، علامتی مشاهده نمی‌شود و هنگامی که دست‌اندازی می‌کند تشخیص داده می‌شود (1). بر مبنای آخرین آمار مرکز تحقیقات سرطان کشور، سرطان معده شایع‌ترین سرطان در کشور و سومین سرطان شایع در میان زنان ایرانی پس از سرطان سینه و روده بزرگ است (2). شروع بیماری به طور معمول از دهه چهارم زندگی است و با افزایش سن، شیوع بیماری نیز افزایش می‌یابد به طوری که حداکثر شیوع در دهه هفتم زندگی برای مردان و به خصوص سنین بالاتر در زنان است (4).

در تحلیل بقای کلاسیک، یک مجموعه از افراد تحت مشاهده از زمان‌های مختلف وارد مطالعه شده و تا زمان رخداد مورد نظر یا پایان مطالعه مورد بررسی قرار می‌گیرند که در این میان ممکن است پیامد مورد انتظار برای بعضی از افراد رخ ندهد که زمان بقای آنها به عنوان سانسور شده نامیده می‌شود. داده‌های سانسور شده یکی از مهم‌ترین مشخصه‌های داده‌های بقا است که در اثر گم شدن فرد در

طول مطالعه، خارج شدن از مطالعه و یا پایان مطالعه رخ می‌دهد (5). مدل کاکس یک مدل نیمه پارامتری برای تحلیل داده‌های سانسور شده بقا است. این مدل هیچ پیش فرضی درباره شکل تابعی خطر ندارد، در عوض می‌توان خطر پایه را از داده‌ها برآورد کرد (6). در مدل کاکس همواره فرض می‌شود که اثر متغیرهای کمکی روی لگاریتم تابع خطر، تابعی خطی از متغیرهای کمکی باشد. حال آن‌که متغیرهای پیوسته می‌توانند به صورت شکل‌های غیرخطی پیچیده‌ای بر روی میزان خطر اثر بگذارند در مدل کاکس برای یافتن اثر غیرخطی متغیر پیوسته می‌توان مستقیماً شکل غیرخطی متغیر را وارد مدل کرد و با استفاده از آزمون‌های مرتبط تصمیم گرفت که آیا اثر غیرخطی در مدل باقی بماند یا خیر (7،8).

یک روش جایگزین که در دهه اخیر به وفور در مطالعات اپیدمیولوژیکی مورد استفاده قرار گرفته است، اسپلاین‌ها (Splines) می‌باشند. اسپلاین‌ها توابع چند جمله‌ای تکه‌ای هستند که محدود به ملحق شدن به نقاط کنترل مشخصی به اسم گره (Knot) می‌شوند (9). در حالت کلی، توابع اسپلاین در جاهایی مورد استفاده قرار می‌گیرند که خطر نسبی دارای برازش غیرخطی باشد. چنین روش‌های هموار سازی، در اپیدمیولوژی زیست محیطی و شغلی مورد استفاده قرار می‌گیرد (10). به عنوان مثال، اسپلاین هموار شده در مدل‌های جمعی تعمیم یافته برای تعیین رابطه بین مواجهه سیلیس و سرطان ریه و هم‌چنین در مدل آلودگی هوا و مرگ و میر مورد استفاده قرار گرفت (11-13). اسپلاین جریمه شده نیز اخیراً، در مطالعات مختلفی از جمله خطرات شغلی و عوارض مرتبط با سلامتی مورد استفاده قرار گرفته است (10-16).

این پژوهش در نظر دارد به مقایسه مدل اسپلاین جریمه شده در کاکس که یکی از روش‌های هموار سازی است با مدل مخاطرات متناسب کاکس بپردازد و عوامل موثر بر بقای بیماران مبتلا به سرطان معده ثبت شده در مرکز تحقیقات بیماری‌های گوارش و کبد بیمارستان طالقانی تهران را مورد بررسی قرار دهد.

## مواد و روش‌ها

در این مطالعه گذشته‌نگر، پس از حذف داده‌هایی که مشاهدات گمشده داشتند، تعداد 216 بیمار با تشخیص قطعی سرطان معده در گزارش پاتولوژی که طی سال‌های 1382 تا 1387 در بخش گوارش بیمارستان طالقانی تهران تحت عمل جراحی قرار گرفتند، وارد مطالعه شدند. این بیماران با استفاده از آندوسکوپی و رادیولوژی تشخیص داده شدند که برای برخی از بیماران بیوپسی نیز انجام گردید. اطلاعات این بیماران توسط مرکز تحقیقات بیماری‌های گوارش و کبد دانشگاه علوم پزشکی شهید بهشتی تهران جمع‌آوری گردید، هم‌چنین این پژوهش در کمیته اخلاق مرکز تحقیقات بیماری‌های گوارش و کبد بیمارستان طالقانی تهران با شماره مجوز A2212 به تصویب رسیده است. از طریق بازخوانی پرونده‌های بیماران و پیگیری تلفنی مدت زمان بقای آنها پس از عمل جراحی مورد بررسی قرار گرفت. در این مطالعه بیمارانی که در پایان بررسی زنده بودند و نیز بیمارانی که در زمان پی‌گیری اطلاعاتی در مورد وضعیت بقای آنها وجود نداشت، به‌عنوان سانسور از راست در نظر گرفته شدند. در این مطالعه با استفاده از فرم‌های ثبت سرطان، اطلاعات مربوط به متغیرهایی از قبیل: سن هنگام تشخیص، جنسیت، وجود متاستاز، اندازه تومور، نوع هیستوپاتولوژی، متاستاز غدد لنفاوی به صورت N1، N2 و N3 مبتنی بر سیستم استیجینگ و مرحله پاتولوژی جمع‌آوری شد (17).

یکی از قدیمی‌ترین و پرکاربردترین روش‌های آماری، رگرسیون خطی است. در حالت معمولی مقدار مورد انتظار براساس متغیر پاسخ و براساس رابطه خطی متغیرهای کمکی مدل‌بندی می‌شود و تابع پارامتری مجموعه متغیرهای کمکی بر روی مقدار متغیر پاسخ اثر می‌گذارد. مجموعه داده‌های زیادی وجود دارند که مدل‌های خطی، برازش خوبی روی آنها ندارند. در این حالت رویکرد کلاسیک، توسعه دادن متغیر کمکی با اضافه کردن توابع چندجمله‌ای به آن است. اما اگر تابعی که قرار است تخمین زده شود در محدوده‌های متفاوت، در دامنه دارای شکل‌های متفاوتی باشد، آن‌گاه استفاده از چندجمله‌ای‌های یکسان در

تمام دامنه اثربخش نخواهد بود. اسپلاین‌ها یک جایگزین مناسب برای برازش به داده‌ها و توابع چندجمله‌ای تکه‌ای هستند که محدود به ملحق شدن به یکدیگر در نقاط کنترل یا گره در داده‌ها می‌باشند. بدین صورت که هر قسم از شکل تابع را در محدوده‌های مورد نظر تکه تکه کرده و در هر تکه به هموار سازی تابع مشخص شده می‌پردازند، اتصال اولین تکه به تکه‌ی بعدی یک گره ایجاد می‌کند که در اولین محدوده صورت می‌پذیرد. به طور معمول توابع چند جمله‌ای تکه‌ای درجه 3 هستند (9). اسپلاین جریمه شده (Penalized Spline) زیر مجموعه اسپلاین هموار شده است که مبتنی بر مدل رگرسیونی و برآورد ناپارامتریک می‌باشد. مدل رگرسیونی ناپارامتریک اختصاص دارد به مدلی که در آن رابطه بین متغیر وابسته و متغیرهای مستقل با هیچ تابع مشخصی قابل تصور نیست. اسپلاین جریمه شده یک مدل غیرخطی است که مقدار اثرات خطی را با در نظر گرفتن اثرات غیرخطی متغیرهایش بررسی می‌کند، چون بر این امر معتقد است که از بین بردن اثرات غیرخطی منجر به از دست رفتن اطلاعاتی می‌شود که بسته به اهداف مطالعه ممکن است تاثیرگذار باشد در نتیجه با حضور اثرات غیرخطی، به بررسی اثرات خطی می‌پردازد. نکته قابل توجه این است که مدل‌های اسپلاین از جمله اسپلاین جریمه شده برای متغیرهای پیوسته مناسب می‌باشند (18).

در این پژوهش نخست، تحلیل چند متغیری مدل مخاطرات متناسب کاکس بر روی کلیه متغیرها جهت یافتن متغیرهای موثر در بقای بیماران مبتلا به سرطان معده برآزنده شد. از آنجایی که مدل‌های اسپلاین برای متغیرهای پیوسته کارآمد هستند در نتیجه برای مقایسه مدل مخاطرات متناسب کاکس و اسپلاین جریمه شده در مدل کاکس از متغیرهای موثر پیوسته شناسایی شده در تحلیل چند متغیری مدل کاکس استفاده شد. لازم به ذکر است که برای مقایسه کارایی دو مدل از ملاک آکایکه (Akaike information criterion) و برای مقایسه برازش دو مدل از آزمون نسبت درستنمایی (Likelihood Ratio Test) استفاده گردید. برای تحلیل داده‌ها نیز از نرم افزار R ویراست 2.14.1 استفاده شد.

## یافته‌ها

با توجه به هدف پژوهش بیمارانی که در پایان بررسی زنده بودند و بیمارانی که در زمان پیگیری، اطلاعاتی در مورد وضعیت بقای آنها وجود نداشت، به عنوان سانسور راست در نظر گرفته شدند که از 216 بیمار (36/1 درصد) 78 نفر فوت نمودند و (63/9 درصد) 138 نفر نیز در انتهای مطالعه زنده بودند یا اطلاعات دقیقی از وضعیت بقای آنها موجود نبود (سانسور از راست). میانگین سن هنگام تشخیص بیماران  $50/23 \pm 8/11$  و میزان بقای یک تا پنج ساله بیماران مبتلا به سرطان معده در این مطالعه به ترتیب 80 درصد، 56 درصد، 40 درصد، 35 درصد و 30 درصد بوده است. میانگین مدت زمان پیگیری  $16/60 \pm 15/28$  و میانگین و میانه بقای بیماران به ترتیب 46 و 30 ماه بود.

از 216 بیمار مبتلا به سرطان معده، 156 نفر مرد (72/2 درصد)، 23 نفر دارای متاستاز (10/6 درصد) و 175 نفر (80/1 درصد) دارای اندازه تومور بیش از 35 میلی‌متر بودند. 148 نفر (68/5 درصد) از بیماران بالای 45 سال سن داشتند و 136 نفر (63 درصد) از آنها دارای مرحله پاتولوژی پیشرفته بودند. 131 نفر (60/6 درصد) دارای متاستاز غدد لنفاوی در مرحله N2 و 26 نفر (12 درصد) نیز دارای متاستاز غدد لنفاوی در مرحله N3 بودند. از کل بیماران، هیستوپاتولوژی 164 نفر (75/9 درصد) از نوع آدنوکارسینوما و 27 نفر (12/5 درصد) از نوع سلول‌های نگین انگشتی بود.

در مرحله اول، به تحلیل چند متغیری عوامل مرتبط با بقای بیماران، با استفاده از مدل مخاطرات متناسب کاکس پرداختیم که نتایج آن را در جدول 1 مشاهده می‌کنید. نتایج این تحلیل نشان‌دهنده معنی‌داری متغیرهای سن هنگام تشخیص، اندازه تومور و مرحله پاتولوژی در سطح 0/05 بود. در نتیجه افرادی که سن تشخیص آنها پایین‌تر بود، بقای بیشتری نسبت به افراد با سن تشخیص بالاتر داشتند. جوان‌تر بودن و پیشرفته نبودن بیماری دلیل بر این نتیجه است. هم‌چنین متغیر اندازه تومور نیز معنی‌دار است و افرادی که اندازه تومور کوچکتری دارند، از بقای بالاتری برخوردارند و مخاطره مرگ در آنها کمتر از

سایرین است. مرحله پاتولوژی نیز یکی از عوامل تاثیرگذار در بقای بیماران است، بدین ترتیب افرادی که در مرحله پاتولوژی اولیه هستند دارای بقای بالاتری نسبت به افرادی می‌باشند که در مرحله پیشرفته‌تر قرار دارند. با توجه به این که مدل‌های اسپلین تنها متغیرهای پیوسته را در خود می‌پذیرند، برای مقایسه دو مدل مخاطرات متناسب کاکس و اسپلین جریمه شده در کاکس فقط متغیرهای پیوسته تاثیرگذار در تحلیل چندمتغیری کاکس را وارد اسپلین کرده تا اثر خطی و غیرخطی این متغیرها را با مدل مخاطرات متناسب کاکس بررسی کنیم. نتایج این تحلیل نشان می‌دهد در داده‌های موجود، سن هنگام تشخیص و اندازه تومور دو متغیر تاثیرگذار پیوسته معرفی شده در تحلیل چندمتغیری کاکس بودند. آن‌گاه دو مدل با ملاک آکاییکه مورد مقایسه و ارزیابی قرار گرفتند.

جدول 1. نتایج حاصل از تحلیل چند متغیری عوامل مرتبط با بقای بیماران مبتلا به سرطان معده با استفاده از مدل مخاطرات متناسب کاکس

p	نسبت خطر	خطای معیار	ضریب رگرسیون	
0/001	0/950	0/015	-0/052	سن هنگام تشخیص
-	1	-	-	جنسیت
0/677	0/892	0/274	-0/114	زن† مرد
-	1	-	-	متاستاز به سایر اندام‌ها
0/889	0/953	0/342	-0/048	نداشته† داشته
-	1	-	-	اندازه تومور
0/048	1/730	0/277	0/548	کمتر از 35 میلی‌متر† بیشتر از 35 میلی‌متر
-	1	-	-	نوع هیستوپاتولوژی
0/280	0/553	0/548	-0/592	سایر† سلول‌های نگین
0/332	0/706	0/359	-0/348	انگشتی آدنوکارسینوما
-	1	-	-	متاستاز غدد لنفاوی
0/490	1/335	0/418	0/289	N1†
0/223	1/939	0/548	0/662	N2 N3
-	1	-	-	مرحله پاتولوژی
0/034	0/448	0/379	-0/803	اولیه† پیشرفته

† رده مرجع

در نمودار 1 منحنی‌های بقای بیماران مبتلا به سرطان معده را در دو مدل رسم کرده تا به طور شهودی نیز مدل‌ها را مقایسه و بقای بیماران را بررسی کنیم.

جدول 2. معیارهای مقایسه دو مدل مخاطرات متناسب کاکس و اسپلین جریمه شده در کاکس برای متغیرهای پیوسته

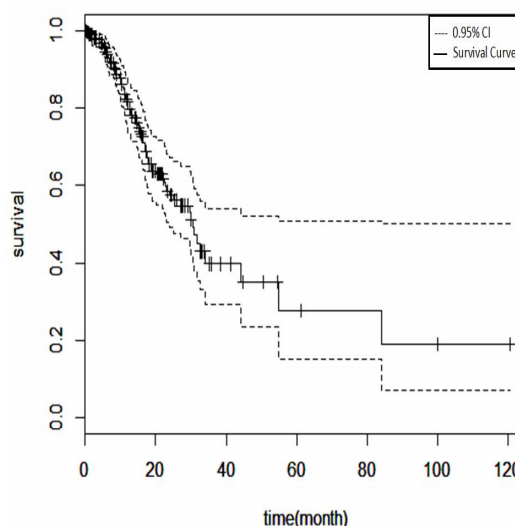
مدل	ضرایب رگرسیون	p	آماره نسبت (p) درست‌نمایی	ملاک آکاییکه
اسپلین جریمه شده				
سن هنگام تشخیص (خطی)	-0/054	0/00028	22/6 (0/00197)	689/4
سن هنگام تشخیص (غیرخطی)		0/94		
اندازه تومور (خطی)	-0/0262	0/04400		
اندازه تومور (غیرخطی)		0/93		
مخاطرات متناسب کاکس				
سن هنگام تشخیص	-0/054	0/00028	22/6 (0/00197)	689/4
اندازه تومور	-0/0262	0/04400		

فوق، مدل خطی به اندازه کافی مناسب و اثر غیرخطی متغیرهای پیوسته موردنظر کاملاً از بین رفته است. منحنی بقای بیماران مبتلا به سرطان معده در نمودار 1 نیز برای دو متغیر سن هنگام تشخیص و اندازه تومور در دو مدل سیری نزولی داشته و با هم برابر شدند، در نتیجه در صورت تشخیص این بیماری در سنین پایین‌تر و مراحل اولیه پیشرفت بیماری، خطر نسبی مرگ کاهش می‌یابد.

### بحث

اسپلین‌ها زیر مجموعه‌ای از تکنیک‌های هموار سازی هستند که به عنوان روش‌های انعطاف‌پذیر برای مدل‌های بقا به کار می‌روند و به‌طور گسترده برای تجزیه و تحلیل داده‌های اپیدمیولوژیک به خصوص در حوزه سلامت زیست محیطی که در آن روابط غیرخطی موجود است، مورد استفاده قرار می‌گیرند. هدف این مطالعه مقایسه مدل اسپلین جریمه شده در کاکس با مدل مخاطرات متناسب کاکس و تشخیص عوامل موثر در بقای بیماران مبتلا به سرطان معده با استفاده از این دو مدل بود.

در این مطالعه، میزان بقای یک، سه و پنج ساله بیماران مبتلا به سرطان معده به ترتیب 80 درصد، 40 درصد و 30 درصد شده است. این در حالی‌ست که یونی و همکاران میزان بقای یک و سه ساله بیماران مبتلا به سرطان معده را 75 درصد و 42 درصد گزارش کردند (19). ژانگ و همکاران طی مطالعه‌ای، میزان بقای یک، سه و پنج ساله این



نمودار 1. نمودار تابع بقای بیماران مبتلا به سرطان معده همراه با فاصله اطمینان 95 درصد برای دو مدل کاکس و اسپلین جریمه شده

با توجه به جدول 2، در مدل اسپلین جریمه شده ضرایب رگرسیونی متغیرهای پیوسته اسپلین شده و p در قسمت خطی، آزمون نسبت درست‌نمایی و همین‌طور ملاک آکاییکه کاملاً برابر با متغیرهای پیوسته مورد نظر در مدل مخاطرات متناسب کاکس شد، در نتیجه دو متغیر سن هنگام تشخیص و اندازه تومور در مدل اسپلین جریمه شده نیز به‌عنوان متغیرهای تاثیرگذار بر بقای بیماران شناسایی شدند. در قسمت غیرخطی نیز برای هر دو متغیر اسپلین شده معنی‌دار نشد که نشان می‌دهد در فرم اسپلین شده متغیرهای

یکسان شد اما ایسن و همکاران در یک مطالعه هم گروهی به بررسی مدل اسپلاین جریمه شده در مدل کاکس پرداختند و به این نتیجه رسیدند که در فقدان یک فرم مناسب پارامتری، مدل‌های اسپلاین جریمه شده می‌توانند اطلاعات مفیدی از تحقیقات بهداشت عمومی و اپیدمیولوژیکی در اختیار ما قرار دهند (15). هم‌چنین استراساکا و همکاران به مقایسه مدل اسپلاین جریمه شده و مدل چند جمله‌ای کسری پرداختند و به انعطاف‌پذیر بودن مدل اسپلاین جریمه شده پی بردند (32). استینلند نیز به بررسی اسپلاین جریمه شده و اسپلاین مکعبی محدود شده در اپیدمیولوژی حرفه‌ای پرداخت (16). گاو در دو مطالعه پیاپی برای تعیین ارتباط بین عوامل خطر متغیر زمان و مرگ بعد از جراحی عروق کرونر از اسپلاین‌های جریمه شده در مدل کاکس استفاده کرد (33، 34). همین‌طور پورت و همکاران نیز برای پیش‌بینی نقش گلوکز خون در مرگ بیماران قلبی - عروقی از اسپلاین‌های مکعبی محدود شده استفاده کردند (35).

در انجام این بررسی با مشکلات عدیده‌ای مواجه بودیم که از جمله می‌توان به ناقص بودن فرم‌های ثبت سرطان، نقص در پرونده‌های پزشکی بیماران، تغییر شماره تلفن و آدرس بیماران و عدم دسترسی به برخی از ایشان و در مواردی معدودی عدم همکاری بیماران یا خانواده‌های آنها در جهت به پاسخگویی به سوالات اشاره کرد. هم‌چنین از محدودیت‌های مطالعه حاضر این بود که تنها از داده‌های یک بیمارستان برای تحلیل مدل‌های اسپلاین جریمه شده در کاکس و مخاطرات متناسب کاکس و عوامل موثر بر بقای بیماران مبتلا به سرطان معده استفاده شد.

### نتیجه‌گیری

مدل‌های مخاطرات متناسب کاکس و اسپلاین جریمه شده در کاکس متغیرهای مهم یکسانی نتیجه دادند به طوری که سن بیمار در زمان تشخیص و اندازه تومور از عوامل مهم مرتبط با بقای بیماران مبتلا به سرطان معده در این دو مدل شناخته شدند. در نتیجه تشخیص این بیماری در سنین پایین‌تر و مراحل اولیه پیشرفت بیماری، خطر نسبی مرگ را کاهش می‌دهد. استفاده از مدل اسپلاین جریمه

بیماران را به ترتیب 87 درصد، 61 درصد و 32 درصد اعلام کردند (20). در مطالعه حاضر، سن هنگام تشخیص از عوامل موثر در بقای بیماران مبتلا به سرطان معده بود. این نتیجه با مطالعات اسپیچ و همکاران و یونی و همکاران، روشنایی و همکاران و بهزادپور و همکاران هم‌خوانی دارد (19-23). سیریوس و همکاران و ژانگ و همکاران طی مطالعه‌ای به ارتباط معنی‌داری بین سن هنگام تشخیص و بقای بیماران مبتلا به سرطان معده نرسیدند (20، 24). اندازه تومور نیز یکی دیگر از عوامل موثر بر بقای بیماران در دو مدل اسپلاین جریمه شده در کاکس و مخاطرات متناسب کاکس معرفی شد. این یافته با مطالعات سابق که خطر بالاتری برای مرگ بیماران دارای تومورهای بزرگ‌تر گزارش کردند، مطابقت دارد (25-28). اسپیچ و همکاران که به بررسی بقای بیماران مبتلا به سرطان معده پرداختند، به ارتباط معنی‌داری بین اندازه تومور و طول عمر بیماران دست نیافتند (21). در این مطالعه ارتباط معنی‌داری بین جنس و طول عمر بیماران یافت نشد. کیتاگاو و همکاران و لیو و همکاران نیز به این نتیجه پی بردند و به ارتباطی بین جنس و طول عمر بیماران نرسیدند (29، 30). در مطالعه‌ای روشنایی و همکاران، مردان دارای طول عمر پایین‌تری از زنان بودند یعنی میزان بقای بهتری برای زنان و خطر مرگ بالاتری برای مردان گزارش شد (22). هم‌چنین متغیر وضعیت متاستاز به لحاظ آماری معنی‌دار نشد. برخی مطالعات انجام شده در گذشته نشان دادند، بیمارانی که دچار متاستاز بودند میزان بقای کمتری نسبت به سایر بیماران داشتند (۲۲، ۳۰). در این مطالعه، متاستاز غدد لنفاوی دارای میزان خطر بالایی بود اما ارتباط معنی‌داری با طول عمر بیماران نداشت، که مطابق با نتایج مطالعات لیو و همکاران و یونی و همکاران (۱۹، ۳۰). ژانگ و همکاران و اسپیچ و همکاران طی مطالعه‌ای به ارتباط معنی‌داری بین متاستاز غدد لنفاوی با طول عمر بیماران مبتلا به سرطان معده دست یافتند (۲۱، ۳۱).

در مطالعه حاضر بین نتایج مخاطرات متناسب کاکس و اسپلاین جریمه شده در کاکس تفاوتی وجود نداشت به طوری که منحنی بقای بیماران در هر دو مدل نیز

7. Cao Y, Lin H, Wu TZ, Yu Y. Penalized spline estimation for functional coefficient regression models. *Computational Statistics & Data Analysis*. 2010;54(4):891-905.
8. Malloy EJ, Spiegelman D, Eisen EA. Comparing measures of model selection for penalized splines in Cox models. *Computational Statistics & Data Analysis*. 2009;53(7):2605-16.
9. Hansen MH, Kooperberg C. Spline Adaptation in Extended Linear Models (with comments and a rejoinder by the authors. *Statistical Science*. 2002;17(1):2-51.
10. Govindarajulu US, Spiegelman D, Thurston SW, Ganguli B, Eisen EA. Comparing smoothing techniques in Cox models for exposure-response relationships. *Statistics in medicine*. 2007;26(20):3735-52.
11. Rice F, Park R, Stayner L, Smith R, Gilbert S, Checkoway H. Crystalline silica exposure and lung cancer mortality in diatomaceous earth industry workers: a quantitative risk assessment. *Occupational and environmental medicine*. 2001;58(1):38-45.
12. Dominici F, McDermott A, Zeger SL, Samet JM. On the use of generalized additive models in time-series studies of air pollution and health. *American journal of epidemiology*. 2002;156(3):193-203.
13. Schwartz J, Laden F, Zanobetti A. The concentration-response relation between PM (2.5) and daily deaths. *Environmental health perspectives*. 2002;110(10):1025-9.
14. Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2006;169(2):179-203.
15. Eisen E, Agalliu I, Thurston S, Coull B, Checkoway H. Smoothing in occupational cohort studies: an illustration based on penalised splines. *Occupational and environmental medicine*. 2004;61(10):854-60.
16. Steenland K, Deddens JA. A practical guide to dose-response analyses and risk assessment in occupational epidemiology. *Epidemiology*. 2004;15(1):63-70.
17. cancerstaging.org [homepage on the internet]. American Joint Committee on Cancer: AJCC Cancer staging Manual. [September 24, 2011]. Available at:

شده به ما در از بین بردن اثرات غیرخطی متغیرهای کمکی و دقیق‌تر شدن برآورد پارامترها یاری می‌رساند، اما از آنجایی که در این مطالعه ملاک آکایکه برای هر دو مدل اسپلاین جریمه شده در کاکس و مخاطرات متناسب کاکس یکسان شد، استفاده از مدل مخاطرات متناسب کاکس در داده‌های پزشکی به دلیل تفسیر راحت‌تر و محدود نبودن در استفاده از انواع متغیرهای کمکی، وقت‌گیر و هزینه‌بر نبودن، در دسترس بودن برنامه‌های نرم‌افزاری این مدل و عدم نیاز به برنامه‌های پیچیده و کامپیوترهای پیشرفته جهت برازش مدل، مناسب‌تر است.

### تشکر و قدردانی

نگارندگان مقاله بر خود لازم می‌دانند که از کلیه همکاران مرکز تحقیقات بیماری‌های گوارش و کبد دانشگاه علوم پزشکی شهید بهشتی تهران، تشکر و قدردانی نمایند. این تحقیق برگرفته از پایان‌نامه کارشناسی ارشد رشته آمار زیستی با عنوان "مقایسه تکنیک‌های هموارسازی در مدل کاکس" و با حمایت مالی دانشکده علوم پزشکی دانشگاه تربیت مدرس انجام شده است.

### منابع

1. Kumar V, Abbas AK, Aster J. Robbins basic pathology: WB saunders company; 2012.
2. Mohagheghi MA. Annual report of Tehran cancer registry 1999. Tehran: The cancer institute publication. 2004.[Persian]
3. Mohagheghi MA, Musavi Jarahi A, Shariat Torbaghan S, Zeraati H. Annual report of Tehran university of medical sciences district cancer registry 1997. Tehran: The cancer institute publication. 1998.
4. Nagayato T. Background data to study of advance gastric cancer. New York: Springer-verlag; 1986.
5. Biswas A, Datta S, Fine JP, Segal MR. Statistical advances in the biomedical sciences: clinical trials, epidemiology, survival analysis, and bioinformatics: Wiley-Interscience; 2008:265-85.
6. Costa M, Shaw JEH. Parametrization and penalties in spline models with an application to survival analysis. *Computational Statistics & Data Analysis*. 2009;53(3):657-70.



- <http://www.cancerstaging.org/products/ajccproducts.html>.
18. Wager C, Vaida F, Kauermann G. Model selection for penalized spline smoothing using Akaike information criteria. *Australian & New Zealand Journal of Statistics*. 2007;49(2):173-90.
  19. Yoney A, Bati Y, Isikli L, Unsal M. Postoperative chemoradiation in patients with localized gastric adenocarcinoma: Single center experience. *Indian Journal of Cancer*. 2011;48(1):24-30.
  20. Zhang YF, Shi J, Yu HP, Feng AN, Fan XS, Lauwers GY, et al. Factors predicting survival in patients with proximal gastric carcinoma involving the esophagus. *World Journal of Gastroenterology: WJG*. 2012;18(27):3602-9.
  21. Spych M, Serbiak B, Rychter A, Jesien-Lewandowicz E, Gottwald L, Fijuth J. Post-operative radiochemotherapy in patients with gastric cancer: one department's experience of 56 patients. *British Journal of Radiology*. 2011;84(1001):457-63.
  22. Roushanaei G, Kazemnejad A, Sedighi S. Postoperative survival estimation of gastric cancer patients in cancer institute of Tehran, Imam Khomeini hospital and its relative factors. *Scientific Journal of Hamadan University*. 2010; 17(57): 13-8.[Persian]
  23. Behzadpoor M, Zeraati H, Mahmoudi M, Rahimi A. Postoperative Survival After Relapsing in Gastric Cancer Patients and Factors Related to it. *Journal of School of Public Health and Institute of Public Health Research*. 2009; 7(1): 27-34.[Persian]
  24. Syrios J, Sougioultzis S, Xynos ID, Kavantzias N, Kosmas C, Agrogiannis G, et al. Survival in patients with stage IV noncardia gastric cancer-the influence of DNA ploidy and Helicobacter Pylori infection. *BMC cancer*. 2012;12(1):1-8.
  25. Maroufizadeh S, Hajizadeh E, Bagestani AR, Fatemi SR. Prognostic factor for patients with gastric cancer using the Aalen's additive hazards model. *Koomesh Jurnal*. 2011; 1:120-6.[Persian]
  26. Baghestani AR, Hajizadeh E, Fatemi SR. Bayesian analysis for survival of patients with gastric cancer in Iran. *Asian Pacific Journal of Cancer Prevention*. 2009;10:823-6.
  27. Biglarian A, Hajizadeh E, Kazemnejad A, Zali MR. Survival analysis of gastric cancer patients using Cox model: a five year study. *Tehran University Medical Journal*. 2009; 67: 317-25.[Persian]
  28. Li C, Oh SJ, Kim S, Hyung WJ, Yan M, Zhu ZG, et al. Risk factors of survival and surgical treatment for advanced gastric cancer with large tumor size. *Journal of Gastrointestinal Surgery*. 2009;13(5):881-5.
  29. Kitagawa M, Ichikawa D, Komatsu S, Okamoto K, Shiozaki A, Fujiwara H, et al. The directionality of lymphatic spreading as a prognostic factor in gastric cancer: uni-or multidirectional lymphatic metastases. *Hepato-gastroenterology*. 2011;58(112):2148-51.
  30. Liu Y, Chen XH, Meng XH, Liu CF, Zhao LL, Han JW, et al. Multivariate prognostic study on node-positive gastric cancer: is tumor size a prognostic indicator? *Hepato Gastroenterology-Current Medical and Surgical Trends*. 2012;59(114):623-6.
  31. Zhang M, Li Z, Zhao B, Wei Y, Zhu G, Ma Y, et al. Predictors of Long-Term Survival in Large Gastric Carcinoma Patients. *Hepato-gastroenterology*. 2011;58(112):2162-5.
  32. Strasak AM, Umlauf N, Pfeiffer RM, Lang S. Comparing penalized splines and fractional polynomials for flexible modelling of the effects of continuous predictor variables. *Computational Statistics & Data Analysis*. 2011;55(4):1540-51.
  33. Gao D, Grunwald GK, Rumsfeld JS, Mackenzie T, Grover FL, Perlin JB, et al. Variation in mortality risk factors with time after coronary artery bypass graft operation. *The Annals of thoracic surgery*. 2003;75(1):74-81.
  34. Gao D, Grunwald GK, Rumsfeld JS, Schooley L, MacKenzie T, Shroyer ALW. Time-varying risk factors for long-term mortality after coronary artery bypass graft surgery. *The Annals of thoracic surgery*. 2006;81(3):793-9.
  35. Port SC, Boyle NG, Hsueh WA, Quiñones MJ, Jennrich RI, Goodarzi MO. The predictive role of blood glucose for mortality in subjects with cardiovascular disease. *American journal of epidemiology*. 2006;163(4):342-51.