

# استفاده از مدل چندجمله‌ای کسری در تعیین عوامل مرتبط با بقای بیماران مبتلا به سرطان معده

هدی نور کجوری<sup>۱</sup>، ابراهیم حاجی زاده<sup>۲</sup>، احمدرضا باغستانی<sup>۳</sup>، محمدامین پور حسینقلی<sup>۴</sup>

<sup>۱</sup>دانشجوی کارشناسی ارشد آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس، تهران، ایران

<sup>۲</sup>دانشیار، گروه آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس، تهران، ایران

<sup>۳</sup>استادیار گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

<sup>۴</sup>استادیار، گروه آمار زیستی، مرکز تحقیقات بیماری‌های گوارش و کبد، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

نویسنده رابط: ابراهیم حاجی زاده آدرس پستی: تهران، جلال آل احمد، پل نصر، دانشگاه تربیت مدرس، تلفن: ۰۲۱-۸۲۸۳۸۱۰؛ نامبر: ۰۲۱-۸۲۸۴۵۵۵ پست الکترونیک:

hajizadeh@modares.ac.ir

تاریخ دریافت: ۹۱/۴/۱۰؛ پذیرش: ۹۱/۱۱/۷

**مقدمه و هدف:** یکی از روش‌های آماری تحلیل داده‌های بقا، مدل رگرسیونی کاکس است. استفاده از تکنیک‌های هموارسازی در مدل کاکس باعث برآورد دقیق‌تر برای پارامترها می‌شود. یکی از تکنیک‌های هموارسازی در مدل کاکس استفاده از مدل چندجمله‌ای کسری است. این مطالعه به منظور تحلیل عوامل مرتبط با بقای بیماران مبتلا به سرطان معده، با استفاده از دو مدل خطرات متناسب کاکس و چندجمله‌ای کسری در مدل کاکس انجام شده است.

**روش کار:** در یک مطالعه گذشته‌نگر طی سال‌های ۱۳۸۲ تا ۱۳۸۷، ۲۱۶ بیمار مراجعه‌کننده با تشخیص قطعی سرطان معده که در بخش سرطان بیمارستان طالقانی تهران تحت عمل جراحی قرار گرفتند، بررسی شدند. در این مطالعه به منظور تحلیل عوامل مرتبط با بقای این بیماران، مدل‌های چندجمله‌ای کسری در کاکس و خطرات متناسب کاکس برازش و نتایج و کارایی آن‌ها توسط ملاک آکاییکه با یکدیگر مقایسه شدند.

**نتایج:** نتایج حاصل از تحلیل مدل خطرات متناسب کاکس و چندجمله‌ای کسری نشان داد که سن هنگام تشخیص و اندازه تومور با طول عمر بیماران ارتباطی معنی‌دار داشتند ( $P < 0.05$ ) و ملاک آکاییکه در هر دو مدل برابر بود.

**نتیجه‌گیری:** در مطالعه حاضر، کارایی دو مدل خطرات متناسب کاکس و چندجمله‌ای کسری یکسان شد. در نتیجه با این که استفاده از روش‌های هموارسازی برای از بین بردن آثار غیرخطی متغیرهای کمکی تأثیرگذار است، ولی مدل خطرات متناسب کاکس به دلیل راحتی تفسیر نتایج در داده‌های پزشکی و محدود نبودن در استفاده از انواع متغیرهای کمکی، مناسب‌تر است.

**واژگان کلیدی:** مدل خطرات متناسب کاکس، هموارسازی، مدل چندجمله‌ای کسری، سرطان معده

## مقدمه

همواره فرض می‌شود اثر متغیرهای کمکی بر لگاریتم تابع خطر، تابعی خطی از متغیرهای کمکی است. حال آن‌که متغیرهای پیوسته می‌توانند به صورت شکل‌های غیرخطی پیچیده‌ای بر میزان خطر اثر بگذارند و در نظر نگرفتن این شکل غیرخطی می‌تواند نتایج را تغییر دهد (۷،۸).

برای رفع این مشکل از هموارسازی استفاده می‌شود که در چند دهه اخیر به‌وفور در داده‌های پزشکی و اپیدمیولوژی به‌کار می‌رود. هموارسازی وسیله‌ای برای توصیف روند متغیر پاسخ، به‌عنوان تابعی از متغیرهای توضیحی است. برآوردگری از روند که با این شیوه به‌دست می‌آید، کمتر از مشاهدات پاسخ نوسان خواهد داشت و به این سبب هموارساز نامیده می‌شود (۹). یکی از

بسیاری از مدل‌های آماری، با هدف تعیین روابط بین متغیرها، تعیین متغیرهای اثرگذار، برآورد و پیش‌بینی استفاده می‌شوند (۵-۱). آشکار است که از این مدل‌ها، برای تحلیل داده‌های مرتبط با سلامت و بهداشت نیز استفاده می‌شود. مدلی که عموماً برای داده‌های بقا به‌کار برده می‌شود، مدل رگرسیونی کاکس است که به نام مدل خطرات متناسب نیز شناخته شده است. از آنجا که این نوع داده‌ها اغلب دارای داده‌های سانسور شده‌اند، برای برآورد بقا از روش‌هایی خاص استفاده می‌شود (۶). به‌رغم حجم زیاد اطلاعاتی که در مورد تحلیل مدل کاکس وجود دارد، بعضی محققان نکته مهمی را در به‌کارگیری این مدل در نظر نمی‌گیرند؛ این مدل شکلی درست از متغیرهای پیوسته مشخص نمی‌کند، چرا که

در این دو مدل بررسی کند.

## روش کار

### داده‌های مورد مطالعه

در این مطالعه گذشته‌نگر، پس از حذف داده‌هایی که مشاهدات گمشده داشتند، تعداد ۲۱۶ بیمار با تشخیص قطعی سرطان معده در گزارش پاتولوژی که طی سال‌های ۱۳۸۲ تا ۱۳۸۷ در بخش گوارش بیمارستان طالقانی تهران تحت عمل جراحی قرار گرفته بودند، وارد مطالعه شدند. این بیماران با استفاده از آندوسکوپی و رادیولوژی تشخیص داده شده بودند و برای برخی از بیماران بیوپسی نیز انجام شده بود. 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 A۲۲۱۲ به تصویب رسیده‌است. از طریق بازخوانی پرونده‌های بیماران و پیگیری تلفنی مدت زمان بقای آن‌ها پس از عمل جراحی بررسی و گردآوری شد. 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 مانند سن هنگام تشخیص، جنسیت (مرد-زن)، وجود متاستاز (داشته - نداشته)، اندازه تومور (بیش‌تر از ۳۵ میلی‌متر - کم‌تر از ۳۵ میلی‌متر)، نوع هیستوپاتولوژی (آدنوکارسینوما - سلول‌های نئگین انگشتری - دیگر)، متاستاز غده لنفاوی به‌صورت N1 (متاستاز در ۱-۶ غده لنفاوی ناحیه‌ای)، N2 (متاستاز در ۷-۱۵ غده لنفاوی ناحیه‌ای) و N3 (متاستاز در بیشتر از ۱۵ غده لنفاوی ناحیه‌ای) مبتنی بر سیستم استیجینگ (۲۹) و مرحله پاتولوژی (اولیه - پیشرفته) جمع‌آوری شد. همچنین، این پژوهش در کمیته اخلاق مرکز تحقیقات بیماری‌های گوارش و کبد بیمارستان طالقانی تصویب شده‌است.

### مدل خطرات متناسب کاکس

رگرسیون کاکس (۳۰) یک مدل شبه‌پارامتری برای برازش تابع مخاطره فرض می‌کند که می‌توان متغیرهای توضیحی یا عوامل خطر را به مدل آن اضافه کرد، ولی همچنان تابع مخاطره پایه را

تکنیک‌های هموارسازی مدل چندجمله‌ای کسری است که یک مدل رگرسیونی چندجمله‌ای و هدف استفاده از آن، رفع آثار غیرخطی متغیرها است (۱۰). این مدل در مطالعه‌های متعدد از جمله، مدل‌های دوز پاسخ الکل و مرگ‌ومیر (۱۱)، تجزیه و تحلیل شیوع ویروس انسانی که به نقص ایمنی منجر می‌شود (۱۲) و همچنین، در مطالعه‌های بهداشت حرفه‌ای (۱۳، ۱۴) به‌کار رفته است.

با توجه به مطالعه‌های انجام‌شده در استفاده از مدل چندجمله‌ای کسری و گسترش بیماران مبتلا به سرطان معده، به‌عنوان چهارمین سرطان شایع و دومین علت مرگ در سراسر دنیا، نیاز به بررسی بیشتر این بیماری و عوامل تأثیرگذار بر آن، بیش از پیش احساس می‌شود. هر ساله در جهان، بیش از ۹۳۴۰۰۰ مورد جدید از سرطان معده گزارش می‌شود و سالانه بیش از ۷۵۰۰۰۰ نفر به‌دلیل سرطان در سطح جهان فوت می‌کنند (۱۵) و تقریباً ۶۰٪ کل موارد در کشورهای در حال توسعه، به‌ویژه در آسیای شرقی رخ می‌دهد (۱۶). در ایران نیز سرطان معده نخستین سرطان شایع در مردان و سومین سرطان شایع در زنان است (۱۷). هر ساله، در حدود ۷۳۰۰ مورد جدید سرطان معده در کشور ما گزارش می‌شود (۱۸) که طبق مطالعه‌های آماری، مرگ‌ومیر ناشی از این نوع سرطان حدود ۳۹٪ کل مرگ‌ومیرهای سرطانی را در ایران تشکیل می‌دهد (۱۹).

مطالعه‌هایی متعدد برای بررسی عوامل مؤثر بر طول عمر بیماران مبتلا به سرطان معده انجام شده‌است. سن هنگام تشخیص بیماری، اندازه تومور، مرحله بیماری، متاستاز غده لنفاوی و نوع عمل جراحی از عوامل مؤثر بر بقا در سرطان معده عنوان شده‌اند (۲۰-۲۳). در ایران نیز، مطالعه‌های گوناگون در مورد بقای بیماران مبتلا به سرطان معده انجام شده و در آن‌ها نقش عوامل پیش‌آگهی‌دهنده در میزان بقای این بیماران، با استفاده از روش‌های ناپارامتری مانند جدول عمر، برآوردگر کاپلان-مایر و آزمون لگ-رتبه، روش نیمه‌پارامتری خطرات متناسب کاکس و روش‌های پارامتری بررسی شده‌است (۲۴-۲۸).

این پژوهش در نظر دارد به مقایسه مدل چندجمله‌ای کسری کاکس که یکی از تکنیک‌های هموار و انعطاف‌پذیر است، با مدل خطرات متناسب کاکس بپردازد تا آثار خطی و غیرخطی متغیرهای پیوسته مورد بررسی را در دو مدل ارزیابی کند و عوامل مؤثر بر بقای بیماران مبتلا به سرطان معده، ثبت‌شده در مرکز تحقیقات بیماری‌های گوارش و کبد بیمارستان طالقانی تهران، را

به‌عنوان تابعی نامشخص، ولی غیرمنفی از زمان ثابت نگاه داشت. چون فرض نشده‌است که این تابع مخاطره فرمی پارامتری دارد، مدل کاکس را یک مدل شبه‌پارامتری برای تابع مخاطره گویند. مدل کاکس کاربردی‌ترین روش برای یافتن ارتباط متغیرهای توضیحی با متغیر پاسخ بقا یا هر متغیر پاسخ دیگری است که از راست سانسور شده باشد. اگرچه این مدل محدودیت‌هایی نیز دارد. یکی از محدودیت‌های آن، نبود شکلی درست از متغیرهای پیوسته موجود در مدل است (۳۱).

### مدل چندجمله‌ای کسری

در مدل کاکس برای یافتن اثر غیرخطی متغیر پیوسته می‌توان به جای رسم کردن الگوی باقی‌مانده‌ها، به‌طور مستقیم 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 (۸،۹). یک روش جایگزین که در دهه اخیر در مطالعه‌های همه‌گیرشناسی استفاده شده، مدل رگرسیونی چندجمله‌ای کسری است. مدل‌های رگرسیون چندجمله‌ای کسری، یکی از انواع تکنیک‌های هموارسازی و جزو روش‌های ناپارامتری به‌شمار می‌روند و حد واسط بین مدل‌های چندجمله‌ای و مدل‌های غیرخطی‌اند. هدف استفاده از توابع چندجمله‌ای کسری در رگرسیون، حفظ مزایای استفاده از چندجمله‌ای خطی و از بین بردن معایب آن است. به‌طور خلاصه، این توابع شبیه چندجمله‌ای متعارفند که در آن از توان‌هایی برای متغیرهای مستقل نیز استفاده می‌کنند، با این تفاوت که توان‌های غیر عدد صحیح و منفی نیز مجاز است. مدل چندجمله‌ای کسری معمولاً برازشی بهتر از چندجمله‌ای‌های متعارف از همان درجه را می‌دهد. توابع این مدل در بسیاری از مدل‌های خطی تعمیم‌یافته و مدل‌های رگرسیون خطرات متناسب کاکس برای داده‌های بقا استفاده می‌شود. برای نمونه، رگرسیون خطی چندجمله‌ای رگرسیون لجستیک و لجستیک چندجمله‌ای کسری است. در تمام این مدل‌ها، متغیر پاسخ  $y$  و متغیر کمکی واحد  $x$  یا دارای چند متغیر کمکی  $x_1, \dots, x_k$  هستند. ما درجه‌ای از مدل چندجمله‌ای کسری را به‌عنوان تعدادی از جملات توان‌دار در  $x$  و دلالت کردن  $m$  (درجه چندجمله‌ای) را در مدل تعریف می‌کنیم.

بنابراین برای نمونه،  $y=b_0+b_1x-1$  با درجه  $m=1$  و  $y=b_0+b_1x-1+b_2x^2$  با درجه  $m=2$  وجود دارد. نیاز مدل به  $m \geq 2$  غیر معمول است، به همین دلیل مدل چندجمله‌ای کسری بر  $m=1$  و  $m=2$  متمرکز شده است. در این مدل، توان‌ها را با

$$y = b_0 + b_1 x^{(p_1)} + b_2 x^{(p_2)}$$

این مدل همانند روش حذف تقریبی، متغیرهای پیوسته بی‌اهمیت را حذف می‌کند و این روند تا زمانی ادامه دارد که در مرحله آخر تأثیرگذارترین متغیر در مدل مشخص شود. استفاده از این مدل هنگامی مفید است که ما بخواهیم با حفظ اثرات متغیرهای کمکی پیوسته که به روابط غیرخطی در آن شک داریم، از یک مدل رگرسیونی استفاده کنیم. الگوریتم این مدل در نرم‌افزار R، تأثیرگذارترین متغیرها را از روی P-value نمایش می‌دهد. سطح معنی‌داری در این مدل همواره ۰/۰۵ در نظر گرفته می‌شود (۳۲).

در این پژوهش، ابتدا عوامل مرتبط با بقای بیماران مبتلا به سرطان معده با استفاده از تحلیل چندمتغیری مدل خطرات متناسب کاکس ارزیابی شد. سپس به این دلیل که برازش مدل چندجمله‌ای کسری تنها متغیرهای پیوسته را در خود جای می‌دهد، ابتدا با استفاده از مدل خطرات متناسب کاکس، معنادارترین متغیرها شناسایی شد و آن‌گاه برای مقایسه مدل خطرات متناسب کاکس با چندجمله‌ای کسری در مدل کاکس، متغیرهای پیوسته شناسایی‌شده را وارد دو مدل کردیم و آن‌ها را از طریق ملاک آکایکه<sup>۱</sup> (AIC) و آزمون نسبت درست‌نمایی<sup>۲</sup> (LRT) مقایسه و ارزیابی کردیم. اطلاعات به‌دست‌آمده با استفاده از نرم‌افزارهای SPSS و ویراست ۱۸.۰ و R و ویراست ۲.۱۴.۱ تحلیل شده‌اند.

### 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، به‌عنوان سانسور راست در نظر گرفته شدند. از ۲۱۶ بیمار (۳۶/۱٪) ۷۸ نفر فوت کردند و (۶۳/۹٪) ۱۳۸ نفر نیز در پایان مطالعه زنده بودند یا اطلاعات دقیقی از وضعیت بقای آن‌ها موجود نبود (سانسور از راست). میانگین سن هنگام تشخیص بیماران ۵۰/۲۳±۸/۱۱ و میزان بقای یک تا پنج ساله بیماران مبتلا به سرطان معده در این مطالعه به‌ترتیب، ۸۰٪، ۵۶٪، ۴۰٪، ۳۵٪ و ۳۰٪ بوده است. میانگین مدت زمان پیگیری ۱۵/۲۸±۱۶/۶۰ و میانگین و میانه بقای بیماران، به‌ترتیب ۴۶ و ۳۰ ماه بود.

<sup>۱</sup> Akaike information criterion

<sup>۲</sup> Likelihood Ratio Test

آن‌گاه دو متغیر پیوسته شناسایی شده در جدول شماره ۱ را وارد مدل چندجمله‌ای کسری و خطرات متناسب کاکس کردیم تا P-value، آماره نسبت درست‌نمایی و ملاک آکاییکه دو مدل را با یکدیگر مقایسه کنیم. در پایان، منحنی بقای این بیماران را که هر یک از این دو مدل قرار گرفتند، رسم کردیم دو مدل خطرات متناسب کاکس و چندجمله‌ای کسری در کاکس را برای بقای بیماران ارزیابی کنیم. در نمودار شماره ۱، رنگ قرمز نشان‌دهنده منحنی بقای بیماران در مدل خطرات متناسب کاکس و رنگ آبی منحنی بقای بیماران در مدل چندجمله‌ای کسری در کاکس تحت تأثیر دو متغیر سن هنگام تشخیص و اندازه تومور است.

همان‌طور که در جدول شماره ۲ دیده می‌شود، نتایج P-value متغیرهای پیوسته در دو مدل، P-value آزمون نسبت درست‌نمایی و منحنی‌های بقای بیماران در دو مدل کاکس و چندجمله‌ای کسری تقریباً مشابه یکدیگرند، ولی ملاک آکاییکه دو مدل کاملاً یکسان است که نشان می‌دهد در این داده‌ها، آثار غیرخطی متغیرهای کمکی پیوسته، از جمله سن هنگام تشخیص و اندازه تومور از بین رفته و معیار مناسب بودن مدل خطرات کاکس با چندجمله‌ای کسری در مدل کاکس برابر شده است.

از ۲۱۶ بیمار مبتلا به سرطان معده، (۷۲/۲٪) ۱۵۶ نفر مرد، (۱۰/۶٪) ۲۳ نفر دارای متاستاز و (۸۰/۱٪) ۱۷۵ نفر دارای اندازه تومور بیش از ۳۵ میلی‌متر بودند. (۶۸/۵٪) ۱۴۸ نفر از بیماران بالای ۴۵ سال سن داشتند و (۶۳٪) ۱۳۶ نفر از آن‌ها دارای مرحله پاتولوژی پیشرفته بودند. (۶۰/۶٪) ۱۳۱ نفر دارای متاستاز غدد لنفاوی در مرحله N2 و (۱۲٪) ۲۶ نفر نیز دارای متاستاز غدد لنفاوی در مرحله N3 بودند. از کل بیماران، هیستوپاتولوژی ۱۶۴ نفر از نوع آدنوکارسینوما و (۱۲/۵٪) ۲۷ نفر از نوع سلول‌های نگین انگشتی بود.

در این مرحله، به تحلیل چندمتغیری عوامل مرتبط با بقای بیماران، با استفاده از مدل خطرات متناسب کاکس پرداختیم که نتایج آن را در جدول زیر مشاهده می‌کنید. نتایج تحلیل چندمتغیری، سه متغیر سن هنگام تشخیص، اندازه تومور و مرحله پاتولوژی را از عوامل مؤثر در بقای بیماران مبتلا به سرطان معده معرفی کرد. در این نتایج بیماران جوان‌تر بقای بهتری داشتند. همین‌طور بیمارانی با اندازه تومور بیش از ۳۵ میلی‌متر خطر مرگ بالاتری را نشان دادند و افرادی که تشخیص بیماری آنان در مرحله اولیه بیماری بود، نسبت به افرادی که بیماری آن‌ها در مرحله پیشرفته تشخیص داده شده بود، بقای بیشتری داشتند.

جدول شماره ۱- نتایج به‌دست‌آمده از تحلیل چندمتغیری عوامل مرتبط با بقای بیماران مبتلا به سرطان معده، با استفاده از مدل خطرات متناسب کاکس

P	(فاصله اطمینان ۰/۹۵) نسبت خطر	خطای معیار	ضریب رگرسیون	سن هنگام تشخیص
۰/۰۴۲*	۲/۱۱۳ (۱/۹۹۴-۴/۴۹۲)	۰/۰۱۵	۰/۰۵۲	جنسیت
-	۱	-	-	زن <sup>†</sup>
۰/۶۷۷	۱/۲۹۲ (۰/۵۲۲-۱/۵۲۶)	۰/۲۷۴	۰/۱۱۴	مرد
-	۱	-	-	وضعیت متاستاز
-	۱	-	-	نداشته <sup>†</sup>
۰/۸۸۹	۱/۴۵۳ (۰/۴۸۸-۱/۸۶۴)	۰/۳۴۲	۰/۴۸۰	داشته
-	۱	-	-	اندازه تومور
-	۱	-	-	کمتر از ۳۵ میلی‌متر <sup>†</sup>
۰/۰۴۸*	۱/۷۳۰ (۱/۰۰۵-۲/۹۷۹)	۰/۲۷۷	۰/۵۴۸	بیشتر از ۳۵ میلی‌متر
-	۱	-	-	نوع هیستوپاتولوژی
-	۱	-	-	دیگر <sup>†</sup>
۰/۲۸۰	۰/۵۵۳ (۰/۱۸۹-۱/۶۱۹)	۰/۵۴۸	-۰/۵۹۲	نگین انگشتی
۰/۳۳۲	۰/۷۰۶ (۰/۳۴۹-۱/۴۲۷)	۰/۳۵۹	-۰/۳۴۸	آدنوکارسینوما
-	۱	-	-	متاستاز غدد لنفاوی
-	۱	-	-	N1 <sup>†</sup>

ادامه جدول شماره ۱ - نتایج حاصل از تحلیل چندمتغیری عوامل مرتبط با بقای بیماران مبتلا به سرطان معده، با استفاده از مدل خطرات متناسب کاکس

P-value	فاصله اطمینان (۰/۹۵) نسبت خطر	خطای معیار	ضریب رگرسیون	
۰/۴۹۰	۱/۳۳۵ (۰/۵۸۸-۳/۰۳۱)	۰/۴۱۸	۰/۲۸۹	N2
۰/۲۲۳	۱/۹۳۹ (۰/۶۶۹-۵/۶۲۲)	۰/۵۴۸	۰/۶۶۲	N3
<b>مرحله پاتولوژی</b>				
-	۱	-	-	اولیه <sup>†</sup>
۰/۰۳۴*	۲/۱۹۸ (۱/۰۷۰-۴/۵۱۳)	۰/۳۷۹	۰/۸۰۳	پیشرفته

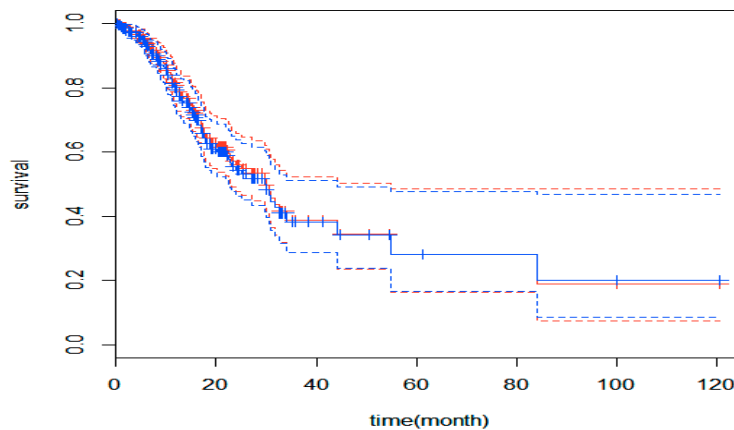
\* معناداری در سطح ۰/۰۵

<sup>†</sup> رده مرجع

جدول شماره ۲ - معیارهای مقایسه دو مدل کاکس و چندجمله‌ای کسری

مدل	ضریب رگرسیونی	خطای معیار	P-value	آماره نسبت درست‌نمایی (P-value)	ملاک آکاییکه
خطرات متناسب کاکس	۰/۰۵۴۵	۰/۰۱۵	۰/۰۰۰۲۸*	۲۲/۶ (۰/۰۰۱۹۷)	۶۸۹/۴
سن هنگام تشخیص	۰/۰۲۶۲	۰/۰۱۳	۰/۰۴۴۰*		
اندازه تومور	۰/۰۵۴۵	۰/۰۱۵	۰/۰۰۰۲۸*	۲۲/۶ (۰/۰۰۱۹۷)	۶۸۹/۴
چندجمله‌ای کسری	۰/۰۲۶۲	۰/۰۱۳	۰/۰۴۴۰*		
سن هنگام تشخیص	۰/۰۵۴۵	۰/۰۱۵	۰/۰۰۰۲۸*		
اندازه تومور	۰/۰۲۶۲	۰/۰۱۳	۰/۰۴۴۰*		

\* معناداری در سطح ۰/۰۵



نمودار شماره ۱- نمودار تابع بقای بیماران مبتلا به سرطان معده، همراه با فاصله اطمینان ۹۵ درصد برای دو مدل کاکس و چندجمله‌ای کسری

## بحث

است تا آثار غیرخطی و خطی متغیرهای پیوسته ارزیابی شوند و همچنین، عوامل مؤثر در بقای بیماران مبتلا به سرطان معده با استفاده از این دو مدل بررسی شوند.

در مطالعه حاضر، مدل چندجمله‌ای کسری دو متغیر سن هنگام تشخیص و اندازه تومور را معنی‌دار شناخت و این گروه از بیماران به‌عنوان گروه تأثیرگذار در مرگ‌ومیر معرفی شدند. در این مطالعه بین نتایج خطرات متناسب کاکس و چندجمله‌ای کسری در کاکس تفاوتی وجود نداشت، به‌طوری‌که منحنی بقای بیماران

روش‌های هموارسازی یک رویکرد انعطاف‌پذیر برای مدل‌های بقا است که روند پاسخ را بر متغیرهای پاسخ برآورد می‌کند و به‌طور گسترده برای تجزیه و تحلیل داده‌های اپیدمیولوژیک استفاده می‌شود، به‌ویژه در حوزه سلامت زیست‌محیطی که در آن روابط غیرخطی موجود است. هدف این مطالعه مقایسه مدل چندجمله‌ای کسری در کاکس با مدل خطرات متناسب کاکس

دارد (۴۸-۴۶، ۲۸، ۲۷، ۲۴). با توجه به نتایج، ارتباطی معنی‌دار بین متاستاز غدد لنفاوی با طول عمر بیماران یافت نشد که با برخی مطالعه‌هایی که در گذشته انجام شده‌است، هم‌خوان است (۲۷، ۲۸، ۳۸). ولی برخی مطالعه‌های دیگری نشان دادند بیمارانی که دچار متاستاز بودند، نسبت به دیگر بیماران میزان بقای کمتری داشتند (۵۱-۴۹، ۴۵، ۴۲، ۴۱).

یکی از مشکلات مهم فرا روی این پژوهش، تغییر شماره تلفن و آدرس بیماران و دسترسی‌ناداشتن به برخی از ایشان و در مواردی محدود همکاری‌نکردن خانواده‌های آن‌ها در پاسخگویی درباره وضعیت فعلی بیماران بود که این افراد در مطالعه حاضر جزو سانسورها انگاشته شدند.

### نتیجه‌گیری

با توجه به اهمیت متغیرهای مؤثر به‌عنوان عوامل مرتبط در بقای بیماران، شناسایی عوامل تاثیرگذار بر بقا، از جمله سن هنگام تشخیص و اندازه تومور دارای اهمیت است. شبیه‌سازی مطالعه‌های گذشته نشان داد که روش‌های هموارسازی در مدل کاکس به برآوردی دقیق از پارامتر منجر می‌شود و آثار غیرخطی متغیرها را در مدل از بین می‌برد. ولی با توجه به یکسان شدن ملاک آکاییکه برای دو مدل در این مطالعه، استفاده از مدل خطرات متناسب کاکس در داده‌های پزشکی، به‌دلیل راحتی در تفسیر نتایج، محدود نبودن در استفاده از انواع متغیرهای کمکی، در دسترس بودن بسته‌های نرم‌افزاری، وقت‌گیر و هزینه‌بر نبودن، مناسب‌تر خواهد بود ولی با این‌حال توصیه می‌شود هر دو روش در داده‌ها اعمال شود و هر کدام که نتیجه بهتر و ملاک آکاییکه کوچک‌تری داد، استفاده شود.

### تشکر و قدردانی

نگارندگان مقاله از همه همکاران مرکز تحقیقات بیماری‌های گوارش و کبد دانشگاه علوم پزشکی شهید بهشتی تهران، قدردانی می‌کنند.

در هر دو مدل نیز یکسان شد. ولی در مطالعه‌ای که Berger و همکاران در سال ۲۰۰۳ به مقایسه مدل چندجمله‌ای کسری با مدل کاکس پویا پرداختند، برآزش چندجمله‌ای کسری مدل کاکس را بهتر دانستند. زیرا عقیده داشتند مدل چندجمله‌ای کسری میزان خطای مدل خطرات متناسب کاکس با متغیرهای کمکی وابسته به زمان را کاهش می‌دهد (۳۳). از کاربردهای مدل چندجمله‌ای کسری استفاده از این روش در مطالعه‌های همه‌گیرشناسی و مداخله پاسخ است، به طوری که Royston و همکارانش طی مطالعه‌ای که در سال ۱۹۹۹ به کاربرد روش چندجمله‌ای کسری در اپیدمیولوژی پرداختند، استفاده از این روش را در این مطالعه‌ها بسیار مفید دانستند (۳۴). در مطالعه‌ای دیگر در سال ۲۰۰۹، Govindarajulu و همکارانش که به مقایسه تکنیک‌های هموارسازی در مدل کاکس با استفاده از روش‌های شبیه‌سازی پرداختند، مدل چندجمله‌ای کسری را یک روش قدرتمند برای مدل‌های مداخله پاسخ و استفاده از این تکنیک‌ها را برای برآورد بهتر مناسب دانستند (۳۵). همچنین در همین سال، Cui و همکارانش در یک مطالعه طولی همه‌گیرشناسی در استرالیا به این نتیجه رسیدند که برآزش مدل چندجمله‌ای کسری بهتر از روش‌های طبقه‌بندی است که در مدل کاکس استفاده می‌شود و این روش انعطاف‌پذیری بسیاری در مدل‌های مداخله پاسخ دارد (۳۶).

مدل خطرات متناسب کاکس نیز، همانند مدل چندجمله‌ای کسری در کاکس سن هنگام تشخیص را از عوامل مرتبط با بقای بیماران معرفی کرد. این نتیجه مشابه برخی مطالعات پیشین است که سن را عاملی معنی‌دار معرفی و بقای بهتری را برای بیماران جوان‌تر گزارش کرده‌اند (۲۸-۲۶، ۴۲-۳۷). در این مطالعه، ارتباطی معنی‌دار بین جنسیت و طول عمر بیماران دیده نشد که با برخی از مطالعه‌ها هم‌خوانی دارد (۲۸، ۲۷، ۲۴)، ولی برخی دیگر از گزارش‌ها نشان‌گر آن است که خطر مرگ بالاتری برای مردان و بقای بهتری برای زنان وجود دارد (۴۵-۴۳). اندازه تومور یکی دیگر از عوامل مؤثر بر بقای بیماران در تحلیل چندمتغیره کاکس بود. این نتیجه با مطالعه‌های پیشین که خطر بیشتری برای مرگ بیماران دارای تومور بزرگ‌تر گزارش کرده‌اند، هم‌خوانی

## منابع

1. Neter J, Wasserman W, Kutner M, Nachtsheim C. Applied linear statistical models. Irwin: 2002.
2. Elisa TL, John WW. Statistical Methods for Survival Data Analysis, John Wiley & Sons, Inc 2003,8-15.
3. Jobson DJ. Applied multivariate data analysis, vol II, Springer, London: 1992, 720-800.
4. Lindsey JK, Jones B. choosing among generalized linear models applied to medical data, *Statistics in Medicine* 1998, 17: 59-68.
5. Kay JW, Titterton DM. (eds), *Statistics and neural networks: Advanced at the interface*, Oxford University Press, Oxford 1999, 75-103.
6. Biswas A, Datta S, P.Fine J, R.Segal. *MStatistical Advances in the Biomedical Science*, John Wiley Series 2007, 265-85.
7. Cao Y, Lin H, Wu TZ, Yu T. Penalized spline estimation for functional coefficient regression models. *Computational Statistics and Data Analysis* 2010; 54: 891-905.
8. Malloy EZ, Spiegelman D, Eisen EA. Comparing measures of model selection for penalized splines in cox models. *Computational Statistics and Data Analysis* 2009; 53: 2605-16.
9. Hastie T, Tibshirani R.J. *Generalized Additive Models*, London, Chapman and Hall: 1990.
10. Royston P, Altman DG. Using fractional polynomials to model curved regression relationship. *STATA Technical Bulletin* No.21, sg26, STATA Corporation, College Station, TX, 1994.
11. Bagnardi V, Zambon A, Quatto P, Giovanni C. Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. *American Journal of Epidemiology* 2004; 159: 1077-86.
12. Wartenberg D, Northridge M. Defining exposure in case-control studies: a new approach. *American Journal of Epidemiology* 1991; 133: 1058-1071.
13. Stone RA, Youk AO, Marsh GM, Buchanich JM, McHenry MB, Smith TJ. Historical cohort study of US manmade vitreous fiber production workers: IV. Quantitative exposure-response analysis of the nested case-control study of respiratory system cancer. *Journal of Occupational and Environmental Medicine* 2001; 43: 779-92.
14. Wijngaarden E. A graphical method to evaluate exposure-response relationships in epidemiologic studies using standardized mortality or morbidity studies. *Dose Response* 2005; 3: 465-473.
15. Parkin DM, Bray F, Ferlay J, P P. Global cancer statistics. *A Cancer Journal for Clinicians* 2002; 55(2): 74-108.
16. Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, et al. *Cancer statistics, 2006*. *CA Cancer J Clin* 2006; 56: 106-30.
17. Sadjadi A, Nouraei M, Mohagheghi MA, Mousavi-Jarrahi A, Malekzadeh R, Parkin DM. Cancer occurrence in Iran in 2002. *Asian Pac J Cancer Prev* 2005; 6: 359-63.
18. Movahedi M, Afsharfard A, Moradi A, Nasermoaddeli A, Khoshnevis J, Fattahia F, et al. Survival rate of gastric cancer in Iran. *Journal of Research in Medical Sciences* 2009; 14: 367-73.
19. Yazdannod A, Arshi S, Derakhshan MH, Sadjadi AR, Malekzadeh R. Gastric cardia cancer; the most common type of upper gastrointestinal cancer in Ardabil. Iran: An endoscopy clinic experience 2001; 4: 76-9.
20. Msika S, Benhamiche AM, Jouve JL, Rat P, Faivre J. Prognostic factors after curative resection for gastric cancer: A population-based study. *Eur J Cancer* 2000; 36: 390-6.
21. Yamamura Y, Nakajima T, Ohta K, Nashimoto A, Arai K, Hiratsuka M, et al. Determining prognostic factors for gastric cancer using the regression tree method. *Gastric Cancer* 2002; 5: 201-7.
22. Meimarakis G, Winter H, Assmann I, Kopp R, Lehn N, Kist M, et al. *Helicobacter pylori* as a prognostic indicator after curative resection of gastric carcinoma: a prospective study. *Lancet Oncol* 2006; 7: 211-22.
23. Kim JP, Hur YS, Yang HK. Lymph node metastasis as a significant prognostic factor in early gastric cancer: analysis of 1,136 early gastric cancers. *Ann Surg Oncol* 1995; 2: 308-13.
24. Pourhoseingholi MA, Hajizadeh E, Moghimi Dehkordi B, Safaee A, Abadi A, Zali MR. Comparing cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pac J Cancer Prev* 2007; 8: 412-16.
25. Moghimi-Dehkordi B, Safaee A, Zali MR. Survival rates and prognosis of gastric cancer using an actuarial life-table method. *Asian Pac J Cancer Prev* 2008; 9: 317-21.
26. 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.
27. Baghestani AR, Hajizadeh E, Fatemi SR. Parametric model to analyse the survival of gastric cancer in the presence of interval censoring. *Tumori* 2010; 96: 433-37.
28. Baghestani AR, Hajizadeh E, Fatemi SR. Bayesian Analysis for Survival of Patients With Gastric Cancer in Iran. *Asian Pacific J Cancer Prev* 2009; 10: 823-826.
29. American Joint Committee on Cancer. *American Joint Committee on Cancer: AJCC Cancer staging Manual* (ed 3). Available at: <http://www.cancerstaging.org/products/ajccproducts.html> (September 22-24, 2011 in Chicago).
30. Cox DR. Regression models and life-table. (With discussion). *Journal of Royal Statistical Society B* 1972; 34: 187-220.
31. Therneau T, Grambsch P. *Modeling Survival Data: Extending the Cox Model*. New York, Springer-Verlag: 2000.
32. Royston P, Altman D.G. Regression using fractional polynomials of continuous covariates: parsimonious parametric modeling. *Applied Statistics* 1994; 43: 429-67.
33. Berger U, Schafer J, Ulm K. Dynamic Cox modelling based on fractional polynomials: time-variations in gastric cancer prognosis. *Statistics In Medicine. Statist Med* 2003; 22: 1163-80.
34. Royston J, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *International Epidemiological Association* 1999; 28: 964-74.
35. Govindarajulu US, Malloy E, Ganguli B, Spiegelman D, Eisen EA. The Comparison of Alternative Smoothing Methods for Fitting Non-Linear Exposure-Response Relationships with Cox Models in a Simulation Study. *The International Journal of Biostatistics* 2009; 5: 1-19.
36. Cui J, Klerk N, Abramson M, Monaco A, Geza Benke, Dennekamp M, et al. Fractional Polynomials and Model Selection in Generalized Estimating Equations Analysis, With an Application to a Longitudinal Epidemiologic Study in Australia. *American J Epidemiol* 2009; 169: 113-21.
37. Liu C, Zhang R, Lu Y, Li H, Lu P, Yao F, et al. Prognostic role of lymphatic vessel invasion in early gastric cancer: A retrospective study of 188 cases. *Surg Oncol* 2010; 19: 4-10.
38. Maroufizadeh S, Hajizadeh E, Baghestani AR, Fatemi SR. Prognostic factor for patients with gastric cancer using the Aalen's additive hazards model. *Koomesh Journal* 2011; 1: 120-6.
39. Saidi RF, Bell JL, Dudrick PS. Surgical resection for gastric cancer in elderly patients: is there a difference in outcome? *J Surg Res* 2004; 118: 15-20.

40. Bucchi L, Nanni O, Ravaioli A, Falcini F, Ricci R, Buiatti E, et al. Cancer mortality in a cohort of male agricultural workers from northern Italy. *J Occup Environ Med* 2004; 46: 249-56.
41. Otsuji E, Yamaguchi T, Sawai K, Sakakura C, Okamoto K, Takahashi T. Regional lymph node metastasis as a predictor of peritoneal carcinomatosis in patients with Borrmann type IV gastric carcinoma. *Am J Gastroenterol* 1999; 94: 434-7.
42. Moghimi Dehkordi B, Rajaeefard A, Tabatabaee H, Zeighami B, Safaee A, Tabeie Z. Modeling survival analysis in gastric cancer patients using the proportional hazards model of Cox. *Iranian Journal of Epidemiology* 2007; 3: 19-24.
43. Curtis RE, Kennedy BJ, Myers MH, Hankey BF. Evaluation of AJC stomach cancer staging using the SEER population. *Semin Oncol* 1985; 12: 21-31.
44. Bako G, Ferenczi L, Hanson J, Hill GH, Dewar R. Factors influencing the survival of patients with cancer of the stomach. *Clin Invest Med* 1985; 8: 22-8.
45. 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: 13-8.
46. Coburn NG, Swallow CJ, Kiss A, Law C. Significant regional variation in adequacy of lymph node assessment and survival in gastric cancer. *Cancer* 2006; 107: 2143-51.
47. 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. *J Gastrointest Surg* 2009; 13: 881-5.
48. Orsenigo E, Carlucci M, Braga M, Tomajer V, Di Palo S, Tamburini A, et al. Prognostic factors of gastric neoplasms: experience with 1,074 cases undergoing surgical treatment at a single center. *Tumori* 2005; 4: 86-7.
49. Sigon R, Canzonieri V, Rossi C. Early gastric cancer: a single-institution experience on 60 cases. *Suppl Tumori* 2003; 2: S23-6.
50. Yagi Y, Seshimo A, Kameoka S. Prognostic factors in stage IV gastric cancer: univariate and multivariate analyses. *Gastric Cancer* 2000; 3: 71-80.
51. Koizumi W, Kurihara M, Tanabe S, Kondo I, Yamazaki I, Nonaka M, et al. Advantages of Japanese response criteria for estimating the survival of patients with primary gastric cancer. *Gastric Cancer* 1999; 2: 14-9.



Original Article

# Application of Fractional Polynomial Model for Determining Prognostic Factors Associated with Survival of Patients with Gastric Cancer

Noorkojuri H<sup>1</sup>, Hajizadeh E<sup>1</sup>, Baghestani AR<sup>2</sup>, MA Pourhoseingholi MA<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

Corresponding author: Hajizadeh E., hajizadeh@modares.ac.ir

**Background & Objectives:** Cox regression model is one of the statistical methods in survival analysis. The use of smoothing techniques in Cox model makes the more accurate estimates for the parameters. Fractional polynomial is one of these techniques in Cox model. The aim of this study was to assess the effects of prognostic factors on survival of patients with gastric cancer using the fractional polynomial in Cox model and Cox proportional hazards.

**Methods:** Information of total of 216 patients with gastric cancer who underwent surgery in the gastroenterology ward of Taleghani Hospital in Tehran between 2003 and 2008 were included in this retrospective study. In this research, fractional polynomial in Cox model and Cox proportional hazards model were utilized for determining the effects of prognostic factors on patients' survival time with gastric cancer. The SPSS version 18.0 and R version 2.14.1 were used for data analysis. These models were compared with Akaike information criterion.

**Results:** The analysis of Cox proportional hazards and fractional polynomial models resulted in age at diagnosis and tumor size as prognostic factors on survival time of patients with gastric cancer independently ( $P < 0.05$ ). Also, Akaike information criterion was equal in both models.

**Conclusion:** In the present study, the Cox proportional hazards and fractional polynomial models led to similar results with equal Akaike information criterions. Using of smoothing methods helped us eliminate non-linear effects but it seemed more appropriate to use Cox proportional hazards model in medical data because of its' ease of interpretation and capability of modeling in both continuous and discrete covariates.

**Keywords:** Cox proportional hazards model, Smoothing, Fractional polynomial model, Gastric cancer