

تحلیل داده‌های بقای بیماران پیوند مغز استخوان با استفاده از مدل‌های ساده و آمیخته وایبل

علی اکبر خادم معبودی^۱، احمد رضا باغستانی^۱، محسن حمیدپور^۱، سعید شیرازی^۲، حامد بازیاد^۲، ناصر احمدی^{۱*}
۱- گروه آمارزیستی، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید بهشتی، تهران، ایران
۲- گروه فناوری اطلاعات، دانشگاه تربیت مدرس، تهران، ایران
(مؤلف مسؤول) تلفن: ۰۹۱۸۹۹۲۷۰۰۱ Naserahmadi300@gmail.com

چکیده

زمینه و هدف: با توجه به اهمیت بیماری سرطان و توجه روزافزون به داده‌های سرطان و همچنین نیاز به انجام تحلیل‌های دقیق‌تر و مناسب‌تر برای بررسی بقای بیماران سرطانی و همچنین ویژگی عودپذیری بودن بیماری سرطان و ناهمگونی میزان خطر برای بیماری سرطان، نیاز به ارائه روش‌های مناسب برای تحلیل داده‌های سرطان همیشه وجود دارد. با توجه به اینکه در سال‌های اخیر مدل آمیخته به عنوان یک مدل مناسب برای برازش به داده‌های بقای غیرهمگن مورد توجه قرار گرفته است؛ بنابراین، در این مقاله تلاش شد که با استفاده از مدل آمیخته وایبل داده‌های بقای پیوند مغز استخوان بررسی و میزان اثر عود سرطان بر روی میزان مرگ و میر افراد برآورد شود.

مواد و روش کار: در این پژوهش ۴۹۲ بیمار مبتلا به سرطان خون که به بیمارستان طالقانی تهران وابسته به دانشگاه علوم پزشکی شهید بهشتی مراجعه نموده و پیوند مغز استخوان دریافت نموده‌اند به صورت سرشماری بررسی شد. این بیماران طی سال‌های ۱۳۸۶ تا ۱۳۹۴ پیوند مغز استخوان دریافت نموده و تا سال ۱۳۹۵ پیگیری شدند. در این مطالعه مدل‌های ساده و آمیخته وایبل در حضور متغیر کمی به داده‌ها برازش و مدل‌های مختلف با هم مقایسه شدند.

یافته‌ها: در این مطالعه میانگین (انحراف معیار) زمان بقا در افرادی که رخداد برای آن‌ها اتفاق افتاده بود ۵۶۳ (۵۷/۴) روز و در افرادی که سانسور شده‌اند ۹۹۹ (۴۵/۴) روز می‌باشد. مدل‌های آمیخته و ساده وایبل به داده‌ها برازش داده شد و با استفاده از شاخص AIC ثابت شد که مدل آمیخته وایبل به داده‌های پیوند مغز استخوان برازش بهتری دارد و می‌تواند نسبت خطر را در دو گروه عود کرده و عود نکرده بعد از پیوند به خوبی تشریح کند.

نتیجه‌گیری: برای تحلیل داده‌های بقای غیرهمگن مانند داده‌های سرطان، استفاده از مدل‌های آمیخته می‌تواند نسبت به مدل‌های ساده پارامتری عملکرد بهتری داشته باشد.

واژه‌های کلیدی: داده‌های بقای پیوند مغز استخوان، مدل آمیخته وایبل، عود، ماکزیمم درست‌نمایی

مقدمه

مطالعات بقا، مجموعه‌ای از روش‌های آماری هستند که متغیر پاسخ در آن‌ها زمان تا وقوع یک رخداد می‌باشد (۱) که به طور عمده در حوزه‌های پزشکی، اقتصادی، بیولوژی و علوم اجتماعی به کار می‌رود. بیشترین مدل‌هایی که در تحلیل بقا مورد استفاده قرار می‌گیرند مدل‌های ناپارامتری مانند مدل کاپلان مایر و مدل‌های نیمه پارامتری مانند مدل کاکس می‌باشند (۲) و همچنین مدل‌های پارامتری مانند گاما، لگ نرمال و وایبل نیز در مقالات و کتاب‌های مربوط به تحلیل بقا مورد استفاده قرار می‌گیرند.

در این میان مدل وایبل بیشترین استفاده را در میان مدل‌های پارامتری دارد. مدل‌های پارامتری معرفی شده در جوامعی که همگن یا تک‌مدی هستند به کار می‌روند (۳) و زمانی که داده‌ها از جامعه غیرهمگن آمده باشند؛ مثلاً، زمانی که برای بیماران از روش‌های مختلف درمانی استفاده شده باشد یا زمانی که موضوع مورد مطالعه، بیماری‌هایی مانند سرطان باشد که خاصیت عود برای آن بیماری متصور است، استفاده از مدل‌های پارامتری ساده مانند وایبل به دلیل اینکه خاصیت غیرهمگنی موجود در داده‌ها را در نظر نمی‌گیرد، برازش مناسبی به داده‌ها نخواهد داشت (۳).

یکی از روش‌هایی که برای حل این مشکل پیشنهاد شده است استفاده از مدل‌های آمیخته برای داده‌های بقای غیرهمگن می‌باشد. این مدل‌ها به دلیل اینکه خاصیت چندمدی بودن داده‌ها را در نظر می‌گیرند، می‌توانند جایگزین مناسبی برای مدل‌های کلاسیک باشند (۴) و به دلیل اینکه مدل‌های کلاسیک هم، حالت خاصی از یک مدل آمیخته هستند، این مدل‌ها می‌توانند داده‌های همگن را نیز پوشش بدهند.

برکسون در سال ۱۹۵۲ برای اولین بار ایده اصلی استفاده از مدل‌های آمیخته را مطرح کرد (۵). چن و

همکاران در سال ۱۹۸۵ مدل آمیخته دو مولفه‌ای را برای تحلیل داده‌های بقای سرطان مورد استفاده قرار دادند (۶). کویانگ در سال ۱۹۹۴ یک مدل مشابه وایبل را که شامل یه جزء وایبل و یک نسبت بقا بود را برای داده‌های سرطان ریه مورد استفاده قرار داد (۷). در سال ۲۰۰۵ مارین و همکاران مدل وایبل را با استفاده از روش برآورد بیزی با تعداد مولفه‌های نامشخص را برای داده‌های بقای ناهمگن دارای سانسور راست به کار برد (۸). اریس اوغلو در سال ۲۰۱۰ مدل آمیخته هندسی-نمایی تعمیم یافته را به عنوان یک مدل جدید برای داده‌های بقا مورد استفاده قرار داد (۴) و در سال ۲۰۱۲ مدل‌های آمیخته گاما، وایبل و لگ نرمال را برای داده‌های سرطان ریه مورد استفاده قرار داد (۳) و برازش مدل‌های مختلف را بر روی داده، مورد بررسی قرار داد و در سال ۲۰۱۴ روش برآورد لگاریتم گشتاوری را که یک روش برآورد جدید برای مدل‌های آمیخته بود، برای برآورد پارامترهای مدل استفاده کرد (۹).

همانطور که گفته شد در مقالات ذکر شده مدل‌های جدید مانند توزیع نمایی هندسی یا توزیع رایلی و روش‌های برآورد جدید مورد بررسی قرار گرفته بودند. از این میان توزیع آمیخته وایبل، روش برآورد ماکزیمم درست‌نمایی، الگوریتم Expectation Maximization و روش برآورد بیزی بیشترین کاربرد را در میان مقالات داشتند. اما نکته مهمی که در مقالات، مورد توجه قرار نگرفته بود استفاده از متغیر کمکی و بررسی اثر متغیرهای کمکی بر روی زمان بقا می‌باشد.

مارین در سال ۲۰۰۵ پیشنهاد کرده است که وارد کردن متغیر کمکی به مدل می‌تواند تعمیم مناسبی به مدل‌های آمیخته باشد (۸). اما با این وجود همچنان این مهم، مورد بررسی قرار نگرفته است به همین دلیل در اینجا تلاش شده تا با استفاده از شبیه‌سازی داده‌ها و

در تحلیل داده‌های بقا، تابع بقا را با $s(t)$ و تابع خطر را با $h(t)$ نشان می‌دهند. رابطه بین تابع بقا و تابع توزیع به صورت

$$= 1 - F(t) s(t) = \Pr(T > t)$$

تعریف می‌شود (۱۰) و رابطه بین تابع خطر، تابع بقا و تابع چگالی هم به صورت

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}$$

می‌باشد (۱۰). مدل مورد استفاده در این مطالعه مدل آمیخته وایبل می‌باشد که تابع چگالی آن به صورت زیر است:

$$f(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{t}{\beta}\right)^\alpha\right), t > 0, \alpha, \beta > 0$$

در این رابطه α پارامتر شکل و β پارامتر مقیاس می‌باشد. مدل آمیخته وایبل به صورت زیر تعریف می‌شود (۱).

$$f(t|g, \pi, \alpha, \beta) = \sum_{k=1}^g \pi_k \frac{\alpha_k}{\beta_k} \left(\frac{t}{\beta_k}\right)^{\alpha_k-1} \exp\left(-\left(\frac{t}{\beta_k}\right)^{\alpha_k}\right)$$

در این رابطه g تعداد مولفه‌های توزیع آمیخته π_k پارامتر ترکیب می‌باشد که عددی در بازه $(0,1)$ است و دارای شرط $\sum_{k=1}^g \pi_k = 1$ می‌باشد (۴).

و همچنین تابع بقای آمیخته به صورت زیر تعریف خواهد شد.

$$S(t|g, \pi, \alpha, \beta) = \sum_{k=1}^g \pi_k \exp\left(-\left(\frac{t}{\beta_k}\right)^{\alpha_k}\right)$$

با فرض بر اینکه تعداد مولفه‌های توزیع برابر با g می‌باشد (۴). داده‌ها با فرض اینکه احتمالاً دارای سانسور راست باشند به صورت $x_i = (t_i, \delta_i)$ نمایش داده خواهند شد. که δ_i یک تابع نشان‌گر می‌باشد که نشان‌دهنده وضعیت سانسور برای مشاهده i می‌باشد و به صورت زیر تعریف خواهد شد.

$$\delta_i = \begin{cases} \text{باشد راست سانسور دارای اگر پذیرد می را 1 مقدار} \\ \text{باشد افتاده اتفاقی رخداد اگر پذیرد می 0 مقدار} \end{cases}$$

روش برآورد ماکزیمم درست‌نمایی، مدل‌های آمیخته در حضور متغیر کمکی بررسی گردد.

با توجه به اینکه داده‌های پیوند مغز استخوان به دلیل ویژگی عود بیماری، میزان خطر در زمان‌های مختلف یکسان نیست، برای تحلیل داده‌ها استفاده از مدل‌های آمیخته می‌تواند منجر به برآوردهای بهتری نسبت به مدل‌های بقای ساده شود، در این مقاله با استفاده از داده‌های پیوند مغز استخوان تمامی بیمارانی که در بیمارستان طالقانی تهران در فاصله سال‌های ۱۳۸۶ تا ۱۳۹۴ پیوند استخوان دریافت کرده بودند مناسبیت مدل‌های آمیخته و ساده برای داده‌های پیوند مغز استخوان بررسی شد.

مواد و روش کار

همانطور که قبلاً ذکر شد یکی از راه‌حل‌های تحلیل داده‌های بقای غیرهمگن استفاده از مدل‌های آمیخته می‌باشد. به عنوان مثال اگر داده‌های مورد استفاده مربوط به روش‌های گوناگون درمان برای یک بیماری باشد یا بیماری مورد مطالعه خاصیت عود داشته باشد داده‌های مورد بررسی می‌تواند غیرهمگن باشد.

اریس اوغلو در سال ۲۰۱۲ (۳) برای تحلیل داده‌های سرطان ریه از مدل‌های آمیخته وایبل، گاما و لگ نرمال استفاده کرده است. به همین دلیل در ادامه به تشریح مدل آمیخته وایبل که بیشترین کاربرد را در میان مدل‌های پارامتریک دارد پرداخته خواهد شد.

تلاش می‌گردد مناسبیت این مدل‌ها را برای تحلیل داده‌های بقای غیرهمگن در حضور متغیر کمکی مورد بررسی قرار داده و نهایتاً این مدل برای بررسی داده‌های واقعی به کار برده شود. تمامی تحلیل‌ها با استفاده از نرم افزار R نسخه ۳,۳,۳ و نرم افزار sas نسخه ۹,۲ انجام شده است.

صورت به‌توان میزان اثربخشی متغیر کمکی را برآورد کرد. دوباره پارامتری کردن پارامتر مقیاس براساس متغیر کمکی به صورت زیر خواهد بود (۱):

$$\beta_k = \exp(\beta_{0k} + \beta_{1k}x)$$

بر همین اساس تابع درست‌نمایی به صورت زیر تعریف خواهد شد.

$$L = \prod_{i=1}^n \sum_{k=1}^g \pi_k \left(\frac{\alpha_k}{\exp(\beta_{0k} + \beta_{1k}x)} \left(\frac{t_i}{\exp(\beta_{0k} + \beta_{1k}x)} \right)^{\alpha_k - 1} \right)^{\delta_i} \exp \left(- \left(\frac{t_i}{\exp(\beta_{0k} + \beta_{1k}x)} \right)^{\alpha_k} \right)$$

مراجعه نموده‌اند و عمل پیوند مغز استخوان برای آن‌ها انجام شده است، به صورت سرشماری مورد بررسی و مطالعه قرار گرفتند.

بر اساس پیگیری‌های انجام شده در سال ۱۳۹۵ تعداد ۱۰۹ نفر از بیماران (۲۲/۲ درصد) فوت کرده بودند و تعداد ۳۸۳ نفر از بیماران هنوز در قید حیات بودند آن‌ها سانسور راست در نظر گرفته شد.

در این مطالعه ۵۵/۴۹٪ افراد مورد بررسی زن بودند و میانگین (انحراف معیار) سن بیماران شرکت‌کننده ۳۸/۲ (۱۴/۰۲) بود.

با این فرضیات می‌توانیم تابع درست‌نمایی به صورت زیر تعریف می‌شود.

$$L = \prod_{i=1}^n \sum_{k=1}^g \pi_k \left(\frac{\alpha_k}{\beta_k} \left(\frac{t_i}{\beta_k} \right)^{\alpha_k - 1} \right)^{\delta_i} \exp \left(- \left(\frac{t_i}{\beta_k} \right)^{\alpha_k} \right)$$

با توجه به اینکه هدف اصلی این مطالعه وارد کردن متغیر کمکی به مدل می‌باشد. باید پارامتر مقیاس بر اساس متغیر کمکی دوباره پارامتری شود تا به این

برای برآورد پارامترهای مدل از روش ماکزیمم درست‌نمایی استفاده خواهد شد (۳).

مقایسه مدل‌ها: برای بررسی میزان اثر بخشی مدل و بررسی نتایج شبیه‌سازی از شاخص *RMSE* استفاده می‌شود که به صورت زیر تعریف خواهد شد (۱۱).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{\beta} - \beta)^2}{n}}$$

یافته‌ها

در مطالعه حاضر پرونده ۴۹۲ بیمار مبتلا به لوسمی یا پیوند مغز استخوان که بین سال‌های ۱۳۸۶ تا ۱۳۹۴ به بیمارستان طالقانی تهران برای پیوند مغز استخوان

جدول ۱: مشخصات دموگرافیک افراد شرکت‌کننده در مطالعه

متغیر	تعداد	درصد
جنس		
مرد	۲۱۹	۴۴٫۵۱
زن	۲۷۳	۵۵٫۴۹
سن		
کمتر از ۱۰ سال	۶	۱٫۲۲
۱۰ تا ۲۰ سال	۳۵	۷٫۱۴
۲۱ تا ۳۰ سال	۱۴۵	۲۹٫۵۹
۳۱ تا ۴۰ سال	۸۵	۱۷٫۳۵
۴۱ تا ۵۰ سال	۹۹	۲۰٫۲۰
۵۱ تا ۶۰ سال	۹۳	۱۸٫۹۸
۶۰ سال به بالا	۲۷	۵٫۵۱

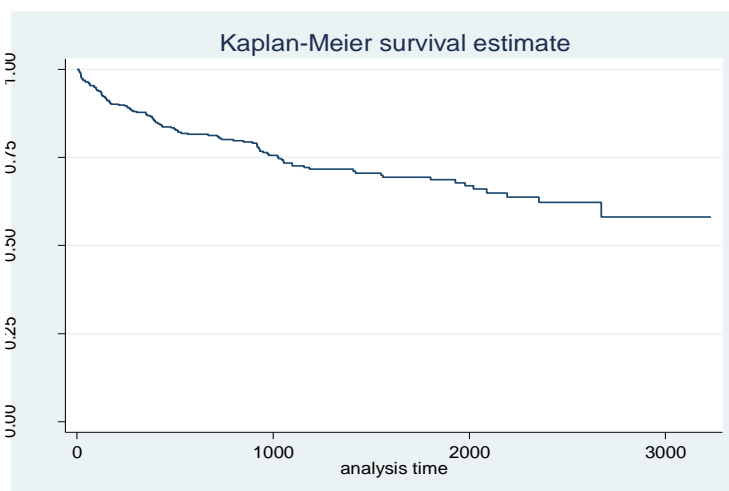
مورد پژوهش در این مطالعه ۲۵/۶ (۵/۲۴) بود.

میانگین (انحراف معیار) شاخص توده‌ی بدنی در افراد

جدول ۲: جدول توصیفی وضعیت عود داشتن بعد از پیوند بیماران شرکت کننده در مطالعه

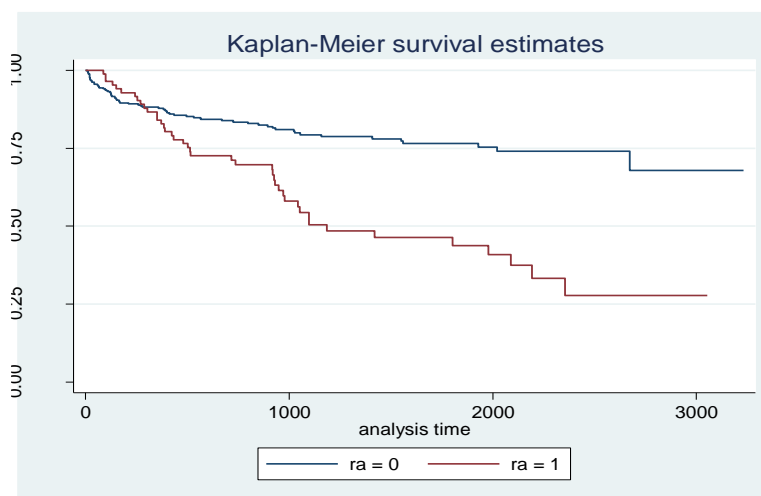
وضعیت عود بعد از پیوند	فراوانی	درصد
دارد	۸۴	۱۷/۰۷
ندارد	۴۰۸	۸۲/۹۳
جمع	۴۹۲	۱۰۰/۰

همانطور که مشاهده می شود ۸۴ (۱۷/۰۷) نفر از بیماران بعد از پیوند دچار عود شده اند و تعداد ۴۰۸ (۸۲/۹۳) نفر از بیماران نیز تا سال ۹۵ دچار عود نشده اند. نمودار میزان بقای کاپلان مایر برای داده های پیوند مغز استخوان به صورت زیر می باشد:



شکل ۱: وضعیت بقای کاپلان مایر افراد شرکت کننده در مطالعه

نتایج تحلیلی: در ابتدا مدل های ساده به داده برازش داده شد و تنها متغیر معنی دار در این مطالعه متغیر عود بعد از پیوند بود که در ادامه به بررسی مدل ها با حضور متغیر کمکی عود بعد از پیوند پرداخته می شود. وضعیت میزان زمان بقای بیماران در گروه دارای عود و گروه فاقد عود بعد از پیوند در شکل ۲ نمایش داده شده است.



شکل ۲: نمودار کاپلان مایر به تفکیک عود داشتن و نداشتن

درصد می‌باشد و در افرادی که عود نداشته‌اند در زمان ۲۵۰۰ احتمال زنده‌ماندن نزدیک به ۰/۷۳ درصد می‌باشد.

نتایج برازش مدل ساده و آمیخته وایبل به داده‌های پیوند مغز استخوان به صورت زیر می‌باشد:

همانطور که در شکل ۲ مشاهده می‌کنید نمودار بقای کاپلان مایر در افرادی که دارای عود بعد از پیوند بوده‌اند به طور قابل ملاحظه‌ای کمتر از افرادی است که عود نداشته‌اند. در زمان ۲۵۰۰ به طور متوسط شانس زنده‌ماندن در افرادی که عود داشته‌اند نزدیک ۰/۲۸

جدول ۳: نتایج برازش مدل‌های ساده و آمیخته وایبل به داده‌های پیوند مغز استخوان

مؤلفه اول	مؤلفه دوم	مؤلفه سوم	مؤلفه چهارم		
1	-	-	-	p	مدل ساده وایبل
0.678	-	-	-	α	
9.245	-	-	-	β_0	
-1.382	-	-	-	β_1	مدل آمیخته وایبل
97.3	2.7	-	-	p	
0.765	6.13	-	-	α	
9.187	2.985	-	-	β_0	با دو مؤلفه
-1.346	3.047	-	-	β_1	
97.0	1.1	1.9	-	p	
0.782	13.977	16.676	-	α	با سه مؤلفه
9.16	2.571	3.058	-	β_0	
-1.33	4.343	1.517	-	β_1	
96.8	1.1	1.8	.3	p	مدل آمیخته وایبل
0.799	14.391	16.25	25.225	α	
9.127	2.572	3.053	1.942	β_0	
-1.310	4.338	1.522	3.997	β_1	با چهار مؤلفه

آمیخته وایبل به صورت زیر می‌باشد:

مقادیر شاخص (12) AIC برای مدل‌های ساده و

جدول ۴: نتایج مقایسه مدل‌های ساده و آمیخته وایبل استفاده از شاخص AIC

گاما ساده	آمیخته گاما دو مؤلفه‌ای	آمیخته گاما سه مؤلفه‌ای	آمیخته گاما چهار مؤلفه‌ای	
1988.7	1980.1	1979.1	1983.9	AIC مقدار شاخص

برازش به داده‌های پیوند مغز استخوان در حضور متغیر کمکی می‌باشد.

بحث و نتیجه‌گیری

همانطور که در شکل ۲ مشاهده می‌کنید میزان زمان بقای افرادی که بعد از عمل عود داشته‌اند به صورت قابل ملاحظه‌ای کمتر از افرادی می‌باشد که عود

مدلی که کمترین مقدار شاخص AIC را دارا باشد بهترین مدل برای برازش به داده می‌باشد. با توجه به مقادیر به دست آمده از جدول ۴ و ۵ که نشان‌دهنده مقادیر شاخص AIC برای مدل‌های مختلف وایبل می‌باشد. مدل آمیخته وایبل با سه مؤلفه بهترین مدل برای

پارامتری دارند. همچنین با توجه به اینکه بررسی میزان اثر متغیرهای کمکی از مهم‌ترین هدف‌های برازش یک مدل می‌باشد، در این مطالعه تلاش شد که متغیرهای کمکی را به مدل اضافه کرده و سپس به بررسی میزان اثر آن متغیرها پرداخته شود.

در مدل‌های ساده پارامتری برای بررسی میزان اثر یک متغیر معمولاً از رابطه

$$HR = \exp(\beta_1)$$

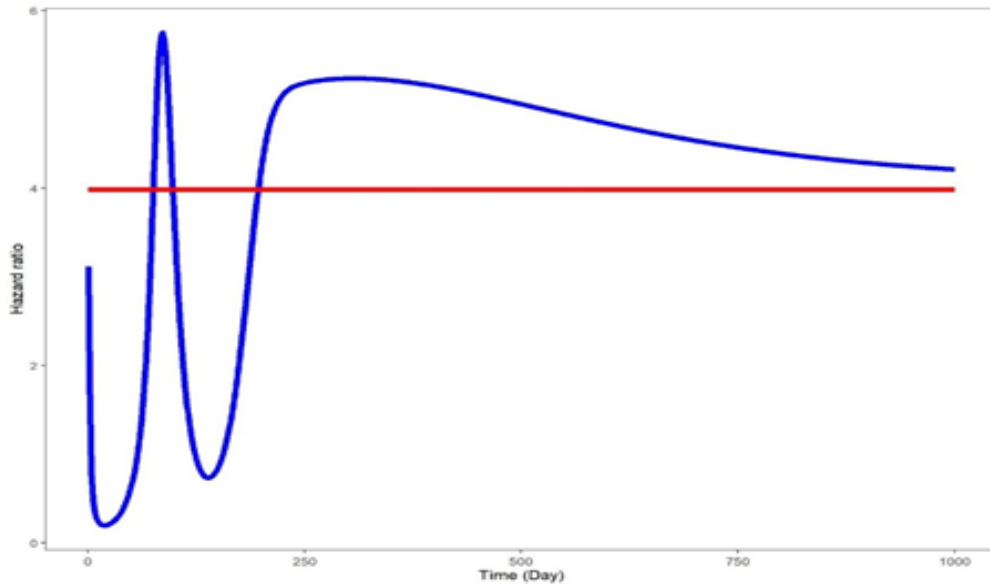
که همان شاخص HR می‌باشد، استفاده می‌شود. ثابت شده است که میزان این تابع در مدل ساده پارامتری ثابت می‌باشد. بنابراین میزان اثر متغیر کمکی در مدل‌های ساده پارامتری ثابت فرض می‌شود. اما در مدل‌های آمیخته همانطور که در شکل ۳ نشان داده شد میزان شاخص HR در طول زمان ثابت نیست و این شاخص می‌تواند کم و یا زیاد شود؛ بنابراین، برای تفسیر میزان شاخص اثر متغیر کمکی باید در زمان‌های خاص میزان این شاخص بررسی شود. به عنوان مثال در زمان $t=150$ میزان شاخص HR نزدیک به ۱ می‌باشد یعنی خطر فوت کردن بیماری که عود داشته است در زمان ۱۵۰ با فردی که عود نداشته است برابر می‌باشد در حالی که در زمان $t=250$ میزان شاخص HR نزدیک ۵ می‌باشد یعنی در زمان ۲۵۰ خطر فوت بیماری که عود داشته است ۵ برابر خطر مرگ بیماری است که عود نداشته است.

نداشته‌اند. به عنوان مثال در زمان ۲۵۰۰ به طور متوسط شانس زنده ماندن افراد که عود داشته‌اند تقریباً یک سوم افرادی است که عود نداشته‌اند.

همانطور که در شکل ۲ مشاهده گردید مقدار بقای بیمارانی که بعد از پیوند عود داشته‌اند به صورت قابل ملاحظه کمتر از افرادی بود که عود نداشته بودند. به همین دلیل از این متغیر به عنوان متغیر کمکی استفاده شد و نتایج برازش با استفاده از روش‌های بیشینه درست‌نمایی ارائه شد. همانطور که در جدول ۴ و ۵ مشاهده شد مدل آمیخته وایبل با سه مؤلفه به عنوان بهترین مدل برازش داده شده به داده‌های سرطان انتخاب شد.

در شرایطی که داده‌های مورد بررسی متعلق به بیمارهایی می‌باشد که میزان خطر در آن‌ها در طول زمان ثابت نیست مانند بیماری سرطان که در آن‌ها خطر عود بیماری وجود دارد استفاده از مدل‌های پارامتری ساده با این پیش‌فرض که داده‌ها در طول زمان نسبت خطر ثابتی دارند، باعث می‌شود که برآوردهایی که به دست می‌آیند برآوردهای درستی نباشد. به همین دلیل در این مطالعه ما سعی کردیم نتایج مدل‌های ساده و آمیخته را با استفاده از داده‌های سرطان خون با هم مقایسه کنیم.

همانطور که قبلاً ذکر شد مدل‌های آمیخته، به صورت کلی برازش بهتری نسبت به مدل‌های ساده



شکل ۳: نمودار تابع مخاطره برای مدل‌های ساده و آمیخته وایبل

- ۲- استفاده از توزیع‌های جایگزین مانند توزیع رایلی به منظور پیدا کردن بهترین مدل برای برازش به داده‌ها
- ۳- بهبود شرایط ثبت داده‌ها در بیمارستان‌ها و طولانی‌کردن زمان پیگیری برای بیماران سرطانی

تشکر و قدردانی

ابتدا از استاد گرانقدرم جناب آقای دکتر علی‌اکبر خادم معبودی که در طول انجام تحقیق و تکمیل این مطالعه زحمت راهنمایی را بر عهده داشتند، کمال سپاس دارم. با تشکر از جناب آقای دکتر احمد رضا باغستانی که با سعه صدر و با راهنمایی‌های عالمانه و بجایشان مرا در تمامی مراحل انجام این پژوهش یاری کردند.

محدودیت‌های پژوهش

یکی از اصلی‌ترین محدودیت‌ها ناقص بودن اطلاعات ثبت شده در پرونده بیماران و نبود اطلاعات مناسب در داده‌ها بود. همچنین با توجه به اینکه زمان انجام پیوند تا زمان اتمام مطالعه در بسیاری از مشاهدات کم بود به طوری که از میان ۴۹۲ بیمار ۱۳۹ بیمار قبل از سال ۱۳۹۰ و تعداد ۳۵۳ بیمار بعد از سال ۱۳۹۰ پیوند دریافت کرده بودند که باعث شد ناهمگنی موجود در داده‌ها که انتظار داشتیم در داده‌های بیماری پیوند مغز استخوان وجود داشته باشد به اندازه کافی مشاهده نشد.

پیشنهادات

- ۱- استفاده از روش‌های برآورد بیزی و یا الگوریتم EM برای برآورد پارامترها به منظور به دست آوردن برآوردهای بهتر.

References

- 1- Kleinbaum DG. Survival Analysis, a Self-Learning Text. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 1998;40(1):107-8.
- 2- Cox DR, Snell EJ. A general definition of residuals. Journal of the Royal Statistical Society Series B (Methodological). 1968;30(2):248-75.
- 3- Erişoğlu Ü, Erişoğlu M, Erol H. Mixture model approach to the analysis of heterogeneous survival time data. Pakistan Journal of Statistics. 2012;28(1):115-30.
- 4- Erişoğlu Ü, Erol H. Modeling heterogeneous survival data using mixture of extended exponential-geometric distributions. Communications in Statistics-Simulation and Computation. 2010;39(10):1939-52.
- 5- Berkson J, Gage RP. Survival curve for cancer patients following treatment. Journal of the American Statistical Association. 1952;47(259):501-15.
- 6- Chen W-C, Hill B, Greenhouse J, Fayos J. Bayesian analysis of survival curves for cancer patients following treatment. Bayesian statistics. 1985;2:299-328.
- 7- Qian J. A Bayesian Weibull survival model [Doctoral dissertation]. Private university in Durham, North Carolina: Duke University; 1994.
- 8- Marin J, Rodriguez-Bernal M, Wiper M. Using weibull mixture distributions to model heterogeneous survival data. Communications in Statistics-Simulation and Computation. 2005;34(3):673-84.
- 9- Erisoglu U, Erisoglu M. L-moments estimations for mixture of Weibull distributions. Journal of data science. 2014;12:69-85.
- 10- Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. Springer Science & Business Media; 2006.
- 11- Barnston AG. Correspondence among the correlation ,RMSE, and Heidke forecast verification measures; refinement of the Heidke score. Weather and Forecasting. 1992;7(4):699-709.
- 12- Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. Dordrecht, The Netherlands: D Reidel. 1986; 81.

Original paper

Bone Marrow Transplant Survival Data Analysis Using Simple Models and Mixture Weibull

Ali Akbar Khadem Maboudi¹, Ahmad Reza Baghestani¹, Mohsen Hamidpoor¹, saeed shiraziz, hamed bazyadz, naser ahmadi*

1- Department of Biostatistics, Faculty of Paramedical Science, Shahid beheshti University of medical Science, Tehran, Iran

2- Department of Information Technology, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran

Abstract

Background and Aim: According to the importance of cancerous diseases and the increasing attention to data about cancers and also the need for more accurate and appropriate analyzes about the survival of cancerous patients, their recurrency, and the heterogeneity of the risk factors, appropriate approaches are needed to be developed. In recent years, the mixture model has been considered as a suitable model for fitting the heterogeneous survival data. In this study, Weibull mixture model was used to assess the bone marrow transplantation survival data and the effect of cancer recurrency on the mortality rate.

Material and Method: In this study, 495 individuals with leukemia who referred to Taleghani hospital in Tehran and received bone marrow transplantation were examined. Individuals had received bone marrow transplantation during 2007 to 2015. They were followed up to 2016. Weibull simple and mixture models were used to fit the data. Finally, different models were compared.

Results: The mean survival lifetime (standard deviation) of individuals who had experienced the event was 563 days (57.4) and 999 (45.4) for censored individuals. The AIC index showed that the Weibull mixture model fitted the data better than the simple model. It could explain the heterogeneity in both recurrent and non-recurrent groups better than the Weibull simple model.

Conclusion: This study showed that the use of mixture models can perform better than simple parametric models to analyze heterogeneous survival data such as cancer data.

Keywords: Heterogeneous survival data, Mixed Weibull distribution, Recurrence, Maximum likelihood estimation