

# بکارگیری روش‌های هموارسازی برای برآورد ضرایب مدل نرخ وابسته به زمان در تحلیل بقا و کاربرد آن در مبتلایان به بیماری پسوریازیس

فرشته عثمانی<sup>۱</sup>، ابراهیم حاجی‌زاده<sup>۲</sup>، پروین منصوری<sup>۳</sup>

<sup>۱</sup> دانشجوی دوره دکتری، گروه آمار زیستی، دانشگاه تربیت مدرس، تهران، ایران

<sup>۲</sup> استاد، گروه آمار زیستی، دانشگاه تربیت مدرس، تهران، ایران

<sup>۳</sup> فوق تخصص پوست، مرکز تحقیقات بیماری‌های پوست، دانشگاه علوم پزشکی تهران، تهران، ایران

نویسنده رابط: ابراهیم حاجی‌زاده، نشانی: تهران، بزرگراه جلال آل‌احمد، پل نصر، دانشگاه تربیت مدرس، دانشکده علوم پزشکی، گروه آمار زیستی، تلفن: ۸۲۸۸۳۸۱

پست الکترونیک: hajizadeh@modares.ac.ir

تاریخ دریافت: ۹۴/۰۸/۱۶؛ پذیرش: ۹۴/۱۲/۰۸

**مقدمه و اهداف:** داده‌های مطالعه‌هایی که در آن هر فرد امکان تجربه متعدد یک پیشامد را در زمان‌های مختلف داشته باشد؛ از نوع حوادث بازگشتی هستند. از رایج‌ترین رویکردهای تحلیل رخدادهای بازگشتی، به‌دست آوردن برآوردی از میانگین نرخ بروز رخدادهای در زمان‌های مختلف است. در این زمینه یکی از مواردی که می‌تواند به شناخت بیشتر اثر فاکتور مورد نظر بر پاسخ کمک کند، مشخص کردن تغییرپذیری اثر متغیرهای کمی روی نرخ وقوع رخدادهای در طول زمان است. در این مطالعه از روش‌های کرنل و بی‌اسپلین برای هموارسازی برآورد ضرایب در مدل نرخ با ضرایب وابسته-زمان و کاربرد آن در داده‌های مبتلایان به پسوریازیس استفاده شد.

**روش کار:** در این مطالعه از داده‌های موجود در پرونده‌های بیماران مبتلا به پسوریازیس که در طی فروردین ۱۳۸۵ تا خرداد ۱۳۹۳ در بخش پوست بیمارستان امام خمینی تهران، عود منجر به بستری داشته‌اند، استفاده شد. برای بررسی میزان عود بیماری در طول سال از مدل نرخ با ضرایب وابسته-زمان استفاده شد و تغییرپذیری اثرات با استفاده از آزمون والد مورد ارزیابی قرار گرفت. برای برآورد ضرایب در مدل نرخ وابسته-زمان از توابع بی‌اسپلین و کرنل کمک گرفته شد. در پایان یافته‌های حاصل از دو روش براساس برآوردهای به‌دست آمده با یکدیگر مقایسه شدند.

**یافته‌ها:** نتایج این مطالعه نشان داد که براساس نتیجه آزمون والد، اثر متغیری مانند فصل عود روی میزان عود پسوریازیس به‌صورت معنی‌داری متفاوت است ( $P < 0.01$ ). همچنین با توجه به برآورد ضرایب از دو روش بی‌اسپلین و کرنل تفاوت چندانی بین این دو روش مشاهده نشد.

**نتیجه‌گیری:** در شرایطی که اثر متغیر مورد بررسی روی وقوع رخدادهای در نقاط مختلف زمانی متفاوت است؛ استفاده از مدل نرخ با ضرایب وابسته-زمان می‌تواند برآورد مناسب‌تری از اثر متغیر مورد نظر بر پاسخ ارائه دهد.

**واژگان کلیدی:** رخدادهای بازگشتی، بی‌اسپلین، کرنل، مدل نرخ با ضرایب وابسته-زمان، پسوریازیس

## مقدمه

به این ترتیب است که ارزیابی اثربخشی داروها ممکن می‌شود. در داده‌های طولی بقا، داده‌های زمان چندگانه مطرح می‌شوند و در این مورد یک فرآیند تصادفی و تعدادی پیشامد برای هر شخص مشاهده می‌شود. معمولاً در مطالعات بیماری‌های مزمن این نوع داده‌ها مشاهده می‌شوند. همچنین مطالعه‌هایی که در آن هر فرد امکان تجربه متعدد یک پیشامد را در زمان‌های مختلف داشته باشد، داده‌ها از نوع طولی بقا هستند. این گونه نوع داده‌ها

تحلیل بقا مجموعه‌ای از روش‌های آماری برای تحلیل داده‌هایی است که متغیر پاسخ آن‌ها زمان لازم تا رخداد یک پیشامد است. عبارت داده‌های بقا<sup>۱</sup> برای توصیف داده‌هایی است که زمان تا وقوع پیشامد<sup>۲</sup> خاصی را اندازه‌گیری می‌کنند. یکی از محورهای اصلی پژوهش در عرصه سلامت و بیماری، بررسی طول عمر یا زمان بازگشت بیماری در بیمارانی است که درمانی را دریافت کرده‌اند و

<sup>۱</sup> Survival Data

<sup>۲</sup> Time to Event

مدل‌های رگرسیونی را برای توابع میانگین و نرخ ارایه کردند (۱۴-۱۱) لین و همکاران (۱۳) فرمول دقیقی را برای مدل‌های نرخ و میانگین حاشیه‌ای با تابع لینک از نوع کاکس به‌دست آوردند و روش‌های استنباط را برای ساختار زمانی پیوسته گسترش دادند. مدل‌های رگرسیونی نیمه پارامتری مذکور فرض می‌کردند که ضرایب رگرسیونی در طول زمان ثابت هستند. در واقع اگرچه پارامترهای رگرسیونی ممکن است در طول زمان تغییر کنند و از طرفی مهم است که اثرهای موقتی این متغیرها روی زمان‌های بازگشت رخداد مشخص شود؛ مدل‌هایی با ضرایب وابسته به زمان باعث استنباط دقیق‌تر درباره اثرات متغیرها می‌شوند. برای داده‌های سانسور از راست برخی از پژوهشگران با استفاده از روش‌های مختلف مانند درست‌نمایی جزئی جرمه شده (۱۵) و روش برآورد یک مرحله ای (۱۶) و باقی مانده شونفلند (۱۷) و تکنیک درست‌نمایی جزئی موضعی (۱۸،۱۹) و اخیراً هم آموریم و همکاران از اسپلاین رگرسیونی برای تحلیل مدل ضرایب وابسته به زمان برای رخدادهای برگشت پذیر استفاده کردند. آن‌ها تعمیمی از مدل لین را ارایه کردند که با استفاده از آن می‌توان اثر متغیرهای کمکی را روی میزان وقوع رخدادهای، در طول زمان بررسی کرد (۲۰) این مدل زمانی مفید است که فرض شود اثر برخی از متغیرهای کمکی روی میزان وقوع رخدادهای به ازای زمان‌های مختلف، متفاوت هستند. این پژوهش بر مبنای این مدل انجام شده است. بنابراین با توجه به مطالب گفته شده در بالا هدف از این مطالعه، مقایسه‌ی روش‌های هموارسازی در برآورد ضرایب مدل نرخ وابسته به زمان و کاربرد این روش در داده‌های بیماران مبتلا به پسروریازیس بود.

## روش کار

این پژوهش که از نوع کاربردی می‌باشد با هدف به‌کارگیری روش‌های هموارسازی در برآورد ضرایب مدل نرخ وابسته به زمان و کاربرد این روش در تعیین میزان بروز عود پسروریازیس در بین ۱۶۰ بیمار مبتلا به پسروریازیس بستری شده در بخش پوست بیمارستان امام خمینی تهران در طی مدت فروردین ۱۳۸۵ تا فروردین سال ۱۳۹۳ انجام گرفت. اطلاعات مورد نیاز با مطالعه پرونده بیماران استخراج شده و داخل چک لیست ثبت شد. برای برازش مدل، ابتدا باید داده‌ها به نرم‌افزار وارد شود. رویکردهای متفاوتی برای وارد کردن داده‌های مربوط به رخدادهای بازگشتی در نرم‌افزار وجود دارند. در این جا از رویکرد فرآیند شمارشی

را حوادث بازگشتی<sup>۱</sup> نیز می‌نامند. طی ۲۰ سال اخیر روش‌ها و مدل‌های متعددی برای مدل‌بندی و تحلیل داده‌های مربوط به این گونه رخدادهای پیشنهاد شده است. یکی از رویکردهای موجود، به‌دست آوردن برآوردی از میانگین/میزان بروز رخدادهای در زمان‌های مختلف است. سادگی و قابل فهم بودن یافته‌های حاصل از این تحلیل آن را در زمره یکی از رایج‌ترین روش‌ها در تحلیل این گونه رخدادهای قرار داده است (۴-۱) دو هدف عمده‌ای که در این زمینه دنبال می‌شوند، عبارت‌اند از: تعیین نوع رابطه‌ی موجود بین متغیرهای کمکی و زمان لازم تا وقوع رخداد و مشخص کردن این که آیا اثر فاکتورهای تحت مطالعه روی وقوع رخدادهای در طول زمان ثابت است یا خیر و در صورت مثبت بودن پاسخ، مشخص کردن نوع این تغییر است. معروف‌ترین و ساده‌ترین مدلی که در زمینه‌ی تحلیل داده‌های بازگردانده مورد استفاده قرار می‌گیرد، مدل اندرسون-گیل یا به اختصار مدل AG نام دارد که در سال ۱۹۸۲ میلادی ارایه شده است. در این مدل، دنباله‌ی رخدادهایی که برای هر یک از واحدهای آزمایشی اتفاق افتاده است؛ به صورت یک فرآیند شمارشی با نمونه‌های مستقل در نظر گرفته می‌شود. مدل AG در شرایطی به کار می‌آید که پژوهشگر تنها بخواهد میزان کلی وقوع رخداد را در زمان‌های مختلف به دست آورد؛ علاوه بر آن، بیش‌تر افراد تحت مطالعه کمتر از دو رخداد را تجربه کرده باشند (۵،۶) بیش‌تر روش‌هایی که برای تحلیل داده‌های بازگشتی به‌کار می‌روند؛ فقط اثرات متغیرهای کمکی مستقل از زمان را در نظر می‌گیرند که این فرض ممکن است گاهی اوقات درست نباشد.

لیونگان سان و همکاران تحلیل رگرسیونی از داده‌های بازگشتی چند متغیره که در آن اثرات برخی از متغیرهای کمکی ممکن بود وابسته به زمان باشد؛ در نظر گرفتند. برای این مشکل از روش مدل‌بندی حاشیه‌ای استفاده کردند (۷).

در چهارچوب مدل‌های پارامتری مانند مدل‌های غیرخطی و خطی حاشیه‌ای و مدل‌های اثرات آمیخته روش‌ها و اصول مختلفی برای برآورد و استنباط، به‌طورگسترده‌ای مورد بحث و مطالعه قرار گرفته است، به عنوان مثال لیانگ و زگر (۸) و روش‌های هموارسازی ناپارامتری مانند برآوردگرهای کرنل و اسپلاین‌های رگرسیونی را برای برآورد میانگین پاسخ بدون حضور متغیرهای کمکی وابسته به زمان ارایه دادند (۱۰) برخی از پژوهشگران

<sup>۱</sup>-Recurrent event

کدورت لوله ۰/۵ مک فارلند تهیه می‌شد و در مرحله بعد، از تعلیق فوق در سطح محیط مولر هینتون آگار توسط سواب استریل کشت شده و دیسک‌های مربوط (N(30 Am (10 µg), F(30 µg), Na(30 µg), St (10 µg), LS(15/200 µg), Tt(30 µg) و Ff(30 µg) در سطح آگار قرار گرفته و پس از ۲۴ ساعت هاله‌های ناشی از عدم رشد باکتری با استفاده از خط‌کش اندازه‌گیری شده و نتایج مورد بررسی قرار می‌گرفت (۱۴، ۱۵).

### تجزیه و تحلیل آماری

ارتباط متغیرهایی چون سن، محل زندگی و جنس با آلودگی گربه‌های روستایی نمونه‌گیری شده در این پژوهش به سالمونلا با استفاده از نرم‌افزار SPSS نسخه ۲۱ و آزمون مربع کای با سطح معنی‌داری  $P \leq 0.05$  مورد بررسی قرار گرفت.

اطلاعات اپیدمیولوژیک اندکی در مورد سایر اشکال بیماری در دسترس است. متأسفانه آمار دقیقی از شیوع پسروریازیس در کشور ما در دست نیست و علاوه بر آن مطالعه‌های انجام شده در این مورد در ایران محدود بوده و در رابطه با عودهای مکرر این بیماری و زمان‌های وقوع آن مطالعه‌های کمی صورت گرفته است. با توجه به اهمیت این بیماری و عدم وجود علاج واقعی برای آن و این‌که این بیماری، فرد مبتلا و همین‌طور دولت را برای ارایه درمان‌های مناسب، تأمین داروهای گران‌بها، بستری در بخش‌های بیمارستانی به مدت طولانی، مراجعه‌های مکرر به درمانگاه‌ها برای پیگیری و صرف هزینه‌های زیاد، دچار مشکلات فراوانی می‌کند؛ بنابراین انجام پژوهش در مورد این بیماری، شناسایی عوامل تشدید و تحریک کننده آن و تلاش برای کمک به طولانی شدن دوره‌ی بهبودی و کاهش عودهای مکرر بیماری و همچنین استفاده از روش‌های تحلیل آماری مخصوص این گونه داده‌ها، مفید به نظر می‌رسد.

روش‌های هموارسازی، یک رویکرد انعطاف‌پذیر برای مدل‌های بقا است، که روند پاسخ را روی متغیرهایی که توسط این تکنیک هموار می‌شوند؛ برآورد می‌کند. هموارسازی وسیله‌ای برای توصیف روند  $\lambda$  به عنوان تابعی از متغیرهای مستقل است. برآوردگری از روند که با این شیوه حاصل می‌شود، کم‌تر از مشاهده‌های  $\lambda$  نوسان خواهد داشت و به این سبب هموارساز نامیده می‌شود. برآوردی که از یک هموارساز حاصل می‌شود؛ هموار می‌نامند.

منحنی‌های بزیر<sup>۱</sup>

استفاده شد. برای بررسی این‌که آیا نرخ عود بیماری در فصل‌های مختلف سال ثابت است یا متغیر، از مدل نرخ با ضرایب وابسته-زمان استفاده شد. در این روش برای برآورد ضرایب وابسته به زمان در مدل نرخ از روش‌های هموارسازی بی-اسپلاین و کرنل استفاده شد تا ضرایب همواری در طول زمان به‌دست آید. برخی از متغیرهای جمع آوری شده در این مطالعه شامل: سن ورود به مطالعه، جنس، سابقه‌ی فامیلی، زمان بستری و ترخیص وضع بیمار در زمان T، محل سکونت (شهر، روستا)، وضع تأهل، فصل عود منجر به بستری، سطح کلسیم خون، بیماری همراه، عامل‌های تشدید کننده بیماری، مصرف سیگار، عفونت و تعداد گلبول‌های سفید خون بود.

پسروریازیس یک بیماری پوستی مزمن عود کننده با منشأ ژنتیکی است که تاکنون درمانی قطعی برای آن یافت نشده است. این بیماری به طور معمول حیات بیمار را مورد تهدید قرار نداده، اما گاهی با درگیری مفصل سبب ناتوانی می‌شود. شیوع بیماری در قسمت‌های مختلف دنیا ۱-۳ درصد گزارش شده است و با توجه به جمعیت کشور ایران، تقریباً ۱-۳ میلیون نفر مبتلا به پسروریازیس وجود دارد. بیماری در هر دو جنس به طور مساوی دیده می‌شود، و سن شروع آن اغلب دهه‌ی ۲-۴ زندگی است. در افرادی که سابقه‌ی فامیلی مثبتی دارند؛ بیماری در سنین پایین‌تری شروع می‌شود. پسروریازیس ممکن است در اثر عفونت از جمله عفونت‌های استرپتوکوک‌های بتاهمولیتیک، قطع ناگهانی مصرف داروهای استروئیدی، استرس، بارداری، گرفتن درمان‌های خاص مانند لیپتیم یا کورتیزون سیستمیک، تشدید شود. بیش‌تر مطالعه‌ها در این زمینه روی پسروریازیس پلاکی مزمن -که شایع‌ترین شکل بیماری هست- انجام شده است (۲۱، ۲۲).

محصولات حاصل از PCR روی ژل آگارز ۱/۲ درصد الکتروفورز شده و پس از رنگ‌آمیزی با اتیدیوم بروماید با دستگاه ژل داکت عکس‌برداری شده و مورد بررسی قرار گرفت.

### تعیین الگوی مقاومت دارویی

از روش کیفی دیسک دیفوزن با روش استاندارد برای تعیین الگوی مقاومت دارویی نسبت به ۸ آنتی‌بیوتیک: آمپی‌سیلین (Am)، نتوماپسین (N)، لینکوسپکتین (Ls)، استرپتوماپسین (St)، نالیدیکسیک اسید (Na)، فورازولیدون (F)، تتراسایکلین (Tt) و فلورفنیکل (Ff) استفاده شد. به این صورت که ابتدا از پرگنه‌های مربوط به جدایه‌های سالمونلا در محیط TSB برای مدت چند ساعت کشت شده و سپس از آن‌ها تعلیقی با کدورت معادل

<sup>۱</sup> Beziér Curve

جمله‌ای‌های برنستین و سپس مجموع این حاصل ضرب‌ها به دست می‌آیند، در واقع از به هم پیوستن چندین منحنی بزیر یک منحنی بی-اسپلاین به دست می‌آید. هر منحنی بی-اسپلاین روی دامنه‌ای از اعداد تعریف می‌شود، برای مثال (0 1). این دامنه متشکل از مجموعه‌ای از زیر فاصله‌ها است که هر کدام از آن‌ها، دامنه منحنی بزیری هستند که آن قسمت از بی-اسپلاین را تشکیل داده است. نقاطی که روی دامنه بی-اسپلاین قرار داشته و موجب تقسیم دامنه به زیر فاصله‌ها می‌شوند، گره نام دارند گره‌ها را با  $u_i$  نشان داده و به مجموعه آن‌ها بردار گره‌ها گفته می‌شود.

$$u_0 \leq u_1 \leq u_2 \leq \dots \leq u_m$$

که در آن  $u_i$  دنباله‌ای غیر نزولی متشکل از  $m+1$  عدد هستند. همان‌طور که منحنی‌های بزیر از چندجمله‌ای‌های برنستین به عنوان تابع وزنی استفاده می‌کنند؛ منحنی‌های بی اسپلاین از توابع پایه‌ای بی-اسپلاین به عنوان وزن نقاط کنترلی بهره می‌جویند. توابع پایه‌ای بی-اسپلاین را می‌توان از روابط بازگشتی زیر به دست آورد:

$$N_{i,0}(u) = \begin{cases} 1 & u_i \leq u \leq u_{i+1} \\ 0 & \text{otherwise} \end{cases}$$

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u)$$

که در آن  $p$  درجه بی-اسپلاین است (۲۳).

### برآوردگر کرنل<sup>۲</sup>:

برآوردگر کرنل تعمیمی از برآوردگر ساده به منظور برطرف کردن بر مشکلات این برآوردگر است. در برآوردگر ساده چنانچه تابع وزن  $w$  به وسیله‌ی تابع  $K$  به نام هسته که در شرایط زیر صدق کند؛ جایگزین شود؛ برآوردگر کرنل با هسته  $K$  حاصل می‌شود.

$$\int_{-\infty}^{\infty} K(x) dx = 1$$

معمولاً اما نه همیشه،  $K$  خود یک تابع چگالی احتمال متقارن مانند چگالی گاوسی می‌باشد.

بسیاری از روش‌های آمار و ریاضی، معطوف به برآزش دادن منحنی‌ای مناسب به مجموعه‌ای از داده‌ها است. توابع چند جمله‌ای از مرسوم‌ترین روش‌هایی هستند که در این زمینه مورد استفاده قرار می‌گیرند:

$$y(t) = \sum_{i=0}^n a_i t^i$$

که در این رابطه  $n$  نشان‌دهنده درجه چندجمله‌ای می‌باشد. با افزایش درجه چندجمله‌ای، گستره وسیع‌تری از منحنی‌ها ساخته می‌شوند. اگرچه بزرگ بودن این درجه مشکلاتی را نیز به همراه دارد. از جمله این‌که محاسبه‌های ریاضی پیچیده‌تر شده، شکل منحنی بسیار غیرقابل کنترل می‌شود. به این معنی که کوچک‌ترین تغییرات در ضرایب، شکل منحنی را به شدت عوض خواهد کرد. بنابراین، هدف پیدا کردن منحنی‌ای است که علاوه بر داشتن حد مطلوبی از انعطاف‌پذیری، محاسبه آن نیز ساده باشد. با نگاهی دوباره به  $y(t) = \sum_{i=0}^n a_i t^i$ ، می‌توان  $a_i$ ها را نقاطی در فضای دو بعدی و  $y(t)$  را مجموع وزنی  $a_i$ ها در نظر گرفته شود. می‌توان از توابع دیگری نیز به عنوان تابع وزنی استفاده کرد (۲۳). منحنی‌های بزیر از چندجمله‌ای‌های برنستین به عنوان تابع وزنی استفاده می‌کنند.

$a_i$  همان وزن‌های تعریف شده در چند جمله‌ای هستند که در منحنی‌های بزیر از تابع وزنی زیر به جای  $a_i$  در چند جمله‌ای بالا استفاده می‌شود.

$$B_{i,n} = \binom{n}{i} u^i (1-u)^{n-i}$$

$$i = 0, 1, \dots, n \quad 0 \leq u \leq 1$$

### منحنی‌های بی-اسپلاین<sup>۱</sup>

دو ویژگی مهم منحنی‌های بزیر عبارت‌اند از: ۱- با افزایش تعداد نقاط کنترلی، درجه منحنی نیز افزایش پیدا می‌کند. ۲- در صورتی که یکی از نقاط کنترلی جابه‌جا شوند، شکل کل منحنی تغییر می‌کند. این تغییر شکل منحنی به دلیل ساختار چند جمله‌ای‌های برنستین اتفاق می‌افتد. از آن جایی که نقاط منحنی بزیر به وسیله ضرب کردن مختصات نقاط کنترلی در چند

<sup>۲</sup> kernel estimator

<sup>۱</sup> B-spline Curve

که در آن  $\beta$  برداری  $(p-1) \times 1$  از ضرایب رگرسیونی ثابت و  $\theta(t)$  ضریب رگرسیونی وابسته به زمان است. متغیرهای کمکی  $Z(t)$  و  $W(t)$  می‌توانند از نوع وابسته به زمان یا مستقل زمان باشند. برای مثال اگر  $W$  متغیری مستقل زمان باشد که مواجهه یا عدم مواجهه با عاملی خطرناک را نشان دهد؛ آن‌گاه:

$$RR(t) = \exp(\theta(t))$$

یعنی نرخ پاسخ در گروه مواجهه یافته  $\exp(\theta(t))$  برابر نرخ پاسخ در گروه مواجه نیافته است. در این مرحله برای برآورد  $\theta(t)$  باید روشی را اتخاذ کرد. از آن جایی که توابع هموارساز، به دلیل انعطاف‌پذیر بودن، در روش‌های آماری بسیار مورد استفاده قرار می‌گیرند، در این پژوهش نیز برای برآورد کردن  $\theta(t)$  از توابع کرنل و بی اسپلاین استفاده می‌شود. همان‌طور که ذکر شد، در استفاده از توابع بی اسپلاین باید درجه و تعداد گره‌های آن مشخص باشند. باید توجه داشت، اگرچه با افزایش درجه بی-اسپلاین، برازش منحنی به داده‌ها بهتر می‌شود، اما علاوه بر پیچیده‌تر شدن محاسبه‌ها، قابلیت تعمیم نتایج نیز پایین می‌آید. در بین توابع بی-اسپلاین با درجه‌های مختلف، استفاده از تابع بی-اسپلاین درجه سوم نسبت به بقیه رایج‌تر است، زیرا علاوه بر دارا بودن حد مطلوبی از انعطاف‌پذیری، محاسبه آن نیز ساده است؛ بنابراین در این مطالعه نیز از توابع بی اسپلاین درجه سوم برای برآورد کردن ضرایب وابسته-زمان در مدل نرخ استفاده می‌شود:

$$\theta(t) = \gamma_0 + \sum_{k=1}^{m+3} \gamma_k \hat{\beta}_k(t)$$

در این رابطه  $\hat{\beta}_k(t)$  توابع پایه‌ای بی-اسپلاین هستند که بر مبنای بردار مربوط به زمان‌های مطالعه ساخته می‌شوند و  $m$  نشان‌دهنده تعداد گره‌های داخلی بی-اسپلاین است؛ بنابراین اگر فرض شود طول بردار  $t$ ، برابر  $n$  باشد  $\hat{\beta}_k(t)$  برداری  $n \times 1$  خواهد بود. حال در تابع نرخ به جای  $\theta(t)$  معادل آن را که با استفاده از بی-اسپلاین محاسبه شده، قرار داده شود:

$$d\mu_i(t) = \exp\{\beta' Z_i(t) + (\gamma_0 + \sum_{k=1}^{m+3} \gamma_k \hat{\beta}_k(t)) W_i(t)\} d\mu_0(t)$$

با ضرب  $W_i(t)$  در عبارت قبل از آن خواهیم داشت:

$$d\mu_i(t) = \exp\left\{\beta' Z_i(t) + \gamma_0 W_i(t) + \sum_{k=1}^{m+3} \gamma_k \tilde{\beta}_k(t) W_i(t)\right\} d\mu_0(t)$$

حال فرض کنید ماتریس  $\tilde{W}_i(t)$  و بردار  $\gamma$  به صورت زیر تعریف

به این ترتیب برآوردگر کرنل با هسته  $K$  به صورت زیر تعریف می‌شود:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

که در آن تابع  $K$  شکل برآمدگی‌ها (مستطیل، مثلث، زنگوله‌ای و سهمی) را مشخص کرده و  $h$  پهنای پنجره، پارامتر هموارسازی یا پهنای باند نامیده می‌شود و مقدار آن میزان هموارسازی در برآورد کرنل را تعیین می‌کند. همانند برآوردگر ساده که آن را می‌توان به عنوان جمع «جعبه»های مرکزی شده در مشاهده‌ها در نظر گرفت؛ برآوردگر کرنل نیز مجموع «برآمدگی»هایی است که مشاهده‌ها در مرکز آن‌ها را مشخص می‌کند. اصول نظری و کاربردی برآوردها بر اساس روش‌های مختلف کرنل، اسپلاین و چندجمله‌ای‌های وزنی موضعی به طور گسترده‌ای در مطالب مربوط به برآورد منحنی ناپارامتری بحث شده است. با انتخاب صحیح پارامترهای هموارسازی، این روش‌های برآورد خواص جانبی خوبی مانند نرخ مطلوب همگرایی و اغلب هم یافته‌های قابل اعتمادی در کاربردهای واقعی می‌دهند؛ بنابراین کاملاً منطقی است که این روش‌ها را برای برآورد برای مشاهده‌های مربوط به مطالعه‌های طولی به کار برده شود؛ یکی از مزیت‌های استفاده از این روش انعطاف‌پذیری شکل است. برآوردگرهای کرنل، برآوردگرهای خطی هستند به این معنی که می‌توان مقدار این برآوردگر در هر نقطه  $t$  به صورت مجموع وزنی از پاسخ‌ها نشان داد (۲۴).

### مدل نرخ با ضرایب وابسته به زمان<sup>۱</sup>:

برای شروع، مدل نرخ متناسب برای فرد  $\mu$  در نظر گرفته می‌شود:

$$d\mu_i(t) = \exp\{\beta' Z_i(t)\} d\mu_0(t)$$

فرض کنید  $W(t)$  متغیری باشد که قرار است اثر آن روی تابع نرخ در طول زمان بررسی شود؛ به عبارت دیگر هدف پیدا کردن برآوردهای متفاوتی از اثر  $W(t)$  روی نرخ پاسخ به ازای زمان‌های مختلف است. از آن جایی که اثر هر متغیر روی تابع نرخ به وسیله ضریب آن متغیر می‌شود، لازم است به ازای زمان‌های مختلف، ضرایب متفاوتی برای متغیر مورد نظر برآورد شود؛ بنابراین می‌توان تابع نرخ را به صورت زیر نوشت:

$$d\mu_i(t) = \exp\{\beta' Z_i(t) + \theta(t) W_i(t)\} d\mu_0(t)$$

<sup>۱</sup> Time dependent Coefficient Rates Model

شود:

$$\tilde{W}_i(t) = (W_i(t), \tilde{B}_1(t) W_i(t), \tilde{B}_2(t) W_i(t), \dots, \tilde{B}_{m+3}(t) W_i(t))'$$

$$\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{m+3})'$$

بنابراین می‌توان مدل نرخ را به شکل زیر نوشت :

$$d\mu(t) = \exp\{\beta'Z(t) + \gamma' \tilde{W}(t)\} d\mu_0(t)$$

با اندکی تأمل در مدل، مشاهده می‌شود که دیگر شامل ضریب وابسته به زمان نیست و تبدیل به مدل نرخ متناسب شده است که در قسمت قبل بررسی شدند. بنابراین پارامترهای آن از روش ماکزیمم کردن تابع درست‌نمایی جزئی به دست آورده می‌شود. حال با استفاده از رابطه‌ی زیر،  $\theta(t)$  یعنی ضرایب وابسته به زمان برای متغیر  $W_i(t)$  محاسبه می‌شود:

$$\hat{\theta}(t) = \hat{\gamma}_0 + \sum_{k=1}^{m+3} \hat{\gamma}_k \tilde{B}_k(t)$$

برای تعیین تعداد گره‌ها برای ساخت توابع بی-اسپلاین، از ملاک AIC کمک گرفته می‌شود. به این ترتیب که ابتدا مدل را به ازای  $m=2,3,\dots,6$  گره برازش داده، سپس AIC را برای هر یک از مدل‌های برازش داده‌شده محاسبه شود و در نهایت مدلی را که دارای کمترین مقدار AIC باشد؛ انتخاب می‌شود. اگرچه مدل به ازای زمان‌های مختلف، برآوردهای متفاوتی از  $\theta$  ارائه می‌کند، لازم است به روشی آزمون شود که آیا تفاوت مقادیر  $\theta(t)$  در زمان‌های مختلف به اندازه‌ای هست که پذیرفته شود اثر متغیر مربوط در طول زمان تغییر می‌کند، یا این که این تفاوت ناچیز است و می‌توان اثر متغیر مربوط در طول زمان را ثابت فرض کرد. اگر در رابطه:

$$\hat{\theta}(t) = \hat{\gamma}_0 + \sum_{k=1}^{m+3} \hat{\gamma}_k \tilde{B}_k(t)$$

جمله دوم صفر باشد،  $\theta(t)$  به ازای تمام زمان‌ها ثابت برآورد می‌شود، یعنی  $\hat{\theta}(t) = \hat{\gamma}_0$ . بنابراین برای اطمینان از ثابت بودن ضرایب در طول زمان، می‌توان آزمون کرد که آیا  $\gamma^* = (\gamma_1, \dots, \gamma_{m+3})'$  مساوی با صفر هست یا خیر.

برای آزمون  $H_0: \gamma^* = 0$ ، می‌توان از آماره زیر استفاده کرد.

$$Q_W = (\hat{\gamma}^*)' (Cov(\hat{\gamma}^*))^{-1} \hat{\gamma}^*$$

که در آن  $Cov(\hat{\gamma}^*)$  ماتریسی  $(m+3) \times (m+3)$  است که در گوشه پایین سمت راست ماتریس  $\hat{\Gamma}$  قرار دارد. این آماره با فرض صفر تقریباً دارای توزیع مربع کای با  $m+3$  درجه آزادی است (۲۲، ۲۳). بنابراین فرض صفر به ازای مقادیر بزرگ  $Q_W$  رد می‌شود. لازم به ذکر است که برای تحلیل داده‌ها از نرم‌افزار R نسخه ۳.۰.۳ برای مدل کرنل از بسته نرم‌افزاری {stats} و برای مدل بی‌اسپلاین از بسته نرم‌افزاری bs,spline استفاده شد.

### یافته‌ها

در این مطالعه تعداد ۲۹۴ عود ثبت شده از ۱۶۰ بیمار حاضر به تفکیک فصل عود منجر به بستری بیماری در جدول شماره ۱ و نمودار شماره ۱ آمده است. همان طور که دیده می‌شود ۴۵ (۱۵/۳ درصد) مورد عود در فصل بهار، ۸۵ (۲۸/۹۱ درصد) مورد در فصل تابستان، ۷۲ (۲۴/۴۸ درصد) مورد در فصل پاییز، و ۹۲ (۳۱/۲۹ درصد) مورد عود منجر به بستری در فصل زمستان رخ داده است.

جدول شماره ۱ نیز فراوانی عودهای منجر به بستری را در فصول مختلف به تفکیک جنس نشان می‌دهد.

برای برازش دادن مدلی که ضریب متغیری خاص از آن وابسته به زمان باشد، از توابع بی-اسپلاین درجه سوم و همچنین توابع کرنل استفاده می‌شود؛ بنابراین اولاً باید تعداد گره‌های به‌کاررفته برای ساخت بی‌اسپلاین را مشخص کرد. برای این کار مدل را به ترتیب با استفاده از ۲ و ۳ و ۴ و ۵ و ۶ گره برازش داده، برای هر حالت ملاک AIC را محاسبه کرده و در آخر مدلی را که دارای حداقل مقدار AIC باشد؛ به عنوان مدل نهایی انتخاب می‌شود. در جدول زیر، تعداد گره‌ها و مقدارهای AIC متناظر با مدل ساخته شده بر مبنای آن‌ها نشان داده شده است.

با توجه به جدول شماره ۳ اگر از ۶ گره استفاده شود؛ مقدار AIC نسبت به حالتی که از سه گره استفاده شده، ۰/۰۹ کاهش پیدا می‌کند؛ یعنی با وجود این که سه پارامتر دیگر به مدل اضافه شده، مقدار AIC تنها ۰/۰۹ واحد کاهش پیدا کرده است. از آن جایی که افزایش پارامترها قابلیت تعمیم مدل را پایین می‌آورند و همچنین تفاوت زیادی بین AIC در دو حالت وجود ندارد، برای ساختن بی-اسپلاین از  $m=3$  گره استفاده می‌شود.

در این مرحله با استفاده از تابع کرنل برای برازش دادن مدلی که ضریب متغیری خاص از آن وابسته به زمان باشد، پس از انتخاب کرنل‌های مختلف با دو پهنای باند ۱۰ و ۵۰ یافته‌های

پسوریازیس در فصل‌های مختلف سال متفاوت است یا خیر از آماره‌ی زیر استفاده می‌شود:

$$Q_w = \tilde{y}' \text{cov}(\tilde{y})^{-1} \tilde{y} \sim \chi^2_{1-\alpha, m+3}$$

با جایگذاری خواهیم داشت:  $Q_w = 43.06, \chi^2_{0.95,6} = 14.07$

$p\text{-value} < 0.01$

بنابراین نتیجه گرفته می‌شود که اثر فصل سال روی رخداد عود پسوریازیس در طول زمان متغیر است.

برای آزمون صفر بودن بردار ضرایب مدل از آماره والد استفاده شد.

$$\chi^2_w = 253.2, \chi^2_{0.95,12} = 21.02, P < 0.001$$

بنابراین فرض صفر بودن ضرایب مدل رد می‌شود.

تقریباً مشابهی به دست آمد که در نهایت از کرنل با پهنای باند ۵۰ استفاده شد و یافته‌ها به صورت زیر به دست آمد:

با توجه به نتایج حاصل از جدول شماره ۴ که از روش کرنل برای دست‌یابی به برآوردی هموار از اثر فصل استفاده شد، ضرایب متغیرها تنها مقدار اندکی نسبت به حالتی که از روش بی اسپلین برای هموار کردن ضرایب‌های متغیر فصل عود استفاده شد، تغییر کرده‌اند. پس می‌توان گفت دقت این دو روش در دست‌یابی به برآوردی هموار از ضرایب وابسته-زمان تقریباً یکسان است.

به طور کلی با توجه به نتایج حاصل از جدول‌های شماره ۳ و ۴ می‌توان گفت در فصل زمستان بیش‌ترین خطر عود نسبت به فصل بهار وجود دارد (حدود ۷ برابر خطر عود بیماران در زمستان بیش‌تر است).

حال برای آزمون این‌که آیا واقعاً اثر فصل سال روی رخداد عود

جدول شماره ۱- فراوانی عودهای منجر به بستری را در فصول مختلف به تفکیک جنس

فصل	جنس		جمع کل	
	مرد	زن		
	فراوانی	درصد	فراوانی	درصد
بهار	۲۷	۱۸/۲۴	۴۵	۱۵/۳
تابستان	۵۵	۳۷/۱۶	۸۵	۲۸/۹۱
پاییز	۲۹	۱۹/۶	۷۲	۲۴/۴۸
زمستان	۳۷	۲۵/۱	۹۲	۳۱/۲۹
جمع کل	۱۴۸	۱۰۰/۰	۲۹۴	۱۰۰/۰

جدول شماره ۲- تعداد گره‌ها و مقادیر AIC متناظر با مدل نرخ با ضرایب وابسته به زمان با هموارسازی بی اسپلین

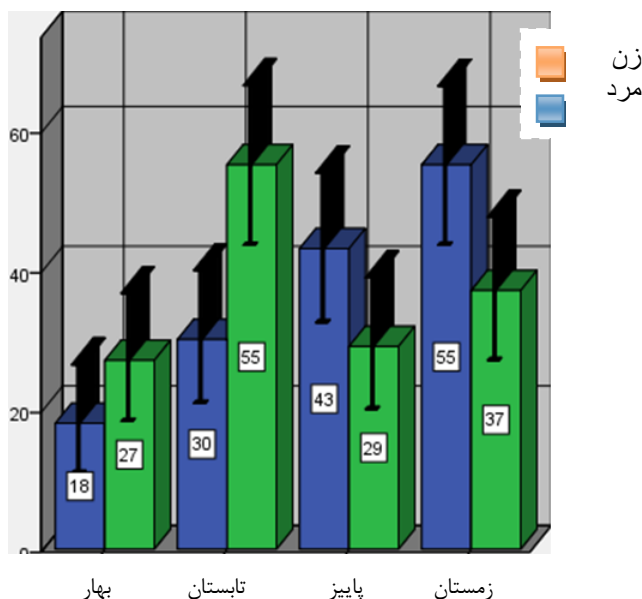
تعداد گره‌ها	۲	۳	۴	۵	۶
AIC	۶۹۲۱/۱۵	۶۹۱۳/۱۸	۶۹۲۸/۲۲	۶۹۳۱/۲۳	۶۹۱۳/۰۹

جدول شماره ۳- برآورد ضرایب متغیر «فصل عود بیماری» در مدل نرخ با ضرایب وابسته زمان با هموارسازی بی اسپلین

فصل عود بیماری	$\beta$	خطای معیار	$Exp(\beta)$	فاصله اطمینان ۰/۹۵ برای نسبت نرخ	P-value
تابستان	۱/۷۹	۰/۵۱	۵/۰۰۷	(۳/۷۵، ۱۰/۰۵)	۰/۰۰۴
پاییز	۱/۳۷	۶/۹۹	۴/۰۲	(۲/۲۲، ۷/۳۳)	۰/۰۴۱
زمستان	۲/۰۳	۱/۳۵	۷/۲۱	(۲/۹۱، ۱۵/۵۶)	۰/۰۰۳

جدول شماره ۴- برآورد ضرایب متغیر «فصل عود بیماری» در مدل نرخ با ضرایب وابسته زمان با هموارسازی کرنل

P-value	فاصله اطمینان ۰/۹۵ برای نسبت نرخ	$Exp(\beta)$	خطای معیار	$\beta$	فصل عود بیماری
۰/۰۰۴	(۳/۲۵،۹/۸۵)	۴/۹۸	۰/۶۲	۱/۷۲	تابستان
۰/۰۳۸	(۲/۰۲،۷/۵۶)	۴/۲۱	۶/۹۹	۱/۳۹	پاییز
۰/۰۰۳	(۲/۶۷،۱۶/۰۶)	۷/۲۵	۱/۳۵	۲/۱۱	زمستان



نمودار شماره ۱- توزیع عودهای بیماران در فصول مختلف بر حسب جنس

## بحث

این مطالعه به منظور بررسی تغییرپذیری اثر متغیر فصل عود بیماری و همچنین بررسی میزان عود بیماری در فصول مختلف سال طراحی و انجام شد. در نهایت نیز از روش‌های هموارسازی برای دست‌یابی به برآوردهایی بهتر استفاده شد.

لین و وی<sup>۱</sup> و همکاران از مدل نرخ متناسب برای بررسی عوامل مؤثر بر عود بیماران عفونی و همچنین از مانده‌های مارتینگل برای بررسی نیکویی برازش مدل استفاده کردند (۲۵). چین تسانگ<sup>۲</sup> و همکاران برای برآورد تابع نرخ از برآوردگر کرنل استفاده کردند و همچنین عملکرد این برآوردگر را از طریق شبیه‌سازی مونت‌کارلو بررسی کردند (۲۶). نلسون<sup>۳</sup> و همکاران از یک مدل نیمه پارامتری با استفاده از اسپلاین رگرسیونی برای توزیع نرخ

رخدادها در طول زمان استفاده کردند و شبیه‌سازی در مدل آن‌ها نشان داد که برآوردهای نرخ به دست آمده با اربیی کمی تقریب شده‌اند (۲۷). می‌چنگ‌وانگ<sup>۴</sup> و همکاران در مطالعه خود از روش‌های حداقل مربعات و گشتاوری با استفاده از برآوردگر کرنل برای برآورد تابع نرخ استفاده کردند و با شبیه‌سازی و به‌کارگیری روش بوت استرپ به این نتیجه رسیدند که روش گشتاوری از روش حداقل مربعات کاراتر است (۲۸). طبق مطالعه‌ای که توسط صالحی و همکاران با عنوان برآورد تابع چگالی احتمال با استفاده از برآوردهای هسته‌ای و بی-اسپلاین انجام شد؛ به این نتیجه رسیدند که میزان دقت برآوردها که از طریق میانگین انتگرال مربعات خطا مورد مقایسه عددی قرار گرفتند؛ تقریباً یکسان است (۲۹). آموریم و همکاران برای نخستین بار از این مدل (مدل نرخ با ضرایب وابسته به زمان)، برای بررسی تأثیر ویتامین A روی رخداد اسهال کودکان استفاده کرده و مصرف این ویتامین را به

<sup>۴</sup> Mei-Cheng Wang

<sup>۱</sup> Lin, Wei

<sup>۲</sup> Chin-Tsang Ching

<sup>۳</sup> Neilson



می‌شود مدل نرخ با ضرایب وابسته-زمان به کار گرفته شود. استفاده از روش‌های هموارسازی اسپلاین‌های رگرسیونی و کرنل در این مدل، این قابلیت را به مدل می‌دهد که برآوردهای همواری از اثر متغیر در نقاط مختلف زمانی ارائه دهد.

علاوه بر این در بسیاری از مطالعه‌های پزشکی با مواردی وجود دارد که اثر متغیر مورد نظر روی میزان بروز رخدادها در طول زمان متفاوت است. برای مثال در بسیاری از بیماری‌های عفونی بازگشتی (عود کننده)، از آن جایی که ممکن است در بدن مقاومتی نسبت به داروی مصرفی ایجاد شود، تمایل به برآورد اثر دارو در نقاط مختلف زمانی وجود دارد. در این‌گونه موارد می‌توان از مدل نرخ با ضرایب وابسته-زمان برای مدل‌بندی اثر متغیر در طول زمان استفاده کرد.

به دلیل ناشناخته بودن بحث ضرایب وابسته-زمان در حوزه‌ی پزشکی، امکان مقایسه و ارزیابی روش به کار رفته در این مطالعه با روش‌های دیگر وجود نداشت.

### تشکر و قدردانی

نویسندگان بر خود لازم می‌دانند مراتب قدردانی خود را از مرکز تحقیقات پوست، به دلیل فراهم آوردن دسترسی به اطلاعات بیماران مبتلا به پسوریازیس عمل آوردند. این مقاله برگرفته از پایان‌نامه کارشناسی ارشد خانم فرشته عثمانی است. بدین وسیله از اعضای هیأت علمی و کارکنان دانشکده پزشکی دانشگاه تربیت مدرس نیز قدردانی می‌شود.

عنوان عاملی مؤثر در کاهش اسهال معرفی کرده‌اند (۲۰). از آن جایی که دوره‌ی اثرگذاری این ویتامین کوتاه بوده، پس از مدتی به صفر می‌گراید، مدل‌های معمول نمی‌توانند آن را به عنوان عاملی معنی‌دار شناسایی کنند. بنابراین توصیه می‌شود در مواردی که ممکن است اثر متغیر مورد نظر به ازای تمام زمان‌ها ثابت نباشد، از مدل نرخ با ضرایب وابسته-زمان استفاده شود.

بسیاری از مواقع در تحلیل رخدادها بازگشتی به‌ویژه آن‌هایی که در حیطه پزشکی هستند، مشخص کردن ساختار همبستگی که بین رخدادها اتفاق افتاده برای هر فرد وجود دارد؛ بسیار پیچیده است. بنابراین استفاده از مدل‌هایی که بر مبنای تابع شدت هستند در این‌گونه موارد چندان مناسب به نظر نمی‌رسد، زیرا یا باید شکل ساختار همبستگی موجود دانسته شود و آن را وارد مدل کرد یا باید فرض شود که این ساختار همبستگی به وسیله متغیرهای کمکی وارد مدل خواهند شد، که در اغلب موارد این فرض قابل قبولی نیست. در مقابل مدل نرخ فرض خاصی را برای شکل ساختار همبستگی در نظر نمی‌گیرد و این امکان را فراهم می‌کند که داده‌ها هرگونه ساختار همبستگی‌ای داشته باشند؛ بنابراین توصیه می‌شود در مواردی که شکل ساختار همبستگی نامعلوم است، از این مدل برای فرمول‌بندی میزان بروز رخدادها استفاده شود (۳۲-۳۰).

### نتیجه‌گیری

در مواردی که حدس زده می‌شود اثر متغیر مورد نظر روی نرخ بروز رخدادها ممکن است در طول زمان تغییر کند، پیشنهاد

### منابع

1. Cook JR, Lawless JF. The Statistical Analysis of Recurrent Events. Gail M, Krickeberg K, Sarmet J, Tsiatis A, Wong W, editors: Springer 2007; 12: 78-86.
2. Clement DY, Strawderman RL. Conditional GEE for Recurrent Event Gap Times. *Biostatistics*. 2009; 10, 3: 17.
3. Odd OAB, Hkon KG. Survival and Event History Analysis. Gail M, Krickeberg K, Samet J, Tsiatis A, New York: Springer 2008; 24: 132-45.
4. Andersen PK, Borgan O, Gill RD, Keiding N. Statistical Models Based on Counting Processes, Springer-Verlag New York 1993; 34: 17-27.
5. Cook RJ, Lawless JF. The Analysis of Recurrent Event Data, Springer-Verlag New York 2007; 12: 24-32.
6. Liuquan Sun LZ, Jianguo Sun. Regression analysis of multivariate recurrent event data with time-varying covariate effects. *Multivariate Analysis*. 2000; 100: 2214-23.
7. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73: 13-22.
8. Clegg L, Cai J, Sen PK. A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics* 1999; 55: 805-12.
9. Chin-Tsang Chiang JAR, Colin O Wu. Smoothing Spline Estimation for Varying Coefficient Models With Repeatedly Measured Dependent Variables. *American Statistical Association*. 2001; 96: 605-19.
10. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *Journal of the American Statistical Association*. 1993; 88: 811-82.
11. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics*. 1995; 37: 58-61
12. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate function of recurrent events. *Journal of the Royal Statistical Society Series B* 2000; 62: 711-73.
13. Lin DY, Wei LJ, Ying Z. Semiparametric transformation models for point processes. *Journal of the American Statistical Association*. 2001; 96: 620-62.
14. Zucker DM, Karr AF. Nonparametric survival analysis with time-dependent covariate effects: a penalized partiallikelihood approach. *The Annals of Statistics* 1990; 18: 329-35.
15. Martinussen T, Scheike TH, Skovgaard IM. Efficient estimation

- of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics*. 2002; 29: 57-74.
16. Winnett A, Sasieni P. Iterated residuals and time-varying covariate effects in Cox regression. *Journal of the Royal Statistical Society, Series B*. 2003; 65: 473-88.
  17. Cai Z, Sun Y. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics*. 2003; 30: 93-111.
  18. Tian L, Zucker D, Wei LJ. On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association*. 2005; 100: 172-81.
  19. Amorim LD, Cai JW, Zeng DL, Barreto ML. Regression splines in the time-dependent coefficient rates models for recurrent event data. *Statistics in Medicine*. 2008; 27: 890-97.
  20. Griffiths CEM, Camp RDR. Psoriasis, In: Burns T, Breathnach S, In: *Rooks Textbook of Dermatology*. 17<sup>th</sup> Ed. Massachusetts: Blackwell; 2004: 35.1-35.69
  21. Habif TP. Psoriasis and other Papulosquamous Diseases. In: Habif. *Clinical Dermatology*. 4<sup>th</sup> Ed. New York; Mosby; 2004: 209-45.
  22. Bsplines. Michigan Technologic University; Available from: <http://www.cs.mtu.edu/~shene/COURSES/cs3621/NOTES/spline/B-spline/>.
  23. Aalen OO. Dynamic analysis of multivariate failure time data. *Biometrics*. 2004; 60: 764-73.
  24. Lin D Y, Wei LJ. Semiparametric regression for the mean and rate function of recurrent events. *J R Statist Soc B*. 2000; 62: 711-30.
  25. Chin-Tsang Chiang, Mei-Cheng Wang. Smoothing Estimation of Rate function for recurrent event data with Informative censoring. *Ann Inst Statist Math*. 2004; 1: 87-100.
  26. Neilson WB. Graphical analysis of system repair data, *Journal of Quality Technology*, 1988; 20: 24-35.
  27. Chiang C, Wang M. Kernel Estimation of Rate Function for Recurrent Event Data. *Journal of Statistics*. 2005; 32: 77-91.
  28. Noorkojuri H, Hajizadeh E, Baghestani A, Pourhoseingholi M. Application of Fractional Polynomial Model for Determining Prognostic Factors Associated with Survival of Patients with Gastric Cancer. *IrJE*. 2013; 9: 50-8.
  29. Saki Malehi A, Hajizadeh E, Ahmadi K, Mansouri P. Assessing Pemphigus Recurrence Rate using the Analysis of the Gap Times between Recurrent Events. *IrJE*. 2014; 10: 1-8.
  30. Osmani F, Hajizadeh E, Mansoori P. Analysis of Risk Factors for Psoriasis Recurrence Using Proportional Rates Model. *J Skin Stem Cell*. 2014; 1: e25808.
  31. Osmani F, Hajizadeh E, Mansoori P. Estimation of seasonal effect on the psoriasis recurrence using time dependent coefficient rates model for recurrent events, *JDC*. 2015; 6: 23-30.

# Use of Smoothing Methods for Estimating the Coefficients of Time Dependent Rate Models in Survival Analysis and Its Application in Psoriasis Disease

Osmani F<sup>1</sup>, Hajizadeh E<sup>2</sup>, Mansoori P<sup>3</sup>

1- PhD Student of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran

2- Professor of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran

3- Professor of Dermatology, Faculty of Medicine, Tehran University of Medical Sciences, Tehran, Iran

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

**Background and Objectives:** In studies in which each person may experience an event at different times, they are recurrent events. One of the most popular approaches in analyzing recurrent event is obtaining an estimate of the means/rate of events at different times. In this context, one of the things that could help to better understand the effect of this factor on the response is determining the variability due to quantitative variables in the rate of events over time. In this study, we applied kernel and B-spline methods to estimate coefficients in the time dependent-coefficients rate model and showed its application in data of psoriasis.

**Methods:** In this study, data of patients with psoriasis who had a relapse leading to hospitalization in the Dermatology Department of Imam Khomeini Hospital, between 2005 and June 2013 were used. To investigate the relapse rate during these years, time-dependent coefficients rate model was used and the variability of these effects was assessed using the Wald test. Both b-Spline and kernel methods were used for estimating time varying coefficients in the time-dependent rate model. Finally, the results of the methods were compared based on estimates obtained.

**Results:** The results of this study showed that according to Wald test, the effect of the variables such as the season on the occurrence of psoriasis was significantly different ( $P$ -value  $< 0.01$ ). Also, according to the estimated coefficients from both methods, there was a little difference between them.

**Conclusion:** When the effect of a variable on the occurrence of the events is different at different time, then time-dependent coefficients rate model may provide a better estimate of the effect of variable on response.

**Keywords:** Recurrent event, B-splines, Kernel, Time-dependent coefficient rates model, Psoriasis