



کاربرد مدل رگرسیون دوجمله ای منفی در تعیین عوامل موثر بر تعداد سلول های $CD4$ در افراد آلوده به HIV مراجعه کننده به مرکز بهداشت استان خراسان رضوی

حبیب الله اسماعیلی^۱، زهرا منتظرآبادی^۲

^۱دانشیار آمار حیاتی گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی مشهد

^۲دانشجوی کارشناسی ارشد آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی مشهد

چکیده:

مقدمه: مطالعات طولی در بسیاری از شاخه های علوم، خصوصاً در علوم پزشکی بسیار رایج است. این مدل ها هم چنین برای بررسی متغیرهای پاسخ مختلف از جمله شمارش که در مطالعات پزشکی کاربرد بسیاری دارند، بسط داده شده اند. در این مقاله به بررسی عوامل موثر بر تعداد سلول های $CD4$ در افراد آلوده به HIV با استفاده از مدل رگرسیون دوجمله ای منفی پرداخته شده است.

مواد و روش ها: جامعه مورد بررسی شامل ۱۰۸ فرد آلوده به ویروس HIV که در مرکز بهداشت استان خراسان از سال ۱۳۷۱ دارای پرونده بوده اند می باشد. پس از جمع آوری اطلاعات عوامل موثر بر تعداد سلول های $CD4$ با استفاده از نرم افزار SAS و مدل رگرسیون دوجمله ای منفی با اثرات تصادفی مورد تجزیه و تحلیل قرار گرفت.

یافته ها: تعداد ۸۲ نفر (۷۵.۹%) مورد مطالعه را مردان و ۲۶ نفر (۲۴.۱%) را زنان تشکیل دادند. میانگین \pm انحراف معیار مقدار $CD4$ اولیه در مردان 359.7975 ± 256.248 و در میان زنان 292.5 ± 186 می باشد که از نظر آماری اختلاف معناداری داشتند ($p=0.0469$). متغیرهای تاثیرگذار در این مدل متغیرهای جنسیت ($p=0.0469$) درمان ($p=0.069$) و زمان ($p=0.001$) می باشد. واریانس جزء تصادفی مدل نیز معنادار شد که نشان دهنده موثر بودن ویژگی های فردی در ابتدا به ویروس HIV می باشد.

نتیجه گیری: در تحلیل این گونه داده ها، استفاده از مدل های اثرات تصادفی در تحلیل داده های طولی شمارشی پیشنهاد می گردد.

واژه های کلیدی: مطالعات طولی، مدل رگرسیون شمارشی، عوامل موثر، $CD4$

^۲زهرا منتظرآبادی * z.abedi1392@yahoo.com

۱ مقدمه

ایدز یا سندروم نقص ایمنی اکتسابی^۱ نوعی بیماری است که در دستگاه ایمنی توسط ویروس نقص ایمنی انسانی ایجاد می شود. (۱) بیماری ناشی از آلودگی یا ویروس HIV خازای سه مرحله اصلی می باشد. در مرحله اول فرد ممکن است برای مدت کوتاهی بیماری شبه آنفلوآنزایی را تجربه کند. به همین دلیل معمولاً این بیماری تا یک دوره طولانی بدون هیچ علائمی دنبال می شود که به این مرحله از بیماری، دوره نهفتگی گفته می شود. هر قدر که بیماری پیشرفت یابد، تداخل بیشتری با دستگاه ایمنی بدن پیدا می کند و باعث می شود که افراد به علوت هایی مانند علوت فرصت طلب و تومور چهار شوند. در نهایت بیماری زمانی وارد مرحله سوم یا ایدز خواهد شد که شمار تعداد سلول های CD۴ به کمتر از ۲۰۰ سلول در هر میکرولیتر برسد. پژوهش های ژنتیکی نشان می دهد که HIV در اصل در اوایل قرن بیستم میلادی در غرب آفریقا جهش و پدید آمده است. (۲) ایدز اولین بار در سال ۱۹۸۱ توسط مرکز کنترل و پیشگیری بیماری^۲ شناخته شد. (۳) ایدز به عنوان همه گیری جهانی شناخته می شود که در حال حاضر حوزه شیوع آن بسیار وسیع و در حال گسترش است. هر چند شیوع آلودگی به HIV در ایران کمتر از کشورهای غربی و حتی بعضی کشورهای منطقه است. اما براساس برآوردهای سازمان بهداشت جهانی در سال ۱۹۹۷ بیش از ۲۱۹۹۹ نفر آلوده به ویروس HIV در ایران وجود داشته که در مقایسه با سال ۱۹۹۲ تقریباً دو برابر شده است. اهمیت نشانگرهای بیماری نیز در تشخیص میزان پیشرفت ایدز و انجام شیوه درمانی مناسب به خوبی شناخته شده است.

پژوهش های بسیاری در زمینه عوامل موثر بر کاهش تعداد سلول های CD۴ انجام شده است.

در مطالعه *Ebonye* و همکاران به منظور ارزیابی عوامل موثر بر کاهش تعداد سلول های CD۴ در افراد آلوده به HIV ۲۱۸ فرد مبتلا را از میان سن های ۲۰ سال به بالا به طور مقطعی مورد ارزیابی قرار داده است و از مدل رگرسیون لجستیک چند متغیره برای تعیین متغیرهایی که مستقلاً بر تعداد سلول های CD۴ اثر می گذارند استفاده کرده است

در مطالعه *Chatt* و همکاران برای تعیین عوامل موثر بر تعداد سلول های CD۴ در مبتلایان به HIV در سنگال ۸۵۰ فرد مبتلا وارد مطالعه طولی شدند. جهت تعیین آماره های تک متغیره از تحلیل واریانس استفاده شده است. تغییرات در تعداد سلول های CD۴ در ۳ مدل چند متغیره با استفاده از رگرسیون خطی انجام شده است. مطالعه حاضر که براساس یک مطالعه طولی انجام شد و هر بیمار در طول مطالعه چندین بار مورد اندازه گیری قرار گرفت، به تعیین عوامل موثر بر تعداد سلول های CD۴ در مبتلایان به HIV پرداخته است. امروزه بخش قابل توجهی از طرح های مطالعاتی در علوم مختلف، از جمله آزمایش های کلیتیکی و اطلاعات وابسته به موجودات زنده، ایجاد می کنند که واحدهای آزمایشی در زمان ها و یا موقعیت های مشخصی، به منظور کنترل تفاوت های بین آزمودنی ها، اندازه گیری شوند. در این نوع آزمایش ها که هر واحد آزمایشی به طور متناوب اندازه گیری می شود اغلب از اصطلاح اندازه های مکرر و در برخی حالات خاص از واژه های طولی استفاده می گردد ویژگی های منحصر به فرد هر یک از بیماران در طول دوره ی پی گیری، اغلب سبب اندازه گیری های مکرر مربوط به هر واحد آزمایشی خواهد شد. به این ترتیب بر خلاف سایر آزمایش ها مانند مطالعات مقطعی، به جای مشاهدات مستقل یا مجموعه ای از مشاهدات وابسته سرورکار خواهیم داشت. در روش های متداول تحلیل داده های طولی دو مطالعات پالینی، در حالت خاصی که تنها دو تکرار برای هر واحد آزمایشی وجود دارد می توان از روش های شناخته شده ای مانند *Paired - Ttest* برای پاسخ های پیوسته و آزمون *McNemars* برای پاسخ های دو مقداری که از توان کمی برای تحلیل ها برخوردارند استفاده کرد. استفاده از روش های تحلیل واریانس با اندازه های مکرر *Repeated*

^۱ Acquired immunodeficiency syndrome

^۲ Centers for Disease Control and Prevention

measures در حالت بیش از دو تکرار، راهبرد دیگری است که برای تحلیل اندازه‌های مکرر مستقل به کار می‌رود (۹). برای بررسی عوامل موثر بر تعداد سلول‌های CD4 در بیماران مبتلا به HIV داده‌ها به صورت طولی اندازه‌گیری شده و هر بیمار در طول مطالعه چندین بار مورد اندازه‌گیری قرار می‌گیرد. ویژگی مهم این نوع داده‌ها مستقل نبودن و همبستگی مشاهدات یک فرد در طول زمان است که در این صورت استفاده از مدل‌های آماری معمولی امکان‌پذیر نیست. رگرسیون طولی یا اثرات تصادفی یکی از روش‌های آماری برای تحلیل داده‌های طولی است که علاوه بر در نظر گرفتن همبستگی بین مشاهدات، ویژگی‌های فرمی را نیز وارد مدل نموده و برآوردهای دقیق‌تری از اثر عوامل موثر بر تعداد سلول‌های CD4 به دست می‌دهد. مزیت استفاده از این روش این است که اولاً تغییرات افراد در طول زمان را مشخص می‌کند، ثانیاً از روش اندازه‌گیری‌های مکرر انصاف‌پذیرتر است زیرا نیاز نیست برای هر فرد فقط تعداد یکسانی از تکرارها را داشته باشیم و برای همه گونه داده‌های چه کسی، کیلی و... لابل کاربرد است. هم چنین برای داده‌های غیر نرمال نیز قابل‌تعمیم است. هدف از این مطالعه تعیین عوامل موثر بر تعداد سلول‌های CD4 در بیماران مبتلا به HIV مراجعه‌کننده به مرکز بهداشت استان خراسان با استفاده از مدل رگرسیون طولی با اثرات تصادفی است.

۲ مواد و روش‌ها

مدل آماری در مطالعات طولی:

یک مطالعه طولی در واقع شامل اندازه‌گیری‌هایی است که از هر فرد تحت مطالعه در طول زمان به طور مکرر انجام می‌گیرد. بنابراین مطالعات طولی ما را قادر می‌سازد تا تغییرات درون فردی در طول زمان را مورد مطالعه قرار دهیم. هدف اصلی یک مطالعه طولی مشخص نمودن تغییر در متغیر پاسخ در طول زمان و بررسی عواملی است که بر این تغییر تاثیر می‌گذرانند (۵).

با توجه به اینکه داده‌های مطالعه طولی هستند و شرط استقلال داده‌ها برقرار نمی‌باشد، بنابراین از مدل اثرات تصادفی جهت بررسی اثر عوامل خطر ساز روی متغیر تعداد سلول‌های CD4 استفاده شد. در مدل اثرات تصادفی، فرض همبستگی در میان پاسخ‌های تکرار شده به وسیله ضرایب رگرسیونی متفاوت در آزمودنی‌های مختلف نشان داده می‌شود. در این مدل، پاسخ به عنوان تابعی از متغیرهای کمکی یا ضرایب رگرسیونی ای فرض شده که از یک آزمودنی به آزمودنی دیگر متفاوت است. این تفاوتها به خاطر عوامل اندازه‌گیری نشده ای است که منتج از عوامل طبیعی و ژنتیکی است. اثر تصادفی می‌تواند به دو صورت در مدل ظاهر شود. نوع ساده این مدل عبارت است از:

$$y_{it} = \alpha_i + \beta_i x_{it} + \epsilon_{it}$$

از نظر آماری تعداد سلول‌های CD4 برای هر فرد یک پاسخ شمارشی است و لذا برای تحلیل این متغیر لازم است از مدل‌های رگرسیون شمارشی استفاده کنیم. مدل‌های رگرسیون شمارشی از نوع مدل‌های غیرخطی هستند و لذا برخی از محدودیت‌های رگرسیون خطی را ندارند. یک مدل رایج برای داده‌های شمارشی مدل پواسن است که محدود به این است که میانگین و واریانس با هم برابر باشند اما اغلب داده‌های شمارشی مشاهده شده واریانس نمونه بطور قابل ملاحظه‌ای از میانگین بزرگتر (کوچکتر) است. وقتیکه واریانس از میانگین بزرگتر باشد، مسأله بیش پراکنش^۳ به وجود می‌آید، عدم توانایی این مدل در کنترل کردن بیش پراکنش باعث بروز مشکلاتی از قبیل برآوردهای ناسازگار، کم برآوردی خطای معیار ضرایب پارامترهای مدل، زیاد شدن آماره‌های آزمون هرکدام از ضرایب متغیرها و باعث

^۳overdispersion

افزایش خطای پذیرش آن متغیر در مدل و باریک شدن فاصله اطمینان و کوچک شدن پی-مقدار می شود. در صورت وجود بیش پراکنش در مدل رگرسیون بواسن، یک راه اصلاح بیش پراکنش براساس برآوردهای توپوند واریانس است و راه دیگر استفاده از دیگر مدل های رگرسیون شمارشی همچون دوجمله ای منفی و یا بواسن تعمیم یافته است. تا سال های اخیر مدل دوجمله ای منفی به منظور توصیف این توزیع که بیش پراکنی فقط به علت ناهمگنی مشاهده نشده است به کار برده شده است (6).

جهت انجام تحلیل های لازم از نرم افزار SAS نسخه 9.4 استفاده شده است.

کاربرد مدل:

مدل فوق را در یک مطالعه با 108 فرد آلوده به ویروس HIV که در مرکز بهداشت استان خراسان از سال 1371 دارای پرونده بوده اند را به کار بریم. معیارهای ورود بیماران به این مطالعه عبارت بودند از: افرادی که طی سالهای 1371 تاکنون به عنوان المراد آلوده به HIV در مرکز بهداشت استان خراسان رضوی دارای پرونده هستند و تمام موارد لازم در پرونده ثبت شده باشد. پس از مراجعه به مرکز بهداشت استان، داده های موجود در پرونده بیماران واجد شرایط در فرم ثبت اطلاع تهیه شده ثبت گردید و سپس برای انجام تحلیل های آماری وارد بانک اطلاعاتی شد.

متغیرهای پژوهش: برای بیمارانی که معیارهای ورود به مطالعه را دارا بودند، اطلاعات زیر از پرونده آنها استخراج گردید: جنسیت، وضعیت تاهل، عفونتهای همراه، سابقه زندان، درمان ضد رتروویروسی، زمان تا مرگ بیمار که تمامی متغیرهای ذکر شده در طول دوره پیگیری اندازه گیری و ثبت شدند.

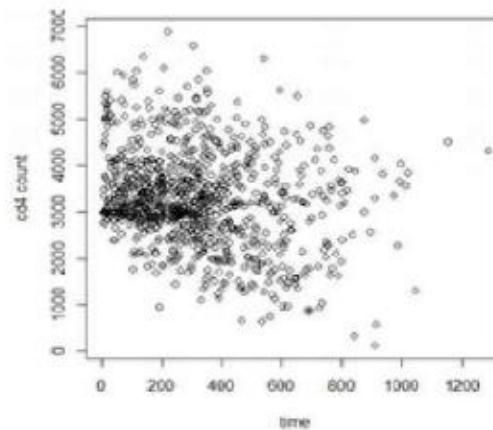
۳ یافته ها

تعداد 82 نفر (92.9%) مورد مطالعه را مردان و 26 نفر (29.1%) را زنان تشکیل دادند. میانگین \pm انحراف معیار مقدار $CD4$ اولیه در مردان 359.29 ± 256.22 و در میان زنان 292.50 ± 186 می باشد که از نظر آماری اختلاف معناداری داشتند ($p=0.049$). اطلاعات مربوط به توزیع فراوانی جنسیت بیماران و سابقه زندان و وضعیت تاهل و ابتلا به بیماری زمینه ای در شکل 1 داده شده است.

متغیر	سطح متغیر	تعداد	درصد
جنسیت	مذکر	82	75.9
	مؤنث	26	24.1
سابقه زندان	دارد	72	66.7
	ندارد	36	33.3
HCV Ab	دارد	62	63.9
	ندارد	46	36.1
HBS Ag	دارد	8	8/3
	ندارد	100	91/7
HTLV-1 Ag	دارد	10	12/2
	ندارد	98	87/8

شکل 1- توزیع فراوانی افراد تحت مطالعه برحسب متغیرهای مورد بررسی.

رود تغییرات تعداد سلول های $CD4$ در طول دوره پیگیری با شکل ۲ نشان داده شده است. تعداد ۹۴ نفر (۸۲٪) بیماران برای نشان درمان شروع شده است و ۱۴ نفر (۱۳٪) درمان شروع نشده است. میانگین $\pm 326CD4 \pm 359$ و در کسانی که درمان شروع نشده، $177/522 \pm 224/154$ می باشد که بصورت متغیر تاثیر گذار در مدل وارد شدند.



شکل ۲: پراکنش تعداد سلول های $CD4$ طی دوره درمان در افراد تحت مطالعه.

متغیرهای تاثیرگذار در این مدل متغیرهای جنسیت (۰.۰۴۶۳)، درمان (۰.۰۶۹۳) و زمان (۰.۰۰۰۱) می باشد. برای تعیین اثر عوامل فردی مربوط به هر فرد در بروز بیماری ایندز، یک جزء تصادفی که نشان دهنده اثر همه عوامل فردی است به مدل افزوده شد و یک مدل رگرسیون دوجمله ای متنی با اثرات تصادفی به داده ها برآزش داده شد. نتایج حاصل از برآزش این مدل و برآورد هر یک از عوامل در شکل ۳ ارائه شده است.

ضریب مثبت متغیر درمان در این روش به این معنی است که با تغییر هر سطوح متغیر درمان، شانس بیشتر شدن تعداد سلول های

پارامتر	ضریب متغیر	خطای معیار	فاصله اطمینان
جنسیت	0/4864	0/2444	(-0/0024, 0/975)
درمان	0/154	0/056	(0/042, 0/265)
زمان	-0/036	0/006	(-0/048, -0/024)
اثر تصادفی	0/2761	0/04982	(0/176, 0/375)
بیش پراکنی	0/278	0/01428	(0/249, 0/306)

شکل ۳: پراکنش تعداد سلول های $CD4$ طی دوره درمان در افراد تحت مطالعه.

$CD4$ افزایش می یابد و تقریباً ۰-۲ برابر است. هم چنین خانم ها $CD4$ بالاتری نسبت به آقایان دارند. یعنی متوسط مورد انتظار $CD4$ در خانم ها ۱.۶۲ برابر آقایان است و ضریب متنی برای برآورد زمان نشاندهنده کاهش تعداد سلول های $CD4$ به ازای هر سال در طول زمان می باشد.

۴ بحث و نتیجه گیری

با توجه به ماهیت داده ها در مطالعات طولی یا پاسخ شمارشی، در تحلیل اطلاعات باید از روش های پیچیده تری استفاده کرد. یکی از رایج ترین شیوه های تحلیل داده های طولی، استفاده از مدل رگرسیون طولی با اثر تصادفی است. که در این مقاله با استفاده از این روش به تحلیل پاسخ های شمارشی پرداختیم. ما در این مقاله از مدل رگرسیون موجهه ای منفی با اثر تصادفی استفاده کردیم. با توجه به این که افراد به طور تصادفی وارد مدل شده اند و دارای ویژگی هایی هستند که از دسترس ما خارج بوده و کنترل آن ها کار ساده ای نیست و از سوی وابستگی اطلاعات مربوط به هر فرد و استقلال هر بیمار از بیمار دیگر، راهی بهتر تحلیل مدل با اثر تصادفی را پائی نمی گذارند. با استفاده از نرم افزار SAS و دویه GLIMMIX به تحلیل این داده ها پرداختیم. با توجه به یافته های این مطالعه می توان نتیجه گرفت که حساسیت درمان و زمان در شمار تعداد سلول های CD۴ موثر هستند و البته بالارفتن سطح آگاهی افراد از چگونگی ابتلا به ویروس HIV در حالت کلی نیز می تواند در کاهش بروز این بیماری موثر باشد.

مراجع

- [1] Sepkowitz KA. AIDS—the first 20 years. *New England Journal of Medicine*. 2001;344(23):1764-72.
- [2] Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harbor perspectives in medicine*. 2011;1(1):a006841.
- [3] Gallo RC. A reflection on HIV/AIDS research after 25 years. *Retrovirology*. 2006;3(1):72.
- [4] Timm N.H. *Applied Multivariate Analysis*. Springer-Verlag, New York. Laird, N. M. and Ware, J. H., *Random Effects Models for Longitudinal Data*. *Biometrics* 1982; 38: 963-74.
- [5] Diggle P, Liang K-Y, Zeger SL. *Longitudinal data analysis*. New York: Oxford University Press; 1994.
- [6] Akbarzadeh Baghiban A, Pourhoseingholi A, Zayeri F, Jafari AA, Alavian SM. Application of Zero-Inflated Poisson Mixed Models in Prognostic Factors of Hepatitis C. *BioMed research international*. 2013;2013.