مجله تخصصی اپیدمیولوژی ایران؛ زمستان ۱۳۹۶، دوره ۱۳، شماره ٤: صفحات ۳٦۱-۳٥٦.

مقاله مروری

برآورد نسبت خطر علیتی حاشیه‌ای در داده‌های بقا با استفاده از روش وزن‌دهی عکس احتمال درمان

ندا محمدی[۱]، محمدعلی منصورنیا[۲]، سید علیرضا سجادی[۳]، معصومه علی محمدیان[٤]، حسین پوستچی[٥]، مهدی یاسری[۲]

[۱]دانشجوی کارشناسی ارشد آمار زیستی، گروه اپیدمیولوژی و آمارزیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

[۲]استادیار، گروه اپیدمیولوژی و آمارزیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

[۳] پزشک، پژوهشکده بیماری های گوارش و کبد، دانشگاه علوم پزشکی تهران، تهران، ایران

[۴] مربی، پژوهشکده بیماری های گوارش و کبد، دانشگاه علوم پزشکی تهران، تهران، ایران

[٥]دانشیار، پژوهشکده بیماری های گوارش و کبد، دانشگاه علوم پزشکی تهران، تهران، ایران

نویسنده رابط: سید علیرضا سجادی، نشانی: مرکز تحقیقات سرطان پژوهشکده بیماری های گوارش و کبد، دانشگاه علوم پزشکی تهران، تهران، ایران

تلفن: ۰۲۱۸۲۴۱۵۱۰۴، پست الکترونیک: n.mohammadi.24@gmail.com

تاریخ دریافت: ۹۵/۰۳/۲۲؛ پذیرش: ۹۵/۰۹/۰۶

از جمله روش‌های سنتی مورد استفاده در آنالیز داده‌های بقا، روش رگرسیون کاکس است که این روش شاخص نسبت مخاطره شرطی (Conditional Hazard ratio) را محاسبه می‌کند. زمانی که هدف مطالعه، برآورد اثر مواجهه روی پیامد در سطح کلی جامعه (Total population) باشد، استفاده از این روش‌های شرطی مناسب نیستند. ضمناً شاخص نسبت مخاطره دارای معایبی از جمله non-collapsible بودن، تورش انتخاب ساختاری(Structural selection bias) و متغیر بودن در زمان است. با توجه به این محدودیت‌ها استفاده از نسبت خطر حاشیه‌ای که متوسط اثر علیتی مواجهه را در سطح جامعه محاسبه می‌کند، توصیه می‌شود.

در این تحقیق به معرفی روش «وزن‌دهی عکس احتمال درمان» به عنوان یکی از روش‌های برآورد اثر حاشیه‌ای علیتی پرداخته می‌شود و با استفاده از آن نسبت خطر علیتی حاشیه‌ای برآورد می‌گردد. در انتها نیز با استفاده از این روش در داده‌های مطالعه‌ی کوهورت گلستان به برآورد اثر علیتی حاشیه‌ای بین مصرف سیگار و زمان تا مرگ افراد در اثر سرطان دستگاه گوارش فوقانی (مری و معده) پرداخته می‌شود.

واژگان کلیدی: اثر حاشیه‌ای، اثر علیتی، وزن‌دهی عکس احتمال درمان

مقدمه

مطالعه‌های مشاهده‌ای شامل مقایسه گروه‌هایی که سطح‌های مواجهه یا درمان متفاوت را دریافت می‌کنند، است. در این نوع مطالعه‌ها، پژوهشگر در مجموعه‌ی تحت بررسی برای تخصیص درمان به افراد، هیچ مداخله‌ای نکرده و صرفاً به مشاهده می‌پردازد. در این مطالعه‌ها به نسبت مطالعه‌های مداخله‌ای تصادفی شده، شانس تأثیرگذاری مخدوشگرها افزایش می‌یابد (۱, ۲)، بنابراین نیاز است تا با روش‌هایی تأثیر مخدوشگرها کنترل شود. در داده‌های بقا که پیامد مورد بررسی، زمان تا رخداد پیامد مورد نظر است (۳)، می‌توان ضمن تعدیل اثر مخدوشگرها، به بررسی اثر مواجهه بر بقای افراد پرداخت. از جمله روش‌های سنتی مورد استفاده در داده‌های بقا، روش کاکس است، اما این روش دارای معایبی است که عبارت‌اند از:

۱- روش رگرسیونی کاکس، اثر مواجهه را به صورت شرطی برآورد می‌کند، یعنی اثر مواجهه روی پیامد، در لایه‌های مختلف مخدوش کننده‌ها گزارش می‌شود. مثلاً اگر سن و جنس به عنوان مخدوشگر رابطه‌ی مواجهه بر پیامد تعدیل شود، در این‌صورت نسبت مخاطره در گروه‌های سنی و جنسی مشخص مقدارهای متفاوتی دارند و جداگانه برآورد می‌شوند. زمانی که هدف مطالعه برآورد اثر به صورت حاشیه‌ای و در سطح جامعه باشد، این روش مناسب نیست (۴).

۲- شاخص قابل برآورد در روش رگرسیونی کاکس، شاخص نسبت مخاطره[۱] است. این شاخص غیرقابل آمیختن[۲] است، یعنی نسبت مخاطره در لایه‌های مختلف عامل خطر بیماری زمانی که مخدوش‌کننده وجود ندارد، با نسبت مخاطره‌ی حاشیه‌ای برابر نیست (۵, ۶)

۳- مخاطره در طول زمان تغییر می‌کند، اما در مدل کاکس اثر متقابل مواجهه با زمان نادیده گرفته می‌شود. که در واقع این نسبت به‌دست آمده ازمدل یک میانگین وزنی از نسبت مخاطره در

[۱] Hazard ratio
[۲] Non-collapsible

[Downloaded from journals.tums.ac.ir on 2024-05-16]

زمان‌های مختلف مطالعه است. بنابراین تفسیر آن در همه زمان‌ها درست نیست (۲, ۷).

۴- هم‌چنین نسبت مخاطره یک زمان مشخص، دارای اریبی انتخاب ساختاری است، زیرا نسبت مخاطره در هر زمان در افرادی محاسبه می‌شود که هنوز پیامد در آن‌ها رخ نداده است، در نتیجه متغیرهای شکنندگی اندازه‌گیری نشده در افراد باقی‌مانده، می‌توانند باعث تورش انتخاب شوند (۷, ۸).

با توجه به مطالب بیان شده، زمانی که هدف مطالعه برآورد اثر حاشیه‌ای مواجهه روی پیامد در سطح کلی جامعه است، از روش‌های حاشیه‌ای که معایب روش‌های شرطی را ندارند، استفاده می‌شود.

از جمله روش‌های برآورد اثر حاشیه‌ای علیتی، روش «وزن‌دهی عکس احتمال» (۹, ۱۰) و روش استانداردسازی مبتنی بر مدل[۱] (۱۱) است.

هدف این مقاله معرفی روش «وزن‌دهی عکس احتمال» و ارایه یک مثال با استفاده از داده‌های کوهورت گلستان است.

در آنالیز داده‌های بقا با استفاده از این دو روش از داده‌های شخص- زمان[۲] استفاده می‌شود. منظور از داده‌های با فرمت شخص- زمان این است که داده‌ها روی زمان بسط داده می‌شود و به بازه‌ی کوچکی مثل ماه یا روز (با توجه به پیامد مورد بررسی) شکسته می‌شود. مثلاً اگر بازه‌های شکستن زمان به صورت ماه انتخاب شود، سطر اول داده‌ها در این فرمت مربوط به فرد اول در ماه نخست پیگیری یعنی k=1، سطر دوم مربوط به فرد اول در ماه دوم بررسی k=2، و به همین ترتیب تا آخرین زمان بررسی فرد اول و پس از آن برای فرد دوم تا nام مورد بررسی به تعداد ماه‌های پیگیری شده، سطرهای جداگانه وجود دارد.

برای هر فرد در ماه kام بررسی خواهیم داشت:

$$\begin{cases} D_k = 1 & \text{if } T \le k \\ D_k = 0 & \text{if } T > k \end{cases}$$

مخاطره در زمان k را به صورت $P(D_k = 1 | D_{k-1} = 0)$ تعریف کرده، در نتیجه احتمال بقا در زمان k با استفاده از فرم شرطی زیر که در واقع، حاصل‌ضرب یک منهای مخاطره در تمام زمان‌های قبلی آن است، قابل محاسبه است:

$$P(D_k = 0) = \prod_{m=1}^{k} P(D_m = 0 | D_{m-1} = 0)$$

مقدار مخاطره در زمان k یعنی $P(D_k = 1 | D_{k-1} = 0)$ را می‌توان به صورت ناپارامتری یعنی تقسیم تعداد افراد در فاصله k به تعداد افراد زنده در پایان فاصله زمان k-1 محاسبه کرد که درصورت محاسبه یک منهای مخاطره و جای‌گذاری در فرمول بالا، به $P(D_k = 0)$ برآوردگر کاپلان مایر گفته می‌شود، اما از آن‌جا که تعداد بازه‌های زمانی زیاد و تعداد موارد در هر فاصله زمانی کم است، مشکل داده پراکنده و ناپایدار[۳] بودن این برآورد به‌وجود می‌آید. به همین دلیل برای برآورد پارامتری و کاراتر برای مخاطره، از فرمول زیر استفاده می‌شود:

(۱)

$$\text{logit } P(D_{k+1} = 1 | D_k = 0, Treatment) = \theta_{0,k} + \theta_1 k + \theta_2 Treatment * k$$

این مدل لجستیک، تقریبی از یک مدل مخاطره است و k نشان دهنده زمان است(۲). از آن‌جا که $\theta_{0,k}$ عرض از مبدأ وابسته به زمان است، مخاطره به صورت وابسته به زمان برآورد می‌شود و عبارت اثر متقابل بین درمان A و زمان k نیز به تغییر نسبت مخاطره[۴] در طول زمان اجازه می‌دهد.

سپس با ضرب عبارت‌های $P(D_{k+1} = 0 | D_k = 0, Treatment)$ که از مدل لجستیک محاسبه می‌شود، می‌توان بقا یعنی $P(D_{k+1} = 0 | Treatment = a)$ و درنهایت مکمل آن یعنی خطر را به‌طور جداگانه برای افراد درمان را گرفته و افراد بدون درمان را برآورد نمود و منحنی بقا را رسم کرد. در این مقاله برای تعدیل اثر مخدوشگرها از روش وزن‌دهی عکس احتمال استفاده می‌شود.

روش وزن‌دهی عکس احتمال

در روش وزن‌دهی عکس احتمال، وزن‌هایی پایا[۵] برای هر فرد ایجاد می‌شود. با استفاده از این وزن‌ها، جامعه کاذبی ساخته می‌شود، که در آن، متغیرهای مخدوشگر مدل از متغیر مواجهه مورد بررسی، مستقل می‌شود و در نهایت اثر مخدوشگرها از مدل حذف شده و اثر حاشیه‌ای مواجهه برآورد می‌شود (۲).

برای محاسبه صورت کسر وزن‌های پایا، با استفاده از مدل لجستیک، مواجهه را بدون حضور مخدوشگرها، برای دو گروه مواجهه یافته و بدون مواجهه محاسبه می‌شود و برای محاسبه مخرج کسر این وزن‌ها، با استفاده از مدل لجستیک، مواجهه در برابر مخدوشگرها که با نماد L نشان داده، مدل می‌شود. بنابراین برای محاسبه وزن‌ها خواهیم داشت:

$$SW = \begin{cases} \frac{\hat{P}[A=1]}{\hat{P}[A=1|L]} & \text{برای افراد مواجهه داشته} \\ \frac{1-\hat{P}[A=1]}{1-\hat{P}[A=1|L]} & \text{برای افراد بدون مواجهه} \end{cases} \quad (۲)$$

---

۱ Hazard ratio
۲ Stabilized weight(SW)
۳ Person-time
۴ Unstable
۲ Model-base standardization

[Downloaded from journals.tums.ac.ir on 2024-05-16]

با توجه به صورت و مخرج محاسبه شده، وزن‌های مدل ساخته می‌شود. شرط لازم برای وزن‌های پایای ساخته شده این است که میانگین وزن‌ها حدوداً برابر با عدد یک باشد.

در گام دوم از داده‌های با فرمت شخص- زمان که در ابتدای مقدمه گفته شد، استفاده می‌شود و در نهایت با استفاده از مدل لجستیک فرمول شماره (۱)، یک مدل مخاطره برای پیامد، در حالی که افراد مورد بررسی توسط وزن‌هایی که ساخته می‌شود، وزن داده شده‌اند، برازش داده می‌شود.

در این مدل، مخاطره وابسته به زمان، برای دو گروه درمان گرفته و بدون درمان محاسبه می‌شود و برآورد $Logit\ P[D^a_{k+1}=0 \mid D^a_k=0]$ را می‌توان در زمان‌های متفاوت در هم ضرب کرد تا به برآوردی از خطر یعنی $P[D^a_{k+1}=0]$ دست یافته شود.

در نهایت در این داده‌ها با استفاده از احتمالات محاسبه شده از مدل، بقا و ریسک و در نهایت نسبت ریسک که هدف این پژوهش بوده، قابل محاسبه است.

ارایه یک مثال

برای نشان دادن کاربرد روش وزن‌دهی عکس احتمال، با استفاده از داده‌های واقعی به بررسی اثر حاشیه‌ای مصرف سیگار بر زمان تا مرگ در اثر سرطان دستگاه گوارش فوقانی و برآورد نسبت خطر[۱] گروه مصرف کننده سیگار به عدم مصرف سیگار با استفاده از نرم‌افزار SAS پرداخته می‌شود. کدهای این روش در کتاب استنتاج علیتی موجود است (۲).

اطلاعات این مطالعه متعلق به مطالعه کوهورت استان گلستان است که در حدود ۵۰۰۰۰ نفر از افراد، طی سال‌های ۸۶-۱۳۸۲ وارد مطالعه شده‌اند و تا سال ۱۳۹۵ به طور سالانه مورد پیگیری قرار گرفتند. روش جمع‌آوری اطلاعات این مطالعه به‌طور گسترده در پروفایل کوهورت گلستان ارائه شده است (۱۲).

در این مطالعه مواجهه مورد بررسی را مصرف سیگار (مصرف کننده- عدم مصرف) و پیامد مورد بررسی را زمان تا مرگ در اثر سرطان دستگاه گوارش فوقانی (مری و معده) در نظر گرفته می‌شود. متغیرهای مخدوش کننده‌ی این ارتباط بر اساس معنی‌داری علیتی آن‌ها، انتخاب شده‌اند که عبارت‌اند از: جنس (دو گروه مرد و زن)- سن افراد در هنگام ورود به مطالعه (متغیر پیوسته)، نژاد (دو گروه ترکمن و غیر ترکمن)و محل زندگی (شهر- روستا).

یافته‌ها

در داده‌های کوهورت گلستان، از میان ۴۹۹۴۶ فردی که به این مطالعه وارد شدند، در حدود ۸۶۲۲ نفر سیگار مصرف می‌کردند. در طول دوره پیگیری، ۴۸۷ نفر بر اثر سرطان دستگاه گوارش فوقانی (مری و معده) فوت کرده‌اند که از آن‌ها، ۱۳۱ نفر مصرف‌کننده سیگار بوده و ۳۵۶ نفر مصرف‌کننده سیگار نبوده‌اند.

روش وزن‌دهی عکس احتمال به عنوان روش حاشیه‌ای برآورد اثر سیگار روی زمان تا مرگ افراد در اثر سرطان دستگاه گوارش فوقانی به‌کار گرفته شد و فاصله اطمینان برآوردها بر اساس روش بوت استرپ با حجم ۵۰۰ محاسبه شده است.

در جدول شماره (۱) مشاهده می‌شود که با استفاده از این روش در داده‌های بقا و تعدیل اثر مخدوشگرها، بین وضعیت مصرف سیگار و زمان تا مرگ در اثر سرطان گوارش فوقانی ارتباط علیتی وجود دارد. نسبت ریسک تعدیل شده برابر با ۲/۲ است. (فاصله اطمینان ۹۵ درصد: (۵/۲- ۰/۶۱)

در شکل شماره (۱) مشاهده می‌شود بقای تعدیل شده برای افرادی که سیگار مصرف نمی‌کردند بیش‌تر از بقای افرادی است که سیگار مصرف می‌کردند.

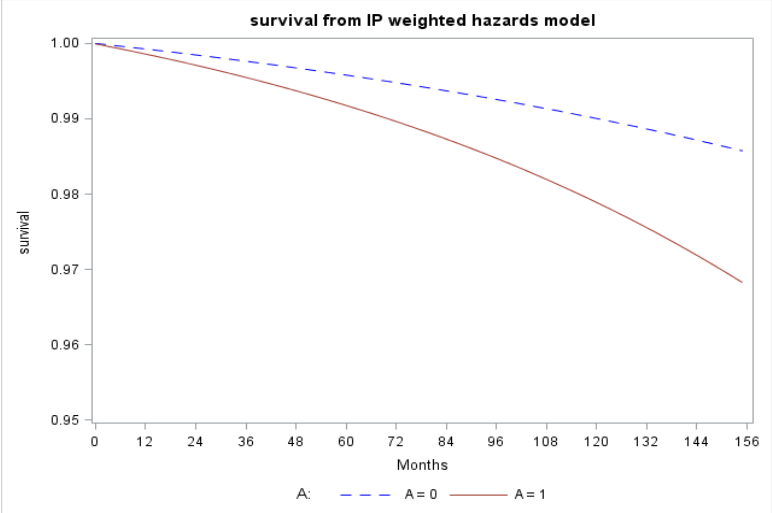
[۳] Risk ratio

[Downloaded from journals.tums.ac.ir on 2024-05-16]

جدول شماره ۱- نسبت خطر تعدیل شده با روش وزن‌دهی عکس احتمال و نسبت خطر بدون تعدیل مخدوشگرها

	نسبت خطر (فاصله اطمینان ۹۵٪)	
وضعیت مصرف سیگار	تعدیل نشده	تعدیل شده با روش وزن‌دهی عکس احتمال*
عدم مصرف	Ref	Ref
مصرف کننده	1.76(1.18-2.58)	2.2(0.61-5.2)

*تعدیل شده برای جنس، قومیت، وضعیت اقتصادی- اجتماعی، شهر-روستا، سن، زمان پیگیری



نمودار شماره ۱- نمودار بقای تعدیل شده با استفاده از روش وزن‌دهی عکس احتمال

بحث

با توجه به محدودیت‌های شاخص نسبت مخاطره از جمله متغیر بودن در زمان، غیرقابل آمیختن و تورش انتخاب و هم‌چنین به دلیل اینکه برآورد اثر علیتی حاشیه‌ای نسبت به برآورد اثرات شرطی برای مداخلات در جامعه مناسب‌تر می‌باشند، برآورد نسبت خطر حاشیه‌ای توصیه می‌شود. نسبت خطر حاشیه‌ای با استفاده از داده‌های شخص - زمان، و استفاده از روش وزن‌دهی عکس احتمال درمان، قابل محاسبه است و فاصله اطمینان ۹۵درصد برای این شاخص با استفاده از روش بوت استرپ برآورد می‌شود.

منابع

1 Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. European journal of epidemiology. 2015; 30: 1101-10.

2 Hernán MA, Robins J. Causal inference: Boca Raton: Chapman and Hall / CRC,forthcorning; 2017.

3 Kleinbaum DG, Klein M. Survival analysis: a self-learning text: Springer Science & Business Media; 2006.

4 Mohammad K, Hashemi-Nazari SS, Mansournia N, Mansournia M. Marginal versus conditional causal effects. Journal of Biostatistics and Epidemiology. 2015; 1: 121-8.

[Downloaded from journals.tums.ac.ir on 2024-05-16]

5 Mansournia MA, Greenland S. The relation of collapsibility and confounding to faithfulness and stability. Epidemiology. 2015; 26: 466-72.
6 Greenland S .Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. Epidemiology. 1996: 498-501.
7 Hernán MA. The hazards of hazard ratios. Epidemiology (Cambridge, Mass). 2010; 21: 13.
8 Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004; 15: 615-25.
9 Mansournia MA, Altman DG. Inverse probability weighting. BMJ. 2016; 352: i189.
10 Mansournia MA, Danaei G, Forouzanfar MH, Mahmoodi M, Jamali M, Mansournia N, et al. Effect of physical activity on functional performance and knee pain in patients with osteoarthritis: analysis with marginal structural models. Epidemiology. 2012; 23: 631-40.
11 Gharibzadeh S, Mohammad K, Rahimiforoushani A, Amouzegar A, Mansournia M. Standardization as a Tool for Causal Inference in Medical Research. Archives of Iranian Medicine. 2016; 19: 666.
12 Pourshams A, Khademi H, Malekshah AF, Islami F, Nouraei M, Sadjadi AR, et al. Cohort profile: the Golestan Cohort Study—a prospective study of oesophageal cancer in northern Iran. International journal of epidemiology. 2010; 39: 52-9.

[Downloaded from journals.tums.ac.ir on 2024-05-16]

Iranian Journal of Epidemiology Winter 2018; 13 (4): 356-361.

Review Article

Estimation of the Marginal Causal Risk Ratio in Survival Data Using Inverse Probability Treatment Weighting (IPTW)

Mohammadi N[1], Mansournia MA[2], Sadjadi SA A[3], Alimohammadian M[4], Poustchi H[5], Yaseri M[2].

1- MSc Student in Biostatistics, Department of Epidemiology and Biostatistics, Tehran University of Medical Sciences,Tehran, Iran
2- Assistant Professor, Department of Epidemiology and Biostatistics, Tehran University of Medical Sciences,Tehran, Iran
3- Physician, Digestive Diseases Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran
4-Instructor, Digestive Diseases Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran
5- Associate Professor, Digestive Diseases Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran
Corresponding author: Sadjadi SA A, n.mohammadi.24@gmail.com
(Received 11 June 2016**; Accepted** 26 November 2016**)**

One of the traditional methods used for the analysis of survival data is the Cox regression technique. This method calculates the conditional risk ratio. However, when the aim of the study is to estimate the effect of exposure in the total population level, using these conditional methods is not apposite. Furthermore, the hazard ratio has disadvantages of its own such as being non-collapsible, having the risk of structural selection bias and variability in time. Given the limitations, it is recommended to use the marginal hazard ratio, which estimates the average causal effect of exposure in the total population level.

This study introduces the inverse probability treatment weighting (IPTW) as a method of estimating the marginal causal effect. Finally, to illustrate IPTW method, we used Golestan Cohort Study and estimated the marginal causal effect of smoking on time to death due to the upper gastrointestinal cancer (esophageal-gastric).

Keywords: Marginal effect, Causal effect, Inverse probability treatment weighting

[Downloaded from journals.tums.ac.ir on 2024-05-16]