



بررسی عوامل موثر بر متاستاز سرطان پستان با مقادیر گمشده با استفاده از مدل لجستیک با درستنمایی توانیده نیمرخ

فروغ جمشیدی^۱، عنایت الله بخشی^{۱*}، سمانه حسین زاه^۱، سودابه شهید ثالث^۲

۱- گروه آمار، دانشگاه علوم بهزیستی و توانبخشی، تهران، ایران

۲- گروه رادیوتراپی و انکولوژی، دانشگاه علوم پزشکی مشهد، مشهد، ایران

چکیده	مقاله پژوهشی اصیل
مقدمه سرطان پستان یکی از شایع‌ترین و نگران‌کننده‌ترین مشکلات بهداشتی زنان در دنیاست که با بروز عود موضعی و متاستاز در این بیماری، پاسخ به درمان مشکل‌تر شده و بقای بیماران کاهش می‌یابد. همچنین در چنین مواردی علائم و نشانه‌های ایجاد شده بسیار ناراحت‌کننده و دردناک می‌باشد، به همین دلیل یافتن میزان تاثیر فاکتورهای مختلف بر بروز متاستاز می‌تواند نقش تاثیرگذاری بر درمان و جلوگیری از پیشرفت آن داشته باشد. مواد و روش‌ها در مطالعه حاضر اطلاعات ۱۹۵۹ نفر از زنان مبتلا به سرطان پستان که در فاصله سال‌های ۱۳۹۰-۱۳۸۰ به بیمارستان امید مشهد مراجعه نموده‌اند جمع‌آوری و مورد بررسی قرار گرفت. برای تحلیل داده‌ها پس از جابجایی داده‌های گمشده، از روش درستنمایی توانیده نیمرخ استفاده گردید و اجرای مدل و تحلیل داده‌ها در نرم‌افزار R انجام شد. یافته‌ها نتایج نشان داد، متغیرهای سابقه خانوادگی، مرحله بالینی، پروتئین Her2 و T (اندازه تومور) اثر معنی‌دار بر بروز متاستاز بیماران سرطان سینه دارد. نتیجه‌گیری با توجه به نتایج، متغیرهای مرحله بیماری و پروتئین her2، اندازه تومور و سابقه خانوادگی در تمامی شرایط تحلیل، اثر معنی‌داری بر بروز متاستاز در بیماران مبتلا به سرطان پستان داشتند. کلیدواژه‌ها درستنمایی توانیده نیمرخ، رگرسیون لجستیک، مقادیر گمشده، سرطان پستان، متاستاز	تاریخ دریافت: ۹۶/۱۲/۲۸ تاریخ پذیرش: ۹۷/۲/۳۰ *نویسنده مسئول: عنایت الله بخشی، گروه آمار، دانشگاه علوم بهزیستی و توانبخشی، تهران، ایران تلفن: _____ پست الکترونیک: _____ E.bakhshi@uswr.ac.ir



مقدمه

سرطان پستان یکی از شایع‌ترین و نگران‌کننده‌ترین مشکلات بهداشتی زنان در دنیا است. بروز مرگ و میر این بیماری در چند دهه اخیر دستخوش تغییرات زیادی شده است. کشورهای توسعه یافته از دهه ۱۹۸۰ دچار افزایش بروز این بیماری شدند و روند افزایش شدید آن تا دهه ۱۹۹۰ ادامه یافت که غالباً آن را به انجام ماموگرافی غربالگری منصوب می‌کنند (۱). این سرطان به عنوان اولین علت مرگ و میر زنان ۴۰-۴۴ ساله محسوب می‌شود (۲).

بر اساس گزارش سازمان بهداشت جهانی سالیانه بیش از ۱/۲ میلیون نفر مبتلا به سرطان پستان شناسایی می‌شوند (۳). در ایران ۱۶ درصد از کل سرطان‌ها مربوط به سرطان پستان می‌باشد که رتبه اول را در بین زنان ایرانی داشته است (۴). در بیماری سرطان، برگشت بیماری به صورت متاستاز در همان عضو یا اعضای دیگر بدن بروز پیدا می‌کند. متاستاز نشانه‌ای از غلبه بیماری بر درمان قبلی و نیاز بیمار به درمان مجدد است و سبب کاهش طول عمر بیمار و همچنین پایین آمدن کیفیت زندگی می‌گردد. (۵و۶).

بقای بدون بیماری یکی از متداول‌ترین معیارها، برای ارزیابی درمان بیماران سرطانی است که در مطالعات، عوامل مختلفی برای بقای بدون بیماری شناخته شده است. زمان بقای بدون بیماری، از شروع بیماری تا زمان رویداد متاستاز، یا مرگ بیمار تعریف می‌شود (۷و۸). براساس نتایج حاصل از یک کارآزمایی بالینی در آکسفورد انگلستان، بیش از ۵۰ درصد از عود سرطان پستان در زنان، بعد از ۵ سال پس از تشخیص اولیه رخ می‌دهد که استفاده از درمان هورمونی تاموکسیفن، باعث کاهش خطر عود سرطان پستان در طول ۵ سال اول بعد از جراحی شده است (۹).

در پژوهش‌ها نشان داده شده است که بقای گیرنده، HER2 بیماران سرطان پستان با عوامل گیرنده استروژن و

پروستروژن مثبت که درمان‌های کمکی دریافت کردند بالاتر از بیمارانی بود که درمان کمکی دریافت نکردند (۱۰). آگاهی از عوامل عود و متاستاز بیماران سرطان پستان، نقش مهمی در فرآیند درمان و مراقبت از بیماران، دارد. بر اساس پژوهش‌های پیشین علاوه بر نوع درمان، فاکتورهای دیگری از جمله گیرنده‌های استروژن و گیرنده پروژسترون و پروتئین her2، بر متاستاز موضعی تأثیر دارند. با بروز عود موضعی و متاستاز پاسخ به درمان مشکل‌تر می‌شود و بقای بیماران کاهش می‌یابد، همچنین در چنین مواردی علائم و نشانه‌های ایجاد شده بسیار ناراحت‌کننده و دردناک است، به همین دلیل یافتن میزان تأثیر فاکتورهای مختلف بر بروز متاستاز می‌تواند نقش تأثیرگذاری بر درمان و جلوگیری از پیشرفت آن داشته باشد.

روش آماری مناسب برای رسیدن به این هدف استفاده از مدل رگرسیونی لجستیک^۱ است اما تابع درستنمایی این مدل با افزایش متغیرهای مستقل در مدل، وجود داده‌های گمشده و کمبود رخدادها کارآمد نیست. در نتیجه به عنوان جایگزین می‌توان از تابع درستنمایی تاوانیده نیمرخ استفاده کرد. وجود داده‌های گمشده در داده‌ها امری طبیعی است اما تعداد زیاد داده گمشده ممکن است موجب استنباط‌های اریب در تجزیه و تحلیل آماری شده و دستیابی به نتایج نزدیک به واقعیت را با مشکل مواجه سازد. روش‌های مختلفی برای برخورد با داده‌های گمشده وجود دارد که در این مقاله از روش جانهی چندگانه استفاده شده است. هدف از این پژوهش تعیین عوامل موثر بر متاستاز سرطان پستان با مقادیر گمشده با استفاده از مدل لجستیک با درستنمایی تاوانیده نیمرخ است.

مواد و روش‌ها

^۱Logistic regression



به وسیله حل مربع رابطه زیر بدست می‌آید که $\ell(\beta)$ تابع لگاریتم درست‌نمایی است.

$$\partial \ell(\beta) / \beta_j = U(\beta_j) = 0$$

$$\ell(\beta) = \sum_{i=1}^n y_i \log \pi_i + (1 - y_i) \log (1 - \pi_i)$$

بازه‌های اطمینان برای پارامترها معمولاً به وسیله تقریب نرمال (روش والد) به دست می‌آیند که اگر

$$V(\beta) = I(\beta)^{-1}$$

ماتریس واریانس - کوواریانس برآورد شده بردار پارامترهای β و $I(\beta)$ ماتریس اطلاعات فیشر باشد بازه اطمینان بدست آمده با

$$-\{ \partial \ell(\beta) \}^2 / \{ \partial \beta_j \partial \beta_s \} (j, s = 1, \dots, k)$$

معادل است. در برآورد درست‌نمایی ماکزیمم، فرض کنید β به $\beta = (\gamma, \beta)$ تجزیه شود، که γ ضریب رگرسیون یک متغیر پیشگوی مورد نظر است و سایر ضرایب رگرسیون در بردار δ باشند و به عنوان جمله اضافی نامیده شوند، بازه اطمینان والد $100\% (1 - \alpha)$ برای γ از رابطه زیر بدست می‌آید

$$[\hat{\gamma} + z_{\alpha/2} \hat{\sigma}_{\gamma}, \hat{\gamma} + z_{1-\alpha/2} \hat{\sigma}_{\gamma}]$$

که در آن $\hat{\sigma}_{\gamma}$ ریشه دوم عنصر قطری V متناظر با γ است که از برآورد ماکزیمم درست‌نمایی $\hat{\beta} = (\hat{\gamma}, \hat{\delta})$ بدست می‌آید. روش والد برای پارامترهای برآورد شده فرض توزیع نرمال را در نظر می‌گیرد، اما در درست‌نمایی نیمرخ می‌توان توزیع نامتقارن را هم برای برآوردها فرض کرد (۱۲).

این پژوهش یک پژوهش کاربردی از نوع مقطعی-تحلیلی است. جامعه آماری این پژوهش کلیه زنان مبتلا به سرطان پستان مراجعه کننده به بیمارستان امید مشهد، طی سال-های ۱۳۸۰-۱۳۹۰ می‌باشد. با اخذ مجوز لازم از دانشگاه علوم پزشکی مشهد و مسئولان مربوطه، متغیرهای مورد نظر در این مطالعه از پرونده بیماران پذیرش شده بین سال‌های ۱۳۸۰-۱۳۹۰ در بیمارستان امید مشهد که دارای سرطان پستان بوده و در مرحله اول تا سوم بیماری بوده‌اند، استخراج و در فرم طراحی شده ثبت گردید. لازم به ذکر است که بیمارانی که قبل از گذشت ۳ ماه از شروع درمان (به دلیل تاخیر در تشخیص) فوت نموده‌اند، از این مطالعه حذف گردیده‌اند. در این پژوهش برای جایگزینی داده‌های گمشده از روش جهانی چندگانه استفاده شده است. در مطالعه حاضر متغیرهای متاستاز به عنوان متغیر وابسته و متغیرهای سن، قد، وزن، BMI، BSA، وضعیت تاهل، محل سکونت، سابقه خانوادگی، مرحله بالینی، گیرنده استروژن، گیرنده پروژسترون، پروتئین her2 و اندازه تومور به عنوان متغیرهای پیشگو مورد بررسی قرار گرفتند. تمامی آنالیزهای آماری با استفاده از نرم‌افزار R انجام شد. رگرسیون لجستیک به شکل زیر مدل‌بندی می‌شود:

$$P_r(y_i = 1 | x_i, \beta) = \pi_i = 1 + \exp \left(- \sum_{j=1}^k x_{ij} \beta_j \right)^{-1}$$

که در آن $y_i \in \{0, 1\}$ و $i = 1, \dots, n$ متغیر پاسخ دو حالتی، x_{ij} ($i = 1, \dots, n; j = 1, \dots, k$) مشاهدات در $k-1$ متغیر پیشگو و β ($j = 1, \dots, k$)، بردار پارامترهای رگرسیون را نشان می‌دهد. برآوردهای ماکزیمم درست‌نمایی β_j ($j = 1, \dots, k$) که می‌تواند به عنوان برآوردهای Log نسبت شانس‌ها تعبیر شوند



تابع لگاریتم درست‌نمایی نیم‌رخ γ به صورت
 $\ell_{\gamma}(\gamma) = \max_{\delta} \ell(\gamma, \delta)$ به وسیله ماکزیمم
کردن لگاریتم درست‌نمایی تحت پارامترهای برد δ و در نظر
گرفتن مقدار ثابت برای γ به دست می‌آید بنابراین بازه
اطمینان نیم‌رخ (PL)، $100\%(1 - \alpha)$ برای γ
مجموعه‌ای پیوسته از مقادیر γ برای هر لگاریتم نسبت
درست‌نمایی آماره

$$D_{\gamma}(\gamma) = -2 \{ \ell_{\gamma}(\gamma) - \ell(\hat{\gamma}, \hat{\delta}) \}$$

است که از $100\%(1 - \alpha)$ توزیع کای اسکور تجاوز
نکند. برای به دست آوردن حدود اطمینان یک طرفه و برای
استنباط آسان‌تر، کار کردن با ریشه مثبت که از رابطه زیر
بدست می‌آید راحت‌تر است.

$$r_{\gamma}(\gamma) = \text{sign}(\gamma - \hat{\gamma}) \sqrt{D_{\gamma}(\gamma)}$$

ماکزیمم درست‌نمایی توان‌یافته یا لگاریتم درست‌نمایی توان‌یافته
از رابطه زیر بدست می‌آید:

$$L^*(\beta) = L(\beta) |I(\beta)|^{1/2}$$

$$\ell^*(\beta) = \ell(\beta) + \frac{1}{2} \log |I(\beta)|$$

با $\ell(\beta)$ ، ماتریس اطلاع فیشر ارزیابی شده در β را نشان
می‌دهد که برخلاف لگاریتم درست‌نمایی معمولی، اریبی‌ها را
از برآوردهای ماکزیمم درست‌نمایی حذف می‌کند.

یافته‌ها

متغیرهای مورد نظر در این مطالعه از پرونده بیماران پذیرش
شده در سال‌های ۱۳۸۰-۱۳۹۰، در بیمارستان امید مشهد
که دارای سرطان پستان بوده و در مرحله اول تا سوم

بیماری بوده‌اند، استخراج و در فرم طراحی شده ثبت گردید.
متغیرهای این پژوهش شامل متاستاز به عنوان متغیر پاسخ
و متغیرهای سن، قد، وزن، BMI، وضعیت تاهل، محل
سکونت، سابقه خانوادگی، مرحله بالینی، وضعیت گیرنده
استروژن (ER)، وضعیت گیرنده پروژسترون (PR)، پروتئین
her2 و اندازه تومور به عنوان متغیرهای پیشگو وارد مدل
شدند.

در مطالعه حاضر ۱۹۵۹ نفر حضور داشتند که از این تعداد
۷۵۸ نفر متاستاز داشته و ۱۲۰۱ نفر متاستاز نداشتند. داده-
های حاضر نشان دهنده وجود مقادیر گمشده قابل ملاحظه
در متغیرهای پیشگو می‌باشد. بنابراین لازم است داده‌ها
پیش از برازش مدل به شیوه مناسب جانمایی گردند. روش
جانمایی داده‌ها در پژوهش حاضر روش جانمایی چندگانه است
که با استفاده از بسته mice در نرم‌افزار R انجام شده است.
نتایج توصیفی این مطالعه نشان داد میانگین سن و انحراف
معیار برای بیماران که متاستاز داشته‌اند به ترتیب ۴۸/۵۴ و
۱۱/۶۱۶ و میانگین سن و انحراف معیار بیماران که متاستاز
نداشته‌اند به ترتیب ۴۹/۲۷ و ۱۲/۱۹۴ می‌باشد. آزمون تی
دو نمونه‌ای مستقل نشان داد میانگین سن بیماران که
متاستاز داشته‌اند با میانگین سن بیماران که متاستاز
نداشته‌اند تفاوت معنی‌داری ندارد ($P=0/188$). آزمون تی
دو نمونه‌ای مستقل نشان داد میانگین BMI (شاخص توده
بدنی) بیماران که متاستاز داشته‌اند با میانگین BMI
بیماران که متاستاز نداشتند تفاوت معنی‌دار، ندارد
($P=0/349$).

بررسی وضعیت مرحله بالینی بیماران که شامل مرحله اول
(Stage ۱) یعنی زمانی که سرطان محدود به پستان است و
مرحله دوم (Stage ۲) که سرطان به غدد لنفاوی سرایت
کرده و مراحل سوم و چهارم که شامل سرایت بیماری به
سایر قسمت‌های بدن می‌باشد، بر حسب وضعیت متاستاز با



استفاده از آزمون کای اسکور نشان داد، ارتباط معنی‌داری بین وضعیت متاستاز و مرحله بالینی بیمار وجود دارد. در مراحل بالینی بالاتر تعداد افرادی که متاستاز داده‌اند بیشتر است ($P < 0.001$).

نتایج برازش مدل لجستیک توانیده نیم رخ پس از جانهی در جدول ۱ آمده است.

با توجه به نتایج مدل، متغیرهای سابقه خانوادگی سرطان، مرحله بالینی، پروتئین Her2 (این پروتئین گیرنده فاکتور رشد اپی‌درمال انسانی است) و اندازه تومور معنی‌دار شدند. تفسیر ضرایب معنی‌دار مدل به شرح زیر است: نسبت شانس متاستاز دادن برای بیمارانی که در مرحله بالینی ۲ هستند نسبت به بیمارانی که در مرحله ۱ هستند، $1/668$ برابر است. نسبت شانس متاستاز دادن برای بیمارانی که در مرحله بالینی ۳ هستند نسبت به بیمارانی که در مرحله ۱ هستند، $2/999$ برابر است. نسبت شانس متاستاز دادن برای افرادی که سابقه خانوادگی سرطان ندارند $30/3$ درصد کمتر است. نسبت شانس متاستاز دادن برای بیمارانی که در سطح سوم اندازه تومور هستند نسبت به بیمارانی که در سطح چهارم اندازه تومور هستند، $22/7$ درصد کمتر است. نسبت شانس متاستاز دادن برای بیمارانی که در سطح دوم اندازه تومور هستند نسبت به بیمارانی که در سطح چهارم اندازه تومور هستند، $42/6$ درصد کمتر است. نسبت شانس متاستاز دادن برای بیمارانی که در سطح اول اندازه تومور هستند نسبت به بیمارانی که در سطح چهارم اندازه تومور هستند، $46/9$ درصد بیشتر است. نسبت شانس متاستاز دادن برای افرادی که Her2 مثبت هستند $1/049$ برابر کسانی است که Her2 منفی دارند.



جدول ۱- برازش مدل رگرسیون لجستیک با تابع درست‌نمایی توانیده نیمرخ پس از جان‌هی چندگانه

متغیر	نسبت شانس	فاصله اطمینان ۹۵٪	P-value
عرض از مبدا	-	۰/۵۸۶	۱/۲۴۳
سن	-	۰/۹۹۵	۱/۰۰۳
BMI	-	۱/۰۱۱	۱/۰۲۷
وضعیت تاهل	مجرد	۱	-
	متاهل	۰/۷۹۸	۱/۳۱۱
	مطلقه	۱/۰۱۷	۰/۹۴۶
سابقه خانوادگی سرطان	بله	۱	-
	خیر	۰/۶۹۷	۰/۸۶۱
محل سکونت	شهر	۱	-
	روستا	۱/۰۷۸	۱/۳۶۶
مرحله بالینی	سطح ۱	۱	-
	سطح ۲	۱/۶۶۸	۲/۲۸۷
	سطح ۳	۲/۹۹۹	۴/۰۸۷
ER سطح مثبت نسبت به منفی	منفی	۱	-
	مثبت	۱/۰۶	۱/۴۱۸
PR سطح مثبت نسبت به منفی	منفی	۱	-
	مثبت	۰/۹۸۸	۱/۳۲
Her2 سطح مثبت نسبت به منفی	منفی	۱	-
	مثبت	۱/۰۴۹	۱/۲۷۶
اندازه تومور	سطح ۴	۱	-
	سطح ۳	۰/۷۷۳	۰/۹۷۹
	سطح ۲	۰/۵۷۴	۰/۷۹۱
	سطح ۱	۰/۵۳۱	۰/۷۷۹

* سطح معنی‌داری ۰/۰۵ می‌باشد.



بحث

سالانه بیش از ۷ میلیون نفر در جهان بر اثر سرطان جان خود را از دست می‌دهند و پیش‌بینی می‌شود که تعداد موارد جدید ابتلا تا سال ۲۰۲۰ سالانه از ۱۰ میلیون نفر به ۱۵ میلیون نفر برسد (۱۴ و ۱۳). در این بین سرطان پستان شایع‌ترین نوع سرطان در بین زنان است (۱۵).

در ایران نیز سرطان به‌طور کلی سومین عامل مرگ و میر است. سالانه بیش از ۳۰,۰۰۰ نفر در اثر سرطان جان خود را از دست می‌دهند. تخمین زده می‌شود سالانه بیش از ۸۰,۰۰۰ مورد جدید سرطانی در کشور اتفاق بیفتد و با افزایش امید به زندگی و درصد سالمندی در جمعیت کشور انتظار می‌رود بروز موارد سرطانی در دو دهه آینده به دو برابر افزایش یابد (۱۶) که در این بین سرطان پستان ۳۲ درصد موارد سرطان زنان را تشکیل می‌دهد (۱۷). بر طبق آمار مرکز مدیریت بیماری‌های وزارت بهداشت، سرطان پستان در ایران از نظر بروز بین تمام سرطان‌ها در زنان همچنان در رتبه اول قرار دارد و با بروز تعدیل شده سنی ۲۷/۱۵ و تعداد ۶۹۷۶ مورد در سال ۸۶ بالاترین موارد بروز بین سرطان‌ها را دارا است (۱۸).

ریسک فاکتورهای متعددی در ارتباط با سرطان پستان بیان شده‌اند که مهمترین آنها عبارتند از: سن، جنس، نژاد و قومیت، بیماری‌های خوش‌خیم پستان، سابقه فردی سرطان پستان، شیوه زندگی و رژیم غذایی، فاکتورهای هورمونی، سابقه خانوادگی و ژنتیک، مواجهه با اشعه، فاکتورهای محیطی و سیگار که سن و جنس به عنوان قویترین ریسک فاکتورها برای ابتلا به سرطان پستان نام برده شده‌اند. میزان شیوع بیماری تا سن ۴۵-۵۰ سالگی با سرعت زیادی افزایش می‌یابد و پس از آن از سرعت شیوع کاسته می‌شود که این تغییر به علت تاثیر تغییرات هورمونی در زنان یائسه می‌باشد.

از سن ۷۵-۸۰ سالگی میزان ابتلا به سرطان پستان کاهش می‌یابد.

با توجه به موارد ذکر شده و درصد زیاد ابتلا به سرطان پستان، جدای از مباحث پیشگیری و تشخیص، شناسایی عوامل افزایش دهنده بروز متاستاز در مبتلایان می‌تواند نقش موثری در پیشگیری از گسترش دامنه بیماری و افزایش مرگ به دلیل این بیماری باشد. برای شناسایی این عوامل به روش‌های پیشرفته‌تر آماری نیاز است که یکی از این روش‌ها، روش درستنمایی تاوانیده نیمرخ است که برآوردهای دقیق‌تری برای پارامترهای مدل لجستیک ارائه می‌دهد از طرفی وجود مشاهدات گمشده در بسیاری از پژوهش‌های بالینی امری اجتناب‌ناپذیر است. وجود مشاهدات گمشده روش‌های معمول آماری را ناکارآمد می‌کند. بنابراین نیاز است درصورت وجود مشاهدات گمشده ابتدا با روش مناسب جانهی داده‌ها صورت پذیرد سپس مدل آماری مناسب بکار گرفته شود.

روش رگرسیون لجستیک تاوانیده توسط بومن^۱ در سال ۱۹۹۳ استفاده شد (۱۹). هر چند این مدل بیشتر در مورد داده‌های مرتبط با ژنتیک بکار رفته است اما صرفاً محدود به این داده‌ها نشده و در زمینه‌های دیگری هم چون مطالعات زیست پزشکی نیز استفاده گردیده است.

جی‌زو و تریوور هستی^۲ از رگرسیون لجستیک تاوانیده برای طبقه‌بندی میکروآرایه‌های ژنی در تشخیص و درمان سرطان استفاده کردند در این مطالعه رگرسیون لجستیک تاوانیده با SVM^۳ (یکی از روش‌هایی موفق در مسائل تشخیصی سرطان) مقایسه شد که دو روش طبقه‌بندی مشابه بودند اما رگرسیون لجستیک تاوانیده برآوردی از احتمال را نیز در

^۱Bouman

^۲Ji Zhu & Trevor Hesti

^۳Support vector machine



اختیار قرار می‌داد (۲۰). هی^۱ و همکاران در یک مطالعه مورد - شاهدی به مقایسه توان کاهش ابعاد چندعاملی و رگرسیون لجستیک توانانیده برای تشخیص اثرات متقابل ژن - ژن پرداختند و به این نتیجه رسیدند که اگر مدل جمع-پذیر باشد رگرسیون لجستیک توانانیده از کاهش ابعاد چندعاملی بهتر عمل می‌کند (۲۱).

کرویزی^۲ و همکارانش در آنالیز داده‌های مربوط به مفاصل روماتیسمی از رگرسیون لجستیک توانانیده استفاده کردند (۲۲). میلی^۳ و همکاران برای اعتبار تکرار در مطالعات اثر متقابل ژن - ژن از روش کاهش ابعاد چندعاملی برای سرطان سینه در مطالعه مورد - شاهدی استفاده کردند و به اثر متقابل ۴ طرفه میان ژن‌ها با $P < 0.006$ دست یافتند (۲۳). ریج^۴ و همکاران از روش کاهش ابعاد چندعاملی برای آشکار کردن اثرات متقابل مرتبه بالا میان ژن‌های استروژن - متابولیسم در سرطان سینه پراکنده پرداختند و بدون یافتن هیچ اثر اصلی، اثرات متقابل مرتبه بالایی میان ۴ پلی مورفیسم از ۳ ژن متفاوت استروژن - متابولیسم شناسایی کردند (۲۴). وایت و همکاران^۵ وجود چندین متغیر پاسخ با داده‌های گمشده با الگوی غیر یکنواخت به وسیله جانهی چندگانه بررسی کرده و پیشنهادهای ارزنده‌ای را در خصوص گسترش مناسب و مفید مدل‌های جانهی ارائه می‌کنند (۱۹).

مقایسه نتایج بالینی این پژوهش با سایر پژوهش‌ها نشان داد در مطالعه‌ای که توسط مقدمی فرد و همکاران که با هدف به دست آوردن عوامل موثر بر متاستاز انجام گرفت، درجه بدخیمی تومور و تعداد گره‌های درگیر، عامل پیش آگهی

دهنده در بروز متاستاز شناخته شدند (۲۵)، همچنین در مطالعه دیگری که توسط گوهری و همکاران انجام گرفت، تعداد گره‌های لنفی درگیر و پروتئین HER2 پیش آگهی دهنده متاستاز تعیین شدند و متغیرهای سن، گیرنده استروژن، گیرنده پروستروژن، درجه بدخیمی تومور و اندازه تومور معنی‌دار نشدند (۲۶). در مطالعه دیگری که توسط حسین زاده و همکاران انجام گرفت، استروژن از عوامل مؤثر بر عود موضوعی و درگیری غدد لنفاوی از عوامل مؤثر بر عود موضوعی و متاستاز بود. اثر وضعیت گیرنده استروژن و درگیری غدد لنفاوی بر وقوع عود از اثر آنها بر متاستاز بیشتر بود (۲۷).

در مطالعه حاضر نیز، با توجه به وجود داده‌های گمشده در متغیرها از روش جانهی چندگانه برای داده‌های گمشده استفاده گردید و از روش درست‌نمایی توانانیده نیمرخ برای براز مدل لجستیک استفاده شد.

پژوهش حاضر دارای محدودیت‌هایی همچون انجام به صورت مقطعی و نقص پرونده برخی بیماران بود که در مطالعات آتی با رفع این مشکلات می‌توان به نتایج مطلوب‌تر دست یافت.

نتیجه‌گیری

مطالعه حاضر با هدف کمک به شناخت عوامل موثر در بروز متاستاز در بیماران مبتلا به سرطان پستان صورت گرفت که با توجه به نتایج مدل، متغیرهای سابقه خانوادگی سرطان، مرحله بالینی، پروتئین Her2 و اندازه تومور تاثیر معنی‌داری بر وضعیت متاستاز فرد از خود نشان دادند. انتظار می‌رود یافته‌های مطالعه حاضر با تاکید بر متغیرهای معنی‌دار در بروز متاستاز بتواند در شناخت بهتر وضعیت درمانی این بیماران و ارائه برنامه‌های مداخله‌ای مناسب‌تر راه‌گشا باشد.

^۱He H

^۲Pascal Croiseau and Hearther J Cordell

^۳Roger L.Milne

^۴Ritchie

^۵White



References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893-917.
2. Jalali Nadoushan MR, Davati A, Bagherzadeh Shahidi SH. Expression of MDM2 gene in primary breast cancer and it's relationship with prognostic factors. *J Gorgan Univ Med Sci* 2008;10(2):27-31. [Persian]
3. Faghani M, Nasiri E, Bahadori MH, Mohammad Ghasemi F. Genetic predisposing of P53 codon 72 on developing of breast cancer in postmenopausal women in Isfahan. *Jour Guilan Univ Med Sci* 2008;17(67):94-100. [Persian]
4. Mousavi M, Gouya M; Ramazani R, Davanlou M, Hajsadeghi N, Seddighi Z. Cancer incidence and mortality in Iran. *Ann Oncol* 2009;20(3):556-63.
5. Gohari MR, Mahmoudi M, Kazem M, Pasha E, Khodabakhshi R. Recurrence in breast cancer analysis with frailty model. *Saudi Med J* 2006;27(8):1187-93.
6. Bernstein L, Ross RK. Endogenous hormones and breast cancer risk. *Epidemiol Rev* 1993;15(1):48-65.
7. Cowper DC, Kubal JD, Maynard C, Hynes DM. A primer and comparative review of major US mortality databases. *Nn Epidemiol* 2002;12:462-8.
8. Lamont EB, Herndon JE, Weeks JC, Henderson IC, Earle CC, Schilsky RL, *et al.* Measuring disease-free survival and cancer relapse using Medicare claims from CALGB breast cancer trial participants (companion to 9344). *J Natl Cancer Inst* 2006;98(18):1335-8.
9. Khodabakhshi R, Gohari MR, Moghadamifard Z, Foadzi H, Vahabi N. Disease-free survival of breast cancer patients and identification of related factors. *Razi Journal of Medical Sciences* 2011;18(89):27-33.
10. Ska N, Murawska M, Pogoda K. Breast cancer brain metastases: differences in survival depending on biological subtype, RPA RTOG prognostic class and systemic treatment after whole-brain radiotherapy (WBRT). *Ann Oncol* 2010;21(5):942-8.
12. Heinz G, Ploner M, Beyea J. Confidence intervals after multiple imputation: combining profile likelihood information from logistic regressions. *Stat Med* 2013;32(29):5062-76.
13. Hasanpoor Dehkordi A, Azari S. Quality of life and related factor in cancer patients. *Behbood* 2006;10(2):110-19. [Persian]
14. Saki A, Hajizadeh E, Tehranian N. Evaluating the risk factors of breast cancer using the analysis of tree models. *Ofogh-e-Danesh* 2011;17(1):60-9. [Persian]
15. Safaee A, Zeighami B, Tabatabaee HR, Moghimi Dehkordi B. Quality of life and related factors in breast cancer patients under chemotherapy. *irje* 2008;3(4):61-6.
16. Fathi F, Pourasghar M, Janbabai Gh. Pain control with hypnosis for acute myeloid leukemia patients under chemotherapy. *Medical Journal of Mashhad University of Medical Sciences* 2016;59(2):71-81. [Persian]
17. Hosseini M, Hassannejad R, Khademolghorani SH, Tabatabaieian M, Mokarian F. Identification of patterns of breast cancer metastasis among women referred to Isfahan Seyedoshohada Center, Iran, between 1999 and 2009 by association rules and ordinal logistic regression scientific research. *Scientific Research Journal of Health System Research (HSR)* 2011;7(6):746-62.
18. Roohparvarzadeh N, Ghaderi M, Parsa A, Allahyari A. Prevalence of risk factors for breast cancer in women (20 to 69 Years old) in Isfahan 2012-2013. *Iranian Journal of Breast Diseases* 2014;7(1):52-61. [Persian]
19. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine* 2011;30(4):377-99.
20. Zhu J, Hastie T. Classification of gene microarrays by penalized logistic regression. *Biostatistics* 2004;5(3):427-43.
21. He H, Oetting WS, Brott Mj, Basu S. Power of multifactor dimensionality reduction and penalized logistic regression for detecting gene-gene Interaction in a case-control study. *BMC Med Genet* 2009;10:127.
22. Croiseau P, Cordell HJ. Analysis of North American rheumatoid arthritis consortium data using a pevalized logistic regression approach. *BMC Proc* 2009;3:S61.
23. Milne RL, Fagerholm R, Vevanlinna H, Benitez J. The importance of replication in gene-gene interaction studies: multifactor dimensionality reduction applied to a two-stage breast cancer case-control study. *Carcinogenesis* 2008;29(6):1215-8.

24. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF , *et al.* Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism gene in sporadic breast cancer. *Am J Hum Genet* 2001;69(1):138-47.
25. Moghadamifard Z, Abolghasemi J, Asadilari M, Mohammadi M, Gohari M, Salehi M. Effective factors in the appearance of metastasis in patients with breast cancer using frailty model. *J Arak Uni Med Sci* 2013;15(8):85-94.
26. Gohari M, Moghadamifard Z, Abolghasemi J, Mohammadi M, Mokhtari P. Prognostic factors of metastases in breast cancer patients using the recurrent Andersen-Gill model. *koomesh* 2013;14(4):483-9.
27. Hosseinzadeh S, Haghighat S. Determining risk factors for local recurrence and metastasis in breast cancer patients using bivariate logistic model. *JHPM* 2016;5(5):55-62.
- .



Applying logistic model on breast cancer metastatic with missing data: penalized profile likelihood

Forugh Jamshidi¹, Enayatollah Bakhshi^{*1}, Samaneh Hoseinzadeh¹, Sudabeh Shahid Sales²

1- Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

2- Department of Radiotherapy and Oncology, Mashhad University of Medical Sciences, Mashhad, Iran

Original Article

Received: Mar 19, 2018

Accepted: May 20, 2018

***Corresponding Author:**

Enayatollah Bakhshi,
Department of Biostatistics,
University of Social Welfare
and Rehabilitation Sciences,
Tehran, Iran

TEL:

Email: E.bakhshi@uswr.ac.ir

ABSTRACT

Introduction

Breast cancer is one of the most common and worrisome diseases among women all over the world which becomes more difficult with localized relapses and metastases in this disease and the chance of survival is reduced. therefore the effect of various factors on the incidence of metastases can have an important role in the treatment and prevention of disease progression.

Materials and Methods

: In this study, 1959 women with breast cancer who referred to Omid Hospital in Mashhad from 2001 to 2011 were selected and evaluated. To analyze the data, after the imputation of missing data, the method of penalized profile likelihood was used and data analysis was done in the software R.

Results

The results of the analysis showed that family background, clinical stage, Her2 protein, and T (tumor size) variables had a significant effect on the incidence of metastasis in breast cancer patients.

Conclusion

According to the results, the stage code and her2 variables, tumor size and family history in all conditions of analysis had a significant effect on the incidence of metastasis in patients with breast cancer.

Keywords

penalized profile likelihood, logistic regression, missing values, breast cancer, metastasis

► **Please cite this article as:** Jamshidi F, Bakhshi E, Hoseinzadeh S, Shahid Sales S. Applying logistic model on breast cancer metastatic with missing data: penalized profile likelihood. J Neyshabur Univ Med Sci 2019;6(4):94-104.