

## عوامل مرتبط با بروز دیابت نوع دو در زنان پیش‌دیابتی با استفاده از مدل متوسط بیزی

مریم مهدوی<sup>۱</sup> (M.Sc.)، یداله محرابی<sup>۲\*</sup> (Ph.D.)، داوود خلیلی<sup>۳</sup> (Ph.D.)، احمدرضا باغستانی<sup>۱</sup> (Ph.D.)، فریده باقرزاده خیابانی<sup>۳</sup> (M.Sc.)، سمانه منصوری<sup>۱</sup> (M.Sc.)

۱- گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

۲- گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

۳- مرکز تحقیقات پیشگیری از بیماری‌های متابولیک، پژوهشکده علوم غدد درون‌ریز و متابولیسم، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

### چکیده

هدف: دیابت، بیماری مزمنی است که معمولاً با اختلال تحمل گلوکز آغاز می‌گردد. این مرحله، اصطلاحاً به عنوان پیش‌دیابت شناخته می‌شود. افراد پیش‌دیابتی در معرض خطر بیش‌تری برای رخداد دیابت هستند، بنابراین شناسایی عوامل خطر ساز دیابت، در افراد پیش‌دیابتی، مورد توجه بیش‌تری است. به طور معمول برای انتخاب متغیر، از روش گام‌به‌گام استفاده می‌شود که عدم قطعیت مدل را در نظر نمی‌گیرد. در این مطالعه، برای برطرف کردن این مشکل، از روش متوسط بیزی (Bayesian Model Averaging, BMA) استفاده شد.

مواد و روش‌ها: جمعیت مورد بررسی، ۷۳۴ زن ۲۰ سال و بالاتر پیش‌دیابتی، در مطالعه کوهورت قند و لیپید تهران (TLGS) است. در مطالعه حاضر، برای انتخاب متغیر در رگرسیون لجستیک از روش‌های گام‌به‌گام و متوسط بیزی استفاده شد؛ سپس AUC دو روش محاسبه و با آزمون دلونگ (Delong) مقایسه شد. برای انجام تحلیل‌ها از بسته BMA نرم‌افزار R نسخه ۳،۱،۳ استفاده شد.

یافته‌ها: روش متوسط بیزی، قندخون ناشتا، قندخون ۲ ساعته، سابقه خانوادگی، نمایه توده بدنی و مصرف آسپرین را در ابتدای مطالعه، به عنوان عوامل خطر ساز بروز دیابت، انتخاب کرد. روش گام‌به‌گام علاوه بر این‌ها، فشارخون دیاستولیک، سابقه بستری ۳ ماه گذشته، استفاده از داروهای تیروئید و تحصیلات را نیز انتخاب کرد. با این‌که تعداد متغیرهای روش گام‌به‌گام (۹ متغیر) بیش‌تر از روش متوسط بیزی (۵ متغیر) است، AUC برای روش‌های متوسط بیزی و گام‌به‌گام تفاوت معنی‌داری نداشت.

نتیجه‌گیری: به نظر می‌رسد مدل متوسط بیزی قابلیت خوبی برای غربالگری بیماری دیابت دارد، زیرا ضمن تشخیص متغیرهای موثر کم‌تر، توانایی پیش‌بینی مدل نیز حفظ شد.

واژه‌های کلیدی: مدل متوسط بیزی، روش گام‌به‌گام، پیش‌دیابت، رگرسیون لجستیک، مطالعه قند و لیپید تهران

### مقدمه

گلوکز آغاز می‌گردد. این مرحله، اصطلاحاً به عنوان پیش‌دیابت شناخته می‌شود [۱]. این حالت، قبل از ابتلای

دیابت، بیماری مزمنی است که معمولاً با اختلال تحمل

روش‌های گام‌به‌گام، به مدل‌های رقیب بی‌توجه هستند و بدین جهت با مشکل عدم قطعیت مدل مواجه هستند [۱۳]. ویگاندر در سال ۲۰۱۰، عمل‌کرد ضعیف الگوریتم‌های گام‌به‌گام چندگانه را برای انتخاب متغیر نشان داد [۱۴].

ترکیب نتایج مدل‌ها، اجازه می‌دهد که از نقاط قوت مدل‌های متفاوت بهره‌گرفت و عدم قطعیت مدل را لحاظ کرد [۱۵، ۱۶]. با تعداد زیاد عوامل خطر ساز، متوسط‌گیری روی همه مدل‌ها نشدنی است. بنابراین، لازم است، مکانیزی برای ترکیب نتایج از مدل‌های متفاوت به کار گرفته شود [۱۷]. در این مطالعه روش متوسط بیزی به کار رفت که با متوسط‌گیری روی زیر مجموعه‌ای از مدل‌های حمایت شده به وسیله داده‌ها، مساله عدم قطعیت مدل را در نظر گرفت.

ژنگ و همکاران در سال ۲۰۰۴، هم‌چنین جنل و همکاران در سال ۲۰۱۰، عمل‌کرد پیش‌بینی بهتر روش متوسط بیزی نسبت به روش گام‌به‌گام را به ترتیب در رگرسیون لجستیک و رگرسیون خطی نشان دادند [۱۷، ۱۸]. هدف این مطالعه، تعیین عوامل خطر ساز مرتبط با دیابت نوع ۲ در زنان پیش‌دیابتی ۲۰ سال و بالاتر TLGS، با استفاده از روش متوسط بیزی و مقایسه نتایج آن با روش‌های گام‌به‌گام است.

## مواد و روش‌ها

جمعیت مورد بررسی، زنان ۲۰ سال و بالاتر مبتلا به پیش‌دیابت منطقه ۱۳ تهران بود که در فازهای ۱ و ۲ وارد مطالعه قند و لیپید تهران (TLGS) شدند. این افراد تا فاز ۴ از نظر رخداد دیابت، پیگیری شدند. افرادی که میزان قندخون ناشتای آنان مساوی یا بیش‌تر از ۱۰۰ و کم‌تر از ۱۲۶ بود یا قند ۲ ساعته آنان مساوی یا بیش‌تر از ۱۴۰ و کم‌تر از ۲۰۰ بود، به عنوان افراد پیش‌دیابتی شناخته شدند. جزییات مطالعه قند و لیپید تهران و چگونگی انجام آزمایشات و ابزارهای مورد استفاده قبلاً چاپ شده است [۱۹].

در مطالعه TLGS در فاز اول ۱۵۰۰۵ نفر ۳ سال و بالاتر شرکت کردند که ۱۰۷۷ نفر از آن‌ها زنان پیش‌دیابتی ۲۰ سال و بالاتر بودند. در فاز دوم نیز ۳۵۰۰ فرد جدید ۳ سال و

جدی به دیابت نوع ۲ بروز می‌کند. در این حالت، قندخون فرد، بالاتر از سطح نرمال است ولی در حدی بالا نیست که بتوان آن را دیابت در نظر گرفت [۲]. حدود ۷ درصد از جمعیت جهان، دیابتی هستند [۳] و از عوارض متعاقب آن رنج می‌برند [۴]. برآورد سازمان جهانی بهداشت، نشان می‌دهد که تعداد دیابتی‌ها در جهان ۳۶۶ میلیون نفر است [۵]. اگرچه میزان وقوع دیابت نوع ۱ و ۲ در سراسر جهان رو به افزایش است، ولی سرعت این افزایش در دیابت نوع ۲ بیش‌تر است [۱]. دیابت نوع ۲، شایع‌ترین نوع دیابت بوده و ۹۰ درصد موارد بیماری را به خود اختصاص داده است. شیوع دیابت نوع ۲، در کشورهای در حال توسعه از جمله ایران، در حال افزایش است [۶].

مطالعات انجام‌شده، شیوع دیابت را در جمعیت شهری اصفهان ۵٫۶ درصد [۷]، در جمعیت روستایی بالای ۳۰ سال استان تهران ۶٫۷ درصد [۸] و در جمعیت شهری و روستایی بالای ۳۰ سال استان کرمانشاه ۸٫۶ درصد نشان می‌دهد [۹]. روچ و همکاران در سال ۲۰۱۴ با استفاده از رگرسیون لجستیک، نشان دادند که عوامل خطر ساز دیابت در زنان و مردان متفاوت است [۵]. با وجود این‌که شیوع دیابت در مردان بیش‌تر از زنان است اما زنان در معرض خطر بیش‌تری از مرگ و میر و بستری شدن در بیمارستان هستند [۱۰]. بنابراین بررسی عوامل خطر ساز دیابت، در مطالعه حاضر محدود به زنان شد.

بررسی با ۹ سال پیگیری، در مطالعه قند و لیپید تهران نشان داده است که از هر ۱۰۰۰ نفر زن ۲۰ سال به بالای سالم در تهران، سالانه ۳۸ نفر دچار حالت پیش‌دیابت می‌شوند [۱۱]. افراد پیش‌دیابتی بدون مداخله طی کم‌تر از ۱۰ سال به سوی دیابتی شدن پیش می‌روند [۵]. بنابراین نیاز به تشخیص عوامل خطر ساز دیابت، برای افراد مبتلا به پیش‌دیابت احساس می‌شود.

روش‌های انتخاب عوامل خطر ساز، باید تعادلی بین بیش‌برازشی و کم‌برازشی ایجاد کنند [۱۲]؛ به طور معمول برای این کار از روش‌های گام‌به‌گام استفاده می‌شود.

عقب و گام به گام رو به جلو که معیار انتخاب متغیر آن‌ها بر اساس AIC است، استفاده شد؛ هم‌چنین روش متوسط بیزی برای انتخاب متغیر به کار رفت.

لیمر در سال ۱۹۷۸ الگوی پایه‌ای مدل متوسط بیزی را ارائه داد [۲۲]. اما به دلیل مشکلات محاسباتی در آن زمان، مورد توجه قرار نگرفت. با پیشرفت فنون محاسباتی در دهه نود، این روش دوباره مطرح و مورد توجه زیادی قرار گرفت. مدیگان و رفتی دو روش برای اجرای این شیوه مدل‌سازی ارائه کردند [۲۳]. مدل متوسط بیزی، یک تکنیک چندمتغیره است که عدم قطعیت مدل را در محاسبات، در نظر می‌گیرد. این روش یک زیرمجموعه  $k$  تایی ( $k=2^q$ ) از همه مدل‌های ممکن را انتخاب می‌کند که در آن  $k$  تعداد مدل‌ها و  $q$  تعداد متغیرها است، یعنی هر یک از متغیرهای مورد بررسی می‌تواند در مدل یا خارج از مدل باشد. سپس روی این مجموعه از مدل‌ها، میانگین‌گیری وزنی انجام می‌دهد که وزن‌ها، احتمالات پسین مدل‌ها هستند. مدل متوسط بیزی از چگالی پسین در معادله (۱) برای استنباط و پیش‌گویی استفاده می‌کند:

$$\Pr(\theta|D) = \sum_{k=1}^K \Pr(\theta|M_k, D) \Pr(M_k|D) \quad (1)$$

که در آن، پارامتر  $\theta$  کمیت مورد علاقه است (به عنوان مثال، پارامتر مورد علاقه می‌تواند لگاریتم نسبت بخت‌ها باشد)،  $D$  نشان‌دهنده داده‌های مورد بررسی است و  $M_k$  نشان‌دهنده  $K$  مدل پیشنهادی برای تحلیل داده‌ها است.  $\Pr(\theta|M_k, D)$  توزیع پسین پارامتر مورد علاقه تحت  $k$  مدل  $M_k$  است که توسط برآورد ماکزیمم درست‌نمایی تقریب زده می‌شود. چگالی پسین پارامتر مورد علاقه، توسط احتمال پسین مدل یعنی  $\Pr(M_k|D)$  وزن‌دار شده است. به معادله (۱)، مدل متوسط بیزی (BMA) گفته می‌شود [۲۴].

از آن‌جا که بررسی تمام مدل‌ها عملی نیست، روش متوسط بیزی از الگوریتم Leaps and bound [۲۵] و روش Occams Window [۲۶] برای کاهش تعداد مدل‌ها استفاده می‌کند. ایده پایه‌ای الگوریتم Leaps and bound بر این اساس است: برای دو مدل  $A$  و  $B$ ، اگر  $A \subset B$  باشد آنگاه

بالاتر در مطالعه وارد شدند که در میان آنان ۱۷۴ زن ۲۰ سال و بالاتر پیش‌دیابتی وجود داشت. در مجموع ۱۲۵۱ زن پیش‌دیابتی ۲۰ سال و بالاتر به این مطالعه وارد شدند و تا فاز ۴ از نظر رخداد دیابت، پیگیری شدند. از این تعداد ۴۴۸ نفر از آنان قبل از پایان مطالعه خارج شدند، بنابراین ۸۰۳ نفر در مطالعه باقی ماندند که مقادیر متغیرهای آنان در شروع مطالعه بررسی شد.

متغیرهای مورد بررسی، دارای مقادیر گمشده بودند که با استفاده از بسته Miss Forest در محیط R، مقادیر گمشده برآورد شد. این روش می‌تواند تحت شرایطی مثل بعد بالا، اثرات متقابل پیچیده و ساختار داده غیرخطی، عمل‌کرد خوبی داشته باشد [۲۰]. سپس برای تشخیص نقاط دورافتاده به دلیل بعد بالای داده‌ها از روش چندمتغیره در بسته mvoutlier استفاده شد [۲۱]. با این کار ۶۹ نفر، مشاهده پرت شناخته شدند که از مطالعه کنار گذاشته شدند، بنابراین ۷۳۴ نفر وارد تحلیل آماری شدند. نحوه ورود افراد در مطالعه در شکل ۱ نشان داده شده است.

در این مطالعه، روش‌های گام به گام و متوسط بیزی برای انتخاب متغیر در رگرسیون لجستیک به کار گرفته و مقایسه شدند. متغیر پاسخ، بروز دیابت در زنان پیش‌دیابتی بالای ۲۰ سال می‌باشد. متغیرهایی که به نظر متخصص بالینی ممکن است با بروز دیابت مرتبط باشند از نظر هم‌خطی و همبستگی بررسی شدند، از میان آن‌ها متغیرهایی که هم‌خطی و همبستگی بالا داشتند و از نظر بالینی دارای اهمیت کم‌تری بودند، کنار گذاشته شدند. در نهایت ۲۹ متغیر، در ابتدای مطالعه مورد بررسی قرار گرفت. این افراد به مدت ۱۲-۱۰ سال برای بروز دیابت پیگیری شده بودند.

در هر فاز افراد دارای قندخون ناشتای ۱۲۶ یا بالاتر، قند ۲ ساعته ۲۰۰ یا بالاتر و یا مصرف داروی کاهنده قندخون، دیابتی تشخیص داده شدند. برای مقایسه شاخص‌های مربوط به افراد دیابتی و غیردیابتی از آزمون‌های تی و مجذور خی استفاده شد. برای انتخاب متغیر از روش‌های گام به گام روبه

مطابق با UIP روی پارامتر  $\beta^k$  است، UIP یک پیشین نرمال چندمتغیره با میانگین برآورد ماکزیمم درست‌نمایی و واریانس مساوی با ماتریس اطلاع مورد انتظار برای یک مشاهده است [۲۸]. سپس روی این مجموعه مدل‌های کاهش یافته، با در نظر گرفتن احتمال پسین مدل به عنوان وزن، متوسط‌گیری وزنی انجام شد.

در روش متوسط بیزی، برآورد نقطه‌ای بیزی (میانگین پسین) و خطای استاندارد (انحراف استاندارد پسین) از  $\beta_j$  به ترتیب به فرم روابط (۴) و (۵) به دست می‌آید.

$$E(\beta_j|D) = \sum_{M_k \in A} \hat{\beta}_j \Pr(M_k|D) \quad (4)$$

$$SE(\beta_j|D) = \sqrt{\sum_{M_k \in A} \{[var(\beta_j|D, M_k) + \hat{\beta}_j^2] \Pr(M_k|D)\} - E(\beta_j|D)^2} \quad (5)$$

در اینجا  $\beta_j$  بردار پارامتر تحت مدل  $M_k$ ،  $P(\beta_j | M_k)$  چگالی پیشین پارامتر و  $P(D|\beta_j, M_k)$  تابع درست‌نمایی تحت مدل  $M_k$  است.  $\hat{\beta}_j$  میانگین پسین  $\beta_j$  تحت مدل  $M_k$  است. در این روش برای فهمیدن این‌که متغیر  $X_j$  مورد بررسی، روی وقوع دیابت اثر دارد یا نه، باید احتمال اثر پسین یعنی  $\Pr[\beta_j \neq 0|D]$  که به صورت رابطه (۶) است، برای متغیر محاسبه شود.

$$\Pr(\beta_j \neq 0|D) = \sum_{M_k \in A} \Pr(M_k|D) * I_k(\beta_j \neq 0) \quad (6)$$

که در آن  $I_k(\beta_j \neq 0) = \begin{cases} 1 & \beta_j \neq 0 \\ 0 & \beta_j = 0 \end{cases}$  است. جمله شامل  $X_j$  است [۲۶]. تفسیر این احتمال پسین بدین صورت است که اگر کم‌تر از ۰/۵۰ باشد متغیر اثرگذار نیست، ۰/۷۵-۰/۵۰ متغیر اثرگذاری ضعیف دارد، ۰/۹۵-۰/۷۵ متغیر اثرگذار است، ۰/۹۹-۰/۹۵ متغیر اثرگذاری قوی دارد و بیش‌تر از ۰/۹۹ متغیر کاملاً بر رخداد دیابت اثرگذار است. متغیرهایی که احتمال اثر پسین بالای ۰,۵ داشته باشند، به عنوان متغیرهای موثر بر بروز دیابت تشخیص داده شدند [۲۹،۲۷].

پس از انتخاب متغیر با روش‌های گام‌به‌گام و متوسط بیزی این دو روش از نظر قدرت پیش‌گویی با هم مقایسه شدند. برای مقایسه مدل متوسط بیزی و روش گام‌به‌گام، ۷۳۴

RSS(A) > RSS(B). با استفاده از این حقیقت می‌توان بهترین مدل‌ها را بدون برازش همه مدل‌ها، پیدا کرد. الگوریتم فضای مدل را به عنوان یک درخت در نظر می‌گیرد که مدل کامل در نوک است. با استفاده از این استراتژی، الگوریتم leaps and bounds شاخه بزرگی از فضای مدل را تکه‌تکه می‌کند و به‌طور سریع بهترین مدل از هر اندازه مدل را، جستجو می‌کند. الگوریتم leaps and bounds به دنبال یافتن بهترین مدل‌های تکی می‌باشد این الگوریتم مانند روش‌های گام‌به‌گام عدم قطعیت مدل را در نظر نمی‌گیرد و بهترین مدل از هر اندازه‌ای را فراخوانی می‌کند. این روش توسط مدیگان و رفتی پیشنهاد شد و بر اساس متوسط‌گیری بر روی یک مجموعه از مدل‌هایی است که به وسیله داده‌ها حمایت می‌شوند [۲۳].

سپس مدل‌های متعلق به leaps and bounds دوباره توسط occam's window کاهش می‌یابد. در occam's window هر مدلی که احتمال پسین بسیار کم‌تری از بهترین مدل دارد، حذف می‌شود. بنابراین اگر یک مدل، قابلیت پیش‌بینی بسیار کم‌تری نسبت به مدل با بالاترین احتمال پسین داشته باشد بی‌اعتبار است و نباید مورد بررسی قرار گیرد. بنابراین مدل‌هایی که متعلق به مجموعه

$$A = \{M_k: \frac{\max_L \{P(M_k|D)\}}{P(M_k|D)} \leq c\} \quad (2)$$

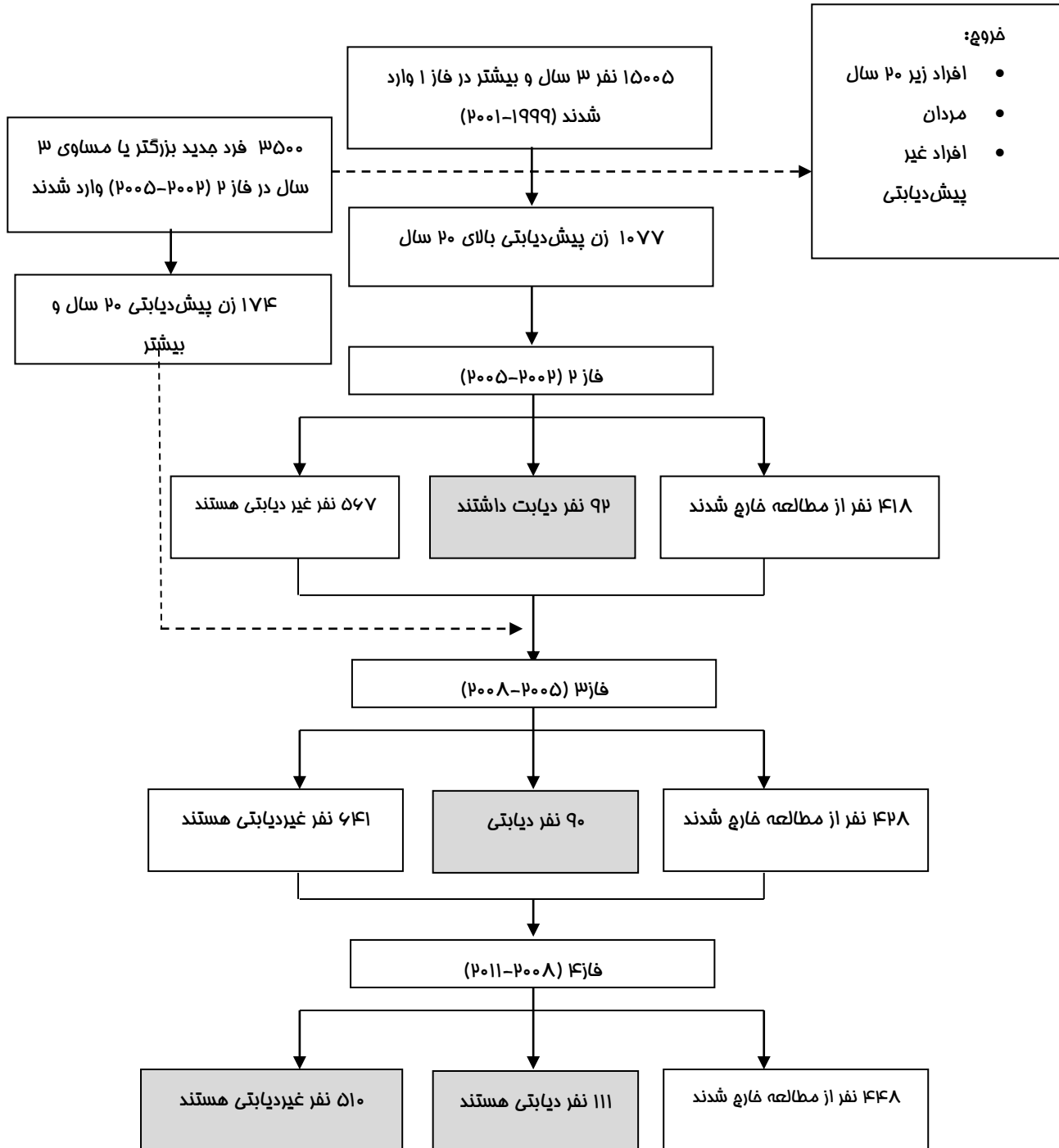
نیستند، از معادله (۱) خارج شد. C عدد ثابتی است که به وسیله تحلیل‌گر داده‌ها انتخاب می‌شود که وابسته به زمینه مورد بررسی است. مدیگان و رفتی در سال ۱۹۹۴ آن را برابر ۲۰ در نظر گرفتند [۲۳]. محاسبه احتمال پسین برای مدل  $M_k$  به صورت معادله (۳) است با این فرض که یکی از مدل‌های در نظر گرفته شده مدل درست است:

$$\Pr(M_k|D) = \frac{\Pr(D|M_k) \Pr(M_k)}{\sum_{j=1}^K \Pr(D|M_j) \Pr(M_j)} \quad (3)$$

در این رابطه  $\Pr(M_k)$  احتمال پیشین مدل است که یک‌نواخت در نظر گرفته شد، یعنی هیچ مدلی بر دیگری برتری ندارد. احتمال پسین مدل با استفاده از BIC تقریب زده شد [۲۷]. کس و واسرمن نشان داده‌اند که استفاده از BIC

مجموعه داده آزمون قرار گرفتند. با استفاده از داده ساخت، هر یک از روش‌های BMA و گام‌به‌گام اجرا شد، سپس با داده آزمون مدل ارزیابی شد.

نمونه کل در مجموعه داده‌ها به طور تصادفی به دو قسمت ۷۰ درصد (داده آزمون) تقسیم شدند. اطلاعات ۵۱۳ نفر در مجموعه داده ساخت و ۲۲۱ نفر در



شکل ۱- فلوجارت افراد مورد بررسی در مطالعه

تعداد را به ۸۳ مدل کاهش داد. از بین این مدل‌ها تعداد ۵ مدل برتر با بالاترین احتمال پسین در جدول ۲ نمایش داده شده است که در مجموع ۲۴/۹۶ درصد از احتمال پسین کل را به خود اختصاص داده‌اند. در این بین، مدل ۱ با بالاترین احتمال پسین تنها ۷/۴ درصد از احتمال پسین کل را بیان می‌کند و این نشان از وجود عدم قطعیت مدل می‌باشد. در این جدول، نیکویی برازش با استفاده از deviance برای ۵ مدل برتر در روش متوسط بیزی و مدل‌های انتخاب شده توسط روش گام‌به‌گام، محاسبه شد. مقدار deviance کم‌تر بهتر بودن برازش را نشان می‌دهد.

جدول ۳ میانگین و انحراف استاندارد پسین بیزی هم‌چنین احتمال اثر پسین برای هر یک از متغیرها را نشان می‌دهد. نتایج روش متوسط بیزی نشان می‌دهد که به عنوان مثال، میانگین پسین بیزی برای فشارخون دیاستولیک ۰/۰۰۳۷ است، یعنی افزایش ۱۰ میلی‌متر جیوه در فشارخون دیاستولیک در شروع مطالعه منجر به افزایش تقریباً ۴ درصدی (۱ - (۰/۰۰۳۷ × ۱۰) exp = ۰/۰۳۷) در بخت بروز دیابت می‌شود.

احتمال اثر پسین بیش از ۰/۵ در جدول ۳ نشان می‌دهد که متغیر بر بروز دیابت موثر است، بنابراین روش متوسط بیزی متغیرهای سابقه خانوادگی دیابت، قندخون ناشتا، قندخون ۲ ساعته، مصرف آسپرین و نمایه توده بدنی را به عنوان متغیرهای موثر بر بروز دیابت تشخیص داد. در روش متوسط بیزی متغیرهای مصرف سیگار در گذشته، نسبت تری‌گلیسرید به HDL، نسبت کلسترول به HDL، سابقه بیماری قلبی، سابقه بستری در بیمارستان، کراتینین، مصرف سیگار و مداخله برای کنترل فشارخون احتمال اثر پسین صفر دارند بنابراین در جدول ۳ نیامده‌اند.

روش گام‌به‌گام رو به جلو علاوه بر متغیرهایی که در روش متوسط بیزی انتخاب شد، متغیرهای فشارخون دیاستولیک، تحصیلات، مصرف داروی تیروئید و سابقه بستری در ۳ ماه گذشته را انتخاب کرد. روش گام‌به‌گام رو به عقب

مقدار احتمال رخداد دیابت تحت دو مدل پیش‌بینی شد و برای مقایسه درستی مدل‌ها، در پیش‌بینی بروز دیابت، سطح زیر منحنی مشخصه عمل‌کرد (AUC) برای هر کدام از آن‌ها محاسبه شد. برای مقایسه AUC در هر یک از روش‌ها از آزمون دلونگ استفاده شد، هم‌چنین از آزمون هاسمر-لمشو برای مناسبت مدل استفاده شد. برای مقایسه روش‌های گام‌به‌گام و متوسط بیزی در رگرسیون لجستیک از نمره بریر، نیز استفاده شد. این شاخص، ملاکی برای ارزیابی عمل‌کرد پیش‌بینی رویداد است. این معیار میانگین مجموع توان دوم واقعیت منهای پیش‌بینی است که به صورت رابطه (۷) است:

$$B(r, i) = \frac{1}{N} * \sum_{j=1}^G (y_j - r_j)^2 \quad (7)$$

وقتی رویداد رخ دهد  $y_j=1$  و در غیر این صورت  $y_j=0$  است،  $r_j$  مقدار پیش‌بینی رویداد برای تحلیل‌های آماری است. نمره بریر کوچک‌تر نشان‌دهنده بهتر بودن روش می‌باشد [۳۰]. از بسته BMA در نرم‌افزار R نسخه ۳،۱،۳ استفاده شد.

## نتایج

۷۳۴ زن ۲۰ سال و بالاتر پیش‌دیابتی وارد مطالعه شدند که در طول مدت پیگیری ۲۵۷ نفر (۳۵ درصد) به دیابت مبتلا شدند، ویژگی‌های پایه این افراد به تفکیک بروز دیابت در جدول ۱ آمده است. مقایسه عوامل خطر ساز مرتبط با دیابت در دو گروه دیابتی و غیر دیابتی نشان داد که همه عوامل خطر ساز غیر از کراتینین، وضعیت تاهل، سابقه بستری در بیمارستان، سابقه فشارخون بارداری، سابقه بستری ۳ ماه گذشته، استفاده از داروهای تیروئید، استفاده از داروهای مسدودکننده کانالهای آنزیمی، داشتن مداخله برای فشارخون، سیگاری بودن در گذشته و سیگار کشیدن فعلی بر بروز دیابت تأثیر معنی‌داری داشتند.

در این مطالعه ۲۹ متغیر مورد بررسی قرار گرفت بنابراین <sup>۲۹</sup> ۲ مدل ممکن وجود دارد که می‌تواند شامل هر یک از متغیرها باشد که از این میان مدل متوسط بیزی با استفاده از الگوریتم Leaps and bound و روش Occam's Window این

متغیرهای متفاوتی نسبت به روش گام به گام رو به جلو انتخاب کرد که در جدول ۳ نشان داده شده است.

جدول ۱. مشخصات عوامل خطر ساز دیابت در زنان پیش‌دیابتی بالای ۲۰ سال مطالعه TLGS در شروع مطالعه به تفکیک موارد جدید دیابتی و غیر دیابتی

P-value	غیردیابتی (تعداد=۴۷۷)	دیابتی (تعداد=۲۵۷)	متغیرهای کمی
<0.001	۱۳۷(۲۶,۵)	*۱۵۱,۱(۲۵,۸)	قند ۲ ساعته (میلی‌گرم بر دسی‌لیتر)
<0.001	۹۷,۶(۹,۱)	۱۰۳,۸(۱۰,۳)	قند ناشتا (میلی‌گرم بر دسی‌لیتر)
0.005	۲۲۰,۱(۴۲,۴)	۲۲۹,۴(۴۴)	کلسترول تام (میلی‌گرم بر دسی‌لیتر)
0.002	۴,۲(۲,۴)	۴,۸(۲,۵)	نسبت تری‌گلیسرید به HDL
0.001	۵,۱(۱,۴)	۵,۵(۱,۴)	نسبت کلسترول به HDL
ns	۰,۹۷(۰,۱۱)	۰,۹۸(۰,۱۱)	کراتینین (میلی‌گرم بر دسی‌لیتر)
0.001	۴۱,۸(۱۳,۴)	۴۵,۳(۱۴)	فشار نبض (میلی‌متر جیوه)
<0.001	۷۹,۸(۹,۹)	۸۲,۶(۹,۹)	فشارخون دیاستولیک (میلی‌متر جیوه)
<0.001	۰,۵۸(۰,۰۷)	۰,۶۲(۰,۰۶)	دور کمر به قد (سانتی متر)
<0.001	۲۸,۶(۴,۱)	۳۰,۸(۴,۴)	نمایه توده بدنی (کیلوگرم بر مجذور قد)
<0.001	۱۶,۲(۰,۹۸)	۱۶,۶(۰,۹۳)	دور مچ دست (سانتی متر)
<0.001	۴۴,۷(۱۱,۷)	۴۸,۱(۱۰,۵)	سن (سال)
متغیرهای کیفی			
<0.001	۱۳۵(۲۸,۳)	**۱۱۳(۴۴)	سابقه خانوادگی دیابت
<0.001	۱۱۹(۲۴,۹)	۹۹(۳۸,۵)	سابقه هایپرلیپیدمی
ns	۴۲۰(۸۸,۱)	۲۲۹(۸۹,۱)	سابقه بستری در بیمارستان
ns	۱۰(۲,۱)	۴(۱,۶)	سابقه بستری در ۳ ماه گذشته
ns	۴۰(۸,۴)	۲۱(۸,۲)	سابقه فشارخون بارداری
0.009	۴۲(۸,۸)	۳۹(۱۵,۲)	سابقه نوزاد بیش از ۴,۵
ns	۵۲(۱۰,۹)	۲۷(۱۰,۵)	سابقه بیماری‌های قلبی
ns	۹۷(۲۰,۳)	۵۲(۲۰,۲)	داشتن گواتر
0.003	۱۲(۲,۵)	۱۸(۷)	استفاده از داروهای ادرار آور
0.050	۲۲(۴,۶)	۲۱(۸,۲)	استفاده از داروهای تیروئید
ns	۱۰(۲,۱)	۶(۲,۳)	استفاده از داروهای مسدودکننده کانال‌های آنزیمی
0.003	۳۶(۷,۵)	۳۷(۱۴,۴)	مصرف آسپرین
ns	۹(۱,۹)	۴(۱,۶)	مصرف سیگار در گذشته
ns	۱۱(۲,۳)	۷(۲,۷)	مصرف سیگار
<0.001	۲۰۴(۴۲,۸)	۱۴۷(۵۷,۲)	تحصیلات کمتر از دبیرستان
ns	۵۷(۱۱,۹)	۳۴(۱۳,۲)	مجرد بودن
ns	۴۴(۹,۲)	۳۱(۱۲,۱)	داشتن مداخله برای فشارخون
<0.001	3315.9(676.4)	(1029.8)۱۷۲۵	زمان پیگیری تا اولین بروز دیابت

\* داده‌ها برای متغیرهای کمی به صورت میانگین (انحراف استاندارد) گزارش شده‌اند

\*\* داده‌ها برای متغیرهای کیفی به صورت تعداد (درصد) گزارش شده‌اند

ns: غیرمعنی دار

جدول ۲. مقایسه مدل‌های بدست آمده توسط روش‌های گام‌به‌گام و روش متوسط بیزی در بروز دیابت روی زنان پیش‌دیابتی مطالعه TLGS

متغیر	مدل متوسط بیزی					گام‌به‌گام روبه جلو	گام‌به‌گام روبه عقب
	۱	۲	۳	۴	۵		
قند ۲ ساعته	T	T	T	T	T	T	T
قند ناشتا	T	T	T	T	T	T	T
فشارخون دیاستولیک							
مصرف آسپرین	T	T		T	T	T	T
سابقه بستری ۳ ماه گذشته							
استفاده از دارو تیروئید							
سابقه خانوادگی دیابت	T		T	T	T	T	T
سابقه فشارخون بارداری							
دور کمر به قد			T				
نمایه توده بدنی	T	T	T			T	T
دور مچ دست							
تحصیلات			T			T	T
تعداد متغیرها	۵	۴	۵	۵	۶	۹	۱۰
Deviance	۷۶۶٫۸۷	۷۷۴٫۳۶	۷۶۷٫۹۲	۷۶۷٫۹۷	۷۶۱٫۴۱	۷۵۱٫۲۵	۷۴۹٫۲۲
احتمال پسین مدل	۰٫۰۷۴	۰٫۰۴۷	۰٫۰۴۴	۰٫۰۴۳	۰٫۰۴۲		

جدول ۳. نتایج روش‌های گام‌به‌گام و متوسط بیزی در بروز دیابت در زنان پیش‌دیابتی مطالعه TLGS

متغیر	مدل متوسط بیزی				گام‌به‌گام روبه جلو	گام‌به‌گام روبه عقب
	میانگین †	انحراف استاندارد پسین	احتمال اثر پسین	ضریب (خطای استاندارد)		
قند ۲ ساعته	۰٫۰۲۶۸	۰٫۰۰۳	۱	۰٫۰۲۶۳(۰٫۰۰۳)	۰٫۰۲۶۱(۰٫۰۰۳)	
قند ناشتا	۰٫۰۸۱۴	۰٫۰۱	۱	۰٫۰۸۱(۰٫۰۱)	۰٫۰۸۱(۰٫۰۱)	
سابقه خانوادگی دیابت	۰٫۴۲۱۸	۰٫۲۸۱	۰٫۷۷۴	۰٫۶۱۴(۰٫۱۸۹)	۰٫۶۳۵(۰٫۱۹)	
مصرف آسپرین	۰٫۵۴۵۱	۰٫۴۴۷	۰٫۶۷۴	۰٫۸۵(۰٫۲۸۸)	۰٫۸۰۸(۰٫۲۹۲)	
نمایه توده بدنی	۰٫۰۵۷۶	۰٫۰۴۸	۰٫۶۲۴	۰٫۰۹(۰٫۰۲۱)	۰٫۰۹(۰٫۰۲۱)	
دور کمر به قد	۱٫۸۵۶	۲٫۷۱۱	۰٫۳۵۴	۴٫۲۰۱(۱٫۵۲۸)	۴٫۲۰۱(۱٫۵۲۸)	
دور مچ دست	۰٫۰۶۴۱	۰٫۱۲۸	۰٫۲۳۸	۰٫۲۶۱(۰٫۱۰۵)	۰٫۲۶۱(۰٫۱۰۵)	
فشارخون دیاستولیک	۰٫۰۰۳۷	۰٫۰۰۸	۰٫۱۸۶	۰٫۰۲(۰٫۰۰۹)	۰٫۰۱۹(۰٫۰۰۹)	
تحصیلات	۰٫۰۸۰۲	۰٫۱۸۹	۰٫۱۸۱	۰٫۳۹(۰٫۱۸۶)	۰٫۳۹(۰٫۱۸۶)	
سن	۰٫۰۰۰۹	۰٫۰۰۴	۰٫۰۵۷			
استفاده از داروی تیروئید	۰٫۰۳۵۶	۰٫۱۷۳	۰٫۰۵۵	۰٫۷۱۲(۰٫۳۸۸)	۰٫۶۰۸(۰٫۳۸۳)	
سابقه بستری ۳ ماه گذشته	-۰٫۰۶۵۱	-۰٫۳۲۵	۰٫۰۵۲	-۱٫۱۸۹(۰٫۷۳۱)	-۱٫۳۲۵(۰٫۷۴۳)	
گواتر	۰٫۰۱۰۲	۰٫۰۰۷	۰٫۰۰۳			
سابقه هایپرلیپیدمی	۰٫۰۰۵	۰٫۰۴۵	۰٫۰۱۹			
تاهل	۰٫۰۰۴۹	۰٫۰۵۲	۰٫۰۱۴			
استفاده از دارو ادرار آور	۰٫۰۰۶۱	۰٫۰۷۵	۰٫۰۱			
استفاده از داروهای مسدودکننده کانال‌های آنزیمی	-۰٫۰۰۷۸	۰٫۱۰۱	۰٫۰۱			
سابقه نوزاد بیش از ۴٫۵	۰٫۰۰۲۸	۰٫۰۳۹	۰٫۰۰۹			
فشارنبض	۰٫۰۰۰۰۴	۰٫۰۰۰۰۷	۰٫۰۰۰۶			
سابقه فشارخون بارداری	-۰٫۰۰۱۹	۰٫۰۳۶	۰٫۰۰۵	-۰٫۴۸۶(۰٫۳۴۴)		
کلسترول تام	۰٫۰۰۰۰۰۸	۰٫۰۰۰۰۱	۰٫۰۰۰۴			

† میانگین‌های پسین، متوسط‌گیری شده روی همه مدل‌های انتخاب شده BMA هستند



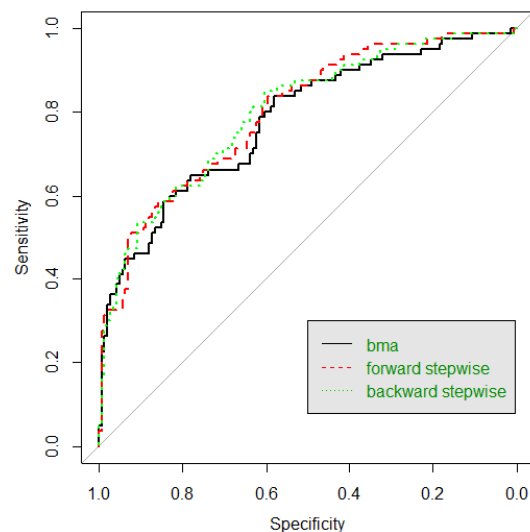
## بحث و نتیجه گیری

از میان متغیرهایی که احتمال می‌رفت با بروز دیابت مرتبط باشند، ۲۹ متغیر انتخاب شدند و تحلیل آماری بر روی این متغیرها انجام شد. نتایج این مطالعه نشان می‌دهد که در روش متوسط بیزی فقط ۵ متغیر در مدل انتخاب شده است ولی روش گام‌به‌گام ۹ متغیر را انتخاب کرد با این حال AUC روش BMA تفاوت معنی‌داری با روش‌های گام‌به‌گام نداشت. از مزیت روش متوسط بیزی نسبت به روش گام‌به‌گام این است که با متوسط‌گیری روی مدل حمایت شده به وسیله داده‌ها می‌تواند بر ضعف مدل‌های تکی غلبه کند و به‌طور مستقیم عدم قطعیت مدل را در نظر داشته باشد، بنابراین انتظار می‌رود که مدل متوسط بیزی قابل اعتمادتر و دقیق‌تر باشد [۱۸، ۲۷]. روش چندمتغیره متوسط بیزی، اثر همه متغیرها را تعدیل می‌کند و اثر همه متغیرها دیده می‌شود [۳۱]، در حالی که در روش گام‌به‌گام تنها اثر متغیرهای موجود در مدل را می‌توان دید. در روش متوسط بیزی برای انتخاب متغیر از احتمال اثرگذاری پسین متغیر استفاده می‌شود که مقداری پیوسته بین صفر تا یک است بنابراین به دلیل ماهیت پیوسته‌اش شدت اثرگذاری را نشان می‌دهد [۳۳، ۳۲]. اما روش گام‌به‌گام تصمیمی دو حالتی از انتخاب یا عدم انتخاب متغیر می‌دهد [۱۴]. الگوریتم Leaps and bounds در روش متوسط بیزی برای بیش از ۳۰ متغیر عمل‌کرد خوبی ندارد و این یکی از محدودیت‌های این روش می‌باشد.

ملاحظه پنج مدل دارای بالاترین احتمال پسین نشان می‌دهد که هر مدل، احتمال پسین کوچکی دارد و هر یک، متغیرهای متفاوتی را انتخاب می‌کنند که دلیلی برای وجود عدم قطعیت مدل است. بنابراین نمی‌توان تنها به یک مدل تکی اکتفا کرد، در حالی که روش گام‌به‌گام منجر به انتخاب یک مدل تکی شد و به همین جهت عدم قطعیت مدل را نادیده گرفت [۳۴].

آنالیز BMA در رگرسیون لجستیک نشان می‌دهد که دلایل قانع‌کننده‌ای وجود ندارد که دور کمر به قدم موثر باشد. با این وجود، دلایل قوی که پارامتر صفر باشد، نیز موجود

روش‌های گام‌به‌گام رو به جلو و رو به عقب شاخص‌های متفاوت تن‌سنجی را انتخاب کردند. سپس برای مقایسه روش‌ها، هر یک از روش‌های گام‌به‌گام و متوسط بیزی بر روی ۵۱۳ مجموعه داده ساخت به کار برده شدند و ۲۲۱ مجموعه داده آزمون را پیش‌بینی کردند. میزان پیش‌بینی بروز دیابت با استفاده از منحنی راک انجام شد که مقدار سطح زیر منحنی راک در روش متوسط بیزی  $0.7793$ ، روش گام‌به‌گام رو به عقب  $0.7981$  و روش گام‌به‌گام رو به جلو  $0.7968$  بود. نمودار ROC برای هر ۳ روش در شکل ۲ آمده است. مقایسه AUC با استفاده از آزمون دلونگ در ۳ روش نشان داد که پیش‌بینی‌ها تفاوت معنی‌داری با هم ندارند ( $P\text{-value} > 0.05$ )، سپس مناسبیت مدل برای همه روش‌ها با استفاده از آزمون هاسمر - لمشو بررسی شد که نشان‌دهنده مناسب بودن برازش است ( $P\text{-value} > 0.05$ ). نمره بریر روش متوسط بیزی  $0.1756$ ، روش گام‌به‌گام رو به جلو  $0.1711$  و گام‌به‌گام رو به عقب  $0.1696$  است. نمره بریر روش متوسط بیزی بیش‌تر از روش‌های گام‌به‌گام رو به عقب و گام‌به‌گام رو به جلو است اما اختلافات بسیار کوچک است و تاییدی بر نتایج سطح زیر منحنی راک است. بنابراین روش متوسط بیزی ضمن انتخاب متغیرهای کم‌تر به عنوان متغیر عامل خطر ساز بروز دیابت، قدرت پیش‌بینی را هم‌چنان حفظ می‌کند.



شکل ۲. نمودارهای ROC برای ارزیابی رگرسیون لجستیک در روش‌های گام‌به‌گام و متوسط بیزی برای بروز دیابت نوع ۲

مدل تکی انتخاب شده در وجود عدم قطعیت مدل دارد. هم‌چنین نشان دادند در مواقعی که چندین پیش‌گو موجود است اما ارتباطی بین آن‌ها و پاسخ وجود ندارد، روش‌های انتخاب متغیر استاندارد، بعضی از زیرمجموعه‌های متغیرها را انتخاب می‌کنند، به این پدیده غیرمتعارف پارادوکس فریدمن گویند. در این حالت، مدل متوسط بیزی با استفاده از روش Occam's Window، مدل صفر به تنهایی یا یک مجموعه کوچک شامل مدل صفر را انتخاب می‌کند، بنابراین تا حد زیادی این پارادوکس را حل می‌کند [۲۶].

به دلیل این‌که مدل متوسط بیزی تا کنون برای پیش‌بینی دیابت استفاده نشده است امکان مقایسه یافته‌های به دست آمده در مطالعه حاضر با مطالعه‌های مشابه وجود نداشت و تنها مقایسه‌ها با مطالعه‌هایی که از مدل‌های پیش‌بینی مرسوم مانند روش گام‌به‌گام استفاده کرده‌اند، انجام شد. هم‌چنین مطالعات خیلی کمی یافت شد که روی زنان پیش‌دیابتی عوامل خطر ساز بروز دیابت را بررسی کنند ولی در مطالعه‌هایی که به بررسی قواعد تصمیم‌گیری در مورد دیابت پرداخته‌اند عوامل خطر ساز مشابه با پژوهش حاضر به دست آمده است.

بر اساس مطالعه‌ای که باقرزاده و همکاران انجام داده‌اند متغیرهای قندخون ناشتا، قندخون ۲ ساعته، سابقه خانوادگی، نمایه توده بدنی و مصرف آسپرین اثرگذارترین متغیرها هستند که با نتیجه مطالعه ما همخوانی دارد [۳۵]. هم‌چنین در مطالعه‌ای توسط محمدی و همکاران سابقه فامیلی و چاقی را از عوامل خطر دیابت نوع ۲ معرفی کردند، در این مطالعه نیز نمایه توده بدنی و سابقه فامیلی به عنوان عوامل خطر برای رخداد دیابت انتخاب شده است [۳۷]. در مطالعه‌ای توسط محرابی و همکاران، با استفاده از رگرسیون لجستیک منطقی، قند ناشتا، قند ۲ ساعته، سابقه خانوادگی دیابت و دورکم بالا، موثر بر بروز دیابت تشخیص داده شد که با نتیجه مطالعه ما همخوانی دارد [۳۸].

در مطالعه انجام شده توسط کیخا و همکاران نشان داده شد که سن و چاقی در ارتباط با پیش‌دیابت هستند اما در

نیست؛ در حالی که روش گام‌به‌گام نمی‌تواند چنین بیانی داشته باشد. مقدار احتمال اثر پسین دور کم به قد نشان می‌دهد که در حضور نمایه توده بدنی، این متغیر به عنوان عامل خطر ساز دیابت شناخته نمی‌شود. اما مقدار احتمال اثر پسین دور کم به قد، صفر نشده، چون می‌تواند در غیاب نمایه توده بدنی، متغیر موثری باشد [۱۸].

در این مطالعه، هر یک از روش‌های گام‌به‌گام منجر به مدل‌های با متغیرهای متفاوت می‌شوند و این باعث سردرگمی در انتخاب متغیرهای موثر می‌شود. مقایسه مدل‌ها نشان می‌دهد که در روش گام‌به‌گام رو به عقب متغیرهای دور میچ دست، دور کم به قد و فشار خون بارداری انتخاب می‌شوند ولی در روش گام‌به‌گام رو به جلو متغیرهای نمایه توده بدنی و تحصیلات انتخاب می‌شوند، در حالی که روش متوسط بیزی هیچ‌یک از این متغیرها را مرتبط با بروز دیابت انتخاب نکرده است.

یافته‌های روش متوسط بیزی نشان می‌دهد که در این روش، از میان متغیرهای تن‌سنجی، نمایه توده بدنی بر بروز دیابت موثر است اما دور میچ دست و دور کم به قد، علی‌رغم این‌که صفر نیستند؛ موثر نمی‌باشند. مقدار احتمال پسین این دو متغیر نشان می‌دهد که این دو متغیر ممکن است بر بروز دیابت اثرگذار باشند اما در حضور متغیر نمایه توده بدنی انتخاب نمی‌شوند.

متغیرهای غیر مرتبط با دیابت نه تنها به عمل‌کرد مدل کمک نمی‌کنند، حتی ممکن است اثرات بدی بر مدل داشته باشند [۳۵]. فلاک و چانگ در سال ۱۹۸۷ با شبیه‌سازی‌هایی نشان دادند که روش‌های گام‌به‌گام گرایش به یافتن متغیرهای غیرمرتبط با پاسخ دارد [۳۶]. در مطالعه حاضر روش متوسط بیزی استفاده از داروی تیروئید، سابقه فشارخون بارداری، سابقه بستری ۳ ماه گذشته، تحصیلات و دور میچ دست را در ارتباط با بروز دیابت تشخیص نداد.

رفتری و همکاران در سال ۱۹۹۳ مدل متوسط بیزی را برای مدل‌های رگرسیون خطی به کار گرفتند و نشان دادند که تکنیک مدل متوسط بیزی عمل‌کرد پیش‌گویی بهتری از هر

## منابع

- [1] Rathmann W, Giani G. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 2004; 27: 2568-2569.
- [2] Azimi-Nezhad M, Ghayour-Mobarhan M, Parizadeh M, Safarian M, Esmaeili H, Parizadeh S, et al. Prevalence of type 2 diabetes mellitus in Iran and its relationship with gender, urbanisation, education, marital status and occupation. *Singapore Med J* 2008; 49: 571.
- [3] Hajihassani A, Bahrpeyma F, Bakhtiari A. Effects of eccentric and concentric exercises on postural sway in type 2 diabetic patients. *Koomesh* 2016; 17: 493-500 (Persian).
- [4] Masaeli N, Attari A, Molavi H, Najafi M, Siavash M. Normative data and psychometric properties of the quality of life questionnaire for patients with diabetes mellitus. *Koomesh* 2010; 11: 263-269 (Persian).
- [5] Roche MM, Wang PP. Factors associated with a diabetes diagnosis and late diabetes diagnosis for males and females. *J Clin Translat Endocrinol* 2014; 1: 77-84.
- [6] Esmail-Nasab N, Abdolrahimzadeh A, Ebrahimi A. The cross-sectional study of effective factors on type 2 diabetes control in a diabetes care center in Sanandaj. *Iran Epidemiol J* 2010; 6: 39-45 (Persian).
- [7] Sajadi F. The prevalence of type 2 diabetes and its relation with CVD risk factors in the population of Isfahan. *Med J Mashhad Univ* 2003; 46: 68-73 (Persian).
- [8] Navaei L. Study of the prevalence of diabetes and impaired glucose tolerance in rural areas of Tehran province. 2001; 4: 92-99. (Persian).
- [9] Mansoori F, Namdaritabar H, Shahrezaee A, Rezaei R, Alikhani A, Montazer MJ, Gazerpoor F. Diabetes mellitus in over-thirty-year-old individuals in Kermanshah province (2002). *J Kermanshah Univ Med Sci* 2004; 8. (Persian).
- [10] Barnett KN, Ogston SA, McMurdo ME, Morris AD, Evans J. A 12- year follow- up study of all- cause and cardiovascular mortality among 10 532 people newly diagnosed with Type 2 diabetes in Tayside, Scotland. *Diabeti Med* 2010; 27: 1124-1129.
- [11] Fahimfar N, Khalili D, Mohebi R, Azizi F, Hadaegh F. Risk factors for ischemic stroke; results from 9 years of follow-up in a population based cohort of Iran. *BMC Neurol* 2012; 12: 117.
- [12] Annett A, Bumgarner RE, Raftery AE, Yeung KY. Iterative bayesian model averaging: A method for the application of survival analysis to high-dimensional microarray data. *BMC Bioinformatics* 2009; 10: 72.
- [13] Noble Jr RB. Multivariate applications of Bayesian model averaging [dissertation]. Citeseer; 2000.
- [14] Wiegand RE. Performance of using multiple stepwise algorithms for variable selection. *Stat Med* 2010; 29: 1647-1659.
- [15] Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. *Stat Sci* 1999; 14: 382-401.
- [16] Hoeting J, Raftery AE, Madigan D. A method for simultaneous variable selection and outlier identification in linear regression. *Computational Statistics & Data Analysis*. 1996; 22(3): 251-270.
- [17] Genell A, Nemes S, Steineck G, Dickman PW. Model selection in medical research: a simulation study comparing Bayesian model averaging and stepwise regression. *BMC Med Res Methodol* 2010; 10: 108.
- [18] Wang D, Zhang W, Bakhai A. Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. *Stat Med* 2004; 23: 3451-3467.

مطالعه ما سن در هیچ کدام از روش‌ها، به عنوان متغیر موثر بر بروز دیابت انتخاب نشد. این بدان جهت است که در مطالعه آن‌ها مقادیر پایه برای میزان قندخون ناشتا و قند خون ۲ ساعته بررسی نشده است، بنابراین در حضور متغیرهای مهم‌تری چون قندخون ۲ ساعته و قندخون ناشتا متغیر سن اثرگذار تشخیص داده نمی‌شود [۳۹].

مطالعه انجام شده توسط سوری و همکاران نشان داد که ژن MTNR1B با میزان ابتلا به دیابت نوع ۲ مرتبط است و افزایش بیان آن خطر دیابت نوع ۲ را افزایش می‌دهد [۴۰]. این مطالعه نشان می‌دهد که علاوه بر متغیرهایی که در این مطالعه بررسی شد، عوامل خطر ساز دیگری نیز برای بروز دیابت وجود دارد. بنابراین مطالعات بیشتر در این زمینه مورد نیاز است.

در مطالعه حاضر، روش متوسط بیزی متغیرهای کم‌تری در مقایسه با روش‌های گام‌به‌گام انتخاب کرد و توان پیش‌بینی‌اش نیز کاهش معنی‌داری نیافت، بنابراین روش بهتری برای ارزیابی عوامل خطر ساز دیابت است. با تامل در متغیرهایی که به عنوان عوامل خطر دیابت شناخته شدند، نتیجه می‌شود که تغییرات شیوه زندگی مربوط به چاقی، رفتارهای تغذیه‌ای و فعالیت فیزیکی نقش اصلی را در پیشگیری و درمان دیابت نوع ۲ بازی می‌کند. هم‌چنین عامل سابقه خانوادگی دیابت علی‌رغم این‌که قابل پیشگیری نیست اما به عنوان هشدار برای افراد تلقی می‌شود.

## تشکر و قدردانی

در این مطالعه، از داده‌های طرح قند و لیپید تهران که توسط پژوهشکده علوم غدد درون‌ریز و متابولیسم دانشگاه علوم پزشکی و خدمات بهداشتی- درمانی شهید بهشتی اجرا شده است، استفاده شد. نویسندگان بر خود لازم می‌دانند از کسانی که در طراحی و جمع‌آوری داده‌های طرح TLGS مشارکت داشتند، قدردانی نمایند. این پژوهش از پایان‌نامه کارشناسی ارشد آمار زیستی استخراج شده است.

Berger, AP Dawid and AFM Smith. Oxford: University Press; 1999; 157-185.

[32] TAY PL. Iterative Bayesian Model Averaging For Patients Survival Analysis [dissertation]. Universiti Teknologi Malaysia; 2010.

[33] Lipkovich IA. Bayesian model averaging and variable selection in multivariate ecological models [dissertation]. Virginia Tech Univ; 2002.

[34] Montgomery JM, Nyhan B. Bayesian model averaging: Theoretical developments and practical applications. *Politic Anal* 2010; 18: 245-270.

[35] Bagherzadeh-Khiabani F, Ramezankhani A, Azizi F, Hadaegh F, Steyerberg EW, Khalili D. A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results. *J Clin Epidemiol* 2016; 71: 76-85.

[36] Flack VF, Chang PC. Frequency of selecting noise variables in subset regression analysis: a simulation study. *Am Stat* 1987; 41: 84-86.

[37] Mohamadi SM, Rashidi M, Afkhami Ardakani M. Risk factors for type 2 diabetes. *J Shahid Sadoughi Unive Med Sci* 2011; 19: 266-280 (Persian).

[38] Mehrabi Y, Khadem-Maboudi A, Hadaegh F, Sarbakhsh P. Prediction of diabetes using logic regression. *Iran J Endocrinol Metab* 2010; 12: 16-24 (Persian).

[39] Kaykha M, Ghorbani MJ, Amini M. The prevalence of type 2 diabetes, pre-diabetes and metabolic syndrome risk factors in first-degree relatives of patients with type 2 diabetes. *J Kerman Univ Med Sci* 2013; 20: 115-128 (Persian).

[40] Soori R, Rashidi M, Choobineh S, Ravasi AA, Baesi K, Rashidy-Pour A. Effects of 12 weeks resistant training on MTNR1B gene expression in the pancreas and glucose and insulin levels in type 2 diabetic rats. *Koomesh* 2017; 19: 46-55 (Persian).

[19] Azizi F, Madjid M, Rahmani M, Emami H, Mirmiran P, Hadjipour R. Tehran lipid and glucose study (TLGS): rationale and design. *Iran J Endoc Metab* 2000; 2: 77-86 (Persian).

[20] Stekhoven DJ. MissForest: nonparametric missing value imputation using random forest. *Astrophysics Source Code Library* 2015; 1: 05011.

[21] Filzmoser P, Gschwandtner M, Filzmoser MP, LazyData T. Package 'mvoutlier'; 2015.

[22] Leamer EE. Regression selection strategies and revealed priors. *J Am Stat Assoc* 1978; 73: 580-587.

[23] Madigan D, Raftery AE. Model selection and accounting for model uncertainty in graphical models using Occam's window. *J Am Stat Assoc* 1994; 89: 1535-1546.

[24] Raftery AE. Bayesian model selection in social research. *Soc method* 1995; 25: 111-163.

[25] Furnival GM, Wilson RW. Regressions by leaps and bounds. *Technometrics* 1974; 16: 499-511.

[26] Raftery A, Hoeting J, Madigan D. Model selection and accounting for model uncertainty in linear regression models: Citeseer; 1993.

[27] Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995; 90: 773-795.

[28] Kass RE, Wasserman L. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *J Am Stat Assoc* 1995; 90: 928-934.

[29] Jeffreys H. The theory of probability: OUP Oxford; 1998.

[30] Thompson WR. Variable selection of correlated predictors in logistic regression: investigating the diet-heart hypothesis [dissertation]. Florida State Univ; 2009.

[31] Clyde M. Bayesian model averaging and model search strategies. *Bayesian Statistics 6 JM Bernardo, JO*

## Factors associated with incidence of type II diabetes in pre-diabetic women using Bayesian Model Averaging

Maryam Mahdavi<sup>1</sup> (M.Sc), Yadollah Mehrabi<sup>2\*</sup> (Ph.D), Davood Khalili<sup>3</sup> (Ph.D), Ahmad Reza Baghestani<sup>1</sup> (Ph.D), Farideh Bagherzadeh-Khiabani<sup>3</sup> (M.Sc), Samaneh Mansouri<sup>1</sup> (M.Sc)

1- Department of Biostatistics, School of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2- Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

3- Prevention of Metabolic Disorders Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

(Received: 25 Jul 2016; Accepted: 8 Apr 2017)

**Introduction:** Diabetes is a chronic disease which usually begins with impaired glucose tolerance. This step is known as pre-diabetes. People with pre-diabetes are at greater risk for diabetes. Typically for the variable selection, stepwise approach is used which does not take into account model uncertainties. In this study, Bayesian Model Averaging (BMA) method was used to sort out the above shortcoming.

**Materials and Methods:** The study population was 734 pre-diabetic women with 20 years and older participated in Tehran Lipid and Glucose Study (TLGS). In this study, the stepwise and BMA variable selection methods were employed in logistic regression. Then area under curve (AUC) for both methods was computed and compared with Delong test. All analyses was done using R version 3.1.3.

**Results:** BMA selected the fasting plasma glucose, 2 hours' blood glucose, and family history of diabetes, body mass index and aspirin use at baseline as risk factors for diabetic. In addition to these factors, stepwise method selected diastolic blood pressure, history of past 3 months' hospitalization, thyroid drug use and education. Although the number of variables selected by BMA (5 variables) was less than that of stepwise (9 variables), AUC for the two methods was not significant.

**Conclusion:** It seems that the BMA provide better model for screening of diabetes because with selecting fewer variables, prediction ability of the model is preserved.

**Keywords:** Bayesian Model Averaging, Stepwise Method, Pre-diabetics Women, Logistic Regression, Tehran Lipid and Glucose Study

---

\* Corresponding author. Tel: +98 21 22432041

mehrabi@sbmu.ac.ir