مجله‌ی غدد درون‌ریز و متابولیسم ایران
دانشگاه علوم پزشکی و خدمات بهداشتی ـ درمانی شهید بهشتی
دوره‌ی بیست و یکم، شماره‌ی ۴، صفحه‌های ۲۰۲ ـ ۱۹۵ (مهر ـ آبان ۱۳۹۸)

مقاله‌ی پژوهشی

مقایسه روش‌های بیزی و فراوانی‌گرا در برآورد شاخص‌های بهبود بازطبقه‌بندی برای ارزیابی مدل‌های پیش‌بینی: مطالعه قند و لیپید تهران

صدیقه نریمان[۱]، دکتر داوود خلیلی[۲]، دکتر احمدرضا باغستانی[۱]، فرزانه احمدی[۱]، مریم مهدوی[۳]، دکتر یداله محرابی[۲،٤]

۱) گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران، ۲) مرکز تحقیقات پیشگیری از بیماری‌های متابولیک، پژوهشکده علوم غدد درون‌ریز و متابولیسم، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران، ۳) مرکز تحقیقات پیشگیری و درمان چاقی، پژوهشکده علوم غدد درون‌ریز و متابولیسم، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران،٤) گروه اپیدمیولوژی، دانشکده بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران. **نشانی مکاتبه نویسنده مسئول:** گروه اپیدمیولوژی، دانشکده بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران، دکتر یداله محرابی، e-mail: mehrabi@sbmu.ac.ir

چکیده

مقدمه: برای برآورد شاخص‌های بهبود بازطبقه‌بندی خالص (**NRI**) و تشخیصی ادغام شده (**IDI**) که برای برآورد مقدار تاثیر افزودن نشانگرهای جدید در مدل‌های آماری به‌کار می‌رود، به‌طور معمول از روش فراوانی‌گرا استفاده می‌شود. این روش در بعضی موارد به‌ویژه در نمونه‌های کوچک، عملکرد ضعیفی دارد. در این مطالعه کارایی دو روش بیزی و فراوانی‌گرا مقایسه شده است. **مواد و روش‌ها:** جمعیت بررسی شده، ۷۳۴ زن با سن ۲۰ سال و بالاتر پیش‌دیابتی، در مطالعه کوهورت قند و لیپید تهران (**TLGS**) بود. برای برآورد احتمال رخداد بیماری دیابت، مدل رگرسیون لوجستیک مشابه مدل پیش‌بینی بیماری دیابت در مطالعه **ARIC** استفاده شد که شامل نشانگرهای اندازه دور کمر، پرفشاری خون، قد، وزن، سن، سابقه خانوادگی دیابت، سیگاری بودن و ضربان نبض بود. روش‌های فراوانی‌گرا و بیزی برای برآورد شاخص‌های بهبود بازطبقه‌بندی خالص و تشخیصی ادغام شده، استفاده شد و سپس این دو روش در حجم نمونه کوچک مقایسه شدند. برای انجام تحلیل‌ها از نرم‌افزار **R** نسخه ۳،۱،۳ استفاده شد. **یافته‌ها:** برآورد شاخص‌های بهبود بازطبقه‌بندی خالص و تشخیصی ادغام شده به روش‌های بیزی و فراوانی‌گرا نتایج یکسانی داشتند که در هر دو روش افزودن نشانگرهای جدید اندازه دور مچ دست، سابقه تولد نوزاد بیش از ۴/۵ کیلوگرم، سابقه فشارخون بارداری به مدل **ARIC**، تاثیر معناداری در شاخص‌های فوق نداشت. در حجم نمونه کوچک، برآورد این دو شاخص در روش بیزی در مقایسه با روش فراوانی‌گرا، عملکرد بهتری نشان داد. **نتیجه‌گیری:** روش بیزی در حجم نمونه کوچک عملکرد معتبر و قابل اعتمادتری در مقایسه با روش فراوانی‌گرا داشت.

واژگان کلیدی: شاخص بهبود بازطبقه‌بندی خالص، شاخص تشخیصی ادغام شده، روش بیزی، دیابت، مدل پیش‌بینی

دریافت مقاله: ۹۸/٤/۳۰ـ دریافت اصلاحیه: ۹۸/۸/۸ ـ پذیرش مقاله: ۹۸/۹/۱٦

مقدمه

دیابت نوع ۲، بیماری مزمنی است که با اختلال تحمل گلوکز آغاز می‌شود. این مرحله، به عنوان پیش‌دیابت شناخته می‌شود.[۱] اگرچه میزان وقوع دیابت نوع ۱ و ۲ در سراسر جهان رو به افزایش است، ولی انتظار می‌رود که سرعت این افزایش در دیابت نوع ۲ بیشتر باشد.[۱،۲] شیوع این بیماری در کشورهای درحال توسعه از جمله ایران، در حال افزایش است.[۳] روچ[i] و همکاران در سال ۲۰۱٤ در مطالعه‌ای نشان دادند که عوامل خطرساز دیابت در زنان و مردان متفاوت است. با وجود این‌که شیوع دیابت در مردان بیشتر از زنان است اما زنان در معرض خطر بیشتری از مرگ‌ومیر هستند.[٤] بنابراین پیش‌بینی بیماری دیابت در مطالعه حاضر، محدود به زنان پیش‌دیابتی شد.

i - Roch

[Downloaded from ijem.sbmu.ac.ir on 2026-11-07]

[DOR: 20.1001.1.16834844.1398.21.4.1.7]

در علوم پزشکی و اپیدمیولوژی، پیش‌بینی بیماری‌ها زمینه مناسب برای کنترل یا پیش‌گیری از آن‌ها را مهیا می‌کند. از این رو در چند دهه اخیر مدل‌های پیش‌بینی بر اساس مدل‌های رگرسیونی به‌خصوص رگرسیون لوجستیک توسعه پیدا کرده است.[۵] یکی از روش‌هایی که برای ارزیابی توان پیش‌بینی مدل‌های صحت تشخیصی به کار گرفته می‌شود منحنی مشخصه عملکرد است.[۶،۷] اما سطح زیر منحنی مشخصه عملکرد (AUCROC)، حساسیت کافی برای تعیین مقداری که یک نشانگر زیستی جدید، سبب بهبود توان پیش‌بینی مدل می‌شود، را ندارد.[۸،۹] از طرفی پنسینا[i] و همکاران تاکید کردند که با افزودن یک نشانگر جدید به مدل پیش‌بینی، فقط نمی‌توان به معناداری آماری آن تکیه کرد، چون ممکن است نشانگر جدیدی که با پاسخ ارتباط خیلی ضعیف یا متوسطی دارد با یک حجم نمونه زیاد معنادار شود.[۱۰،۱۱] برای حل این مشکل، در سال ۲۰۰۸ پنسینا و همکاران ایده بازطبقه‌بندی را با شرطی کردن پاسخ (احتمال رخداد بیماری) کامل‌تر کردند[۱۰] و دو روش جدید به نام شاخص‌های بهبود بازطبقه‌بندی خالص (NRI[ii]) و تشخیصی ادغام شده (IDI[iii]) برای ارزیابی توان پیش‌بینی مدل‌های طبقه‌بندی توسط یک نشانگر جدید پیشنهاد دادند.[۱۰] پنسینا و همکاران در سال ۲۰۰۸، برای برآورد این دو شاخص روش فراوانی‌گرا را به‌کار گرفتند. این روش برآوردهای نقطه‌ای برای این دو شاخص دارد و چون توزیع احتمال مشخصی ندارند، استفاده از توزیع مجانبی را برای برآورد فاصله‌ای پیشنهاد دادند. از آنجا که برآورد فاصله‌ای به‌وسیله توزیع مجانبی ممکن است معتبر نباشد، در مطالعه‌های بعدی از روش بوت استرپ برای برآورد فاصله‌ای استفاده کردند.[۱۰-۱۲] در سال ۲۰۱۶ هانگ[iv] و همکاران روش بیزی را برای برآورد این دو شاخص به‌کار گرفتند و با استفاده از شبیه‌سازی عملکرد بهتر روش بیزی را نسبت به‌روش فراوانی‌گرا[v] در حجم نمونه کوچک نشان دادند.[۱۴]

در این مطالعه، ارزش بالینی افزودن برخی نشانگرها به مدل‌های آماری موجود برای بهبود پیش‌بینی بیماری دیابت نوع ۲ در زنان پیش دیابتی ۲۰ سال و بالاتر مطالعه قند و لیپید تهران با استفاده از شاخص‌های NRI و IDI، به‌روش بیزی و مقایسه نتایج آن با روش فراوانی‌گرا بررسی شد.

مواد و روش‌ها

جمعیت بررسی شده، زنان با سن ۲۰ سال یا بالاتر مبتلا به پیش دیابت منطقه ۱۳ تهران بود که در فازهای ۱ و ۲ وارد مطالعه قند و لیپید تهران شدند. این افراد تا فاز ٤ مطالعه از نظر رخداد دیابت، پیگیری شدند. افرادی که میزان قند خون ناشتای آنان مساوی یا بیشتر از ۱٤۰ و کم‌تر از ۲۰۰ میلی‌گرم بر دسی‌لیتر بود، به عنوان پیش‌دیابتی شناخته شدند. جزییات مطالعه قند و لیپید تهران و چگونگی انجام آزمایش‌ها و ابزارهای استفاده شده، پیشتر چاپ شده است.[۱۵] در مجموع ۱۲۵۱ زن پیش‌دیابتی ۲۰ سال یا بالاتر در فازهای اول و دوم وارد مطالعه قند و لیپید تهران شدند و تا فاز ٤ از نظر رخداد دیابت، پیگیری شدند. از این تعداد، ٤۸۸ نفر قبل از پایان فاز ٤ از مطالعه خارج شدند. همچنین ٦۹ نفر نیز دارای مشاهده پرت بودند. در مجموع ۷۳٤ نفر وارد تحلیل آماری شدند که مقادیر متغیرهای آنان در شروع مطالعه بررسی شد. جزییات دقیق چگونگی برآورد مقادیر گمشده و تشخیص نقاط دوره افتاده متغیرهای بررسی شده، در مطالعه باقرزاده و همکاران آمده است.[۱۶] کد اخلاق: (IR.SBMU.RETECH.REC.1398.048)

در این مطالعه، روش‌های فراوانی‌گرا و بیزی برای برآورد شاخص‌های NRI و IDI به‌کار گرفته و مقایسه شدند. متغیر پاسخ، بروز دیابت در زنان پیش‌دیابتی بالای ۲۰ سال بود. متغیرهایی که از نظر بالینی ممکن است با بروز دیابت مرتبط باشند از نظر هم‌خطی و همبستگی بررسی شدند. از میان آن‌ها متغیرهایی که هم‌خطی و همبستگی بالایی داشتند و از نظر بالینی نیز دارای اهمیت کمتری بودند، کنار گذاشته شدند. این افراد به مدت ۱۰ تا ۱۲ سال برای بروز دیابت پیگیری شده بودند. در هر فاز، افراد دارای قندخون ناشتای ۱۲٦ میلی‌گرم بر دسی‌لیتر یا بالاتر، قند دو ساعته ۲۰۰ میلی‌گرم بر دسی‌لیتر یا بالاتر یا مصرف داروی کاهنده قندخون، دیابتی تشخیص داده شدند.

تجزیه و تحلیل آماری:

فرض کنید $Y=(0,1)$ متغیر پاسخ دو حالته و Ω مجموعه‌ای از K متغیر پیش‌گو، $\Omega=(X_1,...,X_k)$ باشد. اگر Ω_1 و Ω_2 دو زیر مجموعه از متغیرهای پیش‌گو با شرط $\Omega_1 \subset \Omega_2 \subset \Omega$ باشند و M_1 و M_2 مدل‌های

i - Pencina
ii -Net Reclassification Improvement
iii- Integrated Discrimination Improvement
iv - Huang
v - Freqentist

[Downloaded from ijem.sbmu.ac.ir on 2026-11-07]
[DOR: 20.1001.1.16834844.1398.21.4.1.7]

ساخته شده به‌ترتیب روی مجموعه Ω_1 و Ω_2 باشند (مدل M_2 شامل متغیرهای پیش‌گوی بیشتری نسبت به مدل M_1 است). در این صورت شاخص بهبود بازطبقه‌بندی خالص به‌صورت زیر تعریف می‌شود:[۱۴]

$$NRI = \frac{1}{2}\{P_r[P(M_2) \geq e \mid Y=1] - P_r[P(M_1) \geq e \mid Y=1]\}$$
$$+ \frac{1}{2}\{P_r[P(M_2) < e \mid Y=0] - P_r[P(M_1) < e \mid Y=0]\}$$

که در آن $P(M_i) = P(Y=1 \mid \Omega_i)$ ، i=1,2 نشان‌دهنده احتمال موفقیت برای مدل M_i، است و $P_r[P(M_i) \geq e \mid Y=1]$ و $P_r[P(M_i) < e \mid Y=0]$ به‌ترتیب، نشان‌دهنده احتمال‌های طبقه‌بندی درست بر اساس مدل M_i با نقطه برش احتمالی e برای بیماری دیابت است که این مقدار بر اساس اهمیت بالینی انتخاب می‌شود. شاخص بهبود تشخیصی ادغام شده به‌صورت زیر محاسبه می‌شود:[۱۴]

$$IDI = [IS(M_2) - IS(M_1)] - [IP(M_2) - IP(M_1)]$$

در رابطه بالا

$$IS(M_i) = \int p_r(P(M_i) > e \mid Y=1)\, de \quad i=1,2$$

و

$$IS(M_i) = \int p_r(P(M_i) < e \mid Y=1)\, de \quad i=1,2$$

که در آن $e_k, k = 1,2,...,l$: نقطه برش‌های مختلف به اندازه l است که از توزیع یکنواخت $U(0,1)$ به‌دست می‌آید. برای محاسبه این دو شاخص، باید احتمال موفقیت مدل M_i به‌دست آید که از طریق برازش مدل رگرسیون لوجستیک با رهیافت فراونی‌گرا و بیزی و متغیر پاسخ Y و مجموعه متغیرهای پیش‌گو Ω_i حاصل می‌شوند. در روش بیزی توزیع پیشین ضرایب رگرسیونی، β، به‌صورت نرمال چند متغیره ناآگاهی بخش، در نظر گرفته شد:[۱۴]

$$\beta \sim MN(0, 10^8 I_k)$$

برآورد پارامترها با روش زنجیر مارکوف مونت کارلو (MCMC[i]) محاسبه شد که تعداد تکرارها برابر ۱۰هزار و دوره سوخت ۵ هزار فرض شد. همچنین در روش فراوانی‌گرا برای برآورد فاصله‌ای، روش بوت استرپ با ۳۰۰ تکرار به‌کار برده شد. در مطالعه (ARIC[ii]) خطر ۱۰ ساله بیماری دیابت، با استفاده از امتیاز خطر مدل پیش‌بینی دیابت روی جمعیت فاقد بیماری دیابت در ابتدای مطالعه، برآورد شده است،[۱۷] مطالعه قند و لیپید تهران (TLGS[iii]) نیز تا فاز چهارم دارای پیگیری ۱۰ ساله است، بنابراین در این پژوهش نیز متغیرهای مشابه در مطالعه ARIC که شامل متغیرهای اندازه دور کمر، پرفشاری خون، قد، وزن، سن، سابقه خانوادگی دیابت، سیگاری بودن و ضربان نبض است، برای برازش مدل پیش‌بینی دیابت به‌کار گرفته شد. در مطالعه ARIC مدل رگرسیون وایبل برای برآورد احتمال رخداد دیابت استفاده شد، اما در اینجا مدل رگرسیون لوجستیک برای برآورد احتمال رخداد بیماری دیابت، برای تشخیص افراد سالم از بیمار، استفاده شد. سپس هر یک از نشانگرهای اندازه دور مچ دست، فشارخون بارداری، سابقه نوزاد بیش از ۴/۵ کیلوگرم به صورت جداگانه، به مدل لوجستیک پایه افزوده شدند که مدل پایه به مدل توسعه یافته، مدل دارای نشانگر(های) جدید، تبدیل می‌شود. به‌وسیله شاخص‌های بهبود بازطبقه‌بندی خالص و تشخیصی ادغام شده، قدرت تشخیصی هر یک از نشانگرها برای پیش‌بینی بیماری دیابت محاسبه شد. عملکرد دو روش بیزی و فراوانی‌گرا در حجم نمونه کوچک بررسی شدند؛ به این‌صورت که در یک نمونه تصادفی بدون جای‌گذاری با حجم ۳۰ از ۷۳۴ زن پیش‌دیابتی، برآورد نقطه‌ای و فاصله‌ای شاخص‌های بهبود بازطبقه‌بندی خالص و تشخیصی ادغام شده محاسبه و مقایسه شد.

یافته‌ها

تعداد 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 کمی خطرساز مرتبط با دیابت در دو گروه ابتلا به دیابت و مبتلا نبودن به دیابت در شروع مطالعه نشان داد که به جز کراتینین، سایر عوامل در افراد دیابتی به‌طور معناداری بیشتر از افراد غیردیابتی بوده است. همچنین مقایسه عوامل کیفی خطرساز دیابت در شروع مطالعه نشان داد که فراوانی مصرف آسپیرین، مصرف داروهای ادرارآور، سابقه خانوادگی دیابت، سابقه هایپرلیپیدمی، سابقه نوزاد بیش از ۴/۵ کیلوگرم و پرفشاری خون و سیگاری بودن در افراد دیابتی به‌طور معناداری

i- Monte Carlo Markov Chain
ii- Atherosclerosis Risk in Communities
iii- Tehran Lipid and Glocose Study

[Downloaded from ijem.sbmu.ac.ir on 2026-11-07]
[DOR: 20.1001.1.16834844.1398.21.4.1.7]

بیشتر از افراد غیردیابتی است. میزان بهنجاری دادهها، با آزمون کولموگروف-اسمیرنوف سنجیده شد و تمامی متغیرهای مطالعه شده از توزیع نرمال برخوردار بودند.

جدول ۱- عوامل خطرساز دیابت در زنان پیش‌دیابتی بالای ۲۰ سال در شروع مطالعه

متغیر	ابتلا به دیابت (تعداد=۲۵۷)	مبتلا نبودن به دیابت (تعداد=۴۷۷)	P-value
سن (سال)	۴۸/۱(۱۰/۵)	۴۴/۷(۱۱/۷)	<۰/۰۰۱
قد (سانتی‌متر)	۱۵۶ (۵)	۱۵۵ (۵)	<۰/۰۰۱
وزن (کیلوگرم)	۷۴ (۱۲)	۶۹ (۱۰)	۰/۰۱۹
اندازه دور کمر (سانتی‌متر)	۹۶ (۱۰/۰۶)	۹۰ (۱۰)	<۰/۰۰۱
ضربان نبض	۸۱ (۱۲)	۸۳/۰۹ (۱۲)	<۰/۰۰۱
قندخون دو ساعته (میلی‌گرم بر دسی‌لیتر)	۱۵۱/۱(۲۵/۸)	۱۳۷(۲۶/۵)	<۰/۰۰۱
قند خون ناشتا (میلی‌گرم بر دسی‌لیتر)	۱۰۳/۸(۱۰/۳)	۹۷/۶(۹/۱)	<۰/۰۰۱
کلسترول تام (میلی‌گرم بر دسی‌لیتر)	۲۲۹/۴(۴۴)	۲۲۰/۱(۴۲/۴)	۰/۰۰۵
نسبت کلسترول به HDL	۵/۵(۱/۴)	۵/۱(۱/۴)	۰/۰۰۱
نسبت تری‌گلیسرید به HDL	۴/۸(۲/۵)	۴/۲(۲/۴)	۰/۰۰۲
کراتینین (میلی‌گرم بر دسی‌لیتر)	۰/۹۸(۰/۱۱)	۰/۹۷(۰/۱۱)	۰/۵۲۴
فشارخون دیاستولیک (میلی‌متر جیوه)	۸۲/۶(۹/۹)	۷۹/۸(۹/۹)	<۰/۰۰۱
دور کمر به قد	۰/۶۲(۰/۰۶)	۰/۵۸(۰/۰۷)	<۰/۰۰۱
دور مچ دست (سانتی‌متر)	۱۶/۶(۰/۹۳)	۱۶/۲(۰/۹۸)	<۰/۰۰۱
سابقه خانوادگی دیابت	۱۱۳(۴۴)	۱۳۵(۲۸/۳)	<۰/۰۰۱
سابقه هایپرلیپیدمی	۹۹(۳۸/۵)	۱۱۹(۲۴/۹)	<۰/۰۰۱
سابقه فشارخون بارداری	۲۱(۸/۲)	۴۰(۸/۴)	۰/۹۲۰
سابقه نوزاد بیش از ۴/۵ کیلوگرم	۳۹(۱۵/۲)	۴۲(۸/۸)	۰/۰۰۹
پرفشاری خون	۱۰۳(۴۰/۱)	۱۳۳(۲۷/۸)	۰/۰۰۱
داشتن گواتر	۵۲(۲۰/۲)	۹۷(۲۰/۳)	۰/۹۷۴
داشتن ندول تیروییدی	۲۲(۸/۶)	۳۳(۶/۹)	۰/۴۲۰
استفاده از داروهای تیروئید	۲۱(۸/۲)	۲۲(۴,۶)	۰/۰۵۰
مصرف آسپیرین	۳۷(۱۴/۴)	۳۶(۷/۵)	۰/۰۰۳
مصرف سیگار	۷(۲/۷)	۱۱(۲/۳)	<۰/۰۰۱

داده‌ها برای متغیرهای کمی به صورت میانگین (انحراف استاندارد) و برای متغیرهای کیفی به صورت تعداد (درصد) گزارش شده‌اند.

نتایج توزیع نشانگرهای مدل پایه TLGS در بروز دیابت در زنان پیش‌دیابتی در جدول ۲ نشان می‌دهد که به جز سیگاری بودن سایر نشانگرهای موجود در این مدل معنادار بودند (P.value<۰/۰۵). مدل پایه ARIC در واقع سهم هر کدام از نشانگرها را به‌وسیله امتیاز خطر، در پیش‌بینی بیماری دیابت بیان کرده و ضرایب و معناداری آن‌ها را در بروز بیماری دیابت گزارش نداده است. مدل پایه TLGS که شامل عوامل خطر مشابه بیماری دیابت در مطالعه ARIC بود به داده‌ها برازش داده شد. سپس با استفاده از معیارهای AUC[i] (تشخیص صحیح افراد بیمار از سالم)، آزمون هاسمر-لمشو[ii] (توافق بین احتمال رخداد بیماری توسط مدل و مشاهده‌ها)، AIC[iii] و BIC[iv] مناسبت مدل پایه TLGS بررسی و سپس با مدل پایه ARIC مقایسه شد (جدول ۳). نتایج حاصل از مقایسه معیارهای برازش مدل پایه ARIC و

i -Area Under the Curve
ii - Hosmer-Lemsho
iii- Akaike Information Criterion
iv- Bayesian Information Criterion

[Downloaded from ijem.sbmu.ac.ir on 2026-11-07]
[DOR: 20.1001.1.16834844.1398.21.4.1.7]

مدل پایه TLGS نشان می‌دهد که هر دو مدل میزان جداسازی افراد سالم و بیمار را به‌خوبی انجام داده‌اند و سطح زیر منحنی مشخصه عملکرد مدل TLGS برابر ۰/۶۹۷ (P-value<۰/۰۰۱) است، که فاصله اطمینان برآورد شده برای آن به‌صورت (۰/۷۳۵- ۰/۶۵۸) است. نتایج حاصل از آزمون هاسمر-لمشو نشان داده است که خطر برآورد شده بروز دیابت توسط دو مدل پایه ARIC و TLGS با بروز خطر مشاهده شده توافق دارد که نشان‌دهنده مناسب بودن برازش است (P-value>۰/۰۵).

جدول ۲- باز برآوردی مدل پیش‌بینی دیابتARIC در داده‌های مربوط به زنان پیش‌دیابتی بالای ۲۰ سال مطالعه TLGS

متغیر	برآورد نقطه‌ای	P-value
سن	۰/۰۲٤	۰/۰۰٦
قد	-۰/۰٤٤	۰/۰۱۲
وزن	۰/۰۲۰	۰/۰۱۱
اندازه دور کمر	۰/۰۳۳	۰/۰۱۲
ضربان نبض	-۰/۰۱٤	۰/۰۳٦
پرفشاری خون	۰/۰۳۵	۰/۰٤۵
سابقه خانوادگی دیابت	۰/۰۷۱٦	<۰/۰۰۱
سیگاری بودن	-۰/۲۲۵	۰/٦٦۲

سطح معناداری P<۰/۰۵ است.

جدول ۳- نتایج معیارهای برازش مدل پیش‌بینی بیماری دیابت ARIC روی داده‌های TLGS

روش	AUC	آزمون هاسمر-لمشو P-value (chi 2)	AIC	BIC
مدل پایه TLGS	۰/٦۹٦	۰/٦۲۵ (٦/۲۰)	۸۸٦/٦۱	۹۲۳/٤۰
مدل پایه ARIC	۰/۸۲۹	۰/۰۷۳ (۱۵ /۷)	۱۹٤۰/٦	۱۹٤٦/۳

نتایج حاصل از برآورد دو شاخص بهبود بازطبقه‌بندی خالص و تشخیصی ادغام شده به‌روش فراوانی‌گرا و بیزی در حجم نمونه ۷۳٤ نشان می‌دهد که نتایج هر دو روش یکسان است و در هر دو روش افزودن نشانگرهای اندازه دور مچ دست، سابقه فشارخون بارداری، سابقه نوزاد بیش از ٤/۵ کیلوگرم، به مدل پایه، سبب بهبود توان پیش‌بینی آن نمی‌شوند(جدول٤). با اینکه افزودن نشانگرهای اندازه دور مچ دست و سابقه فشارخون بارداری به مدل پایه، برآورد نقطه‌ای دو شاخص بهبود بازطبقه‌بندی خالص و تشخیصی ادغام شده مقادیری بزرگ‌تر از صفر شده، اما معنادار نیستند. وقتی نشانگر سابقه نوزاد بیش از ٤/۵ کیلوگرم به مدل پایه افزوده می‌شود، برآورد نقطه‌ای دو شاخص بهبود بازطبقه‌بندی خالص و تشخیصی ادغام شده منفی است و به این معناست که این نشانگر توان پیش‌بینی مدل را کاهش می‌دهد و عامل مناسبی برای بهبود توان مدل پیش‌بینی بیماری دیابت نیست. جدول ۵ مقایسه دو روش بیزی و فراوانی‌گرا را در حجم نمونه ۳۰ تایی نشان می‌دهد که نتایج شاخص‌های بهبود بازطبقه‌بندی خالص و تشخیصی ادغام شده به‌روش بیزی در حجم نمونه کوچک به نتایج این دو شاخص در حجم نمونه بزرگ نزدیک است، بنابراین روش بیزی در مقایسه با روش فراوانی‌گرا در حجم نمونه کم قابل قبول‌تر و معتبرتر است.

[Downloaded from ijem.sbmu.ac.ir on 2026-11-07]

[DOR: 20.1001.1.16834844.1398.21.4.1.7]

جدول ۴- نتایج روش فراوانی‌گرا و بیزی برای برآورد شاخص بهبود باز طبقه‌بندی خالص و تشخیصی ادغام شده در مدل پایه TLGSبرای بروز دیابت در زنان پیش‌دیابتی مطالعه TLGS با حجم ۷۳۴

روش	متغیر	شاخص بهبودباز طبقه‌بندی خالص (NRI)		شاخص بهبود تشخیصی ادغام شده (IDI)	
فراوانی‌گرا		برآورد نقطه‌ای	فاصله اطمینان ۹۵ درصد	برآورد نقطه‌ای	فاصله اطمینان ۹۵ درصد
	دور مچ دست	۰/۰۱۰۱	(-۰/۰۳۰۷، ۰/۰۰۸۶-)	۰/۰۰۳۵	(-۰/۰۰۰۱، ۰/۰۱۵۰)
	نوزاد بیش از ٤/٥ کیلوگرم	-۰/۰۰۲۰	(-۰/۰۱۲۹، ۰/۰۰۶۲-)	۰/۰۰۲۵	(-۰/۰۰۰۰۵، ۰/۰۱۲۸)
	فشارخون بارداری	۰/۰۰۶۲	(-۰/۰۱۸۹، ۰/۰۰۲۰-)	۰/۰۰۲۰	(-۰/۰۰۰۰۷، ۰/۰۰۳)
بیزی		برآورد نقطه‌ای	فاصله معتبر	برآورد نقطه‌ای	فاصله معتبر
	دور مچ دست	۰/۰۰٤۰	(-۰/۰٤۰۲، ۰/۰٥۲۸)	۰/۰۰۳۱	(-۰/۰۰۱۳، ۰/۰۰۷٦)
	نوزاد بیش از ٤/٥ کیلوگرم	-۰/۰۰۱۰	(-۰/۰٤۱۹، ۰/۰٤۳۱)	۰/۰۰۲٤	(-۰/۰۰۱٤، ۰/۰۰٦۳)
	فشارخون بارداری	۰/۰۰٤۱	(-۰/۰۳۸۳، ۰/۰٤٥٥)	۰/۰۰۲۸	(-۰/۰۰۱٥، ۰/۰۰٥٤)

اعداد جدول مقدار افزایشی هریک از نشانگرها را به‌وسیله دو شاخص بهبود بازطبقه‌بندی خالص و تشخیصی ادغام شده برای پیش‌بینی بیماری دیابت نشان می‌دهد.

جدول ۵- نتایج روش فراوانی‌گرا و بیزی برای برآورد شاخص بهبود باز طبقه‌بندی خالص و تشخیصی ادغام شده در مدل پایه TLGS برای بروز دیابت در زنان پیش‌دیابتی مطالعه TLGS با حجم ۳۰

فراوانی‌گرا		برآورد نقطه‌ای	فاصله اطمینان ۹۵ درصد	برآورد نقطه‌ای	فاصله اطمینان ۹۵ درصد
	دور مچ دست	۰/۱۱۸۵	(-۰/۰٦۲۰، ۰/۳٤٤)	۰/۰٤٤٥	(-۰/۰۰٦۲، ۰/۰۷٤۷)
	نوزاد بیش از ٤/٥ کیلوگرم	۰/۰۳۸٤	(-۰/۰۳۲۳، ۰/۲٦۹۰)	۰/۰۱۷۰	(-۰/۰۰۱۳، ۰/۳۳۷٤)
	فشارخون بارداری	۰	(-۰/۰۳۷۷، ۰/۳۱۱٦)	۰/۰۱۲۹	(-۰/۰۰٥۰، ۰/۲۷٥۹)
بیزی		برآورد نقطه‌ای	فاصله معتبر	برآورد نقطه‌ای	فاصله معتبر
	دور مچ دست	۰/۰۳۲۰	(-۰/۱۷٥، ۰/۱۸٥)	۰/۰۰٥۰	(-۰/۰۲۸٥، ۰/۱۳۰٤)
	نوزاد بیش از ٤/٥ کیلوگرم	-۰/۰۰۳۲	(-۰/۱٥۷، ۰/۱٦۰)	۰/۰۰۱٤	(-۰/۰۳۳۱، ۰/۰٦۲۱)
	فشارخون بارداری	۰/۰۰٥۰	(-۰/۱۸۹، ۰/۱۸۰)	۰/۰۰٥٥	(-۰/۰۲۸۱، ۰/۰٤۳۱)

عددهای جدول، مقدار افزایشی هریک از نشانگرها را به‌وسیله دو شاخص بهبود بازطبقه‌بندی خالص و تشخیصی ادغام شده برای پیش‌بینی بیماری دیابت نشان می‌دهند

بحث

در این پژوهش، روش‌های فراوانی‌گرا و بیزی، برای برآورد نقطه‌ای و فاصله‌ای دو شاخص NRI و IDI به‌کار گرفته شد و برای بررسی تفاوت عملکرد، این دو روش در حجم نمونه بزرگ و کوچک با هم مقایسه شدند. در روش فراوانی‌گرا از روش حداکثر درست‌نمایی برای برآورد ضریب‌ها و فاصله اطمینان آن‌ها استفاده شد، ولی در بعضی موارد ضرایب رگرسیونی در روش فراوانی‌گرا در نمونه‌های کوچک همگرا نمی‌شوند یا اینکه در نمونه‌های کوچک این روش عملکرد ضعیفی دارد. برای برطرف کردن این مشکل از روش بیزی استفاده شد. مطالعه ARIC روی افراد بزرگسال در فاصله سنی ٤٥ تا ٦٤ سال که در ابتدای مطالعه فاقد بیماری دیابت بوده‌اند، انجام شد و ارزیابی مدل پایه ARIC روی دو گروه زنان و مردان انجام شده است. از آنجاکه مطالعه حاضر فقط روی زنان پیش‌دیابتی انجام شده است،

[Downloaded from ijem.sbmu.ac.ir on 2026-11-07]

[DOR: 20.1001.1.16834844.1398.21.4.1.7]

مقایسه با مدل پایه ARIC روی گروه زنان انجام شد. نتایج معیارهای AUCROC، آزمون هاسمر لمشو، AIC و BIC روی مدل پایه TLGS نشان داد که این مدل عملکرد خوبی دارد و برای پیش‌بینی بیماری دیابت مناسب است. مقایسه برآوردهای نقطه‌ای و فاصله‌ای دو شاخص NRI و IDI در حجم نمونه ۷۳٤ نشان داد که نتایج این دو شاخص به‌روش‌های فراوانی‌گرا و بیزی مشابه و در حجم نمونه ۳۰ نتایج برآورد نقطه‌ای و فاصله‌ای دو شاخص NRI و IDI در روش بیزی، به نتایج برآوردهای این دو شاخص در حجم نمونه ۷۳٤ نزدیک‌تر هستند. مقایسه مدل پایه ARIC و مدل پایه TLGS با استفاده از معیارهای برازش مدل، نشان داد که مدل پایه TLGS نیز برای پیش‌بینی بیماری دیابت مناسب‌تر است. طبق مطالعه‌هایی که بررسی شد، برای گزارش مقدار افزایشی[i] یک نشانگر(های) جدید باید هر دو شاخص با هم گزارش داده شوند. پنسینا و همکاران در سال ۲۰۱۲ با مطالعه‌ای روی ویژگی‌های این دوشاخص، هم اهمیت این موضوع را ذکر و تاکید کردند که نمی‌توان گفت کدام یک از دو شاخص NRI و IDI نسبت به دیگری عملکرد بهتری دارد و برای معناداری مقدار افزایشی (بهبودی) نشانگر(های) جدید مکمل یکدیگر هستند و در صورتی که نتایج دو شاخص NRI و IDI برای نشانگر(های) جدید، یکسان بود و مقداری بزرگ‌تر از صفر و برآورد فاصله‌ای معناداری داشتند، می‌توان معناداری مقدار افزایشی نشانگر(های) جدید را نتیجه گرفت.[11] برتری روش بیزی برای برآورد نقطه‌ای و فاصله‌ای دو شاخص NRI و IDI نسبت به روش فراوانی‌گرا در مطالعه‌های مشابه نیز نتیجه شده است. به عنوان مثال، در سال ۲۰۱٦ هانگ و همکاران، روش بیزی و فراوانی‌گرا را برای برآورد این دو شاخص مقایسه کردند و در حجم نمونه بزرگ مشابه بودن برآوردهای نقطه‌ای فراوانی‌گرا و بیزی را نتیجه گرفتند. هم‌نین با شبیه‌سازی نشان دادند که روش بیزی به حجم نمونه حساسیت ندارد و عملکرد بهتری نیز دارد.[14] در مطالعه حاضر نیز هدف مقایسه روش‌های بیزی و فراوانی‌گرا برای برآورد دو شاخص NRI و IDI بود، که در حجم نمونه بزرگ نتایج دو روش مشابه بود و در حجم نمونه کوچک، روش بیزی برتری داشت. به دلیل اینکه روش بیزی برای پیش‌بینی بیماری دیابت استفاده نشده است، امکان مقایسه یافته‌های به‌دست آمده در مطالعه حاضر با مطالعه‌های مشابه وجود نداشت و تنها مقایسه با مطالعه‌هایی که از روش‌های مرسوم مانند فراوانی‌گرا استفاده کرده‌اند، انجام شد. همچنین بر اساس جست‌وجوی محققان مطالعه‌ای یافت نشد که روی زنان پیش‌دیابتی عملکرد و ارزیابی یک نشانگر یا نشانگرهای جدید و توان پیش‌بینی مدل‌ها را بررسی کند، ولی در مطالعه‌هایی که به بررسی نشانگرهای جدید مشابه با پژوهش حاضر برای پیش‌بینی بیماری‌های قلبی - عروقی و مرگ‌های ناشی از بیماری‌های قلبی و همچنین پیش‌بینی بیماری دیابت پرداخته‌اند، به‌دست آمده است. همچنین به دلیل اینکه می‌توان از مدل‌های رگرسیونی متفاوتی مانند رگرسیون کاکس، وایبل و رگرسیون لوجستیک، برای ارزیابی نشانگرهای جدید و تحلیل پیش‌بینی وقوع بیماری‌ها استفاده کرد، مقایسه با مطالعه‌هایی که از مدل‌های رگرسیونی ذکر شده استفاده کرده‌اند، انجام شد. در سال ۲۰۱۸ چاتپاتاهی[ii] و همکاران در یک مطالعه کوهورت، به بررسی مقدار افزایشی نشانگر قند خون دو ساعته برای پیش‌بینی بیماری‌های قلبی پرداختند. این افراد با افزودن نشانگر قندخون دو ساعته به مدل پایه، معناداری بهبود توان پیش‌بینی آن را به‌وسیله دو شاخص NRI و IDI نتیجه گرفتند[18] اما در این مطالعه قندخون دو ساعته سبب بهبود توان پیش‌بینی مدل پایه نشده است و این به آن دلیل است که در مطالعه آن‌ها مقادیر پایه برای تعریف قندخون دوساعته متفاوت در نظر گرفته شده و همچنین ویژگی‌های افراد بررسی شده در مطالعه متفاوت است. در سال ۲۰۱۸ لی[iii] و همکاران در مطالعه‌ای به بررسی مقدار افزایشی نشانگر قندخون ناشتا برای پیش‌بینی بیماری دیابت پرداختند و به‌وسیله معناداری دو شاخص NRI و IDI بهبود در توان مدل پیش‌بینی پایه را نتیجه گرفتند.[19] در مطالعه حاضر نیز افزودن قندخون ناشتا سبب بهبود توان پیش‌بینی مدل پایه شده است. روش بیزی نسبت به حجم نمونه حساسیت نشان نداد و در نمونه‌های کوچک عملکرد بهتری نسبت به روش فراوانی‌گرا داشت. برای پژوهش‌های بعدی استفاده از تابع ربط‌های probit و complementary log-log با متغیر پاسخ دو حالته و همچنین استفاده از متغیر پاسخ چند حالته مانند رگرسیون لوجستیک چند جمله‌ای پیشنهاد می‌شود. در مجموع، مطالعه حاضر نیز دارای محدودیت بود، اگرچه

i -Added Value

ii -Chattopadhyay

iii -Lee

[Downloaded from ijem.sbmu.ac.ir on 2026-11-07]

[DOR: 20.1001.1.16834844.1398.21.4.1.7]

بـزرگمنش و همکـاران، در زمینـه بررسـی مقـدار افزایشـی شاخص‌های مربوط به انـدازه‌گیـری بـدن انسـان در مطالعـه کوهـورت قنـد و لیپیـد تهـران، مطالعـه‌هـای بسـیاری بـرای پیش‌بینی بیماری دیابت انجام داده‌انـد، در ایـن پـژوهش ایـن شاخص‌ها در مدل پایه به‌کار گرفتـه شـده بـود، بنـابراین بـه دلیل هم‌خطی مقدار افزایشی آن‌ها محاسبه نشد.

سپاسگزاری: در این مطالعه، از داده‌های مطالعه قند و لیپید تهـران که توسط پژوهشکده علـوم غـدد درون‌ریـز و متابولیسـم دانشـگاه علـوم پزشـکی شـهید بهشـتی اجـرا شـده اسـت، اسـتفاده شـد.

نویسـندگان بـر خـود لازم مـی‌داننـد از کسـانی کـه درطراحـی و جمع‌آوری داده‌های طرح TLGS مشارکت داشـتند، قـدردانی کننـد.

این پژوهش از پایان‌نامه کارشناسی ارشد آمـار زیسـتی اسـتخراج شده است.

نویسندگان اعلام می‌کنند که هیچ‌گونه تضاد منافعی در پژوهش حاضر وجود ندارد.

References

1. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004; 27: 1047-53.
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: 571-6.
3. Esmailnasab N, Afkhamzadeh A, Ebrahimi A. Effective factors on diabetes control in Sanandaj diabetes center. Iranian Journal of Epidemiology 2010; 6: 39-45.
4. Roche MM, Wang PP. Factors associated with a diabetes diagnosis and late diabetes diagnosis for males and females. Journal of Clinical & Translational Endocrinology 2014; 1: 77-84.
5. Barkhordari M, Padyab M, Hadaegh F, Azizi F, Bozorgmanesh M. Stata Modules for Calculating Novel Predictive Performance Indices for Logistic Models. Int J Endocrinol Metab 2016; 14: e26707.
6. Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. Ann Intern Med 2008; 149: 751-60.
7. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology 2010; 21: 128-38.
8. Kerr KF, McClelland RL, Brown ER, Lumley T. Evaluating the incremental value of new biomarkers with integrated discrimination improvement. Am J Epidemiol 2011; 174: 364-74.
9. Vickers AJ, Pepe M. Does the net reclassification improvement help us evaluate models and markers? Ann Internal Med 2014; 160: 136-7.
10. Pencina MJ, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in Medicine 2008; 27: 157-72.
11. Pencina MJ, D'Agostino RB, Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. Statistics in medicine 2012; 31: 101-13.
12. Li J, Jiang B, Fine JP. Multicategory reclassification statistics for assessing improvements in diagnostic accuracy. Biostatistics 2012; 14: 382-94.
13. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk-prediction instruments: A critical review. Epidemiology 2014; 25: 114-21.
14. Huang Z, Li J, Cheng CY, Cheung C, Wong TY. Bayesian reclassification statistics for assessing improvements in diagnostic accuracy. Statistics in medicine 2016; 35: 2574-92.
15. Azizi F, Madjid M, Rahmani M, Emami H, Mirmiran P, Hadjipour R. Tehran Lipid and Glucose Study (TLGS): rationale and design. Iranian Journal of Endocrinology and Metabolism 2000; 2: 77-86.
16. 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.
17. Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW. Two risk-scoring systems for predicting incident diabetes mellitus in US adults age 45 to 64 years. Ann Intern Med 2009; 150: 741-51.
18. Chattopadhyay S, George A, John J, Sathyapalan T. Two-hour post-challenge glucose is a better predictor of adverse outcome after myocardial infarction than fasting or admission glucose in patients without diabetes. Acta Diabetologica 2018; 55: 449-58.
19. Lee JW, Lim NK, Park HY. The product of fasting plasma glucose and triglycerides improves risk prediction of type 2 diabetes in middle-aged Koreans. BMC Endocr Disord 2018; 18: 33.

[Downloaded from ijem.sbmu.ac.ir on 2026-11-07]

[DOR: 20.1001.1.16834844.1398.21.4.1.7]

Original Article

Comparison of Bayesian and Frequentist Methods in Estimating the Net Reclassification and Integrated Discrimination Improvement Indices for Evaluation of Prediction Models: Tehran Lipid and Glucose Study

Nariman S[1], Khalili D[2], Baghestani AR[1], Ahmadi F[1], Mahdavi M[3], Mehrabi Y[2,4]

[1]Department of Biostatistics, Faculty of Paramedical Science, Shahid Beheshti University of Medical Sciences, Tehran, Iran.[2]Prevention of Metabolic Disorders Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.[3]Obesity Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.[4]Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, I.R. Iran.

e-mail: mehrabi@sbmu.ac.ir

Received: 21/07/2019 **Accepted:** 07/12/2019

Abstract

Introduction: The Frequency-based method is commonly used to estimate the Net Reclassification Improvement (NRI)- and Integrated Discrimination Improvement (IDI) indices. These indices measure the magnitude of the performance of statistical models when a new biomarker is added. This method has poor performance in some cases, especially in small samples. In this study, the performance of two Bayesian and Frequentist methods were evaluated and compared for the diabetes prediction model in pre-diabetic women. Materials and Methods: A total of 734 pre-diabetic women aged≥20 years, participated in the first and second phases of the Tehran Lipid and Glucose Study (TLGS) were enrolled in this research and the logistic regression model was used with variables of the Atherosclerosis Risk in Communities (ARIC) in order to estimate the probability of diabetes. Predictors of diabetes in the ARIC study were waist circumference, hypertension, height, weight, age, family history of diabetes, smoking and pulse rate. The Frequentist and Bayesian methods were used to estimate the NRI and IDI. In addition these methods were compared in a small sample size selected randomly from the primary sample. Statistical analyses were performed using R- software version 3.1.3. Results: Estimates of the NRI and IDI indices by the Bayesian and Frequentist methods provided almost similar results. Adding new markers of wrist size, history of macrosomia, and history of preeclampsia to the ARIC model did not show significant effect on the improvement indices. In the small sample size, IDI and NRI showed better performance in the Bayesian method compared to the frequentist method. Conclusion: The Bayesian method had a more reliable performance in comparison with the frequency-based method, especially in small samples.

Keywords: Net Reclassification Improvement, Integrated Discrimination Improvement, Bayes method, Diabetes, Prediction models

[Downloaded from ijem.sbmu.ac.ir on 2026-11-07]

[DOR: 20.1001.1.16834844.1398.21.4.1.7]