Journal of Health Promotion Management (JHPM) Original Article

April-May 2018, Volume 7, Issue 2 DOI: 10.21859/jhpm-07024

Application of Logistic Regression with Misclassified Variables in Diabetes Data

Maryam Rastegar [1], Samaneh Hosseinzadeh [2], Enayatollah BAkhshi [2,*]

[1] MS Student, Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

[2] Asssistant Professor, Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

*** Corresponding author:** Enayatollah BAkhshi, 2Asssistant Professor, Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. E-mail: bakhshi@razi.tums.ac.ir

Received: 02 May 2017 **Accepted:** 05 May 2017

Abstract

Introduction: The analysis of classified data in statistics and medical sciences is very important. If the binary response variable is misclassified, the results of fitting the model, will be skewed with false interpretation. The aim of this study was the application of logistic regression with misclassified variables in diabetes data.

Methods: In this descriptive study, data from 819 participants in the diabetes screening program at Zahedan Health Center in 2014 were used. Type 2 diabetes was studied in two ways. At first, by testing normal blood glucose (without fasting), the lack of correlation between type 2 diabetes and blood pressure was determined by logistic regression and odds ratios, and then a fasting blood glucose test was used for validation, which revealed significant results. False classification was considered based on the level of blood glucose due to low sensitivity and specificity of blood glucose test. To correct the classification error, the likelihood ratio method was used to estimate the coefficients. Data analysis was done using the software SAS version 9.1.3 and procedure NLMIXED with a significance level of 0.05.

Results: The correlation coefficient changed the odds ratio of diabetes in the blood pressure variable from 0.227 to 1.20, significantly ($P < 0.001$). In addition, other model variables were modified.

Conclusions: Logistic regression for data with error classification can be used as a suitable method for analyzing data with classification error. Validation using logistic regression for classification error data showed that high blood pressure has a significant effect on diabetes. It is suggested that the logistic regression method should be used in order to correct the odds ratio in view of the probability of classification error in the screening data.

Keywords: Logistic Regression, Misclassification, Odds Ratio, Diabetes

نشریه مدیریت ارتقای سلامت
مقاله پژوهشی
فروردین و اردیبهشت ۱۳۹۷، دوره ۷، شماره ۲
DOI: 10.21859/jhpm-07024

کاربرد رگرسیون لجستیک در داده‌های دیابت دارای خطای طبقه بندی

مریم رستگار [۱]، سمانه حسین زاده [۲]، عنایت اله بخشی [۲،*]

[۱] دانشجوی کارشناسی ارشد، گروه آمارزیستی، دانشگاه علوم بهزیستی و توانبخشی، تهران، ایران
[۲] استادیار، گروه آمارزیستی، دانشگاه علوم بهزیستی و توانبخشی، تهران، ایران
[*] **نویسنده مسئول**: عنایت اله بخشی، استادیار، گروه آمارزیستی، دانشگاه علوم بهزیستی و توانبخشی، تهران، ایران. ایمیل: bakhshi@razi.tums.ac.ir

تاریخ دریافت مقاله: ۱۳۹۶/۰۲/۱۲ **تاریخ پذیرش مقاله:** ۱۳۹۶/۰۲/۱۵

چکیده

مقدمه: تحلیل داده‌های طبقه بندی شده در آمار و علوم پزشکی از اهمیت خاصی برخوردار است. اگر متغیر پاسخ دو حالتی دارای خطای طبقه بندی باشد نتایج برازش مدل، اریب و تفسیر نادرستی خواهد داشت. هدف این مطالعه کاربرد رگرسیون لجستیک در داده‌های دیابت دارای خطای طبقه بندی می‌باشد.

روش کار: در این مطالعه توصیفی، از داده‌های ۸۱۹ نفر از شرکت کنندگان در طرح غربالگری دیابت در مرکز بهداشت زاهدان سال ۱۳۹۳ استفاده شد. ابتلا به دیابت نوع ۲ به دو طریق بررسی شده است. ابتدا بوسیله آزمایش قند خون معمولی (بدون ناشتا) که عدم ارتباط بین ابتلا به دیابت نوع دو با فشار خون بوسیله رگرسیون لجستیک و نسبت شانس مشخص شد و سپس برای معتبر سازی، آزمایش قند خون ناشتا انجام شد، که ارتباط معنادار بود. خطای طبقه بندی غلط بر اساس میزان قند خون با توجه به پایین بودن حساسیت و ویژگی آزمایش قند خون در نظر گرفته شد. برای تصحیح خطای طبقه بندی از روش نسبت درستنمایی برای برآورد ضرایب استفاده شد. در آنالیز داده‌ها از نرم افزار SAS نسخه ۹،۱،۳ و procedure NLMIXED با سطح معنی داری ۰/۰۵ استفاده گردید.

یافته‌ها: ضرایب تصحیح موجب تغییر نسبت شانس ابتلا به دیابت در متغیر فشار خون از ۰/۲۲۷ به ۱/۱۲۰ و معناداری آن گردید ($P < ۰/۰۰۱$). به علاوه، بر آورد سایر متغیرهای مدل نیز تغییر یافتند.

نتیجه گیری: رگرسیون لجستیک برای داده‌های دارای خطای طبقه بندی می‌تواند به عنوان یک روش مناسب در تحلیل داده‌های دارای خطای طبقه بندی مورد استفاده قرار گیرد. معتبرسازی با استفاده از رگرسیون لجستیک برای داده‌های دارای خطای طبقه بندی نشان داد که فشارخون بالا اثر معناداری بر ابتلا به بیماری دیابت دارد. پیشنهاد می‌شود با توجه به اینکه در داده‌های طرح‌های غربالگری احتمال خطای طبقه بندی وجود دارد جهت تعدیل اریبی نسبت شانس از روش رگرسیون لجستیک استفاده شود.

واژگان کلیدی: رگرسیون لجستیک، خطای طبقه بندی، نسبت شانس، دیابت

مقدمه

دیابت یک بیماری مزمن و متابولیکی است که با افزایش سطح قند خون مشخص می‌شود که با گذشت زمان باعث آسیب جدی به قلب، رگ‌های خونی، چشم، کلیه‌ها و اعصاب می‌شود. شایع‌ترین نوع آن دیابت نوع ۲ می‌باشد که معمولاً در بزرگسالان است، وقتی بدن به انسولین مقاوم می‌شود یا به اندازه کافی انسولین را تولید نمی‌کند، اتفاق می افتد. در سه دهه گذشته شیوع دیابت نوع ۲ در تمام کشورها با هر درآمدی به طور چشمگیری افزایش یافته است (۱). برای افراد مبتلا به دیابت، دسترسی به درمان مقرون به صرفه، از جمله انسولین، برای بقای آن‌ها حیاتی است. یکی از اهداف سازمان جهانی بهداشت مشترک توقف رشد دیابت و چاقی تا سال ۲۰۲۵ می‌باشد. افراد دیابتی به دلیل رشد و پیر شدن جمعیت کشورها، شهرنشینی، صنعتی شدن، افزایش شیوع چاقی و بی تحرکتی جسمانی به سرعت در حال افزایش است (۲). بر طبق آمارسازمان جهانی بهداشت در سال ۲۰۱۶ میزان شیوع دیابت در ایران حدود ۱۰/۳ درصد می‌باشد و حدود ۲ درصد از مرگ‌ها در کشور ناشی از دیابت می‌باشد (۳). برآورد اخیر سازمان جهانی بهداشت نشان می‌دهد در سال ۲۰۲۵ تعداد دیابتی‌ها در جهان به ۳۰۰ میلیون افزایش

یابد (۴). عوامل بسیاری مانند سن، جنسیت، چاقی و فشار خون بر دیابت تأثیر دارند (۵-۷). جهت بررسی تأثیر این عوامل بر ابتلای به دیابت از نسبت شانس حاصل از مدل رگرسیون لجستیک استفاده شد. قدرت ارتباط در مورد برخی از بیماری‌های مزمن نظیر بیماری دیابت و فشارخون به اندازه‌ای است که نقش تغییرات فشار خون نه تنها به عنوان یک علامت تشخیصی قوی، بلکه تا حد علت یا حداقل عامل خطر عمده مورد توجه قرار می‌گیرد. فشار خون و اختلالات آن به عنوان عامل زمینه ساز مهم یا یک بیماری مستقل همواره مورد توجه جدی سیاست گزاران بهداشتی است (۸, ۹) توسعه شهرنشینی همراه با تغییر در ساختار سنی جامعه به طرف پیر شدن جمعیت، کشور ایران را هر چه بیشتر با افزایش بروز و شیوع بیماری‌های مزمن نظیر دیابت و پرفشاری خون و عوامل خطرساز آن‌ها رو به رو خواهد کرد. این بیماری‌ها با شیوع بالا و عوارض شدید در اندام‌های حیاتی بدن در زمره اولویت‌های بهداشتی درمانی کشور قرار دارند (۱۰). تقریباً اکثر مطالعاتی که به بررسی عوامل خطر دیابت پرداخته‌اند فشارخون بالا را جزء عوامل مؤثر معرفی نموده‌اند (۱۱, ۱۲). در سوگیری اطلاعات، به دلیل صحیح نبودن اطلاعات جمع آوری شده، محقق دچار نتایج اشتباه می‌شود. طبقه بندی غلط متغیرهای طبقه بندی شده زمانی رخ می‌دهد که وضعیت ثبت یک متغیر برای فرد با وضعیت واقعی متغیر برای آن فرد تفاوت داشته باشد، به این دلیل برخی از افراد هنگام طبقه بندی براساس یک صفت به اشتباه در طبقات نادرست قرار می‌گیرند. گاهی اوقات به دلیل وجود برخی موارد مثل پایین بودن حساسیت و ویژگی روش تشخیصی یا ابزار اندازه گیری، خطای یادآوری و...، هنگام تعیین وضعیت مواجهه یا پی آمد ممکن است وضعیت واقعی به درستی تشخیص داده نشود و برخی از افراد بر اساس متغیر دارای خطا (مواجهه یا پی آمد) به اشتباه در گروه دیگر قرار گیرند و یا به عبارت دیگر، طبقه بندی غلط رخ دهد. برخی مواقع ممکن است دلیل طبقه بندی غلط تشخیص نادرست مواجهه یا پی آمد نباشد و تنها ناشی از یک خطای انسانی هنگام انجام طبقه بندی باشد. اگر میزان حساسیت و ویژگی یک متغیر بصورت شرطی به مقادیر متغیرهای دیگر که دارای خطا می‌باشند وابسته باشد، خطای طبقه بندی افتراقی است. اگر خطای طبقه بندی در متغیر پیش بین و متغیر پاسخ، مستقل از یکدیگر باشند به این نوع خطای طبقه بندی، خطای طبقه بندی غیر افتراقی می گویند (۱۳).

در پزشکی و اپیدمیولوژی مسئله طبقه بندی براساس وضعیت ابتلا به بیماری یا وضعیت مواجهه با عامل خطر از اهمیت ویژه‌ای برخوردار است، در اینگونه مطالعات بعد از طبقه بندی افراد تحت مطالعه بر اساس وضعیت مواجهه و ابتلا، از داده‌های طبقه بندی شده جهت به دست آوردن برخی شاخص‌های آماری (مانند نسبت شانس و نسبت خطر) و سنجش رابطه بین متغیر پیشگو و متغیر پاسخ استفاده می‌شود (۱۴). وجود خطای طبقه بندی غلط می‌تواند موجب انحراف و نتیجه گیری نادرست هنگام تحلیل داده‌ها شود، بنابراین، نیاز به روش‌هایی است که بتوان به وسیله آن‌ها اثر این خطا را کاهش داد و داده‌های غلط طبقه بندی شده را با حداقل خطا تحلیل نمود. برای تعدیل خطای طبقه بندی از جمله روش‌های که وجود دارد عبارتند از روش ماتریسی، روش ماتریس معکوس، روش ماتریس بهبود یافته، روش بیزی و روش رگرسیون لجستیک که البته همه روش‌ها تقریباً نتایج مشابهی می‌دهند (۱۵). در این پژوهش از روش برآورد ضرایب بوسیله تابع حداکثر درستنمایی در رگرسیون لجستیک استفاده شده است. دیابت در مراحل ابتدایی ممکن است بدون علامت باشد. بسیاری از بیماران به طور اتفاقی در یک آزمایش یا در حین غربالگری شناسایی می‌شوند. با بالاتر رفتن قند خون، علائم دیابت آشکارتر می‌شوند. پرادراری، پرنوشی، پرخوری، کاهش وزن با وجود اشتهای زیاد، خستگی و تاری دید از علایم اولیه شایع دیابت است. بسیاری از بیماران در هنگام تشخیص بیماری چندین سال دیابت داشته‌اند و حتی دچار عوارض دیابت شده‌اند و این علایم معمولاً به طور تدریجی بدتر می‌شوند. در نهایت فرد دچار خستگی مفرط و تاری دید شده و دچار عارضه شود (۱۶). مسئله‌ای که در اینجا مهم است این است که اگر فردی واقعاً دیابت داشته باشد اما نتیجه آزمایش آنرا نشان ندهد یا برای فردی به اشتباه دیابت تشخیص داده شود، هر دو نتیجه برای افراد مشکل ساز و هزینه بر خواهند بود. یعنی افرادی که واقعاً بیمارهستند از سالم بودن خود مطمئن می‌شوند. در نتیجه فرد به پدیدار شدن علائم دقت نمی‌کند و در نتیجه تشخیص و درمان وی به تأ خیر می افتد. بنابراین، لازم است از روش‌های غربالگری با مقدار منفی کاذب پایین استفاده شود و همچنین افرادی که واقعاً سالم هستند در سالم بودن خود شک می‌کنند. در نتیجه افراد سالم تحت آزمایشات بعدی برای تشخیص قرار می‌گیرند که در بعضی موارد مشکل، ناراحت کننده، اضطراب آور و گران است. این آزمایشات تا وقتی سالم بودن افراد ثابت شود ادامه خواهد یافت. پس لازم است از روش‌های غربالگری با مقدار مثبت کاذب پایین استفاده شود. هدف این مطالعه کاربرد رگرسیون لجستیک در داده‌های دیابت دارای خطای طبقه بندی می‌باشد (۱۷). خطای اندازه گیری متغیرها با مقیاس اسمی یا رتبه‌ای خطای طبقه بندی نادرست (Misclassification error) نامیده می‌شود. Wachholder و همکاران بیان داشتند که احتمالات طبقه بندی نادرست با حساسیت (sensitivity) و ویژگی (specificity) اندازه گیری و از آن‌ها برای تصحیح طبقه بندی نادرست استفاده می‌شود. به منظور برآورد این احتمالات که ضرایب تصحیح نیز می‌باشند از داده‌های کمکی مانند داده‌های معتبر یا اندازه گیری‌های مکرر استفاده می‌شود تا ضمن تعیین احتمالات طبقه بندی غلط (حساسیت و ویژگی) بتوان برآوردهای اصلاح شده را به دست آورد (۱۸). موضوع طبقه بندی نادرست ابتدا توسط Bross مورد توجه قرار گرفته است (۱۹). توسط Chen مطالعات مروری در این خصوص صورت پذیرفته است (۲۰) و Tenenbein روش برآورد بر اساس نمونه گیری دو مرحله‌ای را با وجود طبقه بندی نادرست ارائه نمود (۲۱). Roy و همکاران یک مدل رگرسیون با پاسخ دو حالتی غلط طبقه بندی شده در نظر گرفتند. آن‌ها فرض کردند برخی از متغیرها قابل مشاهده نمی‌باشند اما متغیر کمکی آن‌ها را می‌توان اندازه گیری کرد. برای برازش مدل یک تحلیل بر پایه درستنمایی انجام دادند و سپس اثر خطای طبقه بندی با خطای اندازه گیری را با استفاده از شبیه سازی تحلیل کردند (۲۲).

Liu& Liang تصحیح خطای طبقه بندی نادرست را براساس مدل‌های خطی عمومی ارائه نموده‌اند (۲۳). Neuhaus یک تحلیل کلی در مورد اندازه اریبی ناشی از خطای طبقه بندی غلط در متغیر پاسخ انجام داد و نشان داد بجز حالتی که حساسیت و ویژگی هر دو

بزرگ هستند چشم پوشی از خطای طبقه بندی غلط منجر به اریبی‌های بزرگ در برآورد اثر کوریت ها می‌شود (۲۴). Davidov و همکاران با استفاده از رگرسیون لجستیک به تصحیح ضرایب مدل دارای خطای طبقه بندی پرداختند (۲۵). Luan و همکاران نشان دادند تعدیل خطای طبقه بندی غلط در یک مدل رگرسیون لجستیک زمانی که متغیر پاسخ آن غلط طبقه بندی شده است همراه کاهش اریبی موجب افزایش واریانس و کمترین خطای برآورد می‌شود (۲۶). Küchenhoff و همکاران یک روش کلی برای برخورد با طبقه بندی غلط در کوریت های گسسته یا متغیر پاسخ رگرسیونی ارائه کردند. آن‌ها روش شبیه سازی و برونیابی را بکار بردند که معمولاً برای برخورد با اندازه گیری خطا مورد استفاده قرار می‌گیرد. آن‌ها نیز در روش خویش از یک مطالعه معتبرسازی استفاده نمودند (۲۷). Duffy و همکاران یک روش و فرم بسته برای تصحیح برآورد رگرسیون لجستیک زمانی که پی آمد دوحالتی غلط طبقه بندی شده شرح دادند. همینطور روش ارائه شده برای دو مطالعه مقطعی و هم گروهی بکار بردند (۱۴). ابدی و همکاران در پژوهشی براساس اطلاعات حاصل از اجرای مرحله اول مطالعه قند و لیپید تهران نشان دادند که طبقه بندی غلط از منابع اریبی است که موجب تحلیل نادرست از روابط بین متغیرها می‌گردد (۱۶). Tang و همکاران با مدل سازی رگرسیون لجستیک بر پایه درستنمایی به تصحیح طبقه بندی نادرست پرداختند (۱۳). در مطالعه حاضر به مقایسه برآورد نسبت شانس (Odds Ratio)، حساسیت، ویژگی و سطح زیر منحنیROC ، در دو مدل رگرسیون لجستیک استاندارد و رگرسیون لجستیک برای داده‌های دارای خطای طبقه بندی است. هدف این مطالعه کاربرد رگرسیون لجستیک در داده‌های دیابت دارای خطای طبقه بندی می‌باشد.

روش کار

داده‌های مورد مطالعه بخشی از داده‌های جمع آوری شده در طرح غربالگری دیابت می‌باشد که در سال ۱۳۹۳ به صورت یک مطالعه مقطعی در شهر زاهدان جمع آوری شد که بعد از حذف نمونه‌ها با اطلاعات ناقص، تعداد نمونه به ۸۱۹ نفر کاهش یافت و هدف تشخیص وجود پره دیابت و دیابت در افراد مورد بررسی بود که با انجام آزمایش قند خون معمولی (sugar test blood) تشخیص داده می‌شد اگر نتیجه آزمایش کمتر از ۲۰۰ میلی گرم بر دسی لیتر بود فرد دیابت ندارد اگر نتیجه بیشتر از ۲۰۰ بود فرد دچار دیابت نوع ۲ است. جهت بررسی ریسک فاکتورهای دیابت، ارتباط دیابت با فشار خون (بر اساس طبقه بندی سازمان بهداشت جهانی پر فشاری خون سیستولیک را بالاتر از ۱۴۰ میلی متر جیوه و پرفشاری خون دیاستولیک را بالاتر از ۹۰ میلی متر جیوه در نظر می‌گیرند)، سن (در این مطالعه افراد بالای ۳۰ سال شرکت داشتند و متغیر سن بصورت کمی در مدل وارد شد)، شاخص توده بدنی (که از تقسیم وزن بر حسب کیلوگرم بر مجذور قد بر حسب متر به دست می‌آید) و جنسیت توسط شاخص نسبت شانس و به دو روش رگرسیون لجستیک استاندارد و رگرسیون لجستیک برای داده‌های دارای خطای طبقه بندی انجام شد، به منظور برازش مدل رگرسیون لجستیک دوتایی، متغیر پاسخ به صورت یک متغیر پاسخ دو حالتی (دارد، ندارد) در نظر گرفته شد. حال مسئله‌ای که در اینجا مهم است این است که اگر فردی واقعاً دیابت داشته باشد اما نتیجه آزمایش آنرا نشان ندهد باعث خطای تشخیص و طبقه بندی می‌شود لذا باید ابتدا وجود خطای طبقه بندی را با استفاده روش‌های تشخیصی استاندارد (Gold standard) بررسی نمود. برخی روش‌های تشخیصی استاندارد وجود دارد که منجر به طبقه بندی صحیح می‌گردد، اما عموماً این روش‌ها یا گران هستند یا امکان بررسی آن برای تمام افراد نمونه وجود ندارد. در این گونه موارد، این روش استاندارد طلایی برای یک زیر گروه کوچک‌تر از افراد مورد مطالعه انجام شده و بر اساس اطلاعات حاصل، نتایج مطالعه اصلی تصحیح می‌گردد (۲۴). به همین جهت زیر نمونه‌ای به اندازه ۲۷۳ نفر از نمونه اصلی بصورت تصادفی انتخاب و با آزمایش قند خون ناشتا (fasting blood sugar) وضعیت ابتلا به دیابت بطور دقیق ثبت گردید. در آنالیز داده‌ها از نرم افزار SAS نسخه ۹,۱,۳ و procedure NLMIXED با سطح معنی داری ۰/۰۵ استفاده گردید. این مطالعه دارای تأییدیه از کمیته اخلاق دانشگاه علوم بهزیستی و توانبخشی می‌باشد.

یافته‌ها

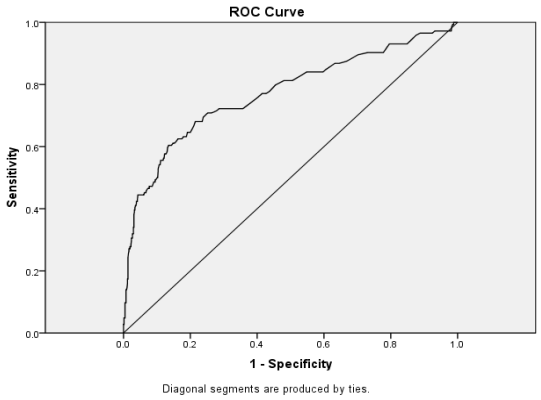
میانگین سنی افراد مورد بررسی ۴۷/۶۵ سال با انحراف معیار ۱۲/۲۳ سال بود. ۴۶۲ مرد (۵۶/۴ در صد) و ۳۵۷ زن (۴۳/۶ در صد) در این مطالعه شرکت کردند. تعداد ۶۲۹ نفر (۷۶/۸ درصد) درصد شاخص توده بدنی کمتر از ۳۰ و تعداد ۱۹۰ نفر (۲۳/۲ درصد) درصد شاخص توده بدنی بیشتر از ۳۰ داشته‌اند که افرادی که شاخص توده بدنی آن‌ها بیشتر از ۳۰ باشد دچار چاقی بودند. ابتدا داده‌ها بر حسب ابتلا به دیابت و فشار خون به یک جدول ۲×۲ تبدیل شد، ۲/۱ درصد هر دو بیماری را داشتند، ۷/۶ در صد افراد دیابت داشتند اما فشار خون نداشتند، ۱۳/۴ درصد افراد دیابت نداشتند اما فشار خون داشتند و ۷۶/۹ درصد افراد هیچکدام از دو بیماری را نداشتند. مقدار نسبت شانس برابر ۱/۵۸ با فاصله اطمینان ۹۵ درصد (۲/۷۸۷ و ۰/۸۸۵) بدست آمد در نتیجه رابطه معناداری بین فشار خون و دیابت مشاهده نشد که با توجه به مطالعات زیادی که در این زمینه انجام شده است مشخص می‌شود که متغیرهای دیابت و فشار خون، یکی از آن‌ها یا هر دو دارای خطای طبقه بندی می‌باشند (۲۸, ۲۹). که در این مطالعه فقط امکان بررسی وجود خطای طبقه بندی در متغیر پاسخ (دیابت) وجود داشت. تصحیح اریبی ناشی از طبقه بندی غلط بر اساس احتمالات طبقه بندی غلط (حساسیت (SE_Y) و ویژگی (SP_Y)) صورت گرفت. میزان حساسیت و ویژگی در حالت خطای طبقه بندی غیر افتراقی برای متغیر پاسخ (ابتلا به دیابت بر اساس قند خون معمولی (Y^*) بر حسب آزمون استاندارد قند خون ناشتا (Y) بصورت ذیل بدست آمد:

$$SE_Y = Pr(Y^* = 1|Y = 1) = \frac{18}{43} = 0.42$$

$$SP_Y = Pr(Y^* = 0|Y = 0) = \frac{204}{230} = 0.89$$

در این مطالعه چون میزان SE_Y خیلی کم اما مقدار SP_Y زیاد شد پس در نتیجه تست غربالگری دارای خطای طبقه بندی می‌باشد. بنابراین، از آنجائی‌که SE_Y این آزمون برای تشخیص دیابت فقط ۴۲ درصد است، ۵۸ درصد موارد شناسائی نمی‌شوند و همین مسئله نتیجه نهائی برنامه‌های

غربالگری را بر روی تشخیص بیماری، به شدت تحت تأثیر قرار می‌دهد. همچنین ۱۱ درصد افراد سالم نیز به اشتباه بیمار تشخیص داده می‌شوند. جهت بررسی ارزش آزمایش قند خون در تشخیص دیابت از منحنی ROC استفاده شد. بدین صورت که داده‌های قند خون ساده بر اساس داده‌های قند خون ناشتا معتبر سازی شد. سطح زیر منحنیROC نیز برای آزمایش قند خون در مطالعه اصلی ۰/۷۷۲ بدست آمد و تصویر ۱ منحنی ROC مربوطه را نشان می‌دهد. این مقدار بدان معنی است که ۷۷ درصد موارد، آزمایش نتیجه درست را تشخیص داده است و اگر تنها تستی ویژگی بالایی داشته باشد اما حساسیت آن کم باشد ملاک مناسبی به عنوان تست غربالگری نبوده و ترجیحاً هر دو میزان ویژگی و حساسیت بالا پذیرفته شده است (۳۰). در این مطالعه چون میزان حساسیت تقریباً کم اما مقدار ویژگی زیاد شد پس در نتیجه تست غربالگری دارای خطای طبقه بندی می‌باشد.



تصویر ۱: منحنی ROC آزمایش قند خون در تشخیص دیابت

همچنین مدل رگرسیون لجستیک استاندارد به داده‌ها برازش داده شد (معادله ۱) که ضرایب رگرسیونی و نسبت شانس در جدول ۱ آمده است. ضریب و نسبت شانس فشار خون در سطح خطای کمتر از ۰/۰۵ معنادار نشده بود. مقدار نسبت شانس برابر ۱/۲۵۵ با فاصله اطمینان ۹۵ درصد (۲/۲۷۸ و ۰/۶۹۱) بدست آمد چون فاصله اطمینان شامل عدد یک است در نتیجه رابطه معناداری بین فشار خون و دیابت را نشان نمی‌داد. در این مطالعه فقط امکان بررسی وجود خطای طبقه بندی در متغیر پاسخ (ابتلا به دیابت) وجود داشت. البته برای متغیرهای سن و چاقی ضرایب معنادار شد. مقدار نسبت شانس برای سن برابر ۱/۰۳۸ با فاصله اطمینان ۹۵ درصد (۱/۰۵۷ و ۱/۰۱۹) بدست آمد در نتیجه رابطه معناداری بین سن و دیابت وجود داشت و با افزایش یک سال سن شانس ابتلا به دیابت حدود ۴ درصد افزایش یافت و مقدار نسبت شانس برای چاقی برابر ۲/۰۷۳ با فاصله اطمینان ۹۵ درصد (۴/۱۳۹ و ۱/۰۳۷) بدست آمد در نتیجه رابطه معناداری بین چاقی و دیابت وجود داشت یعنی چاقی شانس ابتلا به دیابت تقریباً دو برابر افزایش می‌دهد با برازش این مدل متغیرهای جنسیت معنی دار نبود.

$$\text{logit}[(Y=1)] = \beta_0 + \beta_1 X + \beta_2 \text{AGE} + \beta_3 \text{SEX} + \beta_4 \text{OBESITY} \quad (1)$$

تصحیح اریبی ناشی از طبقه بندی غلط براساس احتمالات طبقه بندی غلط (حساسیت و ویژگی) صورت می‌گیرد. میزان حساسیت و ویژگی در حالت خطای طبقه بندی غیر افتراقی در حقیقت نشان دهنده آن هستند که مقادیر حساسیت و ویژگی از باقی متغیرها مستقل هستند. در خطای طبقه بندی غیر افتراقی مستقل همچنین خطای طبقه بندی در متغیر پاسخ مستقل از دیگر از متغیرها است.

مدل رگرسیون لجستیک برای داده‌های دارای خطای طبقه بندی غیر افتراقی برازش داده شد که ضرایب رگرسیونی و نسبت شانس در جدول ۲ آمده است. که ضریب فشار خون و نسبت شانس در سطح خطای کمتر از ۰/۰۵ معنادار شده است. مقدار نسبت شانس ابتلا به دیابت در افراد دارای فشار خون به افرادی که فشار خون ندارند برابر ۳/۰۶۲ با فاصله اطمینان ۹۵ درصد (۱/۲۵۲ و ۷/۴۹۵) بدست آمد در نتیجه رابطه معناداری بین فشار خون و دیابت وجود داشت شانس ابتلا به دیابت در افراد دارای فشار خون بالا ۳/۰۶۲ برابر بیشتر از افرادی بود که فشار خون بالا نداشتند و مقدار نسبت شانس ابتلا به دیابت برای سن برابر ۱/۰۶۶ با فاصله اطمینان ۹۵ درصد (۱/۰۳۱ و ۱/۱۰۳) بدست آمد در نتیجه رابطه معناداری بین سن و دیابت وجود داشت و با افزایش سن شانس ابتلا به دیابت افزایش می‌یافت و مقدار نسبت شانس ابتلا به دیابت برای جنسیت برابر ۰/۵۷۴ با فاصله اطمینان ۹۵ درصد (۰/۲۶۴ و ۱/۲۴۴) بدست آمد در نتیجه رابطه معناداری بین جنسیت و دیابت وجود نداشت. و مقدار نسبت شانس ابتلا به دیابت برای چاقی برابر ۲/۴۷۶ با فاصله اطمینان ۹۵ درصد (۱/۲۳۱ و ۷/۵۶۰) بدست آمد در نتیجه رابطه معناداری بین چاقی و دیابت وجود داشت. در افراد چاق شانس ابتلا به دیابت ۲/۴۷۶ برابر بیشتر است. مقدار شاخص آکائیکه برای این مدل برابر ۵۶۶۶ بدست آمد. در مقابل خطای طبقه بندی غیر افتراقی، خطای طبقه بندی افتراقی زمانی اتفاق می افتد که احتمال خطای طبقه بندی یکی از متغیرها به مقادیر متغیرهای دیگر وابسته است. توجه شود که حساسی و ویژگی متغیر پاسخ بصورت تابعی از متغیرهای مستقل است و در اینجا دوباره از مدل رگرسیون لجستیک استفاده شده است.

جدول ۱: نتایج رگرسیون لجستیک برای دیابت نوع ۲ به عنوان متغیر پاسخ در مطالعه اصلی به تعداد ۸۱۹ نفر

متغیرها	ضریب (β)	انحراف معیار	نسبت شانس (OR)	فاصله اطمینان ۹۵ درصد برای نسبت شانس		مقدار احتمال
				کران پایین	کران بالا	
فشارخون*	۰/۲۲۷	۰/۳۰۴	۱/۲۵۵	۰/۶۹۱	۲/۲۷۸	۰/۴۵۶
سن	۰/۰۳۷	۰/۰۰۹	۱/۰۳۸	۱/۰۱۹	۱/۰۵۷	< ۰/۰۰۱
جنسیت	-۰/۰۲۵	۰/۲۴۵	۰/۹۷۵	۰/۶۰۳	۱/۵۷۶	۰/۹۱۸
چاقی	۰/۷۲۸	۰/۳۵۳	۲/۰۷۳	۱/۰۳۷	۴/۱۳۹	۰/۰۳۹

* رده مرجع عدم فشار خون است

+ رده مرجع جنسیت مرد است

جدول ۲: نتایج رگرسیون لجستیک برای تعدیل ضرایب داده‌های دارای خطای طبقه بندی غیرافتراقی

متغیرها	ضریب (β)	انحراف معیار	نسبت شانس (OR)	فاصله اطمینان ۹۵ درصد برای نسبت شانس		مقدار احتمال
				کران پایین	کران بالا	
فشارخون*	۱/۱۱۹	۰/۴۵۵	۳/۰۶۲	۱/۲۵۲	۷/۴۹۵	۰/۰۱۵
سن	۰/۰۶۴	۰/۰۱۷	۱/۰۶۶	۱/۰۳۱	۱/۱۰۳	۰/۰۰۲
جنسیت+	-۰/۵۵۵	۰/۳۹۴	۰/۵۷۴	۰/۲۵۶	۱/۲۴۴	۰/۱۵۹
چاقی	۰/۹۰۷	۰/۵۶۸	۲/۴۷۶	۱/۲۳۱	۷/۵۶۰	۰/۰۱۱

* رده مرجع عدم فشار خون است

+ رده مرجع جنسیت مرد است

جدول ۳: نتایج رگرسیون لجستیک برای تعدیل ضرایب داده‌های دارای خطای طبقه بندی افتراقی

متغیرها	ضریب (β)	انحراف معیار	نسبت شانس (OR)	فاصله اطمینان ۹۵ درصد برای نسبت شانس		مقدار احتمال
				کران پایین	کران بالا	
فشارخون	۱/۱۲۰	۰/۲۷۱	۳/۰۶۴	۱/۸۰۲	۵/۲۱۳	< ۰/۰۰۱
سن	۰/۰۳۸	۰/۰۰۹	۱/۰۳۹	۱/۰۲۱	۱/۰۵۷	< ۰/۰۰۱
جنسیت	-۰/۴۱۹	۰/۲۱۶	۰/۶۵۷	۰/۴۳۰	۱/۰۰۵	۰/۰۵۳
چاقی	۰/۶۸۹	۰/۲۳۳	۱/۹۹۲	۱/۲۶۰	۰/۱۴۹	۰/۰۰۳

* رده مرجع عدم فشار خون است

+ رده مرجع جنسیت مرد است

جدول ۳ برآورد پارامترها، سطح معنی داری ضرایب متغیرهای توضیحی و نسبتهای شانس را نشان می‌دهد. با توجه به نتایج رگرسیون لجستیک برای تعدیل ضرایب داده‌های دارای خطای طبقه بندی افتراقی، نشان داد که ضریب فشار خون و نسبت شانس در سطح خطای کمتر از ۰/۰۵ معنادار شده است. مقدار نسبت شانس ابتلا به دیابت در افراد دارای فشار خون به افرادی که فشار خون ندارند برابر ۳/۰۶۴ با فاصله اطمینان ۹۵ درصد (۱/۸۰۲ و ۵/۲۱۳) بدست آمد در نتیجه رابطه معناداری بین فشار خون و دیابت وجود داشت. شانس ابتلا به دیابت در افراد دارای فشار خون بالا ۳/۰۶۴ برابر بیشتر از افرادی بود که فشار خون بالا نداشتند و مقدار نسبت شانس ابتلا به دیابت برای سن برابر ۱/۰۳۹ با فاصله اطمینان ۹۵ درصد (۱/۰۲۱ و ۱/۰۵۷) بدست آمد در نتیجه رابطه معناداری بین سن و دیابت وجود داشت و با افزایش یک سال سن شانس ابتلا به دیابت حدود ۴ درصد افزایش می‌یافت و مقدار نسبت شانس ابتلا به دیابت برای جنسیت برابر ۰/۶۵۷ با فاصله اطمینان ۹۵ درصد (۰/۴۳۰ و ۱/۰۰۵) بدست آمد در نتیجه رابطه معناداری بین جنسیت و دیابت وجود نداشت و مقدار نسبت شانس ابتلا به دیابت برای متغیر چاقی برابر ۱/۹۹۲ با فاصله اطمینان ۹۵ درصد (۱/۲۶۰ و ۳/۱۴۹) بدست آمد در افراد چاق شانس ابتلا به دیابت ۱/۹۹۲ برابر بیشتر است. مقدار شاخص آکائیکه برای این مدل برابر ۵۴۳۲/۲ بدست آمد. چون ضرایب برای مدل خطای طبقه بندی افتراقی و همچنین غیر افتراقی معنادار شد و همچنین ضرایب در دو مدل به یکدیگر نزدیک بودند پس برای انتخاب مدل مناسب از شاخص آکائیکه استفاده شد. شاخص آکائیکه برای مدل با فرض افتراقی بودن خطای طبقه بندی برابر ۵۴۳۲/۲ و برای مدل با فرض غیرافتراقی بودن خطای طبقه بندی برابر ۵۶۶۶ شد در نتیجه چون مقدار این شاخص برای مدل افتراقی کمتر شد پس خطای طبقه بندی از نوع افتراقی در نظر گرفته شد.

بحث

در مطالعه اصلی این پژوهش رابطه بین فشارخون با ابتلا به دیابت معنادار مشاهده نشد و حساسیت و ویژگی نیز در هر دو حالت فرض افتراقی و غیر افتراقی کامل نبودند و در حالت فرض افتراقی بودن خطای طبقه متغیر پاسخ در سطوح متغیر فشار خون میزان حساسیت تفاوت داشت که فرض افتراقی بودن خطای طبقه بندی را نشان می‌داد یعنی در گروهی که افراد دارای فشار خون بودند نتیجه آزمایش قند خون آن‌ها دارای خطای بیشتری بود پس از بکار بردن یک روش رگرسیون لجستیک با فرض افتراقی و غیر افتراقی بودن خطای طبقه بندی غلط مشاهده نمودیم نسبت شانس تعدیل یافته از فرض صفر ۱ = OR فاصله گرفته و رابطه بین فشارخون و دیابت معنادار شد. اطلاعات طرح غربالگری دیابت مرکز بهداشت زاهدان دارای حساسیت پایین بودند که برای تعیین وضعیت ابتلا به دیابت قابل استناد نبودند. استفاده از این اطلاعات موجب طبقه بندی غلط داده‌ها و برآوردهای اریب از پارامترهای مورد نظر می‌شد. برای تعیین درست وضعیت ابتلا به دیابت باید از روش‌های دقیق‌تری استفاده نمود و یا اینکه برای تصحیح خطای ناشی از طبقه بندی غلط باید از روش‌های مناسب آماری همانند روش رگرسیون لجستیک برای داده‌های دارای خطای طبقه بندی بهره گرفت. در پژوهش Lyles و همکاران در سال ۲۰۱۱ نیز پس از استفاده از رگرسیون لجستیک رابطه متغیر وابسته دارای خطای طبقه بندی (CLIN: clinically-based diagnosis) با متغیر ابتلا به ایدز معنی دار شد (۳۱) و همچنین Tang و همکاران در مطالعه‌ای در سال ۲۰۱۵ نیز با استفاده از از استفاده از رگرسیون لجستیک برای داده‌های دارای خطای طبقه بندی رابطه متغیر Bacterial Vaginosis (BV) status با متغیر ابتلا به ایدز معنی دار شد (۱۳). Zawistowski و همکاران در پژوهشی در سال ۲۰۱۷ با عنوان تجزیه و تحلیل ROC اصلاح شده برای متغیر پاسخ دو حالتی دارای خطای طبقه بندی با بررسی سطح زیر منحنی

ROC به بررسی وجود خطای طبقه بندی در متغیر پاسخ پرداختند، که در این مطالعه نیز منحنی ROC برای داده‌ها رسم شد و چون سطح زیر منحنی برابر ۰/۷۷۲ بدست آمد و از یک فاصله داشت، احتمال وجود خطای طبقه بندی در متغیر پاسخ وجود داشت (۳۲). نوری و همکاران در سال ۲۰۱۴ پژوهشی با هدف بکارگیری یک روش برای تصحیح اریبی حاصل از خطای طبقه بندی غلط متغیر مستقل گسسته وضعیت مصرف سیگار، در رگرسیون لجستیک برای تحلیل عوامل خطر بیماری سکته قلبی انجام و انواع روش‌ها را برای تصحیح اریبی ناشی از خطای طبقه بندی را انجام داد از جمله از روش ماتریس و روش بیزی و روش رگرسیون لجستیک استفاده کرد که این مطالعه نیز از نظر استفاده از روش رگرسیون لجستیک مشابه این مطالعه بود (۱۵). Jurek و همکاران در سال ۲۰۱۳ در پژوهشی با عنوان تعدیل خطای طبقه بندی برای متغیرهای چند حالته در یک مطالعه با استفاده از گواهی تولد به برر سی ارتباط بین سیگار کشیدن مادران و چاقی و قاعدگی براساس اطلاعات گواهی تولد پرداختند و با تعدیل نسبت شانس به تعدیل خطای طبقه بندی در متغیر پاسخ پرداختند و همچنین به بررسی وجود خطای طبقه بندی هم در متغیر پاسخ و هم در متغیر مواجهه بصورت جداگانه پرداختند که از نظر روش مشابه این مطالعه بود (۳۳). Hill و همکاران بیان داشتند که فشار خون در ۷۰ درصد بیماران دیابتی دیده می‌شود و خطر پیشرفت دیابت در افراد با فشار خون ۲ برابر بیشتر می‌باشد. چاقی، فشار خون و دیابت بطور وسیعی با یکدیگر همراهی می‌کنند، به عبارت دیگر، بیش از ۷۰ درصد افراد دیابتی مبتلا به فشار خون هستند و بیماری‌های قلبی – عروقی در ۷۵ درصد دیابتی‌ها به دنبال فشار خون ایجاد می‌شود. به خوبی مشخص نشده که چند سال قبل از شروع دیابت فشار خون افزایش پیدا می‌کند. فشار خون بالا یک عامل خطر برای عوارض دیابت بوده و وجود آن قبل از شروع دیابت می‌تواند شیوع بالای بیماری‌های قلبی- عروقی در زمان تشخیص دیابت را توضیح دهد. بنابراین فشار خون بالا قبل از شروع دیابت نوع ۲ یک هدف بالقوه در پیشگیری اولیه از عوارض دیابت است (۳۴). در این پژوهش پس تعدیل خطای طبقه بندی ارتباط بین دیابت و فشار خون معنی دار شد و همچنین انجمن دیابت آمریکا بیان می‌کند که فشار خون یکی از بیماری‌های شایع همراه با دیابت است (۳۵). در این پژوهش بدون تعدیل خطای طبقه بندی رابطه بین ابتلا به دیابت و فشار خون بالا معنی دار نشد. Abadi و همکاران نیز نشان دادند که عدم توجه به احتمالات طبقه بندی نادرست منجر به برآوردهای غیر واقع بینانه از نسبت شانس و در نتیجه تحلیل‌های غیر واقعی از ارتباط میان متغیرها می‌گردد. برای تصحیح خطای طبقه بندی نادرست ابتدا لازم است احتمالات طبقه بندی نادرست در نمونه اصلی بررسی و براساس آن از روش تصیح احتمالات طبقه بندی نادرست افتراقی یا غیرافتراقی استفاده شود. عدم حتمیت در مورد ضرایب تصحیح احتمالات طبقه بندی نادرست منجر به برآوردهایی با پراکندگی بیش‌تر می‌شود (۲۸).

نتیجه گیری

اطلاعات طرح غربالگری دیابت مرکز بهداشت زاهدان دارای حساسیت پایین و ویژگی بالایی می‌باشند و برای تعیین وضعیت ابتلا به دیابت قابل استناد نیست. استفاده از این اطلاعات موجب طبقه بندی غلط داده‌ها و برآوردهای اریب از پارامترهای مورد نظر می‌شود. برای تعیین درست وضعیت مواجهه باید از روش‌های دقیق‌تری استفاده نمود و یا اینکه برای تصحیح خطای ناشی از طبقه بندی غلط باید از روش‌های مناسب آماری همانند روش رگرسیون لجستیک برای داده‌های دارای خطای طبقه بندی بهره گرفت. در این مطالعه مشخص شد که بدون توجه به تصحیح خطای طبقه بندی غلط، رابطه معناداری بین فشار خون و دیابت نبود. با تصحیح براساس روش رگرسیون لجستیک، نسبت شانس ابتلا به دیابت در افراد دارای فشار خون بالا ۳/۰۶۴ برابر بیشتر از افرادی بود که فشار خون بالا نداشتند. پیشنهاد می‌شود که از روش‌های دیگری که برای تعدیل خطای طبقه بندی استفاده می‌شوند مانند روشهای ماتریسی و روش‌های بیزی نیز استفاده شود و همچنین روش‌های مذکور با یکدیگر مقایسه شوند. پیشنهاد می‌شود که از روش رگرسیون لجستیگ برای تعدیل خطای طبقه بندی در داده‌های رسته‌ای بخصوص داده‌های مربوط به حوزه بهداشت و درمان و طرح‌های غربالگری استفاده شود. از محدودیت‌های مطالعه این بود که با توجه به اینکه طرح غربالگری بصورت مقطعی و در مدت زمان کوتاه انجام شد در نتیجه امکان خطا در مرحله گرفتن آزمایش و همچنین ثبت داده‌ها وجود داشت که از ۱۵۵۰ نفر مورد بررسی فقط ۸۱۹ نفر دارای اطلاعات کامل بودند.

سپاسگزاری

این مقاله برگرفته از پایان نامه کارشناسی ارشد آمار زیستی دانشجو مریم رستگار به راهنمایی آقای دکتر عنایت اله بخشی مصوب ۱۳۹۵ دانشگاه علوم بهزیستی و توانبخشی با کد اخلاق IR.USWR.REC.1395.238 می‌باشد. از کارکنان مرکز بهداشت زاهدان که در جمع آوری داده‌ها یاری نمودند، تشکر می‌شود.

References

1. WHO. Diabetes Programme: World Health Organization; 2017 [updated 2017; cited 2017]. Available from: http://www.who.int/diabetes/en.
2. Ahanchi N, Slami A, Sharifirad G. [Effects of Family-based Theory of social support on perceived support levels in type 2 diabetic patients]. Health Syst Res. 2012;8(5):757-64.
3. WHO. Diabetes country profiles: World Health Organization; 2016 [updated 2017; cited 2017]. Available from: http://www.who.int/diabetes/country-profiles/irn_en.pdf?ua=1.
4. Delavari A, Mahdavi Hazaveh A. [Planning of diabetes control in Iran]. Tehran: Ministry of Health & Medical Education Undersecretary for Health Disease Management Center; 2004.
5. Rewers M, Ludvigsson J. Environmental risk factors for type 1 diabetes. Lancet. 2016;387(10035):2340-8. DOI: 10.1016/S0140-6736(16)30507-4 PMID: 27302273
6. Althouse AD, Abbott JD, Forker AD, Bertolet M, Barinas-Mitchell E, Thurston RC, et al. Risk factors for incident peripheral arterial disease in type 2

diabetes: results from the Bypass Angioplasty Revascularization Investigation in type 2 Diabetes (BARI 2D) Trial. Diabetes Care. 2014;37(5):1346-52. DOI: 10.2337/dc13-2303 PMID: 24595631
7. Grover SA, Kaouache M, Rempel P, Joseph L, Dawes M, Lau DC, et al. Years of life lost and healthy life-years lost from diabetes and cardiovascular disease in overweight and obese people: a modelling study. Lancet Diabetes Endocrinol. 2015;3(2):114-22. DOI: 10.1016/S2213-8587(14)70229-3 PMID: 25483220
8. Razi S, Sadeghi M, Nasrabadi A, Ebrahimi H, Kazemnejad A. [The effect of family-centered empowerment model on knowledge and metabolic control of patients with type 2 diabetes]. Knowledge Health. 2014;9(1):48-54.
9. Larejani B, Zahedi F. Epidemiology of diabetes mellitus in Iran. Iranian J Diabetes Metab. 2001;1(1):1-8.
10. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med. 2002;346(6):393-403. DOI: 10.1056/NEJMoa012512 PMID: 11832527
11. Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. Arch Intern Med. 1998;158(2):166-72. PMID: 9448555
12. Lee M, Entzminger L, Lohsoonthorn V, Williams MA. Risk factors of hypertension and correlates of blood pressure and mean arterial pressure among patients receiving health exams at the Preventive Medicine Clinic, King Chulalongkorn Memorial Hospital, Thailand. J Med Assoc Thai. 2006;89(8):1213-21. PMID: 17048432
13. Tang L, Lyles RH, King CC, Celentano DD, Lo Y. Binary regression with differentially misclassified response and exposure variables. Stat Med. 2015;34 (9):1605-20. DOI: 10.1002/sim.6440 PMID: 25652841
14. Duffy SW, Warwick J, Williams AR, Keshavarz H, Kaffashian F, Rohan TE, et al. A simple model for potential use with a misclassified binary outcome in epidemiology. J Epidemiol Community Health. 2004; 58(8):712-7. DOI: 10.1136/jech.2003.010546 PMID: 15252078
15. Nouri B, Zare N, Abadi AR, Ayatolahi SMT. Correction the bias of odds ratio of misclassified smoking and myocardial infarction in shahid Moddares hospital. J North Khorasan Univ Med Sci. 2014;6(2):451-8. DOI: 10.29252/jnkums.6.2.451
16. Abadi A, Mohammad K, Meshkani M, Kazemnejad A, Mehrabi Y, Azizi F. Analysis of angina pectoris status based on the probability of misclassifying the risk factor: Tehran lipid and glucose study. J Sch Public Health Inst Public Health Res. 2004;2(1):19-26.
17. Gordis L. Epidemiology. 3rd ed. Philadelphia: W.B Saunders; 2006.
18. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. Am J Epidemiol. 1993;137(11):1251-8. PMID: 8322765
19. Bross I. Misclassification in 2 X 2 Tables. Biometrics. 1954;10(4):478. DOI: 10.2307/3001619
20. Chen TT. A review of methods for misclassified categorical data in epidemiology. Stat Med. 1989; 8(9):1095-106; discussion 107-8. PMID: 2678350
21. Tenenbein A. A Double Sampling Scheme for Estimating from Misclassified Multinomial Data with Applications to Sampling Inspection. Technometrics. 1972;14(1):187-202. DOI: 10.1080/00401706.1972.10488895
22. Roy S, Banerjee T, Maiti T. Measurement error model for misclassified binary responses. Stat Med. 2005; 24(2):269-83. DOI: 10.1002/sim.1886 PMID: 15546132
23. Liu X, Liang K. Adjustment for no differential misclassification error in the generalized linear model: Statistics in Medicine, Wiley Online Library; 1991.
24. Neuhaus J. Bias and efficiency loss due to misclassified responses in binary regression. Biometrika. 1999;86(4):843-55. DOI: 10.1093/biomet/86.4.843
25. Davidov O, Faraggi D, Reiser B. Misclassification in Logistic Regression with Discrete Covariates. Biometr J. 2003;45(5):541-53. DOI: 10.1002/bimj.200390031
26. Luan X, Pan W, Gerberich SG, Carlin BP. Does it always help to adjust for misclassification of a binary outcome in logistic regression? Stat Med. 2005;24(14):2221-34. DOI: 10.1002/sim.2094 PMID: 15889454
27. Kuchenhoff H, Mwalili SM, Lesaffre E. A general method for dealing with misclassification in regression: the misclassification SIMEX. Biometrics. 2006;62(1):85-96. DOI: 10.1111/j.1541-0420.2005.00396.x PMID: 16542233
28. Abadi A, Mohammad K, Moshkani M, Kazem Nejad A, Mehrabi Y. [Analysis of case control studies in the presence of misclassification]. Daneshvar Med. 2004;12(54):1-10.
29. Nouri B, Zare N, Abadi AR, Ayatolahi SMT. [The relationship between myocardial involution and smoking consumption; Oddity adjustment of odds ratio due to misleading exposure]. J North Khorasan Univ Med Sci. 2014;6(2):451-8. DOI: 10.29252/jnkums.6.2.451
30. Rey E, Hudon L, Michon N, Boucher P, Ethier J, Saint-Louis P. Fasting plasma glucose versus glucose challenge test: screening for gestational diabetes and cost effectiveness. Clin Biochem. 2004;37(9):780-4. DOI: 10.1016/j.clinbiochem.2004.05.018 PMID: 15329316
31. Lyles RH, Tang L, Superak HM, King CC, Celentano DD, Lo Y, et al. Validation data-based adjustments for

outcome misclassification in logistic regression: an illustration. Epidemiology. 2011;22(4):589-97. DOI: 10.1097/EDE.0b013e3182117c85 PMID: 21487295

32. Zawistowski M, Sussman JB, Hofer TP, Bentley D, Hayward RA, Wiitala WL. Corrected ROC analysis for misclassified binary outcomes. Stat Med. 2017; 36(13):2148-60. DOI: 10.1002/sim.7260 PMID: 28245528
33. Jurek AM, Maldonado G, Greenland S. Adjusting for outcome misclassification: the importance of accounting for case-control sampling and other forms of outcome-related selection. Ann Epidemiol. 2013;23(3):129-35. DOI: 10.1016/j.annepidem.2012.12.007 PMID: 23332712
34. Golden SH, Wang NY, Klag MJ, Meoni LA, Brancati FL. Blood pressure in young adulthood and the risk of type 2 diabetes in middle age. Diabetes Care. 2003; 26(4):1110-5. PMID: 12663582
35. American Diabetes A. Standards of medical care in diabetes--2011. Diabetes Care. 2011;34 Suppl 1(Supplement_1):S11-61. DOI: 10.2337/dc11-S011 PMID: 21193625