

## تحلیل وضعیت آئزین صدری بر اساس احتمالات طبقه‌بندی نادرست عامل خطر سیگار در مطالعه قند و لیپید تهران ۷۹-۱۳۷۸

علیرضا ابدی\*<sup>۱</sup>، دکتر کاظم محمد<sup>۲</sup>، دکتر محمدرضا مشکانی<sup>۳</sup>، دکتر انوشیروان کاظم نژاد<sup>۴</sup>، دکتر یدالله محرابی<sup>۵</sup>  
و دکتر فریدون عزیزی<sup>۵</sup>

### چکیده:

خطای طبقه‌بندی نادرست بیماری یا عوامل خطر از منابع اصلی ارایی در مطالعات می‌باشد. در مطالعات گذشته‌نگر با تقسیم‌بندی غلط افراد به گروه‌های مواجهه یافته و مواجهه نیافته این خطا در مطالعه و نتایج آن وارد می‌شود. هدف از این بررسی نشان دادن تاثیر این خطا در محاسبه نسبت بخت (نسبت برتری) در مطالعات گذشته‌نگر و ارائه روش تصحیح آن می‌باشد.

بر اساس اطلاعات حاصل از اجرای مرحله اول مطالعه "قند و لیپید تهران" داده‌های مربوط به ۳۳۳۲ مرد بالای ۳۰ سال برای بررسی ارتباط میان وضعیت مصرف سیگار و آئزین صدری مورد استفاده قرار گرفته است. به منظور تعیین وجود آئزین صدری از پرسشنامه استاندارد Rose استفاده شده است. با توجه به دو تعریف مختلف از مصرف سیگار (سیگاری فعال - غیر سیگاری) و (سیگاری و ترک کرده - غیرسیگاری) به بررسی ارتباط میان متغیرها پرداخته شده است.

بر اساس تعریف اولیه از مصرف سیگار که در بسیاری تحقیقات پرسشگری از وضعیت کنونی مصرف سیگار است نسبت بخت ابتلا به آئزین صدری در سیگاریها ۱/۰۹ (۱/۴۱ - ۰/۸۴ CI: ۰/۹۵٪) برابر غیر سیگاریهاست که اختلاف معناداری را نشان نمی‌دهد. پس از تصحیح داده‌ها با توجه به احتمالات طبقه‌بندی نادرست غیر افتراقی نسبت بخت به ۱/۱۳ (۱/۴۳ - ۰/۸۹ CI: ۰/۹۵٪) افزایش یافته که همچنان اختلاف معناداری میان نسبت بخت آئزین صدری برای افراد سیگاری بالنسبه به افراد غیر سیگاری را نشان نمی‌دهد. پس از تصحیح داده‌ها با توجه به احتمالات طبقه‌بندی نادرست افتراقی نسبت بخت به ۱/۴۶ (۱/۸۵ - ۱/۱۵ CI: ۰/۹۵٪) افزایش می‌یابد که موید اختلاف معنادار میان دو گروه است.

طبقه‌بندی غلط از منابع ارایی است که موجب تحلیل نادرست از روابط بین متغیرها می‌گردد. نتایج حاصل از این مطالعه نشان می‌دهد که در صورت وجود احتمالات طبقه‌بندی نادرست افتراقی استفاده از روشهای تصحیح غیر افتراقی کارایی نخواهد داشت.

**واژگان کلیدی:** آئزین صدری، مصرف سیگار، نسبت بخت، احتمالات طبقه‌بندی نادرست

\* (عهده دارمکاتبات)

۱. گروه آمار زیستی دانشکده پزشکی دانشگاه تربیت مدرس
۲. گروه اپیدمیولوژی و آمار زیستی دانشکده بهداشت و انستیتو تحقیقات بهداشتی، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران
۳. گروه آمار دانشکده آمار و ریاضی، دانشگاه شهید بهشتی
۴. گروه بهداشت و پزشکی اجتماعی دانشکده پزشکی دانشگاه علوم پزشکی شهیدبهشتی
۵. مرکز تحقیقات غدد درون ریز و متابولیسم دانشگاه علوم پزشکی شهیدبهشتی

## مقدمه:

تمایل افراد به اعلام حقیقت در مورد در معرض عامل خطر بودن (مثلاً سیگاری بودن) یا تعاریف نادرست از مواجهه داشتن با عامل خطر باعث گروه بندی غلط افراد گردد.

براساس تعریفی که سازمان بهداشت جهانی ارائه کرده کسی که در هفته حداقل ۷ نخ سیگار می کشد سیگاری فعال و کمتر از آنرا سیگاری غیرفعال و کسی را که در گذشته سیگاری بوده و در زمان مصاحبه مصرف نداشته باشد، ترک کرده و کسی را که در گذشته و حال سیگار نمی کشیده را غیر سیگاری نامیده اند. در بسیاری مطالعات بدون توجه به این تقسیم بندی دقیق فقط سیگاری بودن یا نبودن در زمان مصاحبه را مورد توجه قرار می دهند که گروه بندی نادرستی از افراد را به دنبال خواهد داشت.

در این مطالعه تاثیر به کار گیری تعریف نادقیق از مصرف سیگار در تحلیل ارتباط آن با آئزین صدی مورد توجه قرار گرفته است.

احتمالات طبقه بندی نادرست با حساسیت (Sensitivity) و ویژگی (Specificity)، اندازه گیری و از آنها برای تصحیح طبقه بندی نادرست استفاده می شود. به منظور برآورد این احتمالات که ضرایب تصحیح نیز می باشند از داده های کمکی همچون داده های معتبر یا اندازه گیریهای مکرر استفاده نموده تا ضمن تعیین احتمالات طبقه بندی غلط (حساسیت و ویژگی) بتوانیم برآوردهای اصلاح شده را به دست آوریم (Wachholder S. et al. 1993).

موضوع طبقه بندی نادرست ابتدا توسط (Bross I. 1954) مورد توجه قرار گرفت و تاریخچه موضوع توسط (Chen T.T. 1989) ارائه شده است.

## مواد و روشها:

با هدف بررسی عوامل اصلی بروز بیماریهای قلبی-عروقی چون اختلالات متابولیکی، سطح بالای کلسترول، فزونی فشار خون و استعمال دخانیات، مطالعه آینده نگر "قند و لیپید تهران" در مرکز تحقیقات غدد درون ریز و متابولیسم دانشگاه علوم پزشکی شهید بهشتی طراحی و اولین مرحله جمع آوری داده ها در سالهای ۷۸-۷۹

اختلالات شریانی بزرگترین تهدید کننده افراد در اجتماع کنونی ماست. عوامل زیادی باعث ازدیاد این خطر می شود که سیگار کشیدن از مهمترین آنهاست. کشیدن سیگار ترشح کانکولامین ها را از غدد آدرنال افزایش می دهد این عوامل در بروز ترومبوز دخالت دارند و غلظت لیپیدها را افزایش می دهند و تشکیل آتروم را تسریع می کنند و همچنین آرتیمی ها را زیاد می کنند (حلم سرشت و دل پیشه (بی تاریخ)).

در مطالعه ۲۰ ساله بروز حمله قلبی و مرگ و میر ناشی از آن در اسکاتلند مصرف سیگار به عنوان یکی از عوامل تشدید کننده بروز بیماری معرفی شده است (Carole L. and Mart M.A. 2000).

تقریباً در اکثر مطالعاتی که به بررسی عوامل موثر در بروز بیماریهای قلبی پرداخته شده مصرف سیگار را جز عوامل موثر معرفی نموده اند.

آئزین صدی درد فشار دهنده سینه است که به واسطه فعالیت بدنی یا استرس ممکن است ایجاد شود. درد سینه یک علامت غیر عینی است که روش تشخیص قطعی و استاندارد طلایی برای آن وجود نداشته و به همین منظور Rose اول بار در سال ۱۹۶۲ به منظور یکسان سازی اندازه گیریها و استاندارد کردن تشخیص آئزین در مطالعات اپیدمیولوژیک پرسشنامه ای را طراحی کرده که در این مطالعه نیز برای تعیین آئزین صدی و درجات مختلف از آن استفاده شده است.

خطای اندازه گیری متغیرها با مقیاس اسمی یا رتبه ای خطای طبقه بندی نادرست (Misclassification error) نامیده می شود. این خطا در مطالعات گذشته نگر که در دو گروه مورد و شاهد به دنبال تعیین مواجهه با عامل خطر می باشیم با انتساب نادرست افراد به گروه مواجهه یافته یا مواجهه نیافته همراه است. در این مطالعات در معرض عامل خطر بودن غالباً از طریق پرسش از افراد و براساس پاسخ و حافظه آنها تعیین می شود که احتمال طبقه بندی غلط را به دنبال دارد. همچنین ممکن است عدم

در جداول ۲ و ۱ وضعیت آئزین صدی با دو تعریف مختلف از مصرف سیگار که در تعریف اول فقط وضع کنونی افراد بررسی شده و در تعریف دوم گروه ترک کرده نیز همراه افراد سیگاری گروه بندی شده اند بررسی شده با توجه به اطلاعات فوق ملاحظه می شود که  $OR = 1/09$  با فاصله اطمینان  $(CI: 0/84 - 1/41)$  :  $95\%$  مؤید اثر معنی داری بخت آئزین صدی برای افراد سیگاری نسبت به افراد غیر سیگاری نمی باشد. براساس این تعریف اولیه افراد ترک کرده نیز در گروه غیر سیگاری قرار گرفته اند که می تواند به مخدوش شدن نتایج بینجامد. با قرار دادن افراد ترک کرده در گروه سیگاری از آنجا که همانند افراد سیگاری در معرض عامل خطر بوده اند ملاحظه می شود که  $OR = 1/48$  با فاصله اطمینان  $(CI: 1/16 - 1/88)$  :  $95\%$  مؤید ریسک بالاتر افراد سیگاری برای آئزین صدی در مقایسه با افراد غیر سیگاری می باشد. به منظور بررسی دقیقتر این ارتباط جدول ۳ را ملاحظه می کنیم.

با توجه به داده های جدول ۳ ملاحظه می شود که گروه ترک کرده بیشترین سهم را در گروه دارای آئزین صدی دارند. نسبت بخت گروهها نسبت به گروه غیر سیگاری نیز مؤید این معنا می باشد

$$OR(I, III) = 1/29 \quad OR(II, III) = 1/82$$

این نتایج نشان می دهد که تحلیل اطلاعات بدون توجه به سابقه مصرف سیگار نتایج مخدوش کننده به همراه خواهد داشت. در مطالعاتی که فقط وضعیت کنونی مصرف سیگار بررسی شده برای تصحیح اطلاعات نیاز به مطالعه معتبر در نمونه کوچکتر وجود دارد. در این بررسی چون این مطالعه معتبر روی کل افراد نمونه واقع شده در نتیجه ضرایب تصحیح با خطای بسیار کمی ما را به داده های واقعی خواهند رساند، به منظور تعیین ضرایب تصحیح دو روش نسبت به هم در دو گروه بیمار و غیر بیمار از اطلاعات جدول (۱) بعد استفاده می نماییم.

صورت پذیرفته است. در این بررسی ۱۵۰۰۵ نفر مرد و زن ۶۹-۳ سال ساکن در منطقه ۱۳ تهران براساس مصاحبه و معاینه مورد بررسی قرار گرفته اند. در این مقاله براساس داده های حاصل از اجرای مرحله اول ارتباط میان آئزین صدی و مصرف سیگار بررسی شده است. به علت پایین بودن میزان افراد سیگاری در خانمها و پایین بودن میزان آئزین صدی در افراد زیر ۳۰ سال این بررسی روی مردان بالای ۳۰ سال صورت پذیرفته است که تعداد آنها ۳۳۳۲ نفر بود. در این بررسی ۲۷٪ سیگاری فعال و ۱۴٪ ترک کرده و ۵۹٪ غیر سیگاری بوده اند. همچنین ۸/۹٪ براساس ملاک پرسشنامه رز دارای آئزین صدی تشخیص داده شده اند.

به منظور اصلاح خطای طبقه بندی نادرست می توان از روشهای ماتریس، بیزی و مدل سازی استفاده نمود (Armitage P. 1998). در این بررسی از روش ماتریس برای تصحیح استفاده شده است (Morrissey M. and Spicyelmon D. 1999).

در این روش با استفاده از ضرایب تصحیح به دست آمده از مطالعه معتبر اطلاعات به دست آمده از مطالعه اصلی اصلاح شده است. چگونگی اصلاح بدین ترتیب است که اگر  $[X^*]$  را ماتریس مشاهدات و  $[X]$  را ماتریس مقادیر واقعی که از آن اطلاعی نداریم و  $[A]$  را ماتریس ضرایب تصحیح بنامیم خواهیم داشت:

$$[A][X]=[X^*] \rightarrow [X]=[A^{-1}][X^*] \quad (1)$$

و به این صورت برآوردی از وضعیت واقعی خواهیم داشت.

## یافته ها:

این بررسی در گروه مردان بالای ۳۰ سال با میانگین سن ۴۸/۹ و انحراف معیار ۱۳/۴ صورت پذیرفته است. در بررسی ارتباط میان آئزین صدی و مصرف سیگار تعاریف مختلفی از جمله مصرف سیگار در زمان حال، مصرف سیگار با توجه به سوابق و در معرض دود سیگار بودن قابل بررسی است.

همان گونه که ملاحظه می‌شود حساسیت روش ۰/۶۴٪ است و این حساسیت در گروه بیمار ۰/۵۷٪ و در گروه غیربیمار ۰/۶۵٪ است ویژگی روش در هر سه گروه ۰/۱۰۰٪ است.

در صورتی که برای تصحیح از روش خطای طبقه‌بندی غیرافتراقی (Non differential Misclassification) استفاده نماییم ماتریس A با ضرایب  $q = 1$  و  $p = 0/64$  در نظر گرفته می‌شود. در این حالت ماتریس حاصل از روابط بالا به صورت زیر حاصل می‌شود.

$$X = \begin{bmatrix} 141/96 & 153/04 \\ 1371/24 & 1665/76 \end{bmatrix}$$

و نسبت بخت حاصل از آن  $OR = 1/13$  با فاصله اطمینان (۰/۸۹ - ۱/۴۳) CI : ۰/۹۵٪ است. ملاحظه می‌شود که با این تصحیح، نسبت بخت تغییر کمی در جهت نشان دادن اثر سیگار داشته است. با توجه به این که حساسیت دو روش در دو گروه بیمار و غیربیمار اختلاف دارند مناسب است که در این موارد از روش تصحیح خطای طبقه بندی افتراقی (Differential Misclassification) استفاده نماییم. بدین منظور ماتریس ضرایب تصحیح A براساس مقادیر ( $q_0 = 1$  و  $p_0 = 0/65$ ) در گروه غیربیمار و ( $q_1 = 1$  و  $p_1 = 0/57$ ) در گروه بیمار در نظر گرفته می‌شود. در این حالت ماتریس حاصل از روابط (۱) به صورت زیر حاصل می‌شود:

$$X = \begin{bmatrix} 159/25 & 135/75 \\ 1353/66 & 1683/34 \end{bmatrix}$$

و نسبت بخت حاصل از آن  $OR = 1/46$  با فاصله اطمینان (۱/۸۵ - ۱/۱۵) CI : ۰/۹۵٪ است، ملاحظه می‌شود که با این تصحیح، نسبت بخت تغییر بیشتری در جهت نشان دادن اثر سیگار داشته است.

### بحث:

طبقه‌بندی غلط موجب انحراف در برآورد کمیتهای مورد نظر و تحلیل نادرست از روابط بین متغیرها می‌گردد. در این مقاله ضمن تعیین حساسیت و ویژگی دو تعریف مختلف از سیگار براساس دو روش اصلاح طبقه‌بندی

نادرست افتراقی و غیر افتراقی نسبت به تصحیح مشاهدات اقدام و ملاحظه گردید که به لحاظ تفاوت میان ضرایب تصحیح در گروه بیمار با گروه غیر بیمار اصلاح طبقه بندی نادرست افتراقی برآوردهای مناسبتری ارائه می‌نماید. نسبت بخت به دست آمده نشان داد که بخت آئزین صدی داشتن برای گروه سیگاری یا ترک کرده ۱/۴۶ برابر گروه غیرسیگاری است در حالی که بدون در نظر گرفتن تصحیح، اختلافی میان دو گروه مشاهده نمی‌شد از آن جمله در مطالعه ای نشان داده شد که برآورد نسبت بخت با وجود طبقه‌بندی نادرست داده‌ها در بیماری آلزایمر ارتباط درستی میان متغیرها را نشان نمی‌دهد

(Christin.L. et al. 2000). همچنین نسبت بخت ۱/۰۸ با فاصله اطمینان (۱/۱۳ - ۱/۰۳) CI : ۰/۹۵٪ برای مصرف سیگار در رابطه با بیماریهای قلبی در مطالعه ای دیگر نشانگر ارتباط ضعیف میان این دو متغیر است که می‌تواند ناشی از عدم تصحیح اطلاعات غلط طبقه‌بندی شده باشد (Peter.W.F. et al. 1997).

### تشکر و قدردانی:

بدین وسیله از همکاران مرکز تحقیقات غدد درون ریز و متابولیسم دانشگاه علوم پزشکی شهید بهشتی که این مطالعه ارزشمند آینده‌نگر را طراحی و داده‌های جمع‌آوری شده مرحله اول را برای این مطالعه در اختیار این جانب قرار داده‌اند تشکر و قدردانی می‌نمایم.

جدول ۱- وضعیت آئزین صدري بر حسب وضعیت کنونی مصرف سیگار (تعريف اوليه) در مردان بالای ۳۰ سال  
منطقه ۱۳ تهران ۷۹-۱۳۷۸

وضعیت آئزین صدري	تعريف اوليه		جمع
	سیگاری	غیرسیگاری	
دارد	۹۱ (۳۰/۸)	۲۰۴ (۶۹/۲)	۲۹۵
	(۹/۴)	(۸/۶)	(۸/۹)
ندارد	۸۷۹ (۲۸/۹)	۲۱۵۸ (۷۱/۱)	۳۰۳۷
	(۹۰/۶)	(۹۱/۴)	(۹۱/۱)
جمع	۹۷۰ (۲۹/۱)	۲۳۶۲ (۷۰/۹)	۳۳۳۲

جدول ۲- وضعیت آئزین صدري بر حسب سابقه مصرف سیگار (تعريف دقیق) در مردان بالای ۳۰ سال  
منطقه ۱۳ تهران ۷۹-۱۳۷۸

وضعیت آئزین صدري	تعريف دقیق		جمع
	سیگاری و ترک کرده	غیرسیگاری	
دارد	۱۶۰ (۵۴/۲)	۱۳۵ (۴۵/۸)	۲۹۵
	(۱۰/۶)	(۷/۴)	(۸/۹)
ندارد	۱۳۵۱ (۴۴/۵)	۱۶۸۶ (۵۵/۵)	۳۰۳۷
	(۸۹/۴)	(۹۲/۶)	(۹۱/۱)
جمع	۱۵۱۱ (۴۵/۳)	۱۸۲۱ (۵۴/۷)	۳۳۳۲

جدول ۳- وضعیت آئزین صدی بر حسب مصرف سیگار و سابقه آن در مردان بالای ۳۰ سال  
منطقه ۱۳ تهران ۷۹-۱۳۷۸

وضعیت آئزین صدی	سیگاری	ترک کرده	غیرسیگاری	جمع
	I	II	III	
دارد	۹۱ (۹/۴)	۶۹ (۱۲/۸)	۱۳۵ (۷/۴)	۲۹۵ (۸/۹)
ندارد	۸۷۹ (۹۰/۶)	۴۷۲ (۸۷/۲)	۱۶۸۶ (۹۲/۶)	۳۰۳۷ (۹۱/۱)
جمع	۹۷۰	۵۴۱	۱۸۲۱	۳۳۳۲

جدول ۴- توزیع فراوانی افراد مورد بررسی بر حسب وضعیت مصرف سیگار و بیماری آئزین صدی

سیگاری فعال	بیمار		غیر بیمار		جمع	
	سیگاری و ترک کرده	غیرسیگاری	سیگاری و ترک کرده	غیرسیگاری	سیگاری و ترک کرده	غیرسیگاری
بلی	۹۱ (۵۶/۹)	۰	۸۷۹ (۶۵/۱)	۰	۹۷۰ (۶۴/۲)	۰
خیر	۶۹ (۴۳/۱)	۱۳۵ (۱۰۰)	۴۷۲ (۳۴/۹)	۱۶۸۶ (۱۰۰)	۵۴۱ (۳۵/۸)	۱۸۲۱ (۱۰۰)
جمع	۱۶۰	۱۳۵	۱۳۵۱	۱۶۸۶	۱۵۱۱	۱۸۲۱

## منابع:

- حلم سرشت، پریش. دل پیشه اسماعیل (بی تاریخ).  
بهداشت قلب، انتشارات چهر.
- Armitage P. (1998) Encyclopedia of biostatistics. *John Wiley and Sons*.
- Bross I. (1954) Misclassification in 2\*2 tables: *Biometrics*. **10**: 488-495.
- Carole L. and Mart M.A. (2000) Comparison of risk factors for stroke incidence and stroke mortality in the years of follow-up in men and women in Scotland stroke. **31**: 1893-1895. (Available from: [http:// www.strokeaha.org](http://www.strokeaha.org))
- Chen T.T.(1989) A review of methods for misclassified categorical data in epidemiology: *Statistics in Medicine*. **8**:1095-1106.
- Christin L., Emsley E., sujuan G. and Kathleens H. (2000) Estimating odds ratios adjusting for misclassification in Alzheimers disease risk factor assessment: *Statistics in medicine*. **19**: 1523-1530.
- Morrissey M. and Spicyelmon D.(1999) Matrix methods for estimating odds ratios with misclassified exposure data: *Biometrics*. **55**: 338-344.
- Peter. W.F. and Wilson J. (1997) Cumulative effects of high cholesterol levels, high blood pressure and cigarette smoking on carotid stenosis. *The new England journal of medicine*. **337**:516-522.
- Wachholder S., Armstrong B. and Hartge P.(1993) Validation studies using an alloyed gold standard: *American Journal of Epidemiology*. **137**: 1251-1258.

## **ANALYSIS OF ANGINA PECTORIS STATUS BASED ON THE PROBABILITY OF MISCLASSIFYING THE RISK FACTOR: TEHRAN LIPID AND GLUCOSE STUDY**

**Abadi A.\*<sup>1</sup> MSD; Mohammad K.,<sup>2</sup> Ph.D; Meshkani M.R,<sup>3</sup> Ph.D; Kazemnejad A.,<sup>4</sup> Ph.D; Mehrabi Y.,<sup>1</sup>Ph.D; Azizi F.,<sup>5</sup> Ph.D**

Misclassification of disease status and risk factors is one of the main sources of error in studies. Wrong assignment of individuals into exposed and non-exposed groups may seriously distort the results in case-control studies. This study investigates the effect of misclassification error on odds ratio estimates and attempts to introduce a correction method.

Data on 3332 men aged 30-69 years from Tehran Lipid and Glucose Study (TLGS) were considered for investigating the relationship between angina pectoris and smoking. The "Rose" questionnaire was used to evaluate angina status. Two different definitions of smoking were included: I) active smokers versus non-smokers, II) active and ex-smokers versus never smoked. The relation between variables was assessed both with and without considering misclassification in smoking.

Definition I (commonly used in many epidemiologic studies), resulted in a non-significant OR=1.09 (95% CI: 0.84, 1.41). Using the correction method based on the principle of non-differential misclassification, the OR increased to 1.13 (95% CI: 0.89, 1.43), which was still non-significant. However, a correction procedure that included the probability of differential misclassification produced a significant OR=1.46 (95% CI: 1.15, 1.85),  $p<0.05$ .

It is evident that misclassification in risk factors can lead to inaccurate results. This study showed that the relation between variables may have not been discovered if the probability of misclassification was ignored. Moreover, in case of probable

---

\* . Author to whom all correspondence should be addressed.

1. Department of Community Medicine, Shaheed Beheshti University of Medical Sciences, Iran .
2. Department of Biostatistics and Epidemiology, Tehran University of Medical Sciences, Iran .
3. Department of Statistics, Shaheed Beheshti University, Iran .
4. Department of Biostatistics, Tarbiat Modarress University, Iran .
5. Endocrine research center, Shaheed Beheshti University of Medical Sciences, Iran .

differential misclassification, applying correction methods for non-differential misclassification would be inefficient.

**Key words:** *Angina pectoris, Smoking, Odds ratio, Misclassification probability.*