# تعدیل اریبی نسبت شانس حاصل از طبقه‌بندی نادرست مواجهه‌ها با استفاده از روش‌های بیزی در بررسی عوامل محیطی مرتبط با سرطان ریه

علیرضا ابدی[۱]، باقر پهلوان‌زاده[۲*]، کرامت نوری جلیانی[۳]، سید مصطفی حسینی[۴]

۱. دانشیار، گروه آمار زیستی، گروه پزشکی اجتماعی، دانشکده پزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران.
۲. دانشجوی دکتری تخصصی آمار زیستی، گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران.
۳. استاد،'گروه آمار زیستی، دانشکده بهداشت و انستیتو تحقیقات بهداشتی، دانشگاه علوم پزشکی تهران، تهران، ایران.
۴. استاد، گروه آمار زیستی، دانشکده بهداشت و انستیتو تحقیقات بهداشتی، دانشگاه علوم پزشکی تهران، تهران، ایران.

## چکیده

**زمینه و هدف:** عدم توانایی در اندازه‌گیری دقیق مواجهه‌ها در مطالعات اپیدمیولوژیکی مشکلی است که تقریباً در همه مطالعات مخصوصاً مطالعات مورد- شاهدی رخ می‌دهد. روش‌های موجود حل این مشکل، اغلب زمان و هزینه زیادی نیاز دارند و برای بعضی مواجهه ها قابلیت اجرایی ندارند. اخیراً روش‌های جدیدی برای مطالعات مورد-شاهدی دارای همسان‌سازی یک به یک پیشنهاد شده‌اند. در اینجا ما به دنبال تعمیم این روش برای مطالعات مورد-شاهدی دارای همسان‌سازی چندگانه هستیم.

**روش بررسی:** در اینجا توزیع پیشین دریخله استاندارد برای توزیع چندجمله‌ای تعمیم داده شد تا بتوان اطلاعات مربوط به پارامتر ارتباط مواجهه-بیماری (OR) را جدای از اطلاعات مربوط به سایر پارامترها وارد مدل کرد. برای اطلاعات پیشین (OR) از اطلاعاتی که در سایر مطالعات درباره ارتباط مواجهه و بیماری بود استفاده شد. برای تصحیح سوء طبقه‌بندی نیز آنالیز حساسیت انجام شد و نتایج تحت سه مدل بیزی به دست آمد.

**یافته ها:** یافته‌های مدل بیزی خام مشابه با مدل کلاسیک بود، مدل دوم که در آن از اطلاعات OR استفاده شد شدیداً تحت تأثیر این اطلاعات قرار گرفت. مدل پیشنهادی سوم، بیشترین تعدیل اریبی را برای عوامل خطر فلزات سنگین، مصرف دخانیات و مصرف مواد مخدر ایجاد کرد به‌طوری‌که فلزات سنگین را که مدل خام (رگرسیون لجستیک کلاسیک) بر بروز سرطان ریه تأثیرگذار نشان می‌داد را غیر معنی‌دار نشان داد. آنالیز حساسیت نیز نشان داد که مدل در مقابل تغییر مقادیر حساسیت و ویژگی پایدار است.

**نتیجه گیری:** مطالعه حاضر نشان داد که اگرچه نتایج مدل سوم در بیشتر مواجهه‌ها، تفاوت چندانی با مدل دوم نداشت، ولی می‌توان گفت که این مدل می تواند تا حدود زیادی سوء طبقه‌بندی‌ها را اصلاح کند.

**کلمات کلیدی:** سوء طبقه‌بندی، روش‌های بیزی، آنالیز حساسیت، سرطان ریه

**نویسنده مسئول: باقر پهلوان‌زاده**

**آدرس:** ایران، تهران، دانشگاه علوم پزشکی شهید بهشتی، دانشکده پیراپزشکی، گروه آمار زیستی

**ایمیل:** db.pahlavan@gmail.com

**مقدمه**

در بیشتر مطالعات اپیدمیولوژیکی، هدف اصلی این است که ارتباط بین رخداد بیماری ها و مواجهه با بعضی عوامل بالقوه مضر بررسی گردد[۱] برای رخدادهای با شیوع خیلی کم، همانند سرطان‌های کودکان و بیماری‌های تنفسی تنها طرح های مطالعاتی قابل اجرا، مطالعات مورد-شاهدی می باشد [۲] لذا در بحث جمع‌آوری داده‌ها، ممکن است مواجهه‌های یک شخص به‌طور مستقیم قابل‌مشاهده یا اندازه‌گیری نباشند یعنی ممکن است از نظر عملی و یا از نظر هزینه‌ای غیر قابل‌اجرا باشند [۱] درباره مواجهه با مواد شیمیایی، افراد تحت مطالعه ممکن است به طور دقیق موادی را که با آن‌ها کار می‌کردند را به یاد نیاورند و یا اطلاع دقیقی از موادی که با آن‌ها در تماس بودند را نداشته باشند [۳] بنابراین در شرایطی که سطح مواجهه واقعی یک فرد را نتوان به‌طور دقیق به دست آورد، این امکان وجود دارد که افراد به گروه‌های نادرستی از نظر مواجهه تخصیص یابند [۱] که در بحث مواجهه‌های دوگانه، این اشتباه در طبقه‌بندی را سوء طبقه‌بندی می‌نامند [۴]. سوء طبقه‌بندی دارای دو نوع افتراقی (Differential Misclassification) و غیر افتراقی (Non-Differential Misclassification) است، در سوء طبقه‌بندی افتراقی میزان بروز سوء طبقه‌بندی تحت تأثیر متغیرهای دیگر مطالعه است [۵]. اگر در مطالعه‌ای خطر بروز سوء طبقه‌بندی مطرح باشد باید در مرحله طراحی مطالعه و یا تجزیه و تحلیل داده ها تمهیدات لازم برای تصحیح این مشکل اندیشیده شود، زیرا نادیده گرفتن سوء طبقه‌بندی می‌تواند باعث اریبی قابل‌توجهی در برآوردها گردد [۳].

در مطالعات مورد-شاهدی نیز مانند سایر مطالعات دیگر متغیر های مخدوش کننده وجود دارند، که در این مطالعات برای کنترل اثر آنها از همسان سازی استفاده می شود که بر دو نوع فردی و گروهی است. در همسان سازی فردی، هر یک از افراد گروه مورد با شاهد یا شاهد های خود از نظر متغیر های مخدوش گر همسان می شود، و در آمار کلاسیک از تحلیل رگرسیون لجستیک شرطی (Conditional logistic regression) استفاده می گردد [۶]. تفاوت مدل رگرسیون لجستیک (رگرسیون لجستیک شرطی) برای مطالعات مورد- شاهدی دارای همسان‌سازی، با مطالعات مورد-شاهدی بدون همسان‌سازی (رگرسیون لجستیک معمولی) در این است که، در رگرسیون لجستیک شرطی برای هر یک از جفت‌های همسان شده پارامتری وجود دارد که تعداد پارامترهای مدل را افزایش داده و آنالیز این‌گونه داده‌ها را پیچیده‌تر می‌کند [۴].

تلاش برای تصحیح سوء طبقه‌بندی در مرحله تجزیه و تحلیل داده ها، در هر دو شاخه آمار کلاسیک و آمار بیزی انجام شده است. در رویکرد بیزی محققین روش‌های متفاوتی را برای تصحیح سوء طبقه‌بندی در مطالعات مورد-شاهدی بدون همسان‌سازی ارائه داده اند که شامل استفاده از داده‌های اعتبارسنجی داخلی ( Internal Validation

Data)، داده‌ای اعتبار سنجی خارجی ( External Validation Data] [7 هستند. همه روش های پیشنهاد شده این مشکل را دارند که از نظر اقتصادی و زمان مورد نیاز، گران تمام می‌شوند، زیرا مستلزم انجام آزمایش‌های دقیق حداقل برای زیر نمونه‌ای از افراد مطالعه هستند. علاوه بر آن در مطالعات گذشته‌نگر که مواجهه‌ها درگذشته رخ داده‌اند نیز امکان استفاده از این روش‌ها وجود ندارد[۸-۱۱] Rice، Prescott وGarthwaite برای تصحیح سوء طبقه‌بندی در مطالعات مورد- شاهدی دارای همسان‌سازی، روش‌های بیزی را ارائه کردند[۳, ۱۲-۱۵] مدل اثرات تصادفی که پرسکات و گارثویت پیشنهاد دادند، از داده‌ای اعتبار سنجی داخلی استفاده می‌کند. در این روش، برای زیر نمونه‌ای از افراد مورد مطالعه وضعیت درست مواجهه‌ها با استفاده از روش‌های دقیق و پرهزینه به دست می آید، و از حساسیت و ویژگی بدست آمده از این داده‌ها، برای تصحیح سوء طبقه‌بندی سایر داده‌ها استفاده می‌شود[۳].

Hernandez و همکاران تصحیح‌های بیزی را در مطالعات مورد-شاهدی یک‌به‌یک، که در آن متغیر مواجهه پیوسته بوده و دارای خطای اندازه‌گیری است، انجام دادند[۵]. در جستجوی روش‌هایی برای بهبود این فرایند، Liu و همکاران استفاده از اطلاعات متخصصین را پیشنهاد دادند. آن‌ها مدل اثرات تصادفی پرسکات و گارثویت را تغییر دادند، به طوری‌که در آن، به جای اطلاعات داده های اعتبار سنجی از اطلاعات محقق برای تصحیح سوء طبقه‌بندی استفاده کردند. آن‌ها از توزیع دریخله به عنوان توزیع پیشین برای توزیع چندجمله‌ای در مطالعه مورد- شاهدی همسان شده یک‌به‌یک استفاده کردند. آنها توزیع دریخله استاندارد را طوری تغییر دادند که اطلاعات پیشین درباره پارامتر نسبت شانس (OR) و اطلاعات سایر پارامترها را جداگانه وارد مدل کنند[۴]. هدف این مطالعه تعمیم رویکرد پیشنهادی Liu و همکاران برای مطالعه مورد-شاهدی دارای همسان‌سازی یک‌به‌یک به مطالعات مورد-شاهدی دارای همسان‌سازی چندگانه است تا با استفاده از توزیع دریخله، به عنوان توزیع پیشین برای توزیع چند جمله ای در مطالعه مورد شاهدی یک به چند، همچنین افزودن اطلاعاتی درباره میزان بروز سوءطبقه بندی در اندازه گیری مواجهه ها، برآورد های دقیق تری را درباره میزان تاثیر عوامل خطر مختلف بر بروز سرطان ریه به دست آورند.

**روش بررسی:**

داده های این مقاله حاصل یک مطالعه مورد-شاهدی است که برای بیماران مبتلا به سرطان ریه که در فاصله سال‌های ۱۳۸۲ تا ۱۳۸۴ به بیمارستان‌های مسیح دانشوری، خاتم‌الانبیاء (ص)، کسری، لبافی نژاد و امام خمینی مراجعه کرده بودند، جمع‌آوری شده بود. در این مطالعه، به ازای هر بیمار ۲ شاهد در نظر گرفته شده بود. شاهد اول از بیمارستان و شاهد دوم از افراد سالم خویشاوند بیماران انتخاب شده بودند. تعداد ۲۴۲ بیمار به عنوان گروه مورد و ۴۸۴ فرد به‌عنوان گروه کنترل انتخاب شده بودند که این

افراد ازنظر سن، جنسیت و وضعیت تأهل با یکدیگر همسان شدند.

ما حالت ساده با تنها یک مواجهه را در نظر گرفتیم. مدل اصلی شامل یک مواجهه دوتایی E یا که در آن نشان‌دهنده تعداد بیماران و نشان‌دهنده تعداد شاهدهایی هستند که در k امین جفت همسان شده به درستی دارای مواجهه بوده‌اند؛ بنابراین در هر جفت یک بیمار و ۲ شاهد داریم برابر ۰ یا ۱ و برابر ۰،۱ یا ۲ خواهد بود. برای جفت k ام خواهیم داشت.

$$\theta_{ijk} = P[(E_{1k}, E_{2k}) = (i,j)]\ i = 0,1\ j = 0,1,2\ k = 1,\dots,N \quad (1)$$

در این معادله، نشان‌دهنده احتمال واقعی آن‌که در جفت k ام از N جفت وضعیت مواجهه مورد (بیمار) i و وضعیت مواجهه شاهدها j باشد را نشان می‌دهد؛ بنابراین در تمام این جفت‌ها ۶ نوع شامل خواهیم داشت که برای مثال نشان‌دهنده‌ی احتمال واقعی این که در جفت همسان شده‌ای، نه مورد مواجهه داشته باشد و نه هیچ‌یک از شاهدها و احتمال واقعی اینکه در جفت همسان شده‌ای هم مورد و هم هر دو شاهدها مواجهه داشته باشند را نشان می‌دهد.

با توجه به این‌که در مدل‌های لجستیک هر یک از جفت‌ها ازنظر ویژگی‌ها خاص در نظر گرفته می‌شوند به این معنی که در هر یک از جفت‌های همسان شده با توجه به این که بیمار و شاهدها ازنظر متغیرهای مورد همسان‌سازی وضعیت یکسانی دارند بنابراین برای هر یک از جفت‌ها پارامتر اختصاص داده می‌شود که وضعیت آن اعضای آن جفت را ازنظر متغیرهای همسان شده نشان می‌دهد. این موضوع باعث می‌شود تا تعداد پارامترهای مدل زیاد شود و حتی از تعداد جفت‌های همسان شده در مطالعه نیز بیشتر می‌شود و باعث بروز مسأله عدم برآوردپذیری مدل می‌شود. برای حل مشکل باید شرایطی را اعمال کرد که در روش‌های مقدار مورد انتظار روی E بیزی فرض می‌شود که که تمامی جفت‌ها را نشان می‌دهد.

با توجه به مقاله پرسکات و گارثویت [۳] در مدل لجستیک پس از امید گرفتن از احتمال کلی مشاهده وضعیت مواجهه i برای بیمار و j برای شاهدها برابر خواهد بود با:

$$\theta_{ij} = E\left(\theta_{ijk}\right) = \exp(i\delta_1 + j\delta_2)\frac{M!}{j!\,(M-j)!}E[\exp\{(i+j)\beta_k\}]\ i = 0,1\ j = 0,1,2 \quad (2)$$

با جایگذاری برآوردهای حاصل در رابطه زیر که برآوردگر منتل-هنتزل (M-H) در حالت با M شاهد است [۱۶]:

$$OR = \frac{\sum_{j=0}^{M}(M-j)\theta_{1j}}{\sum_{j=0}^{M} j\theta_{0(j+1)}}\ j = 0,\dots,M-1 \quad (3)$$

و با توجه به اینکه برای j=M داریم و همچنین برای j=0، است آنگاه نسبت شانس برای مدل لجستیک برابر خواهد بود با:

$$\exp(\delta_1 - \delta_2) = \frac{M\theta_{10}}{\theta_{01}} = \dots = \frac{(M-j)\theta_{1j}}{(j+1)\theta_{0(j+1)}} = \dots = \frac{\theta_{1(M-1)}}{M\theta_{0M}} = \lambda \quad (4)$$

در اینجا پارامتر احتمال مواجهه با عامل خطر در افراد گروه بیمار و احتمال مواجهه با عامل خطر در افراد گروه مواجهه به (OR) شاهد است؛ و برآوردگر نسبت شانس عامل خطر در بیمار (مورد) به افراد گروه شاهد است.

برای کاهش تعداد پارامترهای و همچنین برای اینکه پارامترهای مدل برای OR از سایر پارامترها تفکیک شود تبدیلات زیر را انجام می‌دهیم:

$$\zeta_j = \theta_{\cdot j} + \frac{j\lambda\theta_{\cdot j}}{(M-j+1)} \quad j = 0, \ldots, M \quad (5)$$
$$\zeta_{M+1} = \theta_{1M}$$

بنابراین خواهیم داشت:

$$\theta_{\cdot j} = \frac{(M-j+1)}{M-j+1+j\lambda}\zeta_j \,،\, \theta_{1(j-1)} = \frac{j\lambda}{M-j+1+j\lambda}\zeta_j$$

،

و همچنین خواهیم داشت:

$$\sum_{j=0}^{M+1}\zeta_j = \sum_{i=0}^{1}\sum_{j=0}^{M}\theta_{ij} = 1$$

برای این که برابری مربوطه به برقرار باشد باید محدودیتی را اعمال کنیم. این محدودیت برابر است با:

$$\frac{j+1}{j}\frac{M-j+2}{M-j+1} \geq \frac{\zeta_j^2}{\zeta_{(j-1)}\zeta_{(j+1)}} \quad (6)$$

با این تبدیلات، پارامترهای $\theta_{ij}$ به دو جز $\zeta$ و $\lambda$ که ازنظر تعداد نیز کمتر هستند، تبدیل می‌شوند که در آن پارامتر $\lambda$ اطلاعات درباره نسبت شانس است و $\bar{\zeta} = (\zeta_0, \ldots, \zeta_{M+1})$ اطلاعات سایر پارامترهای مدل را خواهد داشت. در مطالعه حاضر به‌جای ۶ پارامتر $\theta_{ij}$ مدل، ۴ پارامتر $\zeta$ و یک پارامتر $\lambda$ خواهیم داشت و درمجموع مدل یک پارامتر کمتر خواهد داشت علاوه بر این که پارامترها به دو جزء تفکیک شده‌اند.

مشابه پارامترهایی که برای مواجهه واقعی تعریف شد برای وضعیت مواجهه‌های مشاهده‌شده یا آنچه توسط فرد درباره وضعیت مواجهه‌اش با عامل خطر اظهارشده نیز باید پارامترهایی تعریف شود که با A نشان داده می‌شود به‌طوری‌که اگر فرد وضعیت خود را به‌عنوان مواجهه داشته اعلام کرده باشد A=1 و در غیر این صورت آن را برابر A=0 قرار داده می‌شود. همچنین $\phi_{il}$ را چنان تعریف می‌کنیم که:

$$\phi_{il} = \Pr(A = 1 | E = i, D = l) \quad i = 0,1 \quad l = 0,1 \quad (7)$$

که نشان‌دهنده احتمال اینکه فردی با وضعیت مواجهه واقعی i و وضعیت بیماری l خود را به‌عنوان مواجهه داشته معرفی کند را نشان می‌دهد؛ بنابراین $\phi_{10}$ و $\phi_{11}$ به ترتیب حساسیت (Sensitivity) (SN) در بیمار و حساسیت در شاهدها و $1-\phi_{00}$ و $1-\phi_{01}$ به ترتیب ویژگی (Specifity) (SP) در بیمار و ویژگی در شاهدها را نشان می‌دهد.

در این مطالعه ما سوء طبقه‌بندی را غیر افتراقی در نظر می‌گیریم که در آن میزان سوءطبقه بندی با وضعیت بیماری تغییر نمی‌کند و داریم:

$$SN = \phi_{10} = \phi_{11} \text{ و } SP = (1-\phi_{00}) = (1-\phi_{01})$$

با فرض این که مشاهدات تصادفی و مستقل از همدیگر باشند و همچنین احتمال سوء طبقه‌بندی بیمار یا مورد مستقل از سوء طبقه‌بندی شاهد باشد وضعیت مواجهه واقعی و مشاهده شده را در یک بردار به صورت $(E_1, E_2, A_1, A_2)$ نشان می‌دهیم که در آن $E_1$ و $A_1$ به ترتیب وضعیت مواجهه واقعی و مشاهده‌شده مورد است که برابر خواهند بود با ۰ یا ۱ و $E_2$ و $A_2$ وضعیت مواجهه واقعی و مشاهده‌شده شاهدها است که برابر خواهند بود با ۰، ۱ یا ۲.

ارتباط بین پارامترهای احتمالاتی مشاهده‌شده و واقعی با استفاده از رابطه زیر برقرار می‌گردد[۳].

$$P_{lm}=\sum_{i=0}^{1}\sum_{j=0}^{M}\theta_{ij}P[(A_1=1|E_1=i,D=1)]P[(A_2=1|E_2=i,D=0)]$$
$$=\sum_{i=0}^{1}\sum_{j=0}^{M}\sum_{h=0}^{j}\theta_{ij}\phi_{i1}(1-\phi_{i1})^{l-1}\binom{j}{h}\phi_{10}(1-\phi_{10})^{j-h}\binom{M-j}{m-h}\phi_{00}(1-\phi_{00})^{M-j-m+h} \quad (8)$$

که در آن $l=0,1$ و $m=0,1,\ldots,M$

$P_{lm}$ احتمال آن را که جفت همسان شده‌ای با وضعیت مواجهه واقعی $(i,j)$ به‌عنوان جفت دارای وضعیت مواجهه $(l,m)$ طبقه‌بندی شوند را نشان می‌دهند.

توزیع نمونه‌ای در این مطالعه توزیع چندجمله‌ای (Multinomial) خواهد بود که:

$$(N_{11},N_{10},N_{01},N_{00})\sim Multinomial(n:p_{11},p_{10},p_{01},p_{00}) \quad (9)$$

در اینجا $N_{lm}$ ها تعداد جفت‌های مورد-شاهدی را نشان می‌دهند که وضعیت مواجهه مشاهده‌شده $(l,m)$ داشته‌اند.

با استفاده از تبدیلات ۵ و جایگزینی $\theta_{ij}$ با $\zeta$ و $\lambda$ در رابطه مربوط به $P_{lm}$ تابع درستنمایی تابعی از پارامترهای $\zeta$، $\lambda$ و $\phi_{ij}$ خواهد بود برای این پارامترها توزیع‌های پیشین زیر را در نظر می‌گیریم.

$$(\zeta_0,\ldots,\zeta_{M+1})\sim Dirichlet(c_0,\ldots,c_{M+1}) \quad (10)$$

$$g(OR)\sim Beta(\beta_1,\beta_2) \quad (11)$$

$$SN\sim Beta(\alpha_1,\alpha_2) \quad (12)$$

$$SP\sim Beta(\alpha_3,\alpha_4) \quad (13)$$

توزیع پسین حاصل پیچیده خواهد بود ولی از توزیع پسین می‌توان با استفاده از روش‌های MCMC همان‌طور که در نرم‌افزاری مانند WinBUGS ساخته می‌شود، نمونه تولید کنیم. برای هر یک از پارامترها، زنجیرهای MCMC جداگانه‌ای ساخته می‌شود و برای هر یک از تکرارها یک نمونه از پارامتر ایجاد می‌شود که با استفاده از رابطه شماره ۳، برآوردی برای OR تولید خواهد شد. این فرایند، توالی ئی از OR ها را تولید خواهد کرد که به عنوان برآوردهای پسین ارائه خواهند شد.

## یافته ها

در بررسی عوامل شغلی و محیطی مؤثر بر سرطان ریه ۸ مورد از این مواجهه‌ها را در نظر گرفتیم توزیع وضعیت مواجهه‌های مشاهده‌شده (Observed exposures) جفت‌های همسان شده در این مطالعه در جدول ۱ ارائه گردیده است.

جدول ۱: وضعیت مواجهه‌های مشاهده‌شده در جفت همسان شده

| وضعیت مواجهه متغیر | نه مورد و نه شاهدها | فقط یکی از شاهدها | فقط شاهدها | فقط مورد | مورد و یکی از شاهدها | هم مورد و هم شاهدها |
|---|---|---|---|---|---|---|
| مواد نقاشی[۱۷] | ۲۰۷ | ۱۲ | ۰ | ۱۴ | ۰ | ۱ |
| گردوغبار چوب[۱۷] | ۲۱۹ | ۴ | ۰ | ۸ | ۱ | ۰ |
| گردوغبار پنبه[۱۸] | ۲۰۵ | ۱۸ | ۱ | ۱۷ | ۱ | ۰ |
| سیلیس[۱۹] | ۲۱۹ | ۷ | ۰ | ۷ | ۱ | ۰ |
| مصرف سیگار[۲۰] | ۵۵ | ۱۸ | ۸ | ۴۰ | ۸۱ | ۴۰ |
| مصرف مواد مخدر[۲۱] | ۱۷۴ | ۳ | ۱ | ۳۳ | ۴ | ۰ |
| فلزات سنگین[۲۲] | ۲۰۹ | ۱۳ | ۰ | ۲۰ | ۰ | ۰ |
| قطران زغال‌سنگ[۲۲] | ۲۲۹ | ۴ | ۰ | ۷ | ۰ | ۰ |
| | | | | | | |

برای تعیین پارامترهای توزیع‌های پیشین برای حساسیت و ویژگی و نسبت شانس، از توزیع‌های آگاهی‌بخش (Informative Prior) استفاده شد. برای مدل، دو نوع آگاهی یا اطلاع موردنیاز بود: اطلاعات درباره OR و اطلاعات درباره حساسیت و ویژگی طبقه‌بندی‌ها. برای اطلاعات مورد نیاز درباره نسبت شانس، بروز بیماری برای مواجهه‌های مختلف، مطالعات مشابهی پیدا شد و سعی شد تا جدیدترین مطالعات مرور نظام‌مند ( Systematic Review) انجام‌شده انتخاب گردند. منابع مربوط به این اطلاع در جدول ۱ در کنار هر یک از عوامل خطر ارائه شده است. از این مطالعات صدک های ۲.۵ درصد و ۹۷.۵ درصد برای نسبت شانس که به طور معمول در مطالعات گزارش می‌شوند استخراج گردید. سپس این اطلاعات با استفاده از رابطه$\frac{\lambda}{\lambda+۲}$ با توجه به رابطه ۵ تبدیل شدند و برای مقادیر حاصل با استفاده از نرم‌افزار R و بسته Learn Bayes توزیع بتایی با مقادیر ۹۷/۵ درصد و۲/۵ درصد نظیر به دست آمد. برای اطلاعات موردنیاز درباره سوء طبقه‌بندی، آنالیز حساسیت (Sensitivity Analysis) انجام شد. به طوری که برای هر مواجهه مورد نظر، ۵ بازه به عنوان حساسیت اولیه و ۵ بازه به‌عنوان ویژگی اولیه در نظر گرفته شدند. در اولین مورد فرض شد حساسیت و ویژگی در بازه $(۰.۵,۱)$ قرار داشته باشند، سپس از سمت چپ بازه را ۱۰٪ به سمت مقدار عددی ۱ حرکت داده شد و فرض شد که حساسیت و ویژگی در بازه $(۰.۶,۱)$ باشند. این کار ۳ بار دیگر انجام شد در آخر بازه $(۰.۹,۱)$ در نظر گرفته شد. در اینجا نیز با استفاده از نرم‌افزار R و بسته Learn Bayes توزیع بتای نظیر محاسبه گردید. سپس به ازای مقدار ثابتی از حساسیت، ویژگی تغییر داده شد و مقادیر OR، حساسیت و ویژگی پسین محاسبه گردید. لذا برای بررسی تأثیر هر یک از این اطلاعات بر برآورد OR ۳، مدل در نظر گرفته شد که شامل:

۱-مدل خام (بدون دخیل کردن اطلاعات پیشین درباره OR و فرض عدم رخداد سوء طبقه‌بندی یعنی SN=1 و SP=1)

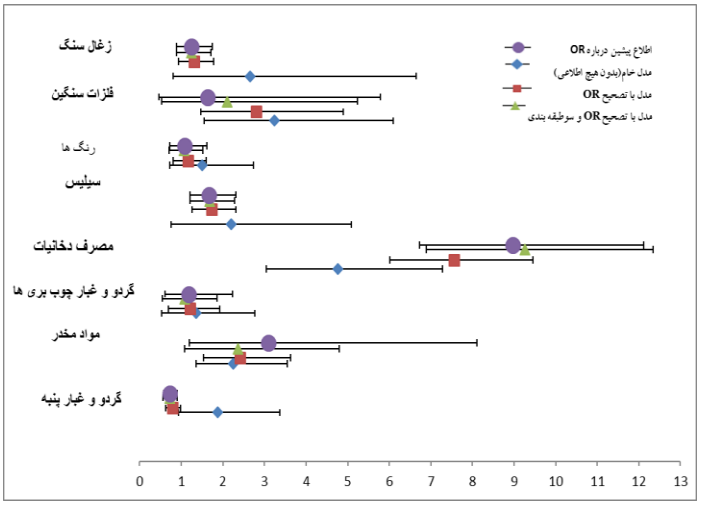
۲-مدلی با دخیل کردن اطلاعات پیشین درباره OR و فرض عدم رخداد سوء طبقه‌بندی (SN=1 و SP=1)

۳-مدلی با دخیل کردن اطلاعات پیشین درباره OR و همچنین اطلاعات درباره سوء طبقه‌بندی

برای پارامتر OR در مدل‌های اول و دوم برآورد نقطه‌ای و فاصله باورمندی (Credible Interval) ۹۵ درصد مقادیر پسین در هر یک از مواجهه‌ها به دست آمد اما در مدل سوم که در آن مقادیر حساسیت و ویژگی تغییر می‌کرد به ازای همه ۲۵ ترکیب حساسیت و ویژگی (۵ مقدار برای حساسیت و ۵ مقدار برای ویژگی ۲۵ = ۵ * ۵) مقادیر OR به دست آمد و درنهایت میانگین این مقادیر به‌عنوان برآورد نقطه‌ای و فاصله باورمندی برای OR تحت مدل سوم در نظر گرفته شد و برای انجام مقایسه مقادیر برآوردها تحت این سه مدل در نمودار ۱ آورده شد.

همان‌طور که در نمودار ۱ مشاهده می‌شود برای همه متغیرها فاصله اطمینان در مدل خام در مقایسه با سایر مدل‌ها به‌طور قابل‌توجهی عریض‌تر است. در نمودار اطلاعاتی که به‌عنوان اطلاع پیشین درباره OR استفاده شد نیز نمایش داده‌شده است. نتایج مدل دوم که در آن تنها اطلاع در مورد OR به مدل قبلی افزوده‌شده است همان‌طور که انتظار می‌رود بین نتایج مدل خام و اطلاع پیشین قرار می‌گیرد. همچنین فاصله باورمندی در مدل دوم در بیشتر متغیرها به فاصله اطمینان استفاده‌شده به‌عنوان اطلاع پیشین نزدیک شده است و اطلاعات اولیه درباره OR نتایج مدل دوم را تحت تأثیر گذاشته است برای مواجهه «گردوغبار پنبه» این تأثیر واضح‌تر است.

در مدل سوم که در آن آنالیز حساسیت به ازای مقادیر مختلف سوء طبقه‌بندی انجام شد فاصله باورمندی به‌دست‌آمده برای بیشتر مواجهه‌ها مشابه مدل دوم بود به‌جز مواجهه‌های «مصرف دخانیات»، «مصرف مواد مخدر» و «فلزات سنگین» که در این موارد طول فاصله باورمندی افزایش‌یافته بود علاوه بر آن در تمامی مواجهه‌ها به‌جز «مصرف دخانیات» برآورد نقطه‌ای OR(میانگین) به مقدار عدم معنی‌داری OR(مقدار عددی ۱) نزدیک‌تر شده است. همچنین در این مدل مشاهده شد که با تغییر حساسیت و ویژگی اولیه که در مدل به‌عنوان اطلاعات درباره سوء طبقه‌بندی‌ها ارائه می‌شود برآورد OR برای بیشتر مواجهه‌ها تأثیری نپذیرفته است در نمودار ۱ در فواصل باورمندی مربوط به مدل سوم (مدل با تصحیح OR و سوء طبقه‌بندی) میانگین ۲۵ برآورد کران‌های پایین، بالا و میانگین به ازای مقادیر مختلف حساسیت و ویژگی قرار داده‌شده است. در تمام موارد انحراف معیار برآوردها برای ۲۵ برآورد کمتر از ۰،۰۱ بود اما در اینجا تنها برای مواجهه «مصرف دخانیات» رفتارها اندکی متفاوت‌تر از بقیه مواجهه‌ها بود و برای میانگین، انحراف معیار ۰/۰۹۷ بود مشاهده شد




نمودار ۱: Forest Plot برای مقایسه مقادیر برآوردهای OR تحت مدل‌های مختلف

که به ازای یک مقدار ثابت از ویژگی مقدار میانگین برآورد پسین برای حالت با عریض‌ترین حساسیت (۱و ۰/۵) بیشترین مقدار را نتیجه می‌داد و با افزایش کران پایین و محدود شدن فاصله

اطمینان حساسیت میانگین برآوردهای OR کمتر می‌شد. ولی به ازای یک مقدار ثابت حساسیت نتایج به‌دست‌آمده تحت ویژگی‌های مختلف خیلی نزدیک به هم بود.

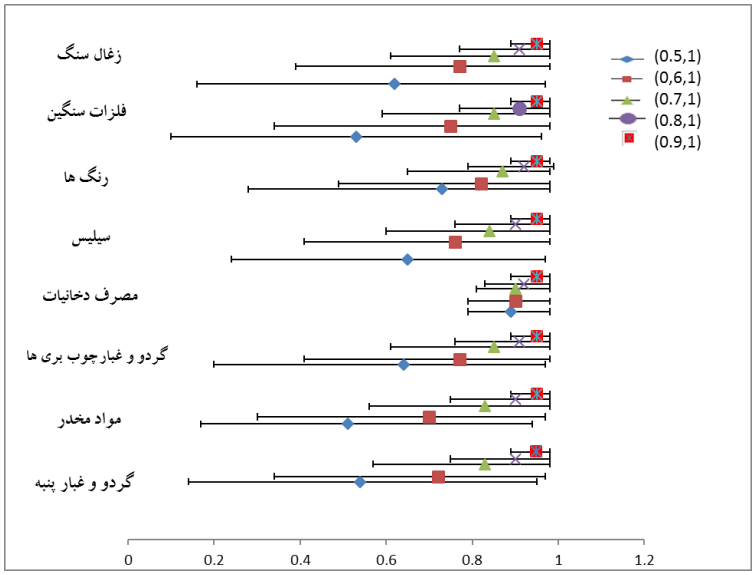
درباره برآوردهای و فواصل باورمندی ارائه شده برای حساسیت و ویژگی مشاهده شد که برآورد ویژگی به ازای مقادیر مختلف ویژگی اولیه کمترین تأثیر را می‌پذیرفت در جدول ۲ میانگین ۲۵ برآورد ویژگی ارائه‌شده است این برآوردها خیلی به همدیگر نزدیک بودند به‌طوری‌که انحراف معیار ۲۵ برآورد میانگین و هم‌چنین کران‌های بالا و پایین برای بیشتر مواجهه‌ها کمتر از ۰/۰۱ بود و تنها برای مواجهه «مصرف دخانیات» این پراکندگی‌ها بیشتر از ۰/۰۱ بود که مقادیر آن در جدول ارائه گردیده است. این برآوردها نه‌تنها برای مقادیر مختلف حساسیت مقاوم بود بلکه حتی برای مقادیر ویژگی نیز مقاوم بود و به ازای مقادیر مختلف ویژگی اولیه تغییرات اندکی داشت.

جدول ۲: میانگین و کران‌های فاصله باورمندی برآوردهای ویژگی

| | میانگین | کران پایین | کران بالا |
|---|---|---|---|
| زغال‌سنگ | ۰/۹۸ | ۰/۹۷ | ۰/۹۹ |
| فلزات سنگین | ۰/۹۷ | ۰/۹۴ | ۰/۹۹ |
| رنگ‌ها | ۰/۹۸ | ۰/۹۷ | ۰/۹۹ |
| سیلیس | ۰/۹۸ | ۰/۹۷ | ۰/۹۹ |
| مصرف | ۰/۸ | ۰/۸۵ | ۰/۹۸ |

| | | | |
|---|---|---|---|
| دخانیات | (۰/۰۷) ( | (۰/۰۴) | (۰/۰۲) |
| گردوغبار چوب | ۰/۹۸ | ۰/۹۶ | ۰/۹۹ |
| مواد مخدر | ۰/۹۶ | ۰/۹۲ | ۰/۹۹ |
| گردوغبار پنبه | ۰/۹۷ | ۰/۹۴ | ۰/۹۹ |

نتایج مشاهده‌شده برای برآوردهای پسین حساسیت اندکی با آنچه برای ویژگی مشاهده شد متفاوت بود در اینجا نیز به ازای یک بازه ثابت برای ویژگی مقادیر حساسیت در ۵ بازه موردنظر به دست آمد. مشاهده شد که با تغییر بازه مربوط به حساسیت میانگین و فاصله باورمندی نیز تغییر می‌کرد و میانگین به مقدار ۱ و فاصله نیست. فواصل گزارش‌شده برای مواجهه «مصرف دخانیات» بار دیگر متفاوت‌تر از سایر مواجهه‌ها است و به ازای همه مقادیر حساسیت و ویژگی نیز همچنان مقدار زیادی را گرفته است. فواصل گزارش‌شده برای سایر متغیرها تقریباً مشابه همدیگر است. باورمندی نیز




کوچک‌تر شده و به سمت ۱ حرکت می‌کرد. نکته قابل‌توجه این بود که مقادیر پسین حساسیت تحت تأثیر مقدار اولیه‌ای که برای ویژگی در نظر گرفته می‌شود نیست. فواصل گزارش‌شده برای مواجهه «مصرف دخانیات» بار دیگر متفاوت‌تر از سایر مواجهه‌ها است و به ازای همه مقادیر حساسیت و ویژگی نیز همچنان مقدار زیادی را گرفته است. فواصل گزارش‌شده برای سایر متغیرها تقریباً مشابه همدیگر است.

## بحث و نتیجه گیری:

در این مطالعه ما به دنبال روشی برای تصحیح سوء طبقه‌بندی در مطالعات مورد-شاهدی دارای همسان‌سازی بودیم که در آن‌یک بیمار با دو شاهد ازنظر متغیرهای مخدوش بالقوه با همدیگر همسان شده‌اند. روش‌های مختلفی برای تصحیح سوء طبقه‌بندی شامل استفاده از داده‌های اعتباربخشی خارجی، استفاده از داده‌های اعتباربخشی داخلی، SIMEX وجود دارند. ما در اینجا از روشی مشابه روشی که لیو و همکاران [۴] استفاده کرده‌اند. مدلی که ما استفاده کردیم از دو منبع اطلاعاتی برای تصحیح برآورد OR استفاده می‌کند که یک جز آن اطلاعات مربوطه OR است در این جز برخلاف کار لیو و همکاران ما از اطلاعاتی که در مطالعات پیشین درباره میزان ارتباط بین مواجهه و بیماری استفاده کردیم. برای جز دوم نیز آن‌ها از اطلاعات متخصصین برای تصحیح سوء طبقه‌بندی استفاده کردند ما در اینجا آنالیز حساسیت را برای این جز از اطلاعات انجام دادیم و با در نظر گرفت ۵ فاصله اطمینان برای حساسیت و ویژگی اولیه به ازای ترکیبات مختلف این ۵ بازه برآوردهای پسین را برای OR، حساسیت و ویژگی به دست آوردیم. مدل خام که درواقع مدل لجستیک شرطی معمولی است مشاهده شد که عوامل خطر فلزات سنگین، دخانیات و مواد مخدر به‌عنوان عواملی تشخیص داده شدند که شانس مواجهه بیماران با آن‌ها

بیشتر از افراد شاهد بود یا به عبارتی این عوامل شانس ابتلا به سرطان ریه را افزایش می‌دهند در مدل دوم که اطلاعات مربوطه به ارتباط مواجهه با بیماری به اطلاعات موجود درداده‌ها افزوده شد. همان‌طور که انتظار می‌رفت در تمامی مواجهه‌های موردبررسی برآورد مربوط به مدل دوم بین دو مقدار OR اولیه و OR مدل خام قرار می‌گیرد. برآوردها شدیداً تحت تأثیر اطلاعات پیشین افزوده‌شده به مدل قرار گرفتند مشاهده شد که با افزودن این اطلاع علاوه بر عوامل خطر قبلی دو عامل خطر سیلیس و گردوغبار پنبه نیز به‌عنوان عوامل تأثیرگذار شناخته شدند البته با توجه به اینکه در مطالعه‌ای که از اطلاعات برای گردوغبار پنبه استفاده شد این عامل به‌عنوان یک عامل محافظت‌کننده در مقابل سرطان شناخته‌شده بود [۲۳] لذا نتیجه مدل دوم نیز تحت این اطلاع قرار گرفت و این عامل به‌عنوان یک عامل محافظت‌کننده در مقابل سرطان تشخیص داده شد می‌داد در توجیه علت اینکه چرا با افزودن اطلاعات درباره OR نتایج مدل به‌جز مصرف سیگار برای بقیه عوامل خطر شدیداً تحت تأثیر این مقادیر قرار می‌گیرند و برآوردها به سمت آن کشیده می‌شوند می‌توان گفت که همان‌طور که در جدول ۱ مشاهده می‌شود برای بیشتر عوامل خطر فراوانی افرادی که در خانه‌های ناهماهنگ (Discordant) صفر یا خیلی کم است که با توجه به اینکه در مدل‌های لجستیک تنها از اطلاعات خانه‌های ناهماهنگ استفاده می‌شود[۳] لذا اطلاعات موجود در جدول در مقایسه با اطلاعات پیشین اندک بوده که باعث می‌شود برآوردها شدیداً تحت تأثیر اطلاعات اولیه (پیشین) قرار گیرند. علاوه بر آن فاصله اطمینانی که برای مدل خام ارائه‌شده در بیشتر مواجهه‌ها عریض است که این مشاهده نیز ناشی از اندک بودن اطلاعات ارائه‌شده درباره ارتباط مواجهه و بیماری است. تنها برای مواجهه "مصرف سیگار" فراوانی این خانه‌ها در جدول قابل‌توجه است. در مدل سوم که در آن‌سوء طبقه‌بندی تصحیح می‌شود مشاهده شد که از عوامل خطی ذکرشده قبلی عامل خطر فلزات سنگین که در هر دو مدل قبلی اثر معنی‌داری بر رخداد سرطان ریه داشت از معنی‌داری خارج شد درباره تصحیح سوء طبقه‌بندی برای دو عامل خطر سیلیس و گردوغبار پنبه که در مدل دوم و تحت تأثیر اطلاع پیشین معنی‌دار دیده شدند نمی‌توان بحث کرد زیرا اطلاع مربوطه به این مواجهه‌ها در جدول اندک بوده و تصحیح سوء طبقه‌بندی زیاد برای آن‌ها مطرح نیست اما برای سه عامل خطر دیگر می‌توان گفت که تصحیح سوء طبقه‌بندی مؤثر بوده به‌طوری‌که بعد از تصحیح نسبت به سوء طبقه‌بندی این عامل دیگر معنی‌دار دیده نشد. برای عامل خطر مصرف مواد مخدر نیز روند مشابهی مشاهده شد و برآورد نقطه‌ای OR همان‌طور که برای بیشتر عوامل خطر تحت مدل سه رخ داد به مقدار عددی ۱ کشیده شد که این یافته با آنچه در مطالعه لیو و همکاران مشاهده شد مطابقت دارد صحت این یافته در مطالعه Diamond و همکاران [۲۴] نیز بررسی شد؛ اما در مورد مواجهه مصرف سیگار این یافته با بقیه مطالعات [۴, ۱۱, ۲۴, ۲۵]که در آن‌ها برآوردها پس از تصحیح سوء طبقه‌بندی به سمت عدم معنی‌داری حرکت می‌کنند مطابقت ندارد. برای توضیح علت شاید لازم باشد که مطالعه بیشتری درباره تأثیر تصحیح سوء طبقه‌بندی در مطالعات همسان شده M:۱ مانند آنچه دایاموند و همکاران انجام دادند انجام گیرد تا بررسی شود که آیا در اینجا نیز باید انتظار داشته باشیم برآوردها پس از تصحیح باید در جهت عدم معنی‌داری حرکت کنند باید به خاطر داشت که در مطالعه حاضر ۴ خانه در جدول ۲*۳ وجود داشت که ناهماهنگ بودند. شاید وجود اطلاع پیشین درباره OR در مدل سوم که قرار است تصحیح سوء طبقه‌بندی برای اطلاعات ارائه‌شده در جدول ۱ ارائه‌شده است مناسب نباشد و توانایی مدل را تصحیح سوء طبقه‌بندی نشان ندهد و تصحیح را در حالتی دیگر و تنها در

حضور اطلاعات جدول و یا همان اطلاعات مورداستفاده در مدل خام لازم باشد که انجام گردد. در مطالعه حاضر این حالت را نیز بررسی شد و برآوردها به دست آمد (در اینجا این اطلاعات ارائه نشده است) ولی همان‌طور که در مطالعه لیو و همکاران نیز مشاهده شد در این مدل به خاطر واریانسی که در برآورد OR وجود دارد (واریانس ناشی از اطلاعات اندک موجود در جدول۱ ) و همچنین واریانسی که درنتیجه عدم قطعیت در مقدار حساسیت و ویژگی مورداستفاده در مدل، برآوردهای این مدل واریانس خیلی زیادی داشتند و درنتیجه کارایی این برآوردها اندک بود. شاید این مدل برای داده‌ای که در آن اطلاعات قابل‌توجهی درباره OR در آن وجود داشته باشد (فراوانی خانه‌های ناهماهنگ بیشتر باشد) بتوان نتایج مفیدی به دست آورد.

در مورد برآوردهای OR در مدل سوم همان‌گونه که در قسمت نتایج اشاره شد برآوردها نسبت به مقادیر اولیه حساسیت و ویژگی مقاوم بودند و کمتر تحت تأثیر قرار می‌گرفتند با توجه به اینکه برآوردهای این مدل در مقایسه با مدل دوم تغییراتی داشته است شاید بتوان گفت که مدل پیشنهادی می‌تواند حتی با داشتن اطلاعات اندکی درباره سوء طبقه‌بندی برآوردهایی قابل قبولی را نتیجه می‌دهد این تغییرات در مواجهه «مصرف دخانیات» بارزتر است هرچند در این برآوردها نتایج مدل سوم نیز تحت تأثیر شدید مقادیر اولیه ارائه‌شده برای OR قرار دارد شاید بتوان با داشتن داده‌های کافی در مطالعه بتوان تصحیح قابل‌توجهی را انجام داد.

با توجه به اینکه نتایج مدل به‌طور قابل‌توجهی تحت تأثیر مقدار اولیه OR است که به‌عنوان اطلاعات پیشین در نظر گرفته‌شده با توجه به آنچه برای عوامل خطر مصرف دخانیات و مواد مخدر که اطلاعات جدول قابل‌توجه بود لازم است این مدل در حالتی که اطلاعات موجود در جدول قابل‌توجه باشند طوری که برآوردها تحت تأثیر شدید اطلاعات اولیه (پیشین) قرار نگیرند نیز بررسی شود. لازم به ذکر است مدل حاضر به‌آسانی قابلیت تعمیم‌پذیری برای مطالعات همسان شده با چندین شاهد ($M \geq 3$)را دارد.

**تشکر و قدردانی**:

از همه مسئولین ذیربط و کلیه افرادی که ما را در انجام این طرح یاری کرده قدردانی و سپاسگزاری به عمل می آید.

**References**:


.۱ Reade-Christopher, S.J. and L.L. Kupper, *Effects of exposure misclassification on regression analyses of epidemiologic follow-up study data.* Biometrics, ۱۹۹۱: p. .۵۴۸-۵۳۵

.۲ Teschke, K., et al., *Occupational exposure assessment in case–control studies :opportunities for improvement.* Occupational and Environmental Medicine, ۲۰۰۲. ۵۹(۹): p. ۵۹۴,-۵۷۵

.۳ Prescott, G.J. and P.H. Garthwaite, *Bayesian analysis of misclassified binary data from a matched case–control study with a validation sub-study*. Statistics in medicine, ۲۰۰۵. ۲۴(۳): p. ۴۰۱,-۳۷۹

.۴ Liu, J., et al., *Bayesian analysis of a matched case–control study with expert prior information on both the misclassification of exposure and the exposure–disease association.* Statistics in medicine, ۲۰۰۹. ۲۸( :(۲۷p. -۳۴۱۱ ۳۴۲۳,

.۵ Espino-Hernandez, G., P. Gustafson, and I. Burstyn, *Bayesian adjustment for measurement error in continuous exposures in an individually matched case-control study.* BMC medical research methodology, ۲۰۱۱. ۱۱(۱): p. ۶۷,

.۶ Clayton, D ,.M. Hills, and A. Pickles, *Statistical models in epidemiology*. Vol. ۱۶۱. ۱۹۹۳: IEA.

.۷ Gustafson, P., *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*. ۲۰۰۳: CRC Press.

.۸ Roeder, K., R.J. Carroll, and B.G. Lindsay, *A semiparametric mixture approach to case-control studies with errors in covariables.* Journal of the American Statistical Association, ۱۹۹۶. ۹۱(۴۳۴): p. ۷۳۲,-۷۲۲

.۹ Müller, P. and K. Roeder, *A Bayesian semiparametric model for case-control studies with errors in variables.* Biometrika, ۱۹۹۷. ۸۴(۳): p. ۵۳۷,-۵۲۳

.۱۰ Gustafson, P., N.D. Le, and R. Saskin, *Case–control analysis with partial knowledge of exposure misclassification probabilities.* Biometrics, ۲۰۰۱. ۵۷(۲): p. ۶۰۹,-۵۹۸

.۱۱ Gustafson, P., N.D. Le, and M. Vallée, *A Bayesian approach to case–control studies with errors in covariables.* Biostatistics, ۲۰۰۲. ۳(۲): p. ۲۴۳,-۲۲۹

.۱۲ Rice, K., *Full-likelihood approaches to misclassification of a binary exposure in matched case-control studies.* Statistics in medicine, ۲۰۰۳. ۲۲(۲۰): p. -۳۱۷۷ ۳۱۹۴,

.۱۳ Rice, K., *Equivalence between conditional and random-effects likelihoods for pair-matched case-control studies.* Journal of the American Statistical Association, ۲۰۰۸. ۱۰۳(.(۴۸۱

.۱۴ Rice, K., *On Bayesian analysis of misclassified data from a matched case-control study with a validation sub-study by Gordon J. Prescott and Paul H. Garthwaite.* Statistics in medicine, ۲۰۰۶. ۲۵(۳): p. ۵۳۹,-۵۳۷

.۱۵ Rice, K.M., *Equivalence between conditional and mixture approaches to the Rasch model and matched case-control studies, with applications.* Journal of the American Statistical Association, ۲۰۰۴. ۹۹(۴۶۶): p. -۵۱۰ ۵۲۲,

.۱۶ Breslow, N.E. and N.E. Day, *Statistical methods in cancer research*. Vol. ۲. ۱۹۸۷: International Agency for Research on Cancer Lyon.

.۱۷ Tse, L.A., et al., *Occupational risks and lung cancer burden for Chinese men: a population-based case-referent study.* Cancer Causes Control, ۲۰۱۲. ۲۳: p. ۱۳۱,-۱۲۱

.۱۸ Lenters, V., et al, *Endotoxin exposure and lung cancer risk: a systematic review and meta-analysis of the published literature on agriculture and cotton textile workers.* Cancer Causes Control, ۲۰۱۰. ۲۱(۴): p. ۵۵۵,-۵۲۳

.۱۹ Vida, S., et al., *Occupational Exposure to Silica and Lung Cancer: Pooled*

*Analysis of Two Case-Control Studies in Montreal, Canada.* Cancer Epidemiol Biomarkers Prevention, ۲۰۱۰. ۱۹: p. ۱۶۱۱,-۱۶۰۲

۲۰. Gandini, S., et al, *Tobacco smoking and cancer: a meta-analysis.* ۲۰۰۸. ۱۲۲: p. ۱۶۴,-۱۵۵

۲۱. MR, M., et al., *Opium could Be Considered an Independent Risk Factor for Lung cancer: A Case-Control Study.* Respiration, ۲۰۱۲,

۲۲. Martin, J.C., et al., *Occupational Risk Factors for Lung cancer in French Electricity and Gas.* American journal of Epidemiology, ۲۰۰۰. ۱۵۱(۹).

۲۳. Tse, L.A., et al ,.*Occupational risks and lung cancer burden for Chinese men: a population-based case–referent study.* Cancer Causes & Control, ۲۰۱۲. ۲۳(۱): p. ۱۳۱,-۱۲۱

۲۴. Diamond, E.L. and A.M. Lilienfeld, *Effects of errors in classification and diagnosis in various types of epidemiological studies.* American Journal of Public Health and the Nations Health, ۱۹۶۲. ۵۲(۷): p. ۱۱۴۴,-۱۱۳۷

۲۵. Gustafson, P. and S. Greenland, *Curious phenomena in Bayesian adjustment for exposure misclassification.* Statistics in medicine, ۲۰۰۶. :(۱)۲۵p. ۱۰۳-۸۷.