

سری آمار: تحلیل جداول توافقی ۳ (معیارهای تاثیر و توافق)

محمد اصغری جعفرآبادی^{۱*}، سیده مومنه محمدی^۲، اکبر سلطانی^۳

چکیده

در مطالعات علوم پزشکی، نیاز است شاخص‌هایی برای بررسی تاثیر مواجهه‌ها و مداخله‌ها و همچنین دقت اندازه‌گیری‌ها به‌کار رود. هدف از این مقاله ارائه شاخص‌های تاثیر و توافق در اندازه‌گیری‌ها و انجام استنباط در مورد آن‌ها بود. مبانی و نحوه انجام محاسبات انواع معیارهای تاثیر و توافق در اندازه‌گیری‌ها و استنباط در مورد آن‌ها برای مطالعات علوم پزشکی، با در نظر گرفتن ملاحظات مربوط، ارائه گردید. برای بررسی تاثیر و تعیین سهم عامل خطر روی پیامدها در مطالعات مورد-شاهدی و هم‌گروهی، شاخص خطر و کسر قابل انتساب در گروه مواجهه و جامعه، به‌همراه فاصله اطمینان آن‌ها استفاده شد. همچنین شاخص‌های نسبی و مطلق کاهش خطر و تعداد مورد نیاز برای درمان (NNT) در مطالعات مداخله‌ای و مخصوصاً کارآزمایی‌ها به‌عنوان اندازه اثر کفایت تاثیر گذاری مداخله به‌کار رفت. در مطالعات تشخیصی و غربال‌گری، فاصله اطمینان حساسیت، ویژگی و شاخص‌های مرتبط به‌همراه فاصله اطمینان آن‌ها محاسبه شد. به‌علاوه نیاز است میزان دقت اندازه‌گیری‌ها برای اندازه‌گیری‌های کمی با استفاده از شاخص‌های خطای معیار اندازه‌گیری، ICC، حدود توافق بلند و آلتمن و ضریب همبستگی توافق لین و برای متغیرهای اسمی و رتبه‌ای به‌ترتیب توسط کاپا و کاپای وزنی محاسبه و در مورد آن‌ها استنباط انجام شود. در هر یک از موقعیت‌های فوق نتایج حاصل از مثال‌های مطالعاتی به‌همراه نحوه انجام محاسبات ارائه شد. برای بررسی تاثیر و توافق در اندازه‌گیری‌ها برای انواع مطالعات علوم پزشکی، تحلیل‌های ارائه شده با توجه به موقعیت و هدف مطالعه توصیه می‌شود.

واژگان کلیدی: جداول توافقی، تاثیر، توافق، خطر قابل انتساب، ارزیابی‌های تشخیصی، NNT، کاپا، حدود توافق

۱- مرکز تحقیقات آموزش علوم پزشکی، گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، تبریز، ایران

۲- گروه علوم تشریحی، دانشکده پزشکی، دانشگاه علوم پزشکی تبریز، تبریز، ایران

۳- مرکز تحقیقات غدد و متابولیسم، پژوهشکده علوم بالینی غدد و متابولیسم، دانشگاه علوم پزشکی تهران، تهران، ایران

* **نشانی:** تبریز، خیابان گلگشت، خیابان عطار نیشابوری، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، گروه آمار و اپیدمیولوژی،

کدپستی: ۵۱۶۶۱۶۷۱۱، تلفن: ۰۲-۰۴۱۱۳۳۵۷۵۸۰، نمابر: ۰۴۱۱۳۳۴۰۶۳۴، پست الکترونیک: asgharimo@tbzmed.ac.ir

تاریخ پذیرش: ۱۳۹۳/۱۰/۲۷

تاریخ درخواست اصلاح: ۱۳۹۳/۱۰/۲۱

تاریخ دریافت: ۱۳۹۳/۰۹/۱۷

مقدمه

در مطالعات علوم پزشکی، برای خلاصه کردن و نمایش دادن متغیرهای مورد بررسی اعم از پیامدها و عوامل خطر (مواجهه) رسته‌ای (Categorical) از فراوانی و درصد در قالب جداول دوطرفه و برای بررسی رابطه بین پیامدها و عوامل خطر (مواجهه) در موقعیت‌های مختلف انواع کای دو معرفی شد [۱]. همچنین به دلیل محدودیت آزمون‌های کای دو در اندازه‌گیری شدت رابطه بین عامل خطر و پیامد شاخص اندازه اثر (Effect Size)، توصیه گردید که میزان اهمیت بالینی یا عملی رابطه یا اثر را به همراه معنی داری آماری آن‌ها نشان دهد [۲، ۳]. در سری قبل فاصله اطمینان برخی از اندازه اثرها محاسبه شد [۴]. در این سری هدف معرفی معیارهای تاثیر و توافق در اندازه‌گیری‌ها و انجام استنباط در مورد آن‌ها می‌باشد.

۱) معیارهای تاثیر (Effect) در اندازه‌گیری‌ها

این معیارها، اندازه‌ای از تاثیر مواجهه یا مداخله روی پیامدهای مطالعه را به دست می‌دهند. در مطالعات هم‌گروهی و مورد-شاهدی، شاخص‌های فوق، اثر حذف مواجهه را روی پیامد اندازه‌گیری می‌کنند. به عبارت دیگر، مشخص می‌نمایند در صورت حذف مواجهه، به چه میزان می‌توان از رخداد پیامد جلوگیری کرد. در مطالعات مداخله‌ای، شاخص‌های فوق، اثر مداخله را روی پیامد اندازه‌گیری می‌کنند. به عبارت دیگر، مشخص می‌نمایند در صورت انجام مداخله، به چه میزان می‌توان سودمندی حاصل کرد و یا به چه میزان می‌توان از رخداد پیامد ناگوار جلوگیری کرد.

۱-۱) محاسبه خطر قابل انتساب (Attributable Risk (AR)) و کسر قابل انتساب (Attributable Fraction (AF)) و استنباط در مورد آن‌ها در مطالعات هم‌گروهی (و مورد-شاهدی)

شاخص‌های خطر قابل انتساب (Attributable Risk (AR)) و کسر قابل انتساب (Attributable Fraction (AF)) معیاری برای اندازه‌گیری خطر ناشی از یک عامل خطر یا مواجهه است که میزان رابطه بین پیامد و مواجهه (که با شاخص‌های RR یا OR) بیان می‌شوند و میزانی از شیوع

پیامد در جامعه را با هم ترکیب نموده و معیاری واحد برای این منظور ارائه می‌کند. شاخص‌های AR و AF به ترتیب معیارهای مطلق و نسبی برای ارزیابی میزان خطر یک عامل خطر در جامعه را فراهم می‌کنند به طوری که با حذف عامل خطر، چه میزانی از رخداد پیامد به طور مطلق و نسبت به جامعه مورد بررسی قابل پیشگیری است. در عمل در هر دو گروه مواجهه یافته و مواجهه نیافته، تعدادی از رخداد پیشامد مورد نظر مشاهده می‌شود (که البته انتظار می‌رود که رخداد پیامد در گروه مواجهه بیشتر باشد).

زمانی که مواجهه حذف گردد، انتظار می‌رود تعداد افرادی که پیامد مورد نظر را تجربه می‌کنند برابر همان تعداد افراد گروه مواجهه نیافته گردد که پیامد مورد نظر را تجربه می‌کنند. بنابراین در جامعه کلی که ترکیبی از گروه‌های مواجهه یافته و نیافته هستند، میانگین وزنی میزان رخداد پیامد در دو گروه مشاهده خواهد شد. اختلاف بین نسبت رخداد پیامد در جامعه کلی و گروه مواجهه نیافته برآوردی از اثر مواجهه، اندازه اثر مطلق مواجهه و یا AR به دست می‌دهد. فرمول آن به صورت زیر می‌باشد:

میزان خطر در گروه مواجهه نیافته - میزان خطر کلی = AR
 که در آن میزان خطر کلی برابر نسبت موارد رخداد پیامد در گروه‌های مواجهه یافته و نیافته می‌باشد و بر اساس جدول ۲×۲ متداول فرمول AR به صورت زیر خواهد بود:

$$AR = [(a+c) \div n] - [c \div (c+d)]$$

 برای محاسبه AF، که معیاری نسبی از میزان کاهش خطر با حذف عامل خطر را در مقایسه با جامعه کلی نشان می‌دهد، AR به میزان خطر کلی تقسیم می‌شود. فرمول آن به صورت زیر می‌باشد:

میزان خطر کلی ÷ (میزان خطر در گروه مواجهه نیافته - میزان خطر کلی) = AF = AR

$$AF = AR \div [(a+c) \div n] = ((a+c) \div n - [c \div (c+d)]) \div [(a+c) \div n]$$

 پس از ضرب نمودن صورت و مخرج در کمیت‌های خاص و پس از ساده‌سازی فرمول‌های زیر بر اساس میزان شیوع عامل مواجهه در جامعه (Pe) و میزان شیوع عامل مواجهه در گروه مورد (Pc) به دست می‌آید:

$$AF = [Pe (RR-1)] \div [1 + Pe (RR-1)] \quad Pe = [(a+b) \div n]$$

$$AF = [Pc (RR-1)] \div [RR-1] \quad Pc = [a \div (a+c)]$$

فرمول‌های فوق زمانی که خطر نسبی و میزان‌های شیوع در دسترس باشند قابل محاسبه هستند.

مقادیر عددی این شاخص‌ها، به‌عنوان اندازه اثر، اهمیت عملی (بالینی) رابطه را نشان می‌دهد. اما برای این که بررسی شود آیا این رابطه از لحاظ آماری معنی‌دار است یا خیر، نیاز است که در برای آن فاصله اطمینان (۹۵ درصدی) محاسبه شود و یا این که برای آن آزمون فرض انجام گردد [۲].

نکته:

شاخص‌های دیگری که در برخی از متون در این راستا ارائه می‌شود کسر قابل انتساب در گروه مواجهه ((AFE) Attributable Fraction for Exposed Group) و ((AFP) Attributable Fraction for Population) می‌باشد.

شاخص AFE سهمی از افراد گروه مواجهه را به‌دست می‌دهد که اگر مواجهه حذف گردد، پیامد مورد نظر را تجربه نخواهند کرد و با فرمول زیر به دست می‌آید:

$$AFE = \frac{[a \div (a+b)] - [c \div (c+d)]}{[a \div (a+b)]} = \frac{(P_1 - P_0)}{P_1}$$

$$AFE = \frac{(P_1 - P_0)}{P_1} = \frac{[RD]}{[a \div (a+b)]}$$

شاخص AFP سهمی از افراد جامعه را به‌دست می‌دهد که اگر مواجهه حذف گردد، پیامد مورد نظر را تجربه نخواهند کرد.

بنابراین اگر N_0 و N_1 تعداد کل افراد در معرض مواجهه و عدم مواجهه باشد، آنگاه $N_1 \times (P_1 - P_0) / P_1$ با حذف مواجهه دیگر پیامد را تجربه نخواهند کرد. بنابراین نسبت به کل

جامعه کسر افراد با این خصوصیت به صورت زیر خواهد بود:

$$\frac{[N_1 \times (P_1 - P_0) / P_1]}{(N_1 + N_0)}$$

و AFP با فرمول زیر به دست می‌آید:

$$AFP = AFE \times \text{percent exposed among cases}$$

نکته:

۱- در مطالعات همگروهی، پس از اعمال تقسیم صورت و مخرج کسر فوق به P_0 یعنی میزان رخداد پیامد در گروه عدم مواجهه، AFE برابر خواهد بود با:

$$AFE = [RR - 1] \div [RR]$$

بر اساس این منطق AFE در مطالعات مورد-شاهدی (با استفاده از OR به جای RR) برابر خواهد بود با:

$$AFE = [OR - 1] \div [OR]$$

۲- پس از محاسبه AFE، AFP نیز با ضرب این شاخص در نسبت افراد مواجهه یافته در گروه مورد به‌دست می‌آید.

نحوه انجام محاسبات AR، AF و AFE در مطالعات همگروهی و انجام استنباط در مورد آن‌ها

در مطالعات هم‌گروهی، انجام محاسبات این شاخص‌ها و انجام استنباط در مورد آن، در قالب یک مثال روشن خواهد شد.

مثال ۱

در مطالعه‌ای با هدف بررسی رابطه میان مصرف سیگار، مصرف الکل با مرگ ناشی از سرطان‌های کولورکتال [۳]، نتایج در جدول ۱ نشان داده شده است:

جدول ۱- فراوانی مرگ و میر افراد مبتلا به سرطان کولورکتال برحسب رده‌های متغیرهای وضعیت کشیدن سیگار

وضعیت کشیدن سیگار	تعداد مرگ و میر	تعداد افراد زنده	تعداد کل	نسبت مرگ و میر
در حال حاضر یا قبلاً مصرف کرده (مواجهه)	۹۷	۱۸۷	۲۸۴	۰/۳۴۲
هرگز مصرف نکرده (عدم مواجهه)	۲۲۵	۶۰۷	۸۳۲	۰/۲۷۰
کل (مجموع مواجهه و عدم مواجهه)	۳۲۲	۷۹۴	۱۱۱۶	۰/۲۸۹

منبع: [۵]

آمد. برای بررسی این موضوع که چه میزان از خطر مرگ ناشی از سرطان کولورکتال را می‌توان به سیگاری بودن افراد نسبت داد نیاز است خطر قابل انتساب و کسر قابل انتساب محاسبه گردد. در مثال حاضر، این کمیت برابر

احتمال (خطر) کلی مرگ ناشی از سرطان کولورکتال در افراد سیگاری و غیر سیگاری (جامعه)، برابر $۰/۲۸۹ (۲۸/۹\%)$ درصداً، و احتمال (خطر) مرگ ناشی از سرطان کولورکتال در افراد غیرسیگاری، برابر $۰/۲۷۰ (۲۷/۰\%)$ به‌دست

برای محاسبه شاخص AFP نیاز است نسبت افراد سیگاری (مواجهه) در میان آن‌هایی که مرگ را تجربه کردند (percent exposed among cases) به دست آید که حاصل تقسیم ۹۷ به ۳۲۲ می‌باشد (۰/۳۰۱). بنابراین AFP برابر ۰/۰۶۳ (= ۰/۳۰۱ × ۰/۲۰۸) خواهد بود که نشان می‌دهد با حذف مصرف سیگار در کل جامعه، ۶/۳ درصد از مرگ و ناشی از سرطان کولون را می‌توان جلوگیری نمود.

نکته

۱- برای انجام محاسبات (فرمول‌ها در ضمیمه ۱) می‌توان از کد `csi` در نرم‌افزار STATA استفاده نمود (607 187 225 97 csi برای مثال ۱ و ۲).

نحوه انجام محاسبات AR، AF و AFE در مطالعات

مورد-شاهدی و انجام استنباط در مورد آن‌ها

در مطالعات مورد شاهدی، انجام محاسبات این شاخص‌ها و انجام استنباط در مورد آن، در قالب یک مثال روشن خواهد شد.

مثال ۳

در یک مطالعه مورد شاهدی برای بررسی رابطه رفتار مصرف سیگار با برخی از فاکتورهای محیطی در نوجوانان دانش‌آموز و غیر دانش‌آموز، تعداد ۱۸۰ مورد (سیگاری) و ۴۷۹ شاهد (غیرسیگاری) انتخاب و سابقه اصرار همسالان در آن‌ها پرسیده شد. نتایج این بررسی در جدول ۲ خلاصه شده است:

است با ۰/۰۱۹ یا ۱/۹٪ (۲۸/۹) منهای ۲۷/۰ درصد). بنابراین ۱/۹ درصد از خطر مرگ ناشی از سرطان کولون را می‌توان به مصرف سیگار نسبت داد (منتسب کرد).

مقدار شاخص AF بر اساس تقسیم AR یعنی عدد ۱/۹٪ به میزان خطر کلی یعنی ۲۸/۹٪ به دست می‌آید که برابر است با ۰/۰۶۳ یا ۶/۳٪ که نشان می‌دهد در مقایسه با کل جامعه می‌توان تقریباً ۶ درصد از مرگ و ناشی از سرطان کولون را به مصرف سیگار منتسب نمود.

همچنین شاخص AFE بر اساس تقسیم RD یعنی عدد ۷/۲٪ (حاصل تفاضل ۰/۳۴۲ و ۰/۲۷۰) به میزان خطر در گروه مواجهه یعنی ۳۴/۲٪ به دست می‌آید که برابر است با ۰/۲۰۸ یا ۲۰/۸٪ که نشان می‌دهد با حذف مصرف سیگار در گروه مواجهه، ۲۱ درصد از مرگ و ناشی از سرطان کولون را می‌توان جلوگیری نمود. در عمل برای AFE فاصله اطمینان ۹۵ درصدی نیز محاسبه می‌شود تا دامنه‌ای را مشخص نمود که با حذف مواجهه می‌توان خطر رخداد پیامد را کاهش داد. به علاوه این شاخص با استفاده از فرمول $(RR-1)/RR$ نیز به دست می‌آید. در این مثال مقدار RR برابر ۱/۲۶۳ [۴] و بر اساس فرمول مقدار شاخص AFE برابر ۰/۲۰۸ به دست می‌آید.

مثال ۲

برای مثال ۱، فاصله اطمینان AFE برابر (۰/۳۵، ۰/۰۴) به دست آمد. بنابراین در جامعه آماری مورد بررسی با اطمینان ۹۵ درصد، نسبت به افراد سیگاری، حداقل و حداکثر کسر مرگ و میر منتسب به مصرف سیگار افراد مبتلا به سرطان کولون در دامنه فوق قرار دارد.

جدول ۲- رابطه رفتار مصرف سیگار و سابقه اصرار همسالان

سابقه اصرار همسالان	مصرف سیگار		تعداد کل	نسبت افراد سیگاری
	تعداد موارد سیگاری	تعداد موارد غیر سیگاری		
دارد	۶۷	۳۳	۱۰۰	۰/۶۷۰
ندارد	۱۱۳	۴۴۶	۵۵۹	۰/۲۰۲
تعداد کل	۱۸۰	۴۷۹	۶۵۹	۰/۲۷۳
نسبت افراد با سابقه اصرار همسالان	۰/۳۷۲	۰/۰۶۹	۰/۱۵۲	

منبع: [۶]

احتمال (خطر) مصرف سیگار در افراد بدون سابقه اصرار همسالان (غیر مواجهه)، برابر ۰/۲۷۳ (۲۷/۳ درصد) و

احتمال (خطر) مصرف سیگار در افراد با و بدون سابقه اصرار همسالان (جامعه)، برابر ۰/۲۷۳ (۲۷/۳ درصد)، و

۰/۳۲۶ (= ۰/۸۷۵ × ۰/۳۷۲) خواهد بود که نشان می‌دهد با حذف سابقه اصرار همسالان در کل جامعه، ۳۲/۶ درصد از مصرف سیگار را می‌توان جلوگیری نمود.
نکته

۱- برای انجام محاسبات (فرمول در ضمیمه ۲) می‌توان از کد cci در نرم‌افزار STATA استفاده نمود (cci 67) 113 33 446 برای مثال ۳ و ۴).

۲-۱) محاسبه ARR و NNT و استنباط در مورد آن‌ها (فاصله اطمینان و آزمون فرض) در مطالعات مداخله‌ای در مطالعات مداخله‌ای (مخصوصاً انواع کارآزمایی) علاوه بر شاخص‌های متداول نظیر RR و OR، بسته به سازوکار تاثیرگذاری مداخله (که سبب بهبودی می‌شود یا از خطر محافظت می‌کند)، اندازه اثر مداخله را با معیارهای مختلفی می‌توان ارائه نمود:

شاخص‌های مطلق نظیر: شاخص مطلق کاهش خطر (Absolute Risk Reduction: ARR) که برابر تفاضل میزان خطر (بهبودی) در گروه مداخله در مقایسه با گروه مقایسه می‌باشد و در واقع سهم مطلق و خالص تاثیر مداخله را نشان می‌دهد.

شاخص‌های نسبی نظیر: شاخص نسبی کاهش خطر (Relative Risk Reduction: RRR) که میزان کاهش خطر یا افزایش بهبودی ناشی از مداخله را در مقایسه با گروه گواه (درمان استاندارد، دارونما یا شاهد) و در واقع سهم نسبی و خالص تاثیر مداخله نسبت به گروه گواه را نشان می‌دهد.

شاخص ARR نشان می‌دهد که اگر مداخله در گروهی از افراد به‌کار رود، فارغ از سایر عوامل موثر، محافظت از خطر یا بهبودی به چه میزان خواهد بود. برای کاربردی‌تر کردن این شاخص، اگر این سوال مطرح شود که مداخله برای چند نفر استفاده شود تا یک مورد محافظت یا بهبودی حاصل گردد، کافی است معکوس این شاخص را محاسبه کرد که به آن تعداد مورد نیاز برای درمان (Number Needed to Treat: NNT) گفته می‌شود تا در نتیجه آن یک مورد سودمندی (محافظت یا بهبودی) حاصل گردد. شاخص NNT یک از پُرکاربردترین شاخص‌های

به‌دست آمد. برای بررسی این موضوع که چه مقدار از مصرف سیگار را می‌توان به سابقه اصرار همسالان نسبت داد نیاز است خطر قابل انتساب و کسر قابل انتساب محاسبه گردد.

برای محاسبه خطر قابل انتساب بر اساس فرمول:

میزان خطر در گروه مواجهه نیافته - میزان خطر کلی = AR

$$AR = [(a+c) \div n] - [c \div (c+d)]$$

در مثال حاضر، این کمیت برابر است با ۰/۰۷۱ یا ۷/۱٪ (۲۷/۳ منهای ۲۰/۲ درصد). بنابراین ۷/۱ درصد از مصرف سیگار را می‌توان به سابقه اصرار همسالان نسبت داد (منتسب کرد).

مقدار شاخص AFE بر اساس تقسیم AR یعنی عدد ۷/۱٪ به میزان خطر کلی یعنی ۲۷/۳٪ به‌دست می‌آید که برابر است با ۰/۲۶۰ یا ۲۶/۰٪ که نشان می‌دهد در مقایسه با کل جامعه ۲۶ درصد از مصرف سیگار را می‌توان به سابقه اصرار همسالان منتسب نمود.

همچنین شاخص AFE با استفاده از فرمول (OR-1)/OR (که در آن OR جایگزین RR در فرمول این شاخص برای مطالعات مورد-شاهدی شده است)، به‌دست می‌آید. در این مثال مقدار OR برابر ۸/۰۱ [۴] و بر اساس فرمول مقدار شاخص AFE برابر ۰/۸۸ به‌دست می‌آید. این شاخص نشان می‌دهد با حذف اثر سابقه اصرار همسالان، ۸/۸ درصد از مصرف سیگار را می‌توان جلوگیری نمود. در عمل برای AFE، فاصله اطمینان ۹۵ درصدی نیز محاسبه می‌شود تا دامنه‌ای را مشخص نمود که با حذف مواجهه می‌توان خطر رخداد پیامد را کاهش داد.

مثال ۴

برای مثال ۳، فاصله اطمینان AFE برابر (۰/۹۲ ، ۰/۸۰) به‌دست آمد. بنابراین در جامعه آماری مورد بررسی با اطمینان ۹۵ درصد، نسبت به افراد با سابقه اصرار همسالان، حداقل و حداکثر کسر افراد سیگاری منتسب به سابقه اصرار همسالان در دامنه فوق قرار دارد.

برای محاسبه شاخص AFP نیاز است نسبت افراد در معرض مواجهه با سابقه اصرار همسالان در افراد سیگاری (percent exposed among cases) به‌دست آید که حاصل تقسیم ۶۷ به ۱۸۰ می‌باشد (۰/۳۷۲). بنابراین AFP برابر

مباحث اقتصادی و بار هزینه‌ای درمان نیز می‌توان استفاده کرد چون میزان تلاش و یا هزینه لازم برای ایجاد یک مورد سودمندی را به دست می‌دهد. از کاربردهای دیگر شاخص‌های فوق مقایسه میان درمان‌های مختلف برای پیشگیری از یک عارضه مشخص می‌باشد (بدیهی است درمانی بهتر است که مقدار NNT آن کوچک‌تر باشد).

معایب ARR و NNT

میزان خطر پایه (خطر رخداد پیامد قبل از انجام مداخله) روی مقادیر این شاخص‌ها تاثیر می‌گذارد به طوری که به ازای مقدار ثابتی از RRR، با کاهش (افزایش) مقادیر خطر پایه، NNT افزایش (کاهش) می‌یابد. به عبارت دیگر **میزان شیوع رخداد پیامد مورد نظر روی مقادیر NNT و ARR تاثیر می‌گذارد.** در استفاده از این شاخص نیاز است این مسئله مد نظر قرار گیرد و یا این که در صورت امکان این شاخص را روی مقادیر پایه تعدیل نمود. نکته دیگر این که میزان رخداد پیامد در گروه مقایسه (شاهد) نیز روی این شاخص تاثیر می‌گذارد به طوری که به ازای مقادیر بزرگ (کوچک) این میزان، حتی مقادیر کوچک (بزرگ) شاخص RRR منجر به مقادیر کوچک (بزرگ) برای NNT می‌گردد. از محدودیت‌های دیگر NNT در این است که مثلاً اگر مقدار ۱۰ برای این شاخص به دست آید به این معنی است که ۱۰ نفر باید درمان را دریافت کنند تا یک نفر محافظت گردد یا بهبود یابد ولی بر اساس این شاخص نمی‌توان مشخص نمود که کدام یک از این ۱۰ نفر در معرض مخاطره بیشتری هستند تا درمان برای آن فرد تجویز گردد.

مثال ۵

در یک مطالعه کارآزمایی بالینی تصادفی کنترل‌دار، با هدف بررسی تاثیر مکمل امگا ۳ بر روی نروپاتی در دو گروه امگا ۳ و دارونما به ترتیب تعداد ۳۰ و ۲۷ نفر شرکت داشتند که از این میان در گروه‌های امگا ۳ و دارونما به ترتیب تعداد ۲۱ و ۱۱ نفر پس از مداخله وضعیت نرمال داشتند [۷] (بنابراین به ترتیب تعداد ۹ و ۱۶ نفر در سایر وضعیت‌های نروپاتی شامل خفیف، متوسط و شدید قرار داشتند) (جدول ۳).

از ریایی سودمندی مداخله است و تعداد افرادی را که باید درمان جدید را در مقایسه با درمان استاندارد، دارونما یا شاهد دریافت کنند، نشان می‌دهد و راهی برای توصیف نتایج یک کارآزمایی بالینی و معرف **تعداد بیمارانی است که باید از درمان جدید استفاده کنند تا در مقایسه با درمان قدیمی به یک نتیجه مطلوب‌تر در زمینه بهبود بیماری مورد بررسی یا نجات جان یک نفر بیشتر برای بیماری کشنده مورد بررسی دست یافت.** در حالتی که بین دو روش درمانی اختلاف معنی‌داری وجود ندارد و یا زمانی که درمان جدید به صورت معنی‌داری بهتر از درمان قدیمی نیست، NNT می‌تواند به عنوان یک معیار بالینی برای این منظور کمک کننده باشد.

حالت ایده‌آل برای مداخله زمانی است که برای هر فردی که مداخله را دریافت می‌کند سودمندی حاصل گردد و در مقابل برای هر فردی که در گروه مقایسه قرار می‌گیرد، سودمندی حاصل نشود. در این صورت مقدار ARR و در نتیجه NNT برابر یک به دست خواهد آمد. در عمل وجود اندازه اثرهای بزرگ برای مداخله مقادیر نزدیک به ۱ (مثلاً ۲ یا ۳) را برای NNT نتیجه می‌دهد. البته ممکن است در موقعیت‌های مختلف مقادیر NNT متفاوتی از نظر بالینی حائز اهمیت تلقی گردند؛ به عنوان مثال، در پیشگیری از پره‌اکلامپسی با درمان آسپیرین، برای این شاخص مقادیر خیلی بزرگ (نظیر ۲۰ تا ۴۰) مشاهده می‌شود ولی بازم این مقدار از نظر بالینی حائز اهمیت است. از طرف دیگر زمانی که مداخله تاثیر نداشته باشد و یا این که تاثیر آن مشابه گروه مقایسه باشد، ARR نزدیک به صفر خواهد بود و در نتیجه NNT مقداری بی‌نهایت خواهد داشت.

مزایای ARR و NNT

شاخص‌های ARR و NNT محاسبات ساده‌ای دارند و اندازه‌ای قابل فهم و با تفسیر ساده برای بیان سودمندی اثر مداخله فراهم می‌کنند. به علاوه این شاخص‌ها سهم یا اثر ویژه یک درمان را نسبت به گروه مقایسه (نظیر درمان استاندارد، دارونما و یا شاهد) ارائه می‌کند و به همین دلیل در سیاست‌گذاری‌های درمانی آن درمان خاص در مراقبت‌های بالینی از آن‌ها استفاده می‌شود. از NNT برای

جدول ۳- بررسی تاثیر مداخله روی بروز نروپاتی

احتمال (خطر) بروز نروپاتی	احتمال بهبودی	تعداد کل	نروپاتی	
			غیر نرمال	نرمال
۰/۳۰	۰/۷۰	۳۰	۹	۲۱
۰/۵۹	۰/۴۱	۲۷	۱۶	۱۱

منبع: [۷]

عبارت دیگر میزان کاهش نسبی نروپاتی ناشی از مداخله در مقایسه با گروه دارونما برابر ۷۱ درصد می باشد.

فاصله اطمینان برای NNT

در عمل مقادیر شاخص های فوق به تنهایی برای تصمیم گیری کافی نیست و فاصله اطمینان ۹۵ درصدی آن محاسبه می شود و زمانی که عدد یک یعنی عدد متناظر با عدم تاثیر در این فاصله قرار گیرد، نتیجه این می شود که مداخله روی پیامد تاثیر معنی دار ندارند. فاصله اطمینان NNT دامنه تاثیر گذاری مداخله را در مقایسه با گروه شاهد نشان می دهد.

مثال ۶

برای مثال ۵، فاصله اطمینان RRR، ARR و NNT به ترتیب برابر (۱/۸۷، ۰/۰۳)، (۰/۵۰، ۰/۰۴) و (۲/۰، ۲۸/۵) به دست آمد. بنابراین مقادیر این شاخص ها، در جامعه آماری مورد بررسی با اطمینان ۹۵ درصد در دامنه های فوق قرار گرفت. به عبارت دیگر بر اساس NNT، تعداد افرادی که باید ۳ امگا کنند تا یک مورد سودمندی حاصل گردد حداقل ۲ و حداکثر ۲۸ نفر می باشند.

بر اساس شاخص RRR چون فاصله اطمینان آن عدد ۱ را در بر داشت، بنابراین مداخله مکمل یاری با استفاده از ۳ امگا در مقایسه با گروه دارونما، تاثیر معنی داری روی پیامد نروپاتی نداشته است. به طور معادل نتایج آزمون فرض نتایج آزمون کای دو نیز در راستای نتیجه حاصل از فاصله اطمینان می باشد (P = ۰/۰۵۱) (ضمیمه ۳).

نکته:

۱- با توجه به پیچیده بودن محاسبات مربوط به فاصله اطمینان RRR، ARR و NNT، برای انجام محاسبات مربوط، به CEBM Statistics Calculator با انتخاب

بنابراین، احتمال (خطر) بروز نروپاتی در گروه ۳ امگا، برابر ۹ بر ۳۰ یعنی ۰/۳۰ (۳۰/۰ درصد)، احتمال (خطر) بروز نروپاتی در گروه دارونما، برابر ۱۶ بر ۲۷ یعنی ۰/۵۹ (۵۹/۰ درصد) به دست آمد (جدول ۳). به طور معادل احتمال بهبودی یا قرار گرفتن در رده نرمال (بر اساس قاعده متمم در احتمال [۳]) برابر ۰/۷۰ و ۰/۴۱ به ترتیب برای گروه ۳ امگا و دارونما به دست آمد. همان طور که ملاحظه می شود تمام ۷۰ درصد بهبودی در گروه ۳ امگا را نمی توان ناشی از این مداخله دانست، چون در گروه دارونما هم میزانی از بهبودی مشاهده شده است. با در نظر گرفتن شرایط یکسان برای دو گروه می توان ادعا کرد که ۴۱ درصد از موارد بهبودی در گروه دارونما ناشی از سایر عواملی است که تحت شرایط مطالعه قابل کنترل نبودند. بنابراین سهم خالص (قابل منتسب) برای بهبودی ناشی از ۳ امگا را می توان با تفاضل این دو عدد به صورت زیر به دست آورد:

$$۰/۷۰ - ۰/۴۱ = ۰/۲۹$$

در نتیجه با استفاده از فقط ۳ امگا می توان ۲۹ درصد از موارد نروپاتی را درمان نمود یا به عبارت دیگر سهم ۳ امگا در درمان نروپاتی ۲۹ درصد است. این شاخص همان ARR می باشد؛ در راستای تفسیر این شاخص می توان گفت: با استفاده از فقط مکمل ۳ امگا می توان خطر ناشی از نروپاتی را تا ۲۹ درصد کاهش داد، یعنی انتظار می رود با استفاده از این مکمل در تقریباً در ۳ نفر از افراد بیمار، یکی از آنها بهبود یابد که این کمیت با معکوس نمودن شاخص ARR حاصل شده است و همان شاخص NNT می باشد.

برای محاسبه RRR نیاز است ۰/۲۹ را به میزان بهبودی در گروه کنترل (۰/۴۱) تقسیم نمود که برابر ۰/۷۱ می شود، به

۲) تغییرات اندازه‌گیری‌های بین ارزیاب‌های مختلف: که در این حالت اندازه‌گیری معینی توسط چند نفر مختلف تکرار می‌شود.

بر اساس روش اول، تغییرپذیری تک تک مشاهدات را می‌توان تفسیر نمود و بر اساس روش دوم ویژگی‌های عمومی روش اندازه‌گیری را می‌توان بررسی نمود. دو روش متفاوت برای اندازه‌گیری موضوع یکسانی را می‌توان با استفاده از حدود ۹۵ درصدی توافق ارزیابی کرد.

در این راستا روش‌های زیر، متداول هستند:

- محاسبه حساسیت و ویژگی در ارزیابی تست‌های تشخیصی به منظور بررسی توافق بین دو اندازه‌گیری اسمی دو حالتی، با در نظر گرفتن یکی از اندازه‌گیری‌ها به عنوان معیار استاندارد (Golden Standard)

- محاسبه ICC و خطای معیار اندازه‌گیری (Standard Error of Measurement: SEM) برای اندازه‌گیری‌های کمی

- حدود توافق بلند و آلتمن (Bland and Altman's limit of agreement) [۸] و ضریب همبستگی توافق (Lin's concordance correlation coefficient) برای اندازه‌گیری‌های کمی [۹]

- محاسبه کاپا و کاپای وزنی برای اندازه‌گیری‌های به ترتیب با مقیاس اسمی و رتبه‌ای

نکات

۱- برای بررسی توافق یک متغیر کمی و یک متغیر اسمی دو حالتی (معیار استاندارد)، با در نظر گرفتن یک نقطه برش برای متغیر کمی، از تحلیل منحنی ROC (Receiver Operating Characteristics) استفاده می‌شود که در سری‌های بعدی ارائه خواهد شد.

۲- شاخص ICC و ملاحظات مربوط به آن در سری همبستگی و رگرسیون ارائه شده است [۱۰].

۱-۲) محاسبه فاصله اطمینان حساسیت، ویژگی، PPV، NPV، LR+ و LR- در مطالعات تشخیصی و غربالگری در مورد ماهیت و نحوه محاسبه و تفسیر شاخص‌های حساسیت، ویژگی، PPV، NPV، قبلاً اشاره شده است [۳]. در قالب یک مثال، ماهیت و نحوه انجام محاسبات سایر

Randomized Controlled Trial (سایت زیر) پیشنهاد

می‌شود:

<http://ktclearinghouse.ca/cebm/practise/ca/calculator/statscalc> (Access date: 2/2014)

۲- به‌طور معادل برای محاسبه مقادیر ARR و NNT، و همچنین فاصله اطمینان برای NNT می‌توان از کد bcii نرم‌افزار STATA (نسخه‌های ۱۱ و بالاتر یا نسخه آپدیت شده ۱۰) برای این منظور استفاده کرد (bcii 21 9 11 16 برای مثال ۵ و ۶).

۳- فاصله اطمینان NNT را باید با احتیاط تفسیر نمود؛ در برخی از موقعیت‌های عملی که اندازه اثر مداخله کوچک باشد، ممکن است فاصله اطمینان این شاخص عدد صفر را در بر داشته باشد و در این صورت یک حد منفی و یک حد مثبت خواهد داشت و در چنین موقعیت‌های دو اشکال اساسی به وجود می‌آید: اول این که NNT مقداری ذاتاً مثبت است و حد منفی برای آن از لحاظ ساختاری اشکال دارد (در دامنه تعریف NNT قرار نمی‌گیرد) و دوم این که در چنین موقعیت‌هایی فاصله اطمینان NNT مقدار برآورد عددی آن را در بر نمی‌گیرد. **حد منفی در واقع مضر بودن مداخله (درمان) را نشان می‌دهد و به تعداد مورد نیاز برای زیان Number Neede to Harm (NNH) تفسیر می‌شود که تعداد مداخله مورد نیاز برای ایجاد یک مورد پیامد ناگوار را نشان می‌دهد.** به همین دلیل برخی از مولفین این موارد عملی را از محدودیت‌های شاخص‌های فوق دانسته و پیشنهاد می‌کنند که این شاخص‌ها بدون فاصله اطمینان ارائه شوند.

۲) معیارهای توافق (Agreemen) در اندازه‌گیری‌ها

تقریباً در همه اندازه‌گیری‌ها در مطالعات (بالینی و غیر بالینی) خطا وجود دارد که آنرا می‌توان به یکی از روش‌های زیر اندازه‌گیری نمود:

۱) تغییرات درونی اندازه‌گیری‌های فردی: که در این حالت اندازه‌گیری معینی توسط یک فرد تکرار می‌شود.

شاخص‌های مورد نیاز و همچنین فاصله اطمینان همه شاخص‌های فوق روشن خواهد شد.

مثال ۷

در مطالعه‌ای برای بررسی ارزش تشخیصی گلوکز پلاسمای ناشتا (Fasting Plasma (FPG) Glucose) در غربال‌گری دیابت بارداری (Gestational Diabetes Mellitus)

(GDM))، ۱۰۰ زن در دوران بارداری برای این مطالعه انتخاب شدند که از میان آن‌ها بر اساس معیار استاندارد ۳۶ نفر دارای GDM و ۶۴ نفر بقیه سالم بودند. همچنین بر اساس FPG، که در این مطالعه به‌عنوان تست تشخیصی در نظر گرفته شد، و با احتساب نقطه برش ۹۱ (mg/dl)، نتایج جدول ۴ حاصل شد:

جدول ۴- ارزش تشخیصی FPG برای GDM

نتیجه FPG با نقطه برش ۹۱	GDM		تعداد کل
	دارند (بیمار)	ندارند (سالم)	
+	۲۳	۱۳	۳۶
-	۱۳	۵۱	۶۴
تعداد کل	۳۶	۶۴	۱۰۰

منبع: [۱۱]

مقایسه با بیمار بدون عارضه مورد نظر می‌باشد. شاخص LR+، شانس رخداد نتیجه مثبت واقعی به مثبت کاذب را در نتایج مثبت تست و LR- شانس رخداد نتیجه منفی کاذب به مثبت واقعی را در نتایج منفی تست نشان می‌دهد. از کاربردهای این شاخص در ارزیابی مناسب بودن آزمون‌های تشخیصی و مقایسه میان آزمون‌های مختلف و انتخاب آزمون بهینه برای یک هدف مشخص می‌باشد. در این راستا مزیت آن‌ها نسبت به شاخص‌های دیگر در این است که نسبت به میزان شیوع پیامد مورد نظر کم‌تر حساس هستند و در نتیجه انتخاب تست تحت تاثیر میزان شیوع یعنی یک خصوصیت خارجی قرار نمی‌گیرد.

مقادیر LR بزرگ‌تر از ۱ (کوچکتر از ۱) به ترتیب نشان می‌دهد نتایج آزمون با حضور (عدم) بیماری مرتبط است. آزمون‌های تشخیصی که مقادیر LR در آن‌ها به یک نزدیک است از لحاظ بالینی اهمیت قابل ملاحظه‌ای ندارند. در عمل مقادیر بین ۱ تا ۲، ۲ تا ۵، ۵ تا ۱۰ و بالاتر از ۱۰ برای شاخص‌های LR+ و LR- به ترتیب قدرت تست تشخیصی را به‌طور مستقیم و اندازه اثر در حد کوچک، متوسط، بزرگ و خیلی بزرگ نشان می‌دهد. مقادیر بین ۰/۵ تا ۱، ۰/۲ تا ۰/۵، ۰/۱ تا ۰/۲ و پایتتر از ۰/۱ برای این شاخص‌ها به ترتیب قدرت تست تشخیصی را به‌طور معکوس و اندازه اثر در حد کوچک، متوسط، بزرگ و خیلی بزرگ نشان می‌دهد.

در این مثال، حساسیت یعنی توان جداسازی افراد بیمار برای تست تشخیصی برابر ۶۴/۱۰۰، ویژگی یا توان جداسازی افراد سالم برای تست تشخیصی برابر ۸۰/۱۰۰ درصد، PPV یا میزان اطمینان به نتایج مثبت تست تشخیصی برابر ۶۴/۱۰۰، NPV یا میزان اطمینان به نتایج منفی تست تشخیصی برابر ۸۰/۱۰۰ درصد محاسبه شد.

نکته

شاخص‌های PPV و NPV به‌میزان شیوع پیامد در جامعه بستگی دارد و مقادیر آن‌ها متناسب با افزایش میزان شیوع به ترتیب افزایش و کاهش می‌یابد. در چنین موقعیت‌های پیشنهاد می‌شود از شاخص‌های درست‌نمایی مثبت (Likelihood Ratio +: LR+) و منفی (Likelihood Ratio -: LR-) استفاده گردد [۳].

شاخص‌های LR+ و LR-، نحوه محاسبه آن‌ها

در مباحث مربوط به آزمون‌های تشخیصی مخصوصاً در حیطه پزشکی مبتنی بر شواهد، شاخص‌های LR+ و LR- برای ارزیابی ارزش انجام یک تست تشخیصی استفاده می‌شود. این شاخص‌ها از حساسیت و ویژگی آزمون برای تعیین اینکه آیا نتیجه تست می‌تواند به‌طور مفیدی احتمال این که مثلاً آیا یک بیماری وجود دارد یا خیر را تغییر دهد؟ شاخص LR درست‌نمایی (شانس) رخداد مورد انتظار نتیجه آزمون در یک بیمار مبتلا به عارضه مورد نظر در

مثال ۸

با مراجعه به مثال ۷، LR+ و LR- تست به صورت زیر محاسبه می شود:

$$LR+ = (23/36) \div (13/64) = 3/15$$

$$LR- = (13/36) \div (51/64) = 0/45$$

بنابراین درست‌نمایی (شانس) مشاهده بیمار واقعی به بیمار کاذب در نتایج مثبت تست تقریباً ۳ برابر (اندازه اثر متوسط) و درست‌نمایی (شانس) مشاهده سالم کاذب به سالم واقعی در نتایج منفی تست تقریباً برابر نصف (اندازه اثر تقریباً متوسط) بوده است.

نکته

در عمل برای نیاز است برای شاخص‌های حساسیت، ویژگی، PPV، NPV، LR+ و LR-، فاصله اطمینان ۹۵ درصدی نیز محاسبه شود.

مثال ۹

با مراجعه به مثال ۷، نتایج محاسبات فاصله اطمینان‌های مربوط به مربوط به شاخص‌های شیوع، حساسیت، ویژگی، PPV، NPV، LR+ و LR- در جدول ۵ خلاصه شده است:

جدول ۵- نتایج محاسبه شاخص‌های شیوع، حساسیت، ویژگی، PPV، NPV، LR+ و LR- به همراه فاصله اطمینان ۹۵ درصدی آنها برای مثال ۷

شاخص	مقدار	L	U
شیوع	٪۳۶/۰	٪۲۷/۰	٪۴۶/۲
حساسیت	٪۶۳/۹	٪۴۶/۲	٪۷۹/۲
ویژگی	٪۷۹/۷	٪۶۷/۸	٪۸۸/۷
PPV	٪۶۳/۹	٪۴۶/۲	٪۷۹/۲
NPV	٪۷۹/۷	٪۶۷/۸	٪۸۸/۷
LR+	۳/۱۵	۱/۸۳	۵/۴۲
LR-	۰/۴۵	۰/۲۹	۰/۷۱

L و U: حد پایین و بالای فاصله اطمینان ۹۵ درصدی

نکته

۱- با توجه به پیچیده بودن محاسبات مربوط به فاصله اطمینان PPV، NPV، LR+ و LR- (ضمیمه ۴)، برای انجام محاسبات مربوط، به CEbm Statistics Calculator با انتخاب Diagnostic Test (سایت زیر) پیشنهاد می شود:

<http://ktclearinghouse.ca/cebm/practise/ca/calculator/statscalc> (Access date: 2/2014)

۲- برای محاسبه مقادیر حساسیت، ویژگی، PPV، NPV، LR+ و LR-، و همچنین فاصله اطمینان برای این شاخص‌ها می توان از کد diagti نرم افزار STATA (نسخه‌های ۱۱ و بالاتر یا نسخه آپدیت شده ۱۰) برای این منظور استفاده کرد (diagti 23 13 13 51) برای مثال ۷).

احتمال‌های پیشین و پسین و نحوه انجام محاسبات آنها

میزان شیوع پیامد ۳۶٪ و با اطمینان ۹۵ درصد در دامنه ۲۷٪ تا ۴۶٪ قرار می گیرد. برای شاخص‌های حساسیت و PPV و NPV، وضعیت در حد متوسط (بالای ۶۰ درصد) به ترتیب برای توان جداسازی افراد بیمار و میزان اعتماد به نتیجه مثبت تست مشاهده می شود همچنین برای شاخص‌های ویژگی و NPV، وضعیت در حد مطلوبی (بالای ۷۰ درصد) به ترتیب برای توان جداسازی افراد سالم و میزان اعتماد به نتیجه منفی تست مشاهده می شود. به علاوه LR+ و LR- وضعیت در حد متوسط تست را به ترتیب برای و میزان اعتماد به نتیجه مثبت و منفی تست نشان می دهند. در هر یک از موارد فوق فاصله اطمینان‌های ۹۵ درصدی محدوده قابل اطمینان حداقل و حداکثری را برای هر یک از شاخص‌ها، نشان می دهد.

نکته

۱- از لحاظ ریاضیاتی ثابت شده است که با در نظر گرفتن میزان شیوع به عنوان احتمال پیشین رخداد مورد نظر و بر اساس فرمول‌های فوق، احتمال پسین بیمار واقعی بودن در تست برابر PPV و احتمال پسین بیمار کاذب بودن در تست برابر $1-NPV$ خواهد بود (مثال ۱۰).

کاربرد احتمال‌های پیشین و پسین

احتمال‌های پیشین و پسین که در مثال فوق محاسبه شدند، کلی هستند. در واقع و به‌طور دقیق‌تر این احتمال‌ها را می‌توان برای تشخیص افرادی در بیماران به‌کار برد یعنی اگر احتمال پیشین بیماری به‌صورت افرادی موجود باشد احتمال پسین بیماری آن‌ها پس از به‌کارگیری تست تشخیصی روی آن‌ها چقدر خواهد بود؟

احتمال‌های پیشین و پسین این امکان را برای تفسیر بهتر نتایج تست تشخیصی و برای پیش‌بینی احتمال مثبت و منفی واقعی نتیجه فراهم می‌کند. به‌عنوان مثال، اگر پزشکی در تشخیص یک بیماری دچار تردید باشد و رخداد آن را ۵۰-۵۰ بدانند در این صورت شانس ۱ به ۱ برای رخداد (در مقابل عدم رخداد) پیامد وجود دارد و در نتیجه شانس پیشین رخداد برابر ۱ خواهد بود. با در نظر گرفتن $LR+$ برابر ۶ برای تست شانس پسین رخداد بیماری برابر $6 = (1 \times 1)$ خواهد شد و در نتیجه احتمال پسین برابر $0.86 = 6 \div (6 + 1)$ به‌دست می‌آید. بنابراین با داشتن یک تست تشخیصی خوب (با $LR+= 6$)، احتمال پسین برابر 0.86 خواهد بود و میزان اطمینان ۸۶ درصدی را برای پزشک در تشخیص بیماری فراهم می‌کند.

نموگرام محاسبه احتمال‌های پسین

نحوه استفاده از این نمودار بدین ترتیب است که احتمال پیشین مورد نظر را با یک خط به مقدار LR تست مورد بررسی در قسمت وسط وصل نموده و امتداد داده می‌شود تا خط احتمال پسین را قطع کند. بدن ترتیب با معلوم بودن احتمال پیشین و LR احتمال پسین به‌دست می‌آید.

احتمال پیشین به‌ترتیب احتمال رخداد پیامد مورد نظر را پیش از اجرای تست تشخیصی و احتمال پسین احتمال رخداد پیامد مورد نظر را بعد از اجرای تست تشخیصی را نشان می‌دهد که احتمال پیشین معادل میزان شیوع پیامد در جامعه و احتمال پسین بر اساس قضیه بیز [۳] به‌صورت ترکیب نتایج احتمال پیشین با درستنمایی رخداد پیامد مورد نظر توسط تست تشخیصی به‌دست می‌آید.

برای محاسبه احتمال‌های پیشین و پسین رخداد پیامد از نسبت درستنمایی تست استفاده می‌شود؛ به عبارت دیگر با داشتن احتمال پیشین مثبت (منفی) بودن تست و درستنمایی مثبت (منفی) تست می‌توان احتمال پسین مثبت (منفی) را به‌صورت زیر به دست آورد:

$$(\text{احتمال پیشین} - 1) \div (\text{احتمال پیشین}) = \text{شانس پیشین}$$

$$LR (+/-) \times \text{شانس پیشین} = \text{شانس پسین}$$

$$(\text{شانس پسین} + 1) \div (\text{شانس پسین}) = \text{احتمال پسین}$$

بر اساس فرمول‌های فوق، زمانی که LR بزرگ‌تر (کوچک‌تر) از یک باشد در این صورت احتمال پسین بزرگتر (کوچک‌تر) از احتمال پیشین خواهد بود. مثلاً زمانی که احتمال پیشین بیماری بین ۳۰ تا ۷۰ درصد باشد با داشتن LR بزرگ احتمال پسین مقداری بزرگ خواهد بود و بیمار بودن فرد را نتیجه می‌دهد. یا زمانی که احتمال پیشین بیماری بزرگ باشد با داشتن LR کوچک احتمال پسین مقداری کوچک خواهد بود و بیمار بودن فرد را نتیجه نمی‌دهد.

مثال ۱۰

با مراجعه به مثال ۷، نتیجه محاسبات روی فرمول‌های فوق به‌صورت زیر خواهد بود:

$$0.36 = 100 \div 36 = \text{احتمال پیشین بیماری (به‌طور کلی برای همه افراد و معادل میزان شیوع در جامعه)}$$

همه افراد و معادل میزان شیوع در جامعه

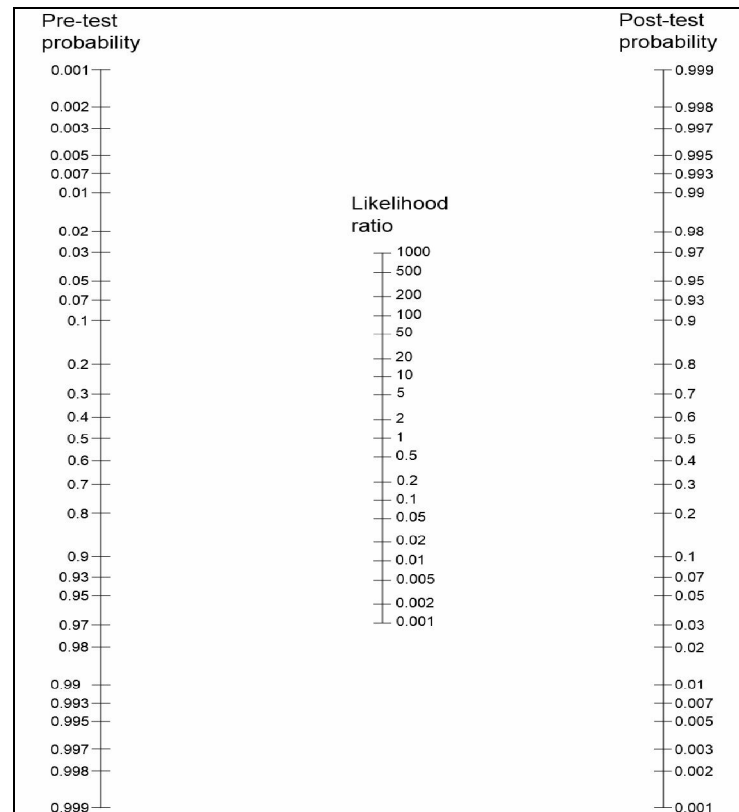
$$0.56 = (1 - 0.36) \div (0.36) = \text{شانس پیشین (بیمار واقعی بودن)}$$

$$1.77 = 0.56 \times 3/15 = \text{شانس پسین (بیمار واقعی بودن)}$$

$$0.64 = (1 + 1.77) \div (1.77) = \text{احتمال پسین (بیمار واقعی بودن)}$$

$$0.26 = 0.56 \times 0.45 = \text{شانس پسین (بیمار کاذب بودن)}$$

$$0.20 = (1 + 0.26) \div (0.26) = \text{احتمال پسین (بیمار کاذب بودن)}$$



نمودار ۱- نمودار محاسبه احتمال‌های پسین بر اساس احتمال پیشین و LR

به‌عنوان مثال اگر احتمال پیشین و LR به ترتیب ۰/۳ و ۱۰ باشند در این صورت احتمال پسین تقریباً برابر ۰/۸۲ خواهد بود.

واقع معیاری از پراکندگی مقادیر اندازه‌گیری شده بین افراد را نشان می‌دهد، در حالی که SEM انحراف معیار اندازه‌گیری‌های تکراری درونی یک فرد خاص را حول اندازه‌گیری واقعی همین فرد نشان می‌دهد. برای محاسبه SEM گام‌های زیر دنبال می‌شود:

- ابتدا ICC برای اندازه‌گیری‌ها به دست می‌آید،
- در گام بعدی انحراف معیار هر یک از اندازه‌گیری‌ها محاسبه می‌شود،
- سپس میانگین ساده آن‌ها به دست می‌آید (SD)،
- در نهایت SEM با استفاده از فرمول: $SEM = SD \times \sqrt{1-ICC}$ به دست می‌آید.

مثال ۱۱

با مراجعه به مثال ارائه شده در سری آمار: همبستگی و رگرسیون (۱۰)، مقدار ICC برابر ۰/۸۱ به دست آمد در صورتی که مقدار انحراف معیار برای اندازه‌گیری‌های اول و دوم به ترتیب برابر ۲۵ و ۳۰ برای اندازه‌گیری‌های مربوط

۲-۲) محاسبه خطای معیار اندازه‌گیری (SEM)

به این دلیل که اندازه‌گیری‌های انجام شده در مطالعات معمولاً با خطا همراه هستند و اندازه‌گیری وجود ندارد که بدون خطا و کامل انجام شود، اگر چند اندازه‌گیری روی یک فرد خاص انجام شود مقادیر این اندازه‌گیری‌ها یکسان نخواهد بود. شاخص SEM میزانی از پراکندگی اندازه‌گیری‌های تکراری درونی یک فرد خاص را حول اندازه‌گیری واقعی همین فرد نشان می‌دهد و به‌طور معکوس با پایایی رابطه دارد، به عبارت دیگر هرچه SEM کوچک‌تر باشد، پایایی بیشتر خواهد بود. ساده‌ترین روش برای معرفی SEM به این صورت است که: فرض کنید روی فرد خاصی اندازه‌گیری‌های تکراری از یک متغیر کمی صورت گیرد که در این صورت انحراف معیار اندازه‌گیری‌های تکراری همان SEM خواهد بود. در مقایسه با انحراف معیار اندازه‌گیری، انحراف معیار داده‌ها زمانی قابل محاسبه و مطرح است که این اندازه‌گیری روی افراد مختلفی برای آن تست خاص صورت گیرد که در

می‌شود و در صورت وجود مشاهدات انفرادی خارج از این محدوده، آن مشاهدات ممکن است عدم توافق را نتیجه دهند.

مثال ۱۲

در مطالعه‌ای با هدف بررسی روایی و پایایی پرسشنامه کیفیت زندگی ویژه افراد دیابتی (Iranian Diabetes Quality of Life: IDQOL)، پایایی ثبات این ابزار با روش آزمون-بازآزمون، توسط شاخص ICC بررسی گردید [۱۰] و نتیجه حاصل، حاکی از تایید پایایی ثبات پرسشنامه بود (مقدار شاخص برابر ۰/۸۱ که بزرگ‌تر از ۰/۷ به دست آمد) [۱۳]. در این بخش به منظور بررسی توافق اندازه‌گیری‌های آزمون و بازآزمون از ضریب همبستگی Lin و حدود توافق بلند و آلتمن استفاده خواهد شد:

نتایج محاسبات ضریب همبستگی توافق Lin

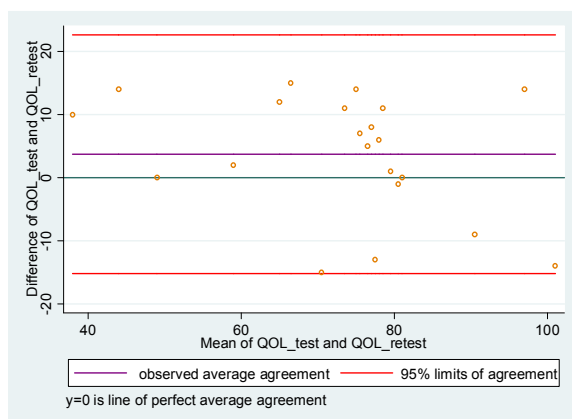
Lin Concordance correlation coefficient = 0.805, 95% CI = (.658 to .951)

توافق بین داده‌های کیفیت زندگی در دو مرحله آزمون-باز آزمون برابر ۰/۸۰۵ (بالای ۰/۷) به دست آمد که توافق بین این دو مرحله را تایید نمود. همچنین بر اساس فاصله اطمینان ۹۵ درصدی این ضریب حداقل و حداکثر توافق به ترتیب برابر ۰/۶۵۸ و ۰/۹۵۱ به دست آمد.

نتایج محاسبات حدود توافق بلند و آلتمن

Difference = qol_test - qol_retest	95% Limits Of Agreement (Bland & Altman)
Average 3.714	Std Dev. 9.644
	(-15.188 to 22.617)

نمودار حدود توافق بلند و آلتمن به صورت زیر می‌باشد:



نمودار ۲- حدود توافق بلند و آلتمن برای داده‌های آزمون-بازآزمون

به مقیاس کیفیت زندگی (نمره بین ۰ تا ۱۰۰) حاصل شده باشد در این صورت:

$$ICC = 0.81$$

$$SD1 = 25 \text{ و } SD2 = 30$$

$$SD = (25 + 30) = 27.5$$

$$SEM = SD \times \sqrt{(1-ICC)} = 27.5 \times \sqrt{(1-0.81)} = 11.99$$

۳-۲ ضریب همبستگی توافق Lin (Lin's concordance correlation coefficient)

و حدود توافق بلند و آلتمن

(Bland and Altman's limit of agreement) برای

اندازه‌گیری‌های کمی

ضریب همبستگی توافق Lin شاخصی برای توافق روی یک اندازه‌گیری کمی فراهم می‌کند که توسط دو روش مختلف یا دو ارزیاب متفاوت انجام شده باشد. این شاخص معیاری از صحت و دقت را به صورت ترکیبی فراهم می‌کند به طوری که مشخص می‌کند که داده‌های مشاهده شده با خط توافق کامل (نیم‌ساز ناحیه اول و سوم) چقدر انحراف دارند. مقدار این شاخص با نزدیک شدن داده‌ها به خط توافق کامل و با فشرده شدن شکل آن‌ها در اطراف این خط، که به ترتیب صحت و دقت را نتیجه می‌دهد، افزایش می‌یابد.

علاوه بر این شاخص می‌توان نمودار توافق به همراه حدود توافق بلند و آلتمن را نیز می‌توان به کار برد. این نمودار به منظور ارزیابی جزئی‌تر وضعیت توافق در مشاهدات انفرادی و میانگین اندازه‌گیری‌های انجام شده در مقابل تفاضل آن‌ها رسم می‌شود. به علاوه با استفاده از فرمول معرفی شده [۱۲]، فاصله اطمینان تفاضل اندازه‌گیری‌های زوجی با فرمول (خطای معیار اختلاف $\pm t \times$ میانگین اختلاف) محاسبه می‌شود که به آن‌ها حدود توافق بلند-آلتمن گفته

نسخه‌های مختلفی از این شاخص ارائه شده است که می‌توان به صورت زیر نام برد:

- کاپای کوهن برای بررسی توافق دو ارزیاب در یک متغیر اسمی
- کاپای وزنی برای بررسی توافق دو ارزیاب در یک متغیر رتبه‌ای

محاسبه کاپای کوهن

برای محاسبه کاپای کوهن از فرمول زیر استفاده می‌شود:

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

(میزان توافق شانسی - ۱) ÷ (میزان توافق شانسی - میزان توافق مشاهده شده) = کاپا

که در آن $\Pr(a)$ میزان توافق مشاهده شده در ارزیاب‌ها و $\Pr(e)$ توافق شانسی بین ارزیاب‌ها است که بر اساس داده‌های مشاهده شده به دست می‌آیند. در این شاخص، اثر توافقی‌هایی که می‌توانند بر اساس شانسی رخ دهند، تعدیل می‌شود. بر اساس این فرمول مقدار K بین صفر (عدم توافق) و یک (توافق کامل) قرار می‌گیرد و مقادیر کم‌تر از $0/4$ توافق ضعیف، بین $0/4$ تا $0/7$ توافق متوسط و بالای $0/7$ توافق خوب بین ارزیاب‌ها را نشان می‌دهد [۱۴].

مثال ۱۳

در مطالعه‌ای به منظور بررسی کیفیت فنی مراقبت‌های دوران بارداری، یکی از اهداف بررسی تطابق بین مشاهدات پرونده زنان باردار و داده‌های حاصل از اظهارات آن‌ها بود که در مورد تعداد ویزیت پزشک، به عنوان یکی از شاخص‌های کیفیت فنی، نتایج در بخشی از داده‌ها بر اساس جدول ۶ به دست آمد:

حدود پایین و بالای توافق ۹۵ درصدی بلند و آلتمن برای داده‌های آزمون-بارآزمون به ترتیب برابر $15/188$ - و $22/617$ به دست آمد. که توسط دو خط بالایی و پایینی در نمودار ۲ نشان داده شده است. همان‌طور که اشاره شد، در این نمودار وضعیت توافق در مشاهدات انفرادی بررسی و میانگین هر یک از اندازه‌گیری‌های انجام شده در مقابل تفاضل آن‌ها رسم می‌شود. همان‌طور که در نمودار فوق مشاهده می‌شود همه نقاط (جز یک مورد که روی حد پایین توافق قرار دارد)، در داخل حدود توافق قرار گرفتند.

نکته:

برای محاسبه محاسبه ضریب همبستگی توافق Lin و حدود توافق بلند و آلتمن و رسم نمودار آن می‌توان از کد نرم افزار STATA (نسخه های ۱۱ و بالاتر یا concord نسخه آپدیت شده ۱۰) برای این منظور استفاده کرد (concord qol_test qol_retest, loa) برای مثال (۱۲). (لازم به ذکر است در این کد داده‌های مربوط در دو متغیر qol_test و qol_retest ذخیره شده بودند).

۲-۴) محاسبه کاپا و کاپای وزنی و فاصله اطمینان آن‌ها در بررسی توافق

شاخص کاپا معیاری برای بررسی توافق درونی (Intra-rater) یا بین (Inter-rater) ارزیاب‌ها برای متغیرهای کیفی است که نسبت به میزان توافق ساده خیلی بهتر عمل می‌کند چون این شاخص توافق شانسی را در محاسبات لحاظ می‌نماید و به عبارت دیگر شاخصی ارائه می‌کند که اثر توافق شانسی در آن تعدیل شده است.

جدول ۶ بررسی توافق ویزیت پزشک بین داده‌های پرونده و اظهارات مشتری

ویزیت پزشک بر اساس پرونده	ویزیت پزشک بر اساس اظهارات مشتری		تعداد کل
	نداشته است	داشته است	
داشته است	۴۵	۱۰	۵۵
نداشته است	۸۰	۵۰	۱۳۰
تعداد کل	۱۲۵	۶۰	۱۸۵

منبع: [۱۵]

بزرگ‌ترین وزن و به موارد عدم توافق نزدیک‌تر به توافق از لحاظ رده متغیر مورد بررسی وزن‌های بزرگ‌تری برای لحاظ نمودن در توافق اختصاص داده می‌شود. مثلاً در یک ارزیابی دو ارزیاب بر اساس مقیاس رتبه‌ای با رتبه‌های خیلی موافقم، موافقم، بی‌نظر، مخالفم و خیلی مخالفم، کوچک‌ترین وزن به ترکیب خیلی "موافقم-خیلی موافقم" از دو ارزیاب داده می‌شود، به ترکیب "خیلی موافقم-موافقم" وزن بزرگ‌تر بعدی و نهایتاً به ترکیب "خیلی موافقم-خیلی مخالفم" بزرگ‌ترین وزن داده می‌شود. در نهایت با فرمول زیر مقدار کاپای وزنی به دست می‌آید:

$$\kappa = 1 - \frac{\sum_{i=1}^k \sum_{j=1}^k w_{ij} x_{ij}}{\sum_{i=1}^k \sum_{j=1}^k w_{ij} m_{ij}}$$

در این صورت موارد فاصله از توافق وزن کم‌تری برای شرکت در محاسبه این کاپا خواهند داشت.

انتخاب نوع وزن

همان‌طوری که در فرمول کاپای وزنی مشاهده شد، نیاز است برای میزان عدم توافق وزن‌هایی تعریف شود. در این راستا دو معمولاً انتخاب دو شکل از وزن‌ها متداول هستند:

وزن ۱: $1 - |i-j| / (k-1)$

وزن ۲: $1 - [(i-j) / (k-1)]^2$

که در آن i و j به ترتیب اندیس‌های ارزیابی برای ارزیاب‌های سطری و ستونی و k بیشینه تعداد ممکن ارزیابی‌هاست. در حالت دوم نسبت به حالت اول وزن بیشتری به موارد نزدیک به توافق داده می‌شود.

البته، لازم به ذکر است وزن‌های دیگری نیز که توسط فرد محقق و بر حسب تشخیص خود فرد تعیین می‌شود می‌تواند در فرمول کاپای وزنی استفاده شود.

مثال ۱۴

در ادامه مثال ۱۳، اگر ارزیابی ویزیت پزشک به صورت رتبه‌ای انجام شده باشد نتایج را می‌توان بر اساس جدول ۷ خلاصه نمود:

در این داده‌ها، محاسبات به صورت زیر خواهد بود:

۱۸۵ = تعداد کل افراد بررسی شده

۹۵ = ۴۵ + ۵۰ = تعداد توافق مشاهده شده

۰/۵۱ = $(۴۵ + ۵۰) \div ۱۸۵$ = میزان توافق مشاهده شده

۵۵ = تعداد کل ویزیت داشته بر اساس پرونده

۰/۳۰ = $۵۵ \div ۱۸۵$ = میزان ویزیت داشته بر اساس پرونده

۰/۷۰ = $۱۳۰ \div ۱۸۵$ = میزان ویزیت نداشته بر اساس پرونده

۱۲۵ = تعداد کل ویزیت داشته بر اساس اظهارات

۰/۶۸ = $۱۲۵ \div ۱۸۵$ = میزان ویزیت داشته بر اساس اظهارات

۰/۳۲ = $۶۰ \div ۱۸۵$ = میزان ویزیت نداشته بر اساس اظهارات

= میزان توافق شانسی

میزان ویزیت نداشته بر اساس اظهارات \times میزان ویزیت نداشته بر اساس پرونده

+

میزان ویزیت داشته بر اساس اظهارات \times میزان ویزیت داشته بر اساس پرونده

۰/۴۳ = $(۰/۳۰ \times ۰/۶۸) + (۰/۷۰ \times ۰/۳۲)$ = میزان توافق شانسی

(توافق شانسی - ۱) \div (توافق شانسی - توافق مشاهده شده) = کاپا

۰/۱۵ = $(۰/۵۱ - ۰/۴۳) \div (۱ - ۰/۴۳)$ = کاپا

بنابراین بین مشاهدات حاصل از پرونده و اظهارات زنان باردار توافق ضعیفی وجود داشت.

نکته

۱- ممکن است تعداد موارد کلی توافق ارزیاب‌ها (یعنی مجموع خانه‌های ۱ و ۴ جدول) در دو ارزیابی یکسان باشد ولی اگر اعداد خانه‌های ۱ و ۴ در دو جدول یکسان نباشد، در این صورت مقدار کاپای آن‌ها یکسان نخواهد بود.

۲- شاخص کاپای کوهن با استفاده از کد `kap` در نرم‌افزار STATA قابل محاسبه است (`kap self doc, t`). (که در آن `self` و `doc` دو متغیر به ترتیب شامل داده‌های اظهارات و پرونده می‌باشد).

کاپای وزنی

همان‌طور که اشاره شد این شاخص به منظور بررسی توافق بین ارزیاب‌ها زمانی استفاده می‌شود که متغیرهای مورد بررسی رتبه‌ای باشند. در این شاخص با موارد عدم توافق به شکل متفاوتی برخورد می‌شود به طوری که به توافق‌ها

جدول ۷- بررسی توافق ویزیت پزشک بین داده‌های پرونده و اظهارات مشتری به صورت رتبه‌ای

ویزیت پزشک بر اساس پرونده	ویزیت پزشک بر اساس اظهارات مشتری			تعداد کل
	کامل	تا حدودی	نداشته است	
کامل	۶	۱۰	۴	۲۰
تا حدودی	۹	۲۰	۶	۳۵
نداشته است	۴۰	۴۰	۵۰	۱۳۰
تعداد کل	۵۵	۷۰	۶۰	۱۸۵

نتیجه‌گیری نهایی

به‌منظور بررسی بررسی تاثیر و توافق در اندازه‌گیری‌ها برای انواع مطالعات علوم پزشکی، با توجه به موقعیت و هدف مطالعه، توصیه می‌شود برای بررسی تاثیر و تعیین سهم عامل خطر روی پیامدها در مطالعات مورد-شاهدی و هم‌گروهی، شاخص خطر و کسر قابل انتساب استفاده شود. همچنین پیشنهاد می‌شود شاخص‌های نسبی و مطلق کاهش خطر و تعداد مورد نیاز برای درمان (NNT) در مطالعات مداخله‌ای و مخصوصاً کارآزمایی‌ها به‌عنوان اندازه اثر کفایت تاثیر گذاری مداخله به کار رود. به‌علاوه نیاز است میزان دقت اندازه‌گیری‌ها برای اندازه‌گیری‌های کمی با استفاده از شاخص‌های خطای معیار اندازه‌گیری، ICC، حدود توافق بلند و آلتمن و ضریب همبستگی توافق لین و برای متغیرهای کیفی اسمی و رتبه‌ای کاپا و کاپای وزنی محاسبه شود. به‌علاوه در هر یک از موارد فوق توصیه می‌شود تا در صورت امکان شاخص‌های مورد استفاده به همراه فاصله اطمینان آن‌ها ارائه شدند.

سپاسگزاری

از همکاری صمیمانه دانشکده بهداشت دانشگاه علوم پزشکی تبریز سپاسگزاریم. این مقاله منبع تأمین مالی نداشته است.

در این مورد بر اساس وزن ۱ مقدار کاپای وزنی برابر ۰/۰۹۴ (SE = ۰/۰۴۶) و بر اساس وزن ۲ مقدار کاپای وزنی برابر ۰/۰۷۰ (SE = ۰/۰۵۶) به دست آمد که در هر دو مورد اندازه‌های ضعیف از توافق را ارائه نمودند.

در این مورد بر اساس وزن ۱، به ترکیب‌های همنام (نظیر کامل-کامل) برابر ۱ "کامل-تا حدودی" و "تا حدودی - نداشته است" برابر ۰/۵ و برابر صفر برای "کامل - نداشته است" وزن اختصاص داده شده است. همچنین در این مورد بر اساس وزن ۲، به ترکیب‌های همنام (نظیر کامل-کامل) برابر ۱ "کامل-تا حدودی" و "تا حدودی - نداشته است" برابر ۰/۷۵ و برابر صفر برای "کامل - نداشته است" وزن اختصاص داده شده است.

نکته

۱- شاخص کاپای وزنی با استفاده از کد kap در نرم‌افزار STATA قابل محاسبه است (= kap self doc [fweight = n], wgt(w) برای وزن ۱ و kap self doc [fweight = n], wgt(w2) برای وزن ۲) (که در آن self و doc دو متغیر به ترتیب شامل داده‌های اظهارات و پرونده و n تعداد افراد در هر رده می‌باشد).

۲- کاپای فلایز برای بررسی توافق بیش از دو ارزیاب در یک متغیر اسمی یا رتبه‌ای [۱۶] به کار می‌رود.

ضمیمه‌ها

ضمیمه ۱

محاسبه فاصله اطمینان AFE در مطالعات هم‌گروهی

با توجه به این که AFE بر اساس فرمول RR به دست می‌آید، برای محاسبه فاصله اطمینان نیاز است ابتدا فاصله اطمینان RR به دست آید (فاصله اطمینان RR = ۱/۵۴)، (۱/۰۴) ((۴)، سپس با استفاده از اعمال فرمول فوق برای حد

پایین و حد بالای RR می‌توان فاصله اطمینان AFE را به دست آورد:

$$\text{AFE Lower Bound (L)} = [L(\text{RR}) - 1] / [L(\text{RR})]$$

$$(1/04 - 1) \div 1/04 = 0/04$$

$$\text{AFE Lower Bound (U)} = [U(\text{RR}) - 1] / [U(\text{RR})]$$

$$(1/54 - 1) \div 1/54 = 0/35$$

Upper limit = 1 / ARR upper limit (rounded to 0 decimal places)

ضمیمه ۴

محاسبه فاصله اطمینان برای .NPV .PPV .SP .SEN

LR- و LR+

فرمول مورد نیاز برای محاسبه حد پایین و بالای فاصله اطمینان به صورت زیر می باشد (۱۷، ۱۸):

$$\begin{aligned} nr1 &= a+b \\ nr2 &= c+d \\ nc1 &= a+c \\ nc2 &= b+d \\ N &= a + b + c + d \\ z &= 1.96 \end{aligned}$$

Sensitivity

$$\begin{aligned} \text{Sensitivity} &= a/nc1 \\ \text{Lower limit} &= ((2 \times a) + z^2 - z \sqrt{(4 \times a \times c / nc1) + z^2}) / ((2 \times nc1) + (2 \times z^2)) \\ \text{Upper limit} &= ((2 \times a) + z^2 + z \sqrt{(4 \times a \times c / nc1) + z^2}) / ((2 \times nc1) + (2 \times z^2)) \end{aligned}$$

Specificity

$$\begin{aligned} \text{Specificity} &= d/nc2 \\ \text{Lower limit} &= ((2 \times d) + z^2 - z \sqrt{(4 \times d \times b / nc2) + z^2}) / ((2 \times nc2) + (2 \times z^2)) \\ \text{Upper limit} &= ((2 \times d) + z^2 + z \sqrt{(4 \times d \times b / nc2) + z^2}) / ((2 \times nc2) + (2 \times z^2)) \end{aligned}$$

Positive Predictive Value (PPV)

$$\begin{aligned} \text{PPV} &= a/nr1 \\ \text{Lower limit} &= ((2 \times a) + z^2 - z \sqrt{(4 \times a \times b / nr1) + z^2}) / ((2 \times nr1) + (2 \times z^2)) \\ \text{Upper limit} &= ((2 \times a) + z^2 + z \sqrt{(4 \times a \times b / nr1) + z^2}) / ((2 \times nr1) + (2 \times z^2)) \end{aligned}$$

Negative Predictive Value (NPV)

$$\begin{aligned} \text{NPV} &= d/nr2 \\ \text{Lower limit} &= ((2 \times d) + z^2 - z \sqrt{(4 \times d \times c / nr2) + z^2}) / ((2 \times nr2) + (2 \times z^2)) \\ \text{Upper limit} &= ((2 \times d) + z^2 + z \sqrt{(4 \times d \times c / nr2) + z^2}) / ((2 \times nr2) + (2 \times z^2)) \end{aligned}$$

Likelihood Ratio +

$$\begin{aligned} \text{LR+} &= \text{Sensitivity} / (1 - \text{Specificity}) \\ \text{Lower limit} &= \exp(\ln((nc2 \times a) / (nc1 \times b)) - z \sqrt{(c / (a \times nc1) + (d / (b \times nc2)))}) \\ \text{Upper limit} &= \exp(\ln((nc2 \times a) / (nc1 \times b)) + z \sqrt{(c / (a \times nc1) + (d / (b \times nc2)))}) \end{aligned}$$

Likelihood Ratio -

$$\begin{aligned} \text{LR-} &= (1 - \text{Sensitivity}) / \text{Specificity} \\ \text{Lower limit} &= \exp(\ln((nc2 \times c) / (nc1 \times d)) - z \sqrt{(a / (c \times nc1) + (b / (d \times nc2)))}) \\ \text{Upper limit} &= \exp(\ln((nc2 \times c) / (nc1 \times d)) + z \sqrt{(a / (c \times nc1) + (b / (d \times nc2)))}) \end{aligned}$$

PosttestProbability

$$\begin{aligned} \text{PosttestProb(+)} &= (\text{PretestOdds} \times \text{LR+}) / (1 + (\text{PretestOdds} \times \text{LR+})) \\ \text{PosttestProb(-)} &= (\text{PretestOdds} \times \text{LR-}) / (1 + (\text{PretestOdds} \times \text{LR-})) \\ \text{where...} & \\ \text{PretestOdds} &= \text{PretestProb} / (1 - \text{PretestProb}) \end{aligned}$$

بنابراین فاصله اطمینان AFE برابر (۰/۳۵ ، ۰/۰۴) به دست می آید (L و U به ترتیب حد بالا و پایین را نشان می دهد).

ضمیمه ۲

محاسبه فاصله اطمینان AFE در مطالعات مورد-شاهدی

با توجه به این که AFE بر اساس فرمول OR به دست می آید، برای محاسبه فاصله اطمینان نیاز است ابتدا فاصله اطمینان OR به دست آید (فاصله اطمینان OR = (۱۳/۱۶ ، ۴/۹۲)) (۴)، سپس با استفاده از اعمال فرمول فوق برای حد پایین و حد بالای OR می توان فاصله اطمینان AFE را به دست آورد.

$$\text{AFE Lower Bound (L)} = [L(OR) - 1] / [L(OR)]$$

$$(4/92 - 1) \div 4/92 = 0/80$$

$$\text{AFE Upper Bound (U)} = [U(OR) - 1] / [U(OR)]$$

$$(13/16 - 1) \div 13/16 = 0/92$$

بنابراین فاصله اطمینان AFE برابر (۰/۳۵ ، ۰/۰۴) به دست می آید (L و U به ترتیب حد بالا و پایین را نشان می دهد).

ضمیمه ۳

محاسبه فاصله اطمینان برای .NNT و .RRR .ARR

فرمول مورد نیاز برای محاسبه حد پایین و بالای فاصله اطمینان به صورت زیر می باشد (۱۷):

Relative Risk Reduction (RRR)

$$\begin{aligned} \text{RRR} &= ((c / (c+d)) - (a / (a+b))) / (c / (c+d)) \\ \text{Lower limit} &= 1 - (\exp(\ln((a \times (c+d)) / (c \times (a+b))) + z \sqrt{(1/c) - (1/(c+d)) + (1/a) - (1/(a+b))})) \\ \text{Upper limit} &= 1 - (\exp(\ln((a \times (c+d)) / (c \times (a+b))) - z \sqrt{(1/c) - (1/(c+d)) + (1/a) - (1/(a+b))})) \end{aligned}$$

Absolute Risk Reduction (ARR)

$$\begin{aligned} \text{ARR} &= (c/(c+d)) - (a/(a+b)) \\ \text{Lower limit} &= \text{ARR} - z \sqrt{(u2 \times (1-u2) / (a+b)) + (w1 \times (1-w1) / (c+d))} \\ \text{Upper limit} &= \text{ARR} + z \sqrt{(u1 \times (1-u1) / (c+d)) + (w2 \times (1-w2) / (a+b))} \end{aligned}$$

where...

$$u1 = ((2 \times c) + z^2 + z \sqrt{(4 \times c \times d / nr2) + z^2}) / ((2 \times nr2) + (2 \times z^2))$$

$$u2 = ((2 \times a) + z^2 + z \sqrt{(4 \times a \times b / nr1) + z^2}) / ((2 \times nr1) + (2 \times z^2))$$

$$w1 = ((2 \times c) + z^2 - z \sqrt{(4 \times c \times d / nr2) + z^2}) / ((2 \times nr2) + (2 \times z^2))$$

$$w2 = ((2 \times a) + z^2 - z \sqrt{(4 \times a \times b / nr1) + z^2}) / ((2 \times nr1) + (2 \times z^2))$$

Number Needed to Treat (NNT)

NNT = 1/ARR (rounded down to 0 decimal places)
Lower limit = 1 / ARR lower limit (rounded to 0 decimal places)

مآخذ

1. Asghari-Jafarabadi M, Soltani A, Mohammadi SM. Statistical Series: Analysis of Contingency Tables 1 (Chi Squared Tests). *Journal of Diabetes and Metabolic Disorders* 2013; Under Press [In Persian].
2. Asghari-Jafarabadi M, Mohammadi SM. Statistical Series: An Introduction to Inferential Statistics (Point Estimation, Confidence Interval and Hypothesis Testing). *Journal of Diabetes and Metabolic Disorders* 2013; Under Press [In Persian].
3. Asghari-Jafarabadi M, Mohammadi SM. Statistical Series: Probability and Distributions. *Journal of Diabetes and Lipid Disorders* 2013; 12(2):101-17 [in Persian].
4. Asghari-Jafarabadi M, Soltani A, Mohammadi SM. Statistical Series: Analysis of Contingency Tables 2 (Measures of Association). *Journal of Diabetes and Metabolic Disorders* 2013; Under Review [In Persian].
5. Asghari-Jafarabadi M, Hajizadeh E, Kazemnejad A, Fatemi SR. A comparative study on the prognostic impact of concurrent smoking and alcohol drinking on colon and rectal cancers: A frailty competing risks survival analysis. *Gastroenterology and Hepatology From Bed to Bench* 2010; 3(1):19-26.
6. Asghari Jafarabadi M, Allahverdipour H, Bashirian S, Jannati A. Modeling the Underlying Predicting Factors of Tobacco Smoking among Adolescents. *Iranian J Publ Health* 2012; 41(5):46-57.
7. Ghoreishi Z, Esfahani A, Djazayeri A, Djalali M, Golestan B, Ayromlou H, et al. Omega-3 fatty acids are protective against paclitaxel-induced peripheral neuropathy: a randomized double-blind placebo controlled trial. *BMC cance* 2012; 12:355.
8. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1(8476):307-10.
9. Lin L. A note on the concordance correlation coefficient. *Biometrics* 2000;56:324-5.
10. Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Correlation and Regression. *Journal of Diabetes and Metabolic Disorders* 2013; 12(6):479-506 [In Persian].
11. Mirfeizi M, Toorzai ZM, Asghari-Jafarabadi M, Shoghi M, Mohammad JG, Tekmehdash AM. Examining Diagnostic Value of the Fasting Plasma Glucose in Screening Gestational Diabetes. *Iranian Journal of Diabetes and Lipid Disorders* 2011; 10:1-5.
12. Asghari-Jafarabadi M, Mohammadi SM. Statistical Series: Tests for Comparing the Means. *Journal of Diabetes and Metabolic Disorders* 2013; 12(4):265-91 [In Persian].
13. Mirfeizi M, Asghari Jafarabadi M, Mehdizadeh Toorzani, Mohammadi SM, Dehghan Azad M, Vizheh Mohammadi A, et al. Feasibility, Reliability and Validity of the Iranian Version of the Diabetes Quality of Life Brief Clinical Inventory (IDQOL-BCI). *Diabetes Research and Clinical Practice* 2012; 96(2):237-47.
14. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York Wiley and Sons; 1986.
15. Tabrizi JS, Gholipour K, Asghari jafarabadi M, Farahbakhsh M, Mohammadzede M. Customer quality and maternity care in Tabriz urban health centers and health posts. *J Clin Res Gov* 2012; 1:12-5.
16. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 1971; 76(5):378-82.
17. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in medicine* 1998; 17(8):857-72.
18. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *Journal of clinical epidemiology* 1991; 44(8):763-70.

**STATISTICAL SERIES: THE ANALYSIS OF CONTINGENCY TABLES 3
(MEASURES OF EFFECT AND AGREEMENT)**

Mohammad Asghari Jafarabadi*¹, Seyede Momeneh Mohammadi², Akbar Soltani³

1. *Medical Education Research Center, Department of Statistics and Epidemiology, Faculty of Health, Tabriz University of Medical Sciences, Tabriz, Iran*
2. *Department of Anatomical Sciences, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran*
3. *Endocrinology and Metabolism Research Center, Endocrinology and Metabolism Clinical Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran*

ABSTRACT

In medical studies, measures are required to reveal the effect of exposures and interventions and also the precision of measurements. This paper aimed to introduce the measures of effect and agreement and inferences about them in these studies. Principles and method of calculating measures of effect and agreement and inference about them were presented for all types of medical studies taking into account the relevant considerations. To assess the effect of risk factors on outcomes in case-control and cohort studies, and to determine the relevant effect, the attributable risk and fraction in the exposed group and population were used along with their confidence intervals. Also the relative risk reduction, absolute risk reduction and number needed to treat were applied as the measures of effect of intervention in the interventional studies especially in trials. The sensitivity, specificity and related measures along with their confidence intervals were computed for diagnostic accuracy and screening studies. In addition it is needed to evaluate the precision of measurements using standard error of measurements, ICC, Altman and Bland's limits of agreement and Lin's concordance correlation coefficient for quantitative variables and using kappa and weighted kappa for nominal and ordinal variables. In each of these situations the results of research based examples were presented along with the methods of their calculations. To assess the measures of effect and agreement, the mentioned analyses are recommended considering the situation and purpose of the study.

Keywords: Contingency Table, Effect, Agreement, Attributable Risk, Diagnostic evaluations, NNT, Kappa, Limits of Agreement

* Iran, Tabriz, Golgasht St, Attar e Neshabouri St., Tabriz University of Medical Sciences, , Faculty of Health, Department of Statistics and Epidemiology, Postal Code: 5166614711 Tel: 0098411-3357580-2 (383), Fax: 0098411-3340634, Mobile: 0098912-4390958. Email: asgharimo@tbzmed.ac.ir