

## مروری بر مفاهیم معنی‌داری آماری و بالینی: استفاده و تفسیر حدود اطمینان

مهدی سپیدارکیش<sup>۱</sup>، زهرا محمدی پیروز<sup>۲</sup>

## مقاله مروری

## چکیده

**مقدمه:** کاربرد و تفسیر معنی‌داری آماری برای اثبات اثربخشی یک مداخله و یا وجود رابطه‌ی بین دو متغیر، یک اصل اساسی و ضروری در مطالعات است. بطور سنتی تجزیه و تحلیل داده‌های یک مطالعه با استفاده از آزمون فرضیه و گزارش p-value انجام می‌شود. بسیاری از پژوهشگران در حیطه‌ی علوم پزشکی، یافته‌های مطالعه‌ی خویش را صرفاً به معنی‌داری و یا غیرمعنی‌داری آماری خلاصه می‌نمایند. آزمون فرضیه و صرف گزارش p-value نمی‌تواند بزرگی اثر و دقت آن را بدست دهد. در حال حاضر قریب به اتفاق مجلات معتبر علوم پزشکی، گزارش تنه‌ای معنی‌داری آماری را نپذیرفته و گزارش همزمان شاخص اندازه‌ی اثر، حدود اطمینان و معنی‌داری بالینی را الزامی کرده‌اند. اما استفاده و گزارش معنی‌داری آماری و بالینی در مجلات علوم پزشکی یک‌دست نبوده، دستورالعمل واحدی در مورد کاربرد عملی، تفسیر آماری و بالینی یافته‌ها وجود ندارد و تناقضات آشکاری در مجلات مشاهده می‌شود. هدف نویسندگان این مقاله، ارائه‌ی یک دستورالعمل صحیح و یکپارچه به پژوهشگران و متخصصین بالینی، به جهت گزارش صحیح معنی‌داری آماری و بالینی یافته‌ها بر اساس اهداف و طراحی مطالعات در علوم پزشکی با رویکرد تخمین (گزارش Confidence interval) بود.

واژگان کلیدی: آمار؛ برآورد؛ اندازه‌ی اثر؛ تحلیل داده‌ها؛ حدود اطمینان

**ارجاع:** سپیدارکیش مهدی، محمدی پیروز زهرا. مروری بر مفاهیم معنی‌داری آماری و بالینی: استفاده و تفسیر حدود اطمینان. مجله دانشکده پزشکی اصفهان ۱۴۰۲؛ ۴۱ (۷۳۱): ۷۰۷-۶۹۶

## مقدمه

بیشتر از یک قرن از معرفی آزمون فرضیه و آزمون‌های معنی‌داری آماری و شاخص p-value می‌گذرد. از ابتدای دهه ی ۱۹۹۰، برخی از متخصصین آمار و متدولوژی، محدودیت‌های معنی‌داری آماری و شاخص p-value را متذکر شدند و مقالات انتقادی متعددی علیه آن در مجلات معتبر چاپ کردند (۱، ۲). هر چند که این انتقادات در بعضی از موارد به نسبت شدید بوده و خواهان حذف معنی‌داری آماری و آزمون فرضیه شدند، اما بسیاری از متخصصین آمار و متدولوژی، گزارش شاخص‌های اندازه‌ی اثر و معنی‌داری بالینی را همراه با معنی‌داری آماری توصیه کردند (۳، ۴). در حال حاضر قریب به اتفاق مجلات معتبر در حیطه‌های پژوهشی مختلف به ویژه علوم پزشکی، گزارش تنه‌ای معنی‌داری آماری را نپذیرفته و گزارش همزمان شاخص‌های اندازه‌ی اثر و حدود اطمینان را الزامی کرده‌اند. با این حال استفاده و گزارش شاخص‌های اندازه‌ی اثر، یک‌دست نبوده و تناقضات آشکاری در گزارش آن‌ها در مجلات مشاهده می‌شود (۵، ۶).

همچنین پژوهشگران و متخصصین بالینی، در مورد چگونگی انتخاب شاخص اندازه‌ی اثر بر اساس اهداف و طراحی مطالعه، بدرستی راهنمایی نشده‌اند و گایدلاین معتبری در مورد کاربرد عملی، تفسیر آماری و بالینی این شاخص‌ها در منابع فارسی وجود ندارد. هدف نویسندگان این مقاله و سری مقالات روش‌شناختی پژوهش‌های کاربردی آتی، ارائه‌ی یک دستورالعمل صحیح و یکپارچه به پژوهشگران و متخصصین بالینی، به جهت انتخاب صحیح شاخص‌های اندازه‌ی اثر بر اساس اهداف و طراحی مطالعه، درک و تفسیر آسان و معنی‌داری آماری و بالینی آن‌ها بود.

## معنی‌داری آماری (Statistical significance)

پژوهشگران دو هدف اصلی را در مطالعات علوم پزشکی پیگیری می‌کنند. هدف اول توصیف بیماری‌ها یا وقایع مرتبط با سلامت و هدف دوم که مهم‌تر محسوب می‌شود، بررسی ارتباط بین متغیرها و یا برآورد اثر متغیر مستقل بر روی متغیر وابسته است (۷). بطور مثال

۱- استادیار، گروه آمار زیستی و اپیدمیولوژی، دانشکده‌ی بهداشت، دانشگاه علوم پزشکی بابل، بابل، ایران

۲- دانشجوی کارشناسی ارشد، گروه آمار زیستی و اپیدمیولوژی، دانشکده‌ی بهداشت، دانشگاه علوم پزشکی بابل، بابل، ایران

نویسنده‌ی مسؤول: مهدی سپیدارکیش: استادیار، گروه آمار زیستی و اپیدمیولوژی، دانشکده‌ی بهداشت، دانشگاه علوم پزشکی بابل، بابل، ایران

Email: mahdi.sepidarkish@gmail.com

(Subjective) بوده و بر اساس نوع مواجهه (یا مداخله)، پیامد، طراحی مطالعه، معیارهای ورود و عدم ورود نمونه‌ها، متفاوت استنباط می‌شود و تعریف دقیقی از آن وجود ندارد. بطور مثال تعریف عملی جامعه‌ی بیماران مبتلا به سندرم تخمدان پلی کیستیک که کورکومین مصرف می‌کنند چیست؟ و یا جامعه‌ی بیماران مبتلا به سندرم تخمدان پلی کیستیک که دارونما مصرف می‌کنند، کدام است؟ ممکن است یافته‌های مطالعه مؤید اثربخشی مداخله (یا وجود رابطه‌ی متغیر مواجهه با پیامد) در نمونه‌ها باشد، اما از آنجایی که همیشه تعداد نمونه کمتر از جامعه‌ی هدف است (نمونه جزئی از جامعه است و تمام افراد جامعه مورد بررسی قرار نمی‌گیرند)، الزاماً این اثربخشی (یا وجود رابطه‌ی متغیر مواجهه با پیامد) در جامعه صادق نیست. بنابراین نتایجی که از مطالعه (نمونه) بدست می‌آید با آنچه که در واقعیت (جامعه) وجود دارد، متفاوت است. یکی از دلایل این تفاوت وجود خطای نمونه‌گیری (Sampling error) (خطای تصادفی یا Random error) است. با افزایش حجم نمونه می‌توان این خطا را کاهش داد و در نهایت این خطا زمانی صفر می‌شود که تعداد افراد نمونه و جامعه برابر شود. از آنجایی که این امر امکان‌پذیر نیست، بصورت قراردادی خطای تصادفی را در سطح قابل قبولی (بطور مثال ۵ درصد) نگه می‌داریم. به مفهوم دیگر باید خطای تصادفی کمی‌سازی (Quantify) شده و یافته‌های مطالعه (شاخص‌های اندازه‌ی اثر) همراه با آن گزارش شوند (۱۳، ۱۴). در مثال قبلی، پژوهشگر با در نظر گرفتن خطای تصادفی، اثربخشی مکمل کورکومین را بررسی و گزارش می‌کند. آیا یافته‌های حاصل از مطالعه (در صورت دیدن اثربخشی مکمل کورکومین در نمونه‌ها)، در واقعیت (جامعه) وجود دارد؟ به مفهوم دیگر یافته‌های مطالعه شانس (خطای تصادفی) است؟

متخصصین آمار و متدولوژی دو رویکرد را برای کمی‌سازی خطای تصادفی پیشنهاد کرده‌اند. این دو رویکرد شامل برآورد (Estimation) (گزارش شاخص اندازه‌ی اثر با حدود اطمینان (Confidence interval) CI) و آزمون فرضیه (Hypothesis testing) (گزارش شاخص p-value) هستند. در هر دو رویکرد، می‌توان معنی‌داری آماری (Statistical significant) را بررسی کرد و نتایج شانس را از واقعیت تمایز داد و با در نظر گرفتن خطای تصادفی مداخله بر پیامدها (و یا رابطه‌ی مواجهه با پیامد) را تأیید و یا رد نمود.

### مفهوم معنی‌داری آماری با رویکرد برآورد

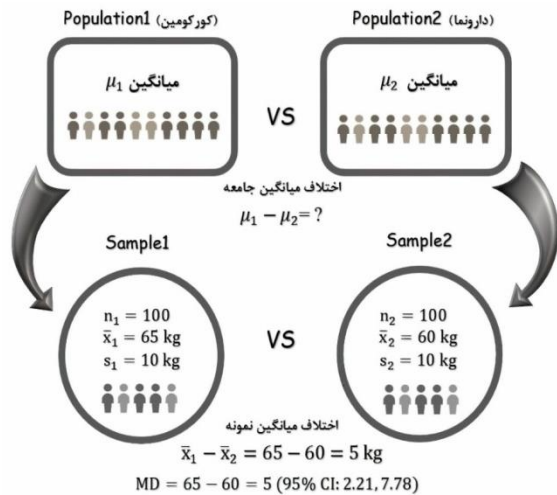
برای روشن شدن مفهوم معنی‌داری آماری با رویکرد برآورد، به مثال ۱ توجه کنید.

مثال ۱: در یک مطالعه‌ی کارآزمایی بالینی تصادفی شده، پژوهشگر به جهت بررسی اثربخشی مکمل کورکومین، بر وزن

در یک مطالعه‌ی مقطعی که در ۱۰۳ بیمارستان استان تهران بر روی ۵۱۷۰ زن باردار انجام شد، هدف پژوهشگران برآورد شیوع چندقلوایی و عوامل مؤثر بر آن بود. در این مطالعه؛ هدف اصلی، توصیفی و جامعه‌ی هدف (Target population) (جامعه‌ای که نتایج نمونه را به آن تعمیم می‌دهند)، تمام زنان باردار استان تهران بود (۸). در مطالعه‌ی کوهورت آینده‌نگر دیگری که در یک بیمارستان تخصصی زنان و زایمان بر روی ۱۹۹ زن باردار انجام شد، هدف پژوهشگران تعیین رابطه‌ی ابتلا به عفونت کووید-۱۹ و رخداد پیامدهای نامطلوب بارداری بود (۹). در این مطالعه، جامعه‌ی هدف می‌تواند وسیع‌تر و حتی تمامی زنان باردار جهان را در بر بگیرد. در یک مطالعه‌ی کارآزمایی بالینی تصادفی شده که بر روی ۶۴ بیمار مبتلا به سکته‌ی مغزی ایسکمیک (Ischemic stroke) انجام شد، هدف پژوهشگران تعیین اثربخشی ژل رویال (Royal jelly) بر روی پیامدهای بالینی، کیفیت زندگی، عملکرد شناختی، سلامت روانی و بیومارکرهای التهابی و استرس اکسیداتیو بیماران مبتلا به سکته‌ی مغزی ایسکمیک بود (۱۰). در این مطالعه جامعه‌ی هدف می‌تواند تمامی بیماران مبتلا به سکته‌ی مغزی ایسکمیک باشد. در یک مطالعه‌ی کارآزمایی بالینی تصادفی شده که بر روی ۷۲ بیمار مبتلا به سندرم تخمدان پلی کیستیک انجام شد، پژوهشگران اثر مکمل کورکومین را بر شاخص‌های قندی و لیپیدی بررسی کردند (۱۱). در این مطالعه، جامعه‌ی هدف می‌تواند تمامی بیماران مبتلا به سندرم تخمدان پلی کیستیک باشد. اما شاید نتوان به جرأت اثربخشی آن را در تمامی بیماران مبتلا به سندرم تخمدان پلی کیستیک بطور مثال زنان مبتلا به سندرم تخمدان پلی کیستیک ناباور تعمیم داد.

در تمامی مثال‌های مذکور؛ هدف پژوهشگران، تعمیم یافته‌های مطالعه (نمونه‌ها) به جامعه‌ی هدف است و منطقی نیست که با صرف هزینه و زمان زیاد، یافته‌ها را صرفاً به نمونه‌ها تعمیم دهیم. بنابراین قصد پژوهشگر، تعمیم یافته‌ها به نمونه‌ها نبوده و همیشه قصد تعمیم‌پذیری به جامعه‌ی بالادستی وجود دارد (۱۲). بطور مثال در مطالعه‌ی کارآزمایی بالینی در بیماران مبتلا به سندرم تخمدان پلی کیستیک، هدف پژوهشگر، تعمیم یافته‌ها به تمامی بیماران مبتلا به سندرم تخمدان پلی کیستیک بوده و منطقی هم نیست که پژوهشگر صرفاً به دنبال اثربخشی مکمل کورکومین در نمونه‌ها باشد. زمانی که یافته‌ها، اثربخشی مداخله بر روی پیامدهای (در مطالعات مشاهده‌ای رابطه‌ی متغیر مواجهه با پیامد) مطالعه را نشان می‌دهند، پژوهشگر به دنبال واقعی بودن این اثربخشی در جامعه‌ی هدف است و در صورتی می‌تواند اثربخشی مداخله (یا وجود رابطه‌ی متغیر مواجهه با پیامد) را تأیید کند که یافته‌های مطالعه نه تنها در نمونه‌ها بلکه در جامعه‌ی هدف نیز بطور یکسان وجود داشته باشد. جامعه‌ی هدف انتزاعی

فاصله‌ای (Interval estimate) گزارش کرد. این فاصله که فاصله‌ای در مورد پارامتر (در اینجا اختلاف میانگین جامعه)، است فاصله‌ای اطمینان (Confidence interval) نامیده می‌شود. اگر خطای نمونه‌گیری را ۵ درصد (البته همیشه ۵ درصد نیست) در نظر بگیریم، اطمینان به این فاصله می‌تواند ۹۵ درصد باشد (۱۵، ۱۶).



شکل ۱. اختلاف میانگین معنادار دو گروه تحت درمان با کورکومین و دارونما با واریانس‌های برابر

این فاصله را می‌توان «به احتمال ۹۵ درصد این فاصله اختلاف میانگین جامعه را در بر می‌گیرد» تفسیر نمود. اما احتمال ۹۵ درصد به چه مفهوم است؟ برای توضیح این مفهوم شکل شماره ۲ را مشاهده بفرمایید. تصور کنید که اختلاف میانگین جامعه ثابت و از توزیع نرمال تبعیت می‌کند. در شکل شماره ۲، توزیع نرمال بالای شکل، توزیع اختلاف میانگین جامعه است. فرض کنید که ما مطالعه‌ی کارآزمایی بالینی اثربخشی کورکومین را ۱۰۰ بار تکرار کنیم. به مفهوم دیگر، ۱۰۰ شاخص اختلاف میانگین نمونه‌ای داریم. اگر توزیعی (توزیع نمونه‌ای) برای این ۱۰۰ مطالعه متصور شویم، می‌توان شکل این توزیع (توزیع نسبتاً نرمال پایین شکل ۲) را مشاهده کرد. اگر از این ۱۰۰ اختلاف میانگین، میانگین می‌گیریم (میانگین اختلاف میانگین‌های نمونه‌ای) و اختلاف میانگین هر مطالعه را از میانگین مذکور کم و متوسط آن را محاسبه می‌کردیم، این شاخص را می‌توان خطای معیار نامید. به مفهوم دیگر خطای معیار متوسط اختلاف هر مطالعه (اختلاف میانگین هر مطالعه) از میانگین مرکزی (میانگین اختلاف میانگین‌های نمونه‌ای) است. همانطور که در شکل شماره ۲ نشان داده شده است، با اضافه و کم کردن دو خطای معیار به اختلاف میانگین مطالعه‌ی واقعی (در عمل یک مطالعه انجام می‌شود و یک اختلاف میانگین داریم)، می‌توان بیان نمود که ۹۵ درصد این

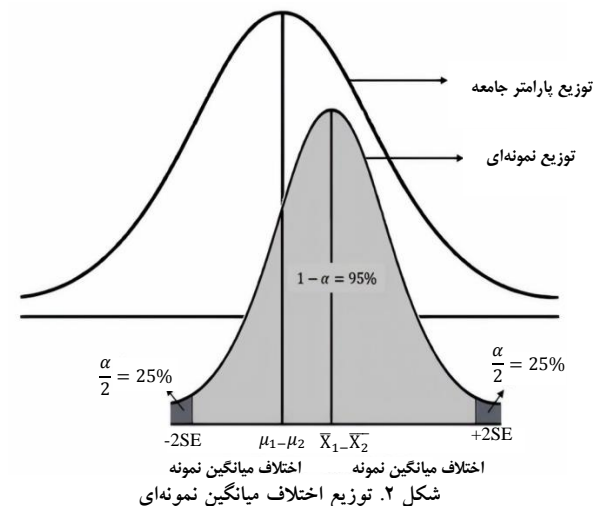
بیماران مبتلا به سرطان معده، آن‌ها را بطور تصادفی به دو گروه مکمل کورکومین (۱۰۰ نفر) و دارونما (۱۰۰ نفر) تقسیم کرد. میانگین و انحراف معیار وزن بعد از مداخله (۲۴ هفته)، در دو گروه مکمل کورکومین و دارونما به ترتیب  $10 \pm 65$  و  $10 \pm 60$  کیلوگرم بدست آمد. اختلاف میانگین وزن بین دو گروه، ۵ کیلوگرم محاسبه شد. سؤال اینجاست که آیا با مشاهده‌ی اختلاف میانگین ۵ کیلوگرم بین دو گروه، می‌توان اثربخشی مکمل کورکومین را تأیید نمود؟ باید به این نکته توجه کرد که این اختلاف میانگین از مقایسه‌ی دو نمونه‌ی ۱۰۰ نفری بدست آمده است و این اختلاف ممکن است به دلیل خطای تصادفی یا خطای نمونه‌گیری باشد. تصور کنید اگر پژوهشگر مکمل کورکومین را به یک بیمار و دارونما را به یک بیمار دیگر می‌داد و وزن این دو نفر به نفع فرد مصرف‌کننده‌ی مکمل کورکومین، ۵ کیلوگرم افزایش می‌یافت، آیا همچنان می‌توانیم مکمل کورکومین را در افزایش وزن اثربخش بدانیم؟ بطور شهودی می‌توان دریافت که نتیجه‌ی حاصل ممکن است تصادفی باشد. در دو نمونه‌ی ۱۰۰ نفری نیز، همین حالت وجود دارد. اگر میانگین وزن افراد گروه کورکومین ۵ کیلوگرم افزایش یافته است، نمی‌توان براحتی مکمل را اثربخش دانست و آن را برای همه‌ی بیماران مبتلا به سرطان معده توصیه کرد. همانطور که در شکل شماره ۱ نشان داده شده است، دو نمونه‌ی مورد مداخله، از دو جامعه‌ی بالادست انتخاب شده‌اند. جامعه‌ی شماره‌ی یک شامل بیماران مبتلا به سرطان معده است که مکمل کورکومین را مصرف می‌کنند. و جامعه‌ی شماره‌ی دو شامل بیماران مبتلا به سرطان معده است که مکمل دارونما مصرف می‌کنند. میانگین وزن جامعه‌ی شماره‌ی یک را با  $\mu_1$  و میانگین وزن جامعه‌ی شماره‌ی دو را با  $\mu_2$  و اختلاف میانگین این دو جامعه را با  $\mu_1 - \mu_2$  نشان می‌دهیم. در واقع صرفاً با دانستن اختلاف میانگین دو جامعه  $(\mu_1 - \mu_2)$ ، می‌توان اثربخشی مکمل کورکومین را تأیید و یا رد کرد. فرض کنید تعداد کل بیماران مبتلا به سرطان معده در سرتاسر جهان یک میلیون نفر باشند و آن‌ها را به دو گروه ۵۰۰ هزار نفری تقسیم کنیم. یک گروه مکمل کورکومین و گروه دیگر دارونما مصرف کنند. میانگین را در هر گروه محاسبه و از هم کم کنیم. این شاخص اختلاف میانگین جامعه بوده و تنها با دانستن آن می‌توان اثربخشی دارو را تأیید و یا رد کرد. اختلاف میانگین نمونه، برآورد نقطه‌ای (Point estimate) بوده و معادل ۵ کیلوگرم بدست آمد، اما برای اختلاف میانگین جامعه نمی‌توان برآورد نقطه‌ای داد. بطور کلی خصوصیات و شاخص‌های جامعه (در اینجا اختلاف میانگین جامعه) را پارامتر (Parameter) می‌نامند و مجهول (Unknown) هستند. با دانستن برآورد نقطه‌ای (اختلاف میانگین نمونه) و در نظر گرفتن خطای تصادفی، برای اختلاف میانگین جامعه، صرفاً می‌توان برآورد

مطالعات، اختلاف میانگین جامعه را در بر گرفته اند. همانطور که در شکل شماره ۲ نشان داده شده است، ۵ درصد مطالعات (۲/۵ درصد مطالعات در دو طرف توزیع نمونه‌ای) نتوانسته‌اند اختلاف میانگین جامعه را در بر بگیرند. این ۵ درصد به مفهوم خطای تصادفی قراردادی در ابتدای متن بود. در عمل، پژوهشگر فقط یک مطالعه را انجام می‌دهد و در واقع ۱۰۰ مطالعه از روی همان یک مطالعه‌ی واقعی شبیه‌سازی می‌شوند. خطای معیار بر اساس متوسط اختلاف، هر اختلاف میانگین مطالعه از اختلاف میانگین مرکزی (میانگین اختلاف میانگین‌ها) بدست می‌آید. بنابراین خطای معیار از روی تکرار تنها مطالعه‌ی انجام شده، ساخته می‌شود. در صورتی که حجم نمونه‌ی تنها مطالعه‌ی انجام شده زیاد باشد، خطای معیار کوچک می‌شود و حدود اطمینان باریک (Narrow) می‌شود (۱۷، ۱۸). تصور کنید که مطالعه‌ی کارآزمایی مثال ۱، بر روی یک میلیون بیمار مبتلا به سرطان معده انجام شود و در هر گروه ۵۰۰ هزار بیمار حضور داشته باشد و به ترتیب به مکمل کورکومین و دارونما تخصیص داده شوند. از آن جایی که به جهت محاسبه‌ی خطای معیار و ساختن حدود اطمینان برای اختلاف میانگین جامعه، تعداد ۱۰۰ مطالعه از روی مطالعه‌ی واقعی (کارآزمایی ۱ میلیونی)، ساخته می‌شود، خطای معیار در کمترین مقدار ممکن خواهد بود و با صفر شدن آن، اختلاف میانگین همان یک مطالعه‌ی واقعی (کارآزمایی ۱ میلیونی)، با اختلاف میانگین جامعه یکی می‌شود. در این حالت می‌توان فرض نمود که به جای مطالعه بر روی نمونه، مطالعه بر روی جامعه انجام می‌شود.

### مفهوم عدم وجود معنی‌داری آماری با رویکرد برآورد (پیامد کمی)

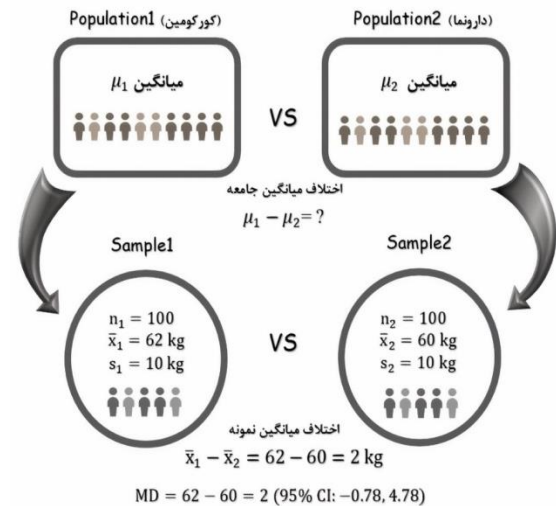
برای روشن شدن مفهوم عدم وجود معنی‌داری آماری، به مثال ۲ توجه کنید.

مثال ۲: در یک مطالعه‌ی کارآزمایی بالینی تصادفی شده، پژوهشگر به جهت بررسی اثربخشی مکمل کورکومین، بر وزن بیماران مبتلا به سرطان معده، آن‌ها را بطور تصادفی به دو گروه مکمل کورکومین (۱۰۰ نفر) و دارونما (۱۰۰ نفر) تقسیم کرد. میانگین و انحراف معیار وزن بعد از مداخله (۲۴ هفته)، در دو گروه مکمل کورکومین و دارونما به ترتیب  $10 \pm 62$  و  $10 \pm 60$  کیلوگرم بدست آمد. همانطور که در شکل شماره ۳ نشان داده شده است، اختلاف میانگین ۲ و حدود اطمینان  $0.7 -$  تا  $4/8$  محاسبه شد. هر چند که بیماران مبتلا به سرطان معده، میانگین وزن بیشتری (۲ کیلوگرم) دارند اما با بررسی حدود اطمینان که عدد صفر را در بر گرفته می‌توان تأیید کرد که احتمالاً اختلاف میانگین جامعه صفر بوده و در واقعیت دارو اثربخش نیست. به مفهوم دیگر حداقل یکی از ۹۵ مطالعه صفر بوده و در اینصورت ممکن است اختلاف میانگین جامعه صفر باشد و در این حالت، مکمل کورکومین اثربخش نیست (۲۱).



از طرفی می‌دانیم که برای دانستن اثربخشی واقعی مکمل کورکومین باید اختلاف میانگین جامعه را دانست و از طرف دیگر نمی‌توان به تنهایی به اختلاف میانگین یک مطالعه، برای اثربخشی

متوسط افزایش ۷۰۰ گرم در بیماران مبتلا به سرطان معده اهمیت دارد؟ البته خصوصیات جمعیت‌شناختی (Demographic characteristics) و بالینی پایه (Baseline clinical characteristics) (مانند مرحله‌ی (Stage) سرطان) نیز باید در نظر گرفته شود. بنظر می‌رسد که افزایش متوسط ۷۰۰ گرم (حدود اطمینان: ۱۴۰ گرم تا ۱۲۵۰ گرم) در بالغین اهمیت نداشته باشد. البته قطعاً متخصص بالینی، علاوه بر خصوصیات جمعیت‌شناختی و بالینی پایه بیماران، باید تمامی مداخلات (سایر داروها) موجود برای بیمار در نظر بگیرد و با در نظر گرفتن این شرایط، اهمیت بالینی مکمل کورکومین را تأیید و یا رد نماید. بنابراین خصوصیات بیماران و سایر شرایط مطالعه، اهمیت بسیار زیادی در تصمیم‌گیری اثربخشی مداخله دارد. تصور کنید که اگر اثربخشی مکمل کورکومین بر وزن نوزادان نارس و متولد شده با وزن کمتر از ۲۵۰۰ گرم، بررسی می‌شد، آیا افزایش متوسط ۷۰۰ گرم (حدود اطمینان: ۱۴۰ گرم تا ۱۲۵۰ گرم) اهمیت بالینی نداشت؟ قطعاً در این حالت به دلیل تفاوت حیطه‌ی پژوهشی و خصوصیات کاملاً متفاوت بیماران در دو مطالعه، اثربخشی مکمل کورکومین در نوزادان نارس با وزن کم هنگام تولد، اهمیت بالینی بالایی دارد. فرض کنید که اثربخشی مکمل کورکومین در بیماران مبتلا به سرطان معده پیشرفته (متاستاز) که در بخش مراقبت‌های ویژه بستری شده‌اند، آیا افزایش متوسط ۷۰۰ گرم (حدود اطمینان: ۱۴۰ گرم تا ۱۲۵۰ گرم) اهمیت بالینی نداشت؟ شاید هر چند که اثربخشی مکمل کورکومین در بالغین بوده اما به دلایل متعدد مانند بستری بودن، مراحل بالای بیماری و سختی تغذیه در بیماران، مکمل کورکومین از نظر متخصص بالینی با اهمیت باشد. این نکته بسیار حائز اهمیت است که در توالی بررسی معنی‌داری آماری و بالینی، اولویت با معنی‌داری آماری است. به مفهوم دیگر اثربخشی مداخله اول باید اتفاقی نبوده و در واقعیت هم باشد و در مرتبه‌ی دوم اثربخشی بالینی آن تعیین شود. همچنین زمانی باید معنی‌داری آماری را با حدود اطمینان تعیین نمود که اولاً سایر خطاها (خطای منظم انتخاب و اندازه‌گیری و مخدوش‌شدگی)، وجود نداشته باشد و ضمناً حجم نمونه کافی باشد. مفهوم کافی بودن حجم نمونه به این معنی است که در زمان طراحی مطالعه و قبل از جمع‌آوری داده‌ها، ابتدا باید برای اختلاف بالینی مهم (بر اساس مرور متون و نظر متخصص)، حجم نمونه تعیین کرد و تنها در اینصورت می‌توان معنی‌داری آماری را بعد از جمع‌آوری داده‌ها و آنالیز داده‌ها از روی حدود اطمینان مشخص نمود. بنابراین ممکن است، حدود اطمینان در مطالعات با حجم نمونه‌ی کم، معنی‌دار آماری نبوده (عدد صفر را در بر می‌گیرد)، در حالی که در واقعیت دارو اثر داشته و مطالعه، تسوان کسافی بسرای دیدن اثربخشی دارو ندارد (۲۰، ۲۲، ۲۳).



شکل ۳. اختلاف میانگین غیرمعتادار دو گروه تهن درمان با کورکومین و دارونما با واریانس‌های برابر

### معنی‌داری بالینی

بعد از آنکه معنی‌داری آماری برای یک شاخص اندازه‌ی اثر (در مثال شماره‌ی ۱، اختلاف میانگین) بررسی و مورد تأیید قرار گرفت، سؤال مهم دیگر، اثربخشی مداخله از نظر بالینی است. به مفهوم دیگر مداخله‌ی مورد نظر از نظر بالینی چقدر اهمیت دارد؟ در مثال شماره‌ی ۱، که اختلاف میانگین وزن ۵ کیلوگرم بدست آمد، با در نظر گرفتن حدود اطمینان ۲/۲ تا ۷/۸، آیا متوسط افزایش ۵ کیلوگرم، از نظر متخصص بالینی اهمیت دارد؟ به این مفهوم معنی‌داری بالینی (Clinical significance) گفته می‌شود و تفسیر آن بر عهده‌ی متخصص بالینی حیطه‌ی پژوهشی مورد نظر است (۲۰، ۲۲).

### وجود معنی‌داری آماری، اما بی‌اهمیت از نظر بالینی

برای روشن شدن مفهوم عدم وجود معنی‌داری آماری، به مثال ۳ توجه کنید.

مثال ۳: در یک مطالعه‌ی کارآزمایی بالینی تصادفی شده، پژوهشگر به جهت بررسی اثربخشی مکمل کورکومین، بر وزن بیماران مبتلا به سرطان معده، آن‌ها را بطور تصادفی به دو گروه مکمل کورکومین (۱۰۰ نفر) و دارونما (۱۰۰ نفر) تقسیم کرد. میانگین و انحراف معیار وزن بعد از مداخله (۲۴ هفته)، در دو گروه مکمل کورکومین و دارونما به ترتیب  $60.7 \pm 2$  و  $62 \pm 2$  کیلوگرم بدست آمد. همانطور که در شکل شماره‌ی ۴ نشان داده شده است، اختلاف میانگین ۰/۷ کیلوگرم (معادل ۷۰۰ گرم) و حدود اطمینان ۰/۱۴ تا ۱/۲۵ محاسبه شد. از آن جایی که در فاصله‌ی حد پایین تا حد بالا، عدد صفر قرار نگرفته است، نتایج از نظر آماری معنی‌دار است. اما آیا از نظر بالینی نیز یافته‌ها، اهمیت دارد؟ آیا از نظر متخصص بالینی

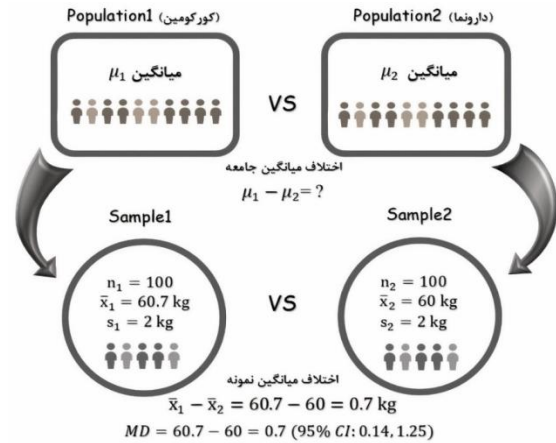
### مفهوم عدم وجود معنی‌داری آماری با رویکرد برآورد (پیامد کیفی)

بسیاری از پیامدهای مطالعات علوم پزشکی شامل متغیرهای کیفی (Qualitative outcomes) می‌شوند. این پیامدهای کیفی می‌توانند دو یا چند حالت باشند. متغیرهای متعددی مانند مرگ یا زندگی، بیمار یا سالم بودن، بارداری یا غیرباردار بودن، تولد زنده یا مرده دو حالت و متغیرهایی مانند شدت بیماری (خفیف، متوسط، شدید)، نمایه‌ی توده‌ی بدنی (کم وزن، نرمال، اضافه وزن، چاقی)، چند حالت (مقیاس رتبه‌ای) در نظر گرفته می‌شوند. همچنین بسیاری از متغیرهای کمی نیز برای درک بالینی بهتر، دسته‌بندی شده و به متغیرهای کیفی تبدیل می‌شوند.

برای بررسی رابطه‌ی بین متغیر مواجهه و پیامد (و یا اثربخشی مداخله)، شاخص‌های اندازه‌ی اثر را می‌توان بر اساس اختلاف مطلق (Absolute effect size) بروز بیماری در دو گروه تحت مطالعه (مواجهه داشته و نداشته/مورد و شاهد) و یا اختلاف‌های نسبی (Relative effect size) (نسبت‌ها) محاسبه کرد. شاخص‌های اندازه‌ی اثر بر اساس اختلاف مطلق وقتی که برنامه‌های پیشگیری و سلامت عمومی مدنظر باشند کارآیی بیشتری دارند. در مقابل مطالعاتی که با هدف بررسی رابطه‌ی بین متغیر مستقل و وابسته (اثربخشی مداخله در مطالعات مداخلاتی) و یا بطور کلی با هدف علت و معلولی طراحی می‌شوند، از شاخص‌های اندازه‌ی اثر نسبی بهره می‌گیرند. نسبت خطر (RR (Risk ratio، نسبت شانس OR (Odds ratio) و نسبت مخاطره HR (Hazard ratio)، سه شاخص از خطر هستند که در مطالعات اپیدمیولوژیک بطور وسیعی استفاده می‌شوند. نسبت خطر یک پیامد در ارتباط با یک مواجهه از نسبت دو احتمال بدست می‌آید و این شاخص صرفاً در مطالعاتی که نمونه‌ها را پیگیری می‌کنند (مانند مطالعات کوهورت یا کارآزمایی بالینی تصادفی شده) قابل محاسبه است. به مفهوم دیگر این شاخص از نسبت خطر یک پیامد در میان افراد مواجهه داشته به خطر همان پیامد در میان افراد بدون مواجهه بدست می‌آید (۴).

نسبت شانس یک پیامد در ارتباط با یک مواجهه از نسبت دو شانس بدست می‌آید و این شاخص در تمام مطالعات علوم پزشکی قابل محاسبه است. این شاخص در مطالعات کوهورت و کارآزمایی بالینی تصادفی شده از نسبت شانس یک پیامد در میان افراد مواجهه داشته (افراد در گروه مداخله) به شانس همان پیامد در میان افراد بدون مواجهه (افراد در گروه عدم مداخله) بدست می‌آید. همچنین این شاخص در مطالعات مورد-شاهدی از نسبت شانس یک مواجهه در میان افراد مورد (بیمار) به شانس همان مواجهه در میان افراد شاهد (سالم) بدست می‌آید (۲۶).

نسبت مخاطره‌ی یک پیامد در ارتباط با یک مواجهه از نسبت دو

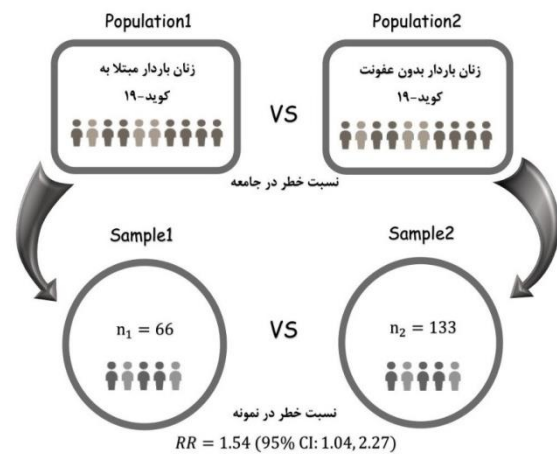


شکل ۴. اختلاف میانگین معنادار اما بدون اهمیت بالینی دو گروه تحت درمان با کورکومین و دارونما با واریانس‌های برابر

### در نظر گرفتن معناداری آماری و بالینی تمامی پیامدهای مطالعه

نکته‌ی دیگر در بررسی اثربخشی مداخله و تجویز آن برای بیمار، در نظر گرفتن معنی‌داری آماری و بالینی تمام پیامدها و نه صرفاً پیامد اولیه است. معمولاً در بررسی اثربخشی مداخلات (به ویژه داروها)، به غیر از پیامد اولیه، پیامدهای دیگر (پیامدهای ثانویه) نیز وجود دارد و بالطبع متخصص بالینی برای تجویز مداخله باید همه‌ی پیامدها و همچنین عوارض را بداند. در مثال ۱، که اثربخشی مکمل کورکومین در افزایش وزن بیماران مبتلا به سرطان معده مورد بررسی قرار گرفت، پیامدهای ثانویه‌ی دیگری مانند کیفیت زندگی، فاکتورهای التهابی و شاخص‌های استرس اکسیداتیو نیز باید در نظر گرفته شوند. به مفهوم دیگر اثربخشی مکمل کورکومین در این مطالعه، چندبعدی است و متخصص با دانستن اثربخشی آماری و بالینی تک تک این پیامدها، مکمل را برای بیمار خود تجویز می‌کند. متخصص بالینی برای درمان بیمار خود، معمولاً چندین جایگزین (سایر داروها) را در نظر دارد و با بررسی معنی‌داری آماری و بالینی تمام پیامدهای ممکن برای تمام داروهای جایگزین، تصمیم به تجویز می‌گیرد. از همین رو بسیاری از مجلات با تکه تکه (Slicing) کردن پیامدهای یک مطالعه و چاپ مقالات متعدد از یک مطالعه‌ی کارآزمایی بالینی (چاپ مقالات جداگانه بر اساس پیامدهای اولیه و ثانویه) مخالفت می‌کنند، تا متخصص بالینی بتواند یکجا در مورد اثربخشی و تجویز مداخله تصمیم‌گیری کند. بر اساس این نکته، بسیاری از متخصصین آمار و متدولوژی بر این باورند که باید محاسبه‌ی حجم نمونه در ابتدای مطالعه برای تمام پیامدهای مطالعه (چه اولیه و چه ثانویه) انجام شود و بالاترین حجم نمونه، به جهت وجود شانس معنی‌داری آماری برای تمام پیامدها (نه تنها پیامد اولیه)، در نظر گرفته شود (۲۴، ۲۵).

دو جامعه (واقعیت) را برآورد کرد. به مفهوم دیگر اگر احتمال (خطر) رخداد سزارین تمام زنان باردار مبتلا به کووید-۱۹ جهان را داشتیم و آن را بر احتمال رخداد سزارین در زنان باردار غیرمبتلا به کووید-۱۹ جهان تقسیم می‌کردیم و مقدار آن بزرگتر از یک می‌شد، می‌توانستیم رابطه‌ی عفونت و رخداد سزارین را تأیید کنیم. همانطور که قبلاً گفته شد نسبت خطر بدست آمده از نمونه (۱/۵۴) برآورد نقطه‌ای بوده و تقریبی از نسبت خطر بین دو جامعه است. با دانستن برآورد نقطه‌ای (نسبت خطر بین نمونه‌ها) و در نظر گرفتن خطای تصادفی، می‌توان برآورد فاصله‌ای برای نسبت خطر بین دو جامعه داد. اگر خطای نمونه‌گیری را ۵ درصد در نظر بگیریم می‌توان این فاصله را بصورت «به احتمال ۹۵ درصد این فاصله نسبت خطر بین دو جامعه را در برمی‌گیرد»، تفسیر نمود. حدود اطمینان ۹۵ درصد معادل ۱/۰۴ تا ۲/۲۷ بدست آمد. بنابراین نسبت خطر دو جامعه در این فاصله قرار دارد.



شکل ۵. نسبت خطر معنادار احتمال رخداد سزارین در زنان مبتلا به کووید-۱۹ در مقایسه با زنان باردار بدون عفونت کووید-۱۹

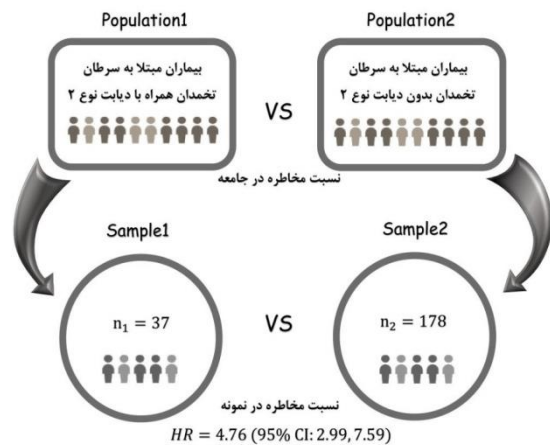
مثال ۵: در یک مطالعه‌ی مورد-شاهدی که بر روی ۹۴ بیمار مبتلا به صرع ایدیوپاتیکی (Idiopathic epilepsy) و ۸۸ شاهد سالم انجام شد، پژوهشگران سابقه‌ی عفونت توکسوکارا را بین دو گروه مذکور مقایسه کردند. شاخص اندازه‌ی اثر گزارش شده در این مطالعه نسبت شانس بود. نسبت شانس معادل ۱/۸۶ بدست آمد. به مفهوم دیگر افراد مبتلا به صرع بطور متوسط ۸۶ درصد شانس بالاتری برای مواجهه با عفونت توکسوکارا در مقایسه با افراد سالم داشتند (۲۸). آیا بر اساس نسبت شانس بدست آمده از نمونه‌ها می‌توان بین عفونت توکسوکارا و صرع ایدیوپاتیکی رابطه‌ی واقعی را تأیید نمود. نسبت شانس بدست آمده از نمونه‌ها (نسبت شانس = ۱/۸۶) برآورد نقطه‌ای بوده و مقدار آن تقریبی از نسبت شانس دو جامعه است. همانطور که در شکل شماره‌ی ۶ نشان داده شده است، این نسبت شانس از

مخاطره (سرعت لحظه‌ای رخداد پیامد) بدست می‌آید. این شاخص صرفاً در مطالعاتی که نمونه‌ها پیگیری می‌شوند قابل محاسبه است و همچنین مدت زمان حضور نمونه‌ها را برای محاسبه‌ی مخاطره باید ثبت نمود. بطور معمول زمانی که حضور افراد در مطالعه متفاوت باشد این شاخص در مقایسه با نسبت خطر ارجحیت دارد و قدرت رابطه‌ی متغیر مواجهه و پیامد را ناریب (Unbiased) (با خطای منظم کمتر) نشان می‌دهد (۲۷).

برای روشن شدن مفهوم معنی‌داری آماری با استفاده از شاخص‌های نسبی به مثال‌های ۴ تا ۶ توجه کنید.

مثال ۴: در یک مطالعه‌ی کوهورت آینده‌نگر که بر روی ۱۹۹ زن باردار انجام شد، هدف بررسی رابطه‌ی عفونت کووید-۱۹ (زنان باردار مبتلا به کووید-۱۹ (۶۶ نفر) و زنان باردار غیرمبتلا به کووید-۱۹ (۱۳۳ نفر)) و پیامدهای نامطلوب بارداری (سزارین، دیابت بارداری، پراکلامپسی، پارگی زودرس کیسه‌ی آب، تأخیر رشد داخل رحمی، مرده‌زایی، پلی‌هیدرامینوس، الیگو‌هیدرامینوس، زایمان زودرس، وزن کم هنگام تولد، بستری نوزاد در بخش مراقبت‌های ویژه نوزادان) بود. تمامی پیامدهای سنجیده شده کیفی و دو حالتی بودند. شاخص اندازه‌ی اثر گزارش شده در این مطالعه به دلیل پیگیری ۹ ماهه‌ی مادران باردار، نسبت خطر بود. نسبت خطر گزارش شده برای رخداد سزارین معادل ۱/۵۴ بدست آمد. به مفهوم دیگر احتمال رخداد سزارین در زنان باردار مبتلا به کووید-۱۹ در مقایسه با زنان باردار بدون عفونت کووید-۱۹، ۵۴ درصد بیشتر بود (۹). بر اساس شاخص گزارش شده (نسبت خطر = ۱/۵۴)، آیا می‌توان رابطه‌ی مبتلا به عفونت کووید-۱۹ و رخداد سزارین را تأیید نمود؟ باید به این نکته دقت کرد که این نسبت خطر و اختلاف ۵۴ درصد از مقایسه‌ی احتمال رخداد سزارین در دو نمونه‌ی زنان مبتلا به عفونت (۶۶ نفر) و زنان غیرمبتلا به عفونت (۱۳۳ نفر) بدست آمده است و این اختلاف ممکن است به دلیل خطای نمونه‌گیری باشد. اگر ما این اختلاف ۵۴ درصد را در نمونه‌های این مطالعه مشاهده کرده‌ایم، آیا براحتی می‌توان این ۵۴ درصد را به تمامی زنان مبتلا به کووید-۱۹ تعمیم داد؟ از طرفی هدف این مطالعه تعمیم یافته‌ها به تمامی زنان باردار مبتلا به کووید-۱۹ است و بر اساس نسبت خطر بدست آمده از مطالعه (نسبت خطر بین نمونه‌ها) می‌خواهیم نسبت خطر واقعی (نسبت خطر بین دو جامعه) را برآورد کنیم. همانطور که در شکل شماره‌ی ۵ نشان داده شده است دو نمونه‌ی زنان باردار مبتلا و غیرمبتلا به کووید-۱۹ از دو جامعه‌ی بالادست انتخاب شده‌اند. جامعه‌ی شماره یک شامل زنان باردار مبتلا به کووید-۱۹ و جامعه‌ی شماره‌ی دو شامل زنان باردار غیرمبتلا به کووید-۱۹ است. برای دانستن رابطه‌ی واقعی بین کووید-۱۹ و سزارین باید نسبت خطر بین

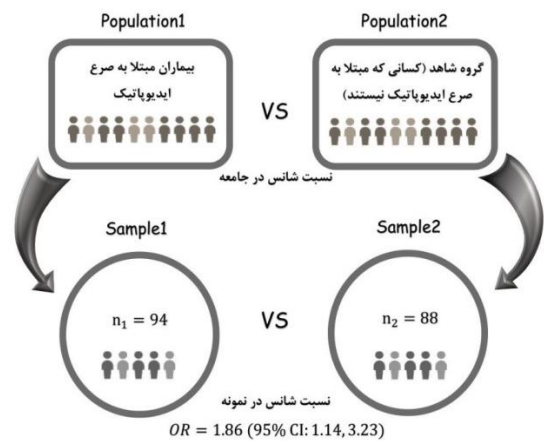
نظر گرفتن خطای نمونه‌گیری می‌توان نسبت مخاطره‌ی واقعی (نسبت مخاطره بین دو جامعه) را حدس زد. اگر خطای نمونه‌گیری را ۵ درصد در نظر بگیریم می‌توان این فاصله را «به احتمال ۹۵ درصد این فاصله نسبت مخاطره بین دو جامعه را در برمی‌گیرد»، تفسیر کرد. حدود اطمینان ۹۵ درصد معادل ۲/۲۹ تا ۷/۵۹ بدست آمد. بنابراین نسبت مخاطره‌ی دو جامعه در این فاصله قرار دارد.



شکل ۷. نسبت مخاطره معنادار مرگ در بیماران مبتلا به سرطان تخمدان همراه با دیابت نوع ۲ در مقایسه با بیماران مبتلا به سرطان تخمدان بدون دیابت نوع ۲

در تمامی مثال‌های مذکور زمانی می‌توان رابطه‌ی بین متغیر مستقل و پیامد را تأیید نمود که شاخص اندازه‌ی اثر نسبی مذکور در واقعیت (بین دو جامعه)، را بدانیم و مقدار آن عدد یک نباشد. همانطور که قبلاً گفته شد، خصوصیات و شاخص‌های جامعه (در اینجا شاخص اندازه‌ی اثر نسبی بین دو جامعه) را پارامتر می‌نامند و مجهول هستند. با دانستن برآورد نقطه‌ای (شاخص اندازه‌ی اثر نسبی بین دو نمونه) و در نظر گرفتن خطای تصادفی، صرفاً می‌توان برای پارامتر (در اینجا شاخص اندازه‌ی اثر نسبی بین دو جامعه) برآورد فاصله‌ای گزارش کرد. این فاصله که فاصله‌ی در مورد پارامتر است، فاصله‌ی اطمینان نامیده می‌شود. اگر خطای نمونه‌گیری را ۵ درصد (البته همیشه ۵ درصد نیست) در نظر بگیریم، اطمینان به این فاصله می‌تواند ۹۵ درصد باشد. به مفهوم دیگر، شاخص نسبی بدست آمده از دو نمونه، به تنهایی نمی‌تواند تعیین‌کننده‌ی رابطه‌ی بین متغیر مستقل و وابسته باشد و با در نظر گرفتن فاصله‌ی اطمینان می‌توان رابطه را تأیید و یا رد کرد. اما احتمال ۹۵ درصد به چه مفهوم است؟ همانطور که در محاسبه‌ی حدود اطمینان اختلاف میانگین توضیح داده شد، در صورتی که مطالعه را صد بار تکرار کنیم می‌توان مدعی شد که ۹۵ مطالعه (۹۵ درصد مطالعات)، شاخص اندازه‌ی اثر نسبی بین

مقایسه‌ی شانس بین دو نمونه بدست آمده و ممکن است تصادفی بوده و در واقعیت رابطه‌ی بین عفونت توکسوکارا و صرع ایدیوپاتیک وجود نداشته باشد. با در نظر گرفتن خطای نمونه‌گیری، می‌توان نسبت شانس واقعی (نسبت شانس بین دو جامعه) را حدس زد. با دانستن برآورد نقطه‌ای (نسبت شانس بین دو نمونه) و در نظر گرفتن خطای تصادفی، صرفاً می‌توان برآورد فاصله‌ای برای نسبت شانس بین دو جامعه گزارش کرد. اگر خطای نمونه‌گیری را ۵ درصد در نظر بگیریم، می‌توان این فاصله را «به احتمال ۹۵ درصد این فاصله نسبت شانس بین دو جامعه را در برمی‌گیرد»، تفسیر نمود. حدود اطمینان ۹۵ درصد معادل ۱/۱۴ تا ۳/۲۳ بدست آمد. بنابراین نسبت شانس دو جامعه در این فاصله قرار دارد.



شکل ۶. نسبت شانس معنادار برای مواجهه با عفونت توکسوکارا در بیماران مبتلا به صرع ایدیوپاتیک در مقایسه با گروه شاهد

مثال ۶: در یک مطالعه‌ی کوهورت گذشته‌نگر که بر روی ۲۱۵ بیمار مبتلا به سرطان اپی‌تلیال تخمدان انجام شد، پژوهشگران به جهت بررسی رابطه‌ی دیابت نوع دو و بقاء (بقاء کلی و بقاء عود مجدد)، نسبت مخاطره را محاسبه کردند. نسبت مخاطره در این مطالعه ۴/۷۶ بدست آمد. نسبت مخاطره در این مطالعه نشان داد که بیماران مبتلا به سرطان تخمدان که همزمان دیابت نوع دو داشتند در مقایسه با بیماران مبتلا به سرطان تخمدان که دیابت نداشتند بطور متوسط مخاطره‌ای معادل ۴/۷۶ برابر داشتند (۲۹). نسبت مخاطره‌ی بدست آمده از نمونه‌ها برآورد نقطه‌ای بوده و نمی‌تواند تأییدی بر رابطه‌ی دیابت نوع دو و بقاء بیماران مبتلا به سرطان تخمدان باشد. همانطور که در شکل شماره‌ی ۷ نشان داده شده است، دو نمونه‌ی بیماران مبتلا و غیرمبتلا به دیابت نماینده‌ی جامعه‌ی خود بوده و ما برای دانستن رابطه‌ی واقعی دیابت نوع دو و بقاء بیماران مبتلا به سرطان تخمدان، باید نسبت مخاطره بین دو جامعه را بدانیم. با در



دو جامعه را در بر گرفته اند. تعداد ۵ مطالعه (۵ درصد خطای تصادفی مفروض)، نمی‌توانند شاخص نسبی بین دو جامعه را در بر بگیرند. برای هر شاخص یک توزیع نمونه‌ای (تکرار صد مطالعه) می‌سازیم و با مفروض کردن توزیع پارامتر (شاخص اندازه‌ی اثر نسبی بین دو جامعه)، می‌توان مدعی شد که ۹۵ درصد توزیع نمونه‌ای (۹۵ درصد مطالعات)، پارامتر مورد نظر را در بر می‌گیرد. برای ساختن توزیع نمونه‌ای، ابتدا برآورد نقطه‌ای (شاخص اندازه‌ی اثر بین دو نمونه) را محاسبه کرده و دو خطای معیار را به آن کم و اضافه می‌کنیم. همانطور که قبلاً توضیح داده شد، خطای معیار متوسط پراکندگی شاخص اندازه‌ی اثر نسبی هر مطالعه از متوسط شاخص اندازه‌ی اثر نسبی کل است. خطای استاندارد، شاخصی از متوسط پراکندگی مطالعات را بدست می‌دهد. از طرفی می‌دانیم که برای دانستن رابطه‌ی واقعی بین متغیر مستقل و وابسته باید شاخص نسبی جامعه را دانست و نمی‌توان به تنهایی به شاخص نسبی نمونه اتکا نمود. اما می‌توان از جانشین آن یعنی ۹۵ درصد مطالعات (۹۵ مطالعه از ۱۰۰ مطالعه/توزیع نمونه‌ای) استفاده کرد. در صورتی که شاخص نسبی هر یک از این ۹۵ مطالعه یک نباشد می‌توان تصور کرد که شاخص نسبی جامعه یک نبوده و بین متغیر مستقل و وابسته رابطه وجود دارد. در صورتی که در فاصله‌ی اطمینان عدد یک قرار بگیرد، شاخص نسبی یکی از ۹۵ مطالعه عدد یک بوده و در اینصورت ممکن است شاخص نسبی جامعه یک باشد و در این حالت بین متغیر مستقل و وابسته رابطه‌ی وجود ندارد. البته باید این نکته را در نظر گرفت که زمانی می‌توان این فاصله را بدرستی استنباط کرد که حجم نمونه‌ی کافی و سایر خطاها (خطای منظم و مخدوش‌شدگی) وجود نداشته باشند.

### پزشکی مبتنی بر شواهد (Evidence-based medicine) و معنی‌داری فرهنگی (Cultural significance)

یکی از مواردی که متخصص بالینی باید در نظر بگیرد، پذیرش مداخله در فرهنگ جامعه‌ی مورد نظر است. به این مفهوم که متخصص بالینی باید با بررسی تمام عوارض (حتی عوارض ناچیز) و جنبه‌های مداخله‌ی مورد نظر، مداخله را تجویز نماید (۳۰). تصور کنید که یکی از عوارض مکمل کورکومین، باد گلو و یا باد معده (بیشتر از معمول) باشد و همچنین این عارضه نیز از نظر آماری و بالینی معنی‌دار باشد. هر چند که مکمل کورکومین پیامد اولیه (افزایش وزن) و سایر پیامدهای ثانویه (کیفیت زندگی، شاخص‌های التهابی و استرس اکسیداتیو) را از نظر آمار و بالینی بهبود می‌بخشد، اما ممکن است تنها با همین عارضه‌ی جانبی نامطلوب، که در فرهنگ جامعه‌ی مورد نظر (بیماران مبتلا به سرطان معده در کشور ایران) پذیرفته نیست، متخصص بالینی از تجویز آن خودداری کند. علاوه بر این

یکی از مهم‌ترین دلایلی که یک متخصص بالینی باید در مورد تجویز مداخله در نظر بگیرد، هزینه‌ی مداخله‌ی مورد نظر است. بدون شک برای هر نوع بیماری، چه به قصد درمان و چه تشخیص، مداخلات متعددی وجود دارد و متخصص بالینی باید از میان یکی از آن‌ها، مداخله‌ی مناسبی را برای بیمارش تجویز کند. با دانستن معنی‌داری آماری، بالینی و فرهنگی مداخلات متعدد بیماری مورد نظر، متخصص باید هزینه‌های آن‌ها را نیز بداند و با توجه به شرایط مالی بیمار و ترجیح (Preference) خود بیمار، یک مداخله را انتخاب کند (۳۱). البته منظور از هزینه-اثربخشی در اینجا، تصمیم‌گیری در سطح فردی است. به مفهوم دیگر اگر پیامدهای دو مداخله (دارو) معنی‌دار آماری و بالینی باشند، اما هزینه‌ی یکی از دو مداخله (دارو) بیشتر باشد، متخصص باید علاوه بر توضیح اثربخشی بالینی دو دارو، هزینه‌ی آن‌ها را نیز بیان کرده، تا بیمار بتواند با توجه به شرایط مالی خویش، انتخاب صحیحی داشته باشد. در نظر گرفتن معنی‌داری فرهنگی، شرایط مالی و ترجیحات بیمار در رویکرد نوین پزشکی یعنی پزشکی شاهد-مدار یا مبتنی بر شواهد مستتر است. پزشکی مبتنی بر شواهد به عنوان (استفاده‌ی آگاهانه، صریح و خردمندانه از بهترین شواهد فعلی در تصمیم‌گیری در مورد مراقبت از بیماران) تعریف می‌شود. در پزشکی مبتنی بر شواهد متخصصین بالینی تصمیمات خود را بر اساس فرایند جستجوی اطلاعات پزشکی و ارزیابی نقادانه کیفیت متدولوژیک، مطابقت شواهد جاری با بیمار و مورد خاص خود می‌گیرد. به مفهوم دیگر مداخله‌ی پزشک با استفاده از بهترین شواهد جاری، مهارت بالینی و توجه به نیازها و خواسته‌های بیماران صورت می‌گیرد. در پزشکی مبتنی بر شواهد، تشخیص، درمان و مراقبت به شکل انفرادی در نظر گرفته می‌شود و این پرسش مدنظر است که (بهترین اقدامات درمانی در این شرایط و برای این بیماری خاص چیست؟ برای پیروی از چنین رویکردی، شواهد بازایی شده بعد از ارزیابی نقادانه و رتبه‌بندی باید با تجربه‌ی بالینی درمانگر تلفیق شده و بر اساس خواسته‌های بیمار که همچنین وابسته به محل زندگی، شرایط اقتصادی و فرهنگی اوست، مداخله شود. این نکته بسیار مهم است که متخصصین بالینی باید بدانند که ممکن است روش‌های درمانی و مراقبت‌های مناسب الزاماً با آنچه بهترین شواهد پیشنهاد می‌کنند مطابقت نداشته باشد (۳۲، ۳۳). ایده‌آل بر آن است که پزشکی شاهد-مدار بر مبنای ارتباط مستحکم بین پزشک و بیمار شکل بگیرد و با رعایت کامل اصول اخلاقی، در نظر گرفتن جنبه‌های انسانی مراقبت و در قالبی قابل درک برای بیمار ارائه شود. ترجیحات و توجه به خواسته‌های بیمار در فرایند درمان و نتایج آن بسیار مهم است. در یک رابطه‌ی درمانی مثبت، یعنی رابطه‌ای که در آن درمانگر احساس همدلی، صداقت و توجه مثبت به بیمار دارد، پیروی از

است (۳۴، ۳۵). نتایج مطالعات متعددی نشان داده‌اند که بهترین شیوه در روابط متقابل پزشک-بیمار و تبادل اطلاعات باید بیمار-محور بوده و الگوی تصمیم‌گیری مشارکتی اتخاذ شود. این الگو مزایای فراوانی مانند افزایش رضایت بیمار، کاهش استرس و نگرانی‌های بیمار و کاهش نتایج مضر مرتبط با تجویزهای نادرست است (۳۶، ۳۷).

درمان، جدی گرفتن توصیه‌های پزشکی از سوی بیمار، ایجاد رابطه‌ی درمانی مطلوب و رضایت بالاتر بیمار به مراتب بالاتر است. مطالعات قبلی نشان داده‌اند که ضعف مهارت‌های ارتباطی درمانگر و بکارگیری شیوه‌های ارتباطی یکسان در برخورد با بیماران بدون توجه به تفاوت‌های فردی و حتی نوع بیماری و جایگاه فردی در نظام فرهنگی-اجتماعی در اغلب موارد از علل اصلی ناموفق بودن درمان

## References

1. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. *Am Stat* 2019; 73(S1): 235-45.
2. Wagenmakers EJ. A practical solution to the pervasive problems of p values. *Psychon Bull Rev* 2007; 14(5): 779-804.
3. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev Camb Philos Soc* 2007; 82(4): 591-605.
4. Ferguson CJ. An effect size primer: a guide for clinicians and researchers. *Prof Psychol Res Pr* 2016; 40(5): 532-8.
5. Figueiredo Filho DB, Paranhos R, Rocha ECd, Batista M, Silva Jr JAd, Santos MLWD, et al. When is statistical significance not significant? *Bras Political Sci Rev* 2013; 7(1): 31-55.
6. Wellek S. A critical evaluation of the current "p-value controversy". *Biom J* 2017; 59(5): 854-72.
7. Frérot M, Lefebvre A, Aho S, Callier P, Astruc K, Aho Glélé LS. What is epidemiology? Changing definitions of epidemiology 1978-2017. *PloS One* 2018; 13(12): e0208442.
8. Samani RO, Almasi-Hashiani A, Vesali S, Shokri F, Cheraghi R, Torkestani F, et al. Tehran survey of potential risk factors for multiple births. *Int J Fertil Steril* 2017; 11(3): 220-5.
9. Pirjani R, Hosseini R, Soori T, Rabiei M, Hosseini L, Abiri A, et al. Maternal and neonatal outcomes in COVID-19 infected pregnancies: a prospective cohort study. *J Travel Med* 2020; 27(7): taaa158.
10. Karimi E, Khorvash F, Arab A, Sepidarkish M, Saadatnia M, Amani R. The effects of royal jelly supplementation on oxidative stress, inflammatory mediators, mental health, cognitive function, quality of life, and clinical outcomes of patients with ischemic stroke: study protocol for a randomized controlled trial. *BMC Nutr* 2023; 9(1): 32.
11. Heshmati J, Moini A, Sepidarkish M, Morvaridzadeh M, Salehi M, Palmowski A, et al. Effects of curcumin supplementation on blood glucose, insulin resistance and androgens in patients with polycystic ovary syndrome: A randomized double-blind placebo-controlled clinical trial. *Phytomedicine* 2021; 80: 153395.
12. Zhao L, Tian L, Cai T, Claggett B, Wei LJ. Effectively selecting a target population for a future comparative study. *J Am Stat Assoc* 2013; 108(502): 527-39.
13. Altman DG, Bland JM. Uncertainty and sampling error. *BMJ* 2014; 349: g7064.
14. Sedgwick PM. What is sampling error? *BMJ* 2012; 344: e4285.
15. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev Camb Philos Soc* 2007; 82(4): 591-605.
16. Borenstein M, Cooper H, Hedges L, Valentine J. Effect sizes for continuous data. In Cooper H, Hedges LV, Valentine JC, editors. *The handbook of research synthesis and meta-analysis*. Manhattan, NY: Russell Sage Foundation; 2009. p. 221-35.
17. Lee DK. Alternatives to P value: confidence interval and effect size. *Korean J Anesthesiol* 2016; 69(6): 555-62.
18. Xie Mg, Singh K. Confidence distribution, the frequentist distribution estimator of a parameter: A review. *Int Stat Rev* 2013; 81(1): 3-39.
19. Fleischmann M, Vaughan B. Commentary: Statistical significance and clinical significance-A call to consider patient reported outcome measures, effect size, confidence interval and minimal clinically important difference (MCID). *J Bodyw Mov Ther* 2019; 23(4): 690-4.
20. Martínez-Ezquerro JD, Riojas-Garza A, Rendón-Macías ME. Clinical significance vs statistical significance. How to interpret the confidence interval at 95%. *Rev Alerg Mex* 2017; 64(4): 477-86.
21. Quertemont E. How to statistically show the absence of an effect. *Psychologica Belgica* 2011; 51(2): 109-27.
22. Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, Walter S. Basic statistics for clinicians: 2. Interpreting study results: confidence intervals. *CMAJ* 1995; 152(2): 169-73.
23. Brahman LE. Confidence intervals assess both clinical significance and statistical significance. *ACP* 1991; 114(6): 515-7.
24. Prentice RL. Discussion: On the role and analysis of secondary outcomes in clinical trials. *Control Clin Trials* 1997; 18(6): 561-7.
25. Ferreira JC, Patino CM. Types of outcomes in clinical research. *J Bras Pneumol* 2017; 43(1): 5.
26. Norton EC, Dowd BE, Maciejewski ML. Odds ratios-current best practice and use. *JAMA* 2018; 320(1): 84-5.
27. Sedgwick P, Joeekes K. Interpreting hazard ratios. *BMJ* 2015; 351: h4631.
28. Alizadeh Khatir A, Sepidarkish M, Rajabalizadeh MR, Alizadeh Moghaddam S, Aghapour S, Mehravar

- S, et al. Case-control study to assess the association between epilepsy and toxocara Infection/Exposure. *Microorganisms* 2021; 9(10): 2091.
29. Akhavan S, Ghahghaei-Nezamabadi A, Modaresgilani M, Mousavi AS, Sepidarkish M, Tehranian A, et al. Impact of diabetes mellitus on epithelial ovarian cancer survival. *BMC Cancer* 2018; 18(1): 1246.
30. Schouten BC, Meeuwesen L. Cultural differences in medical communication: a review of the literature. *Patient Educ Couns* 2006; 64(1-3): 21-34.
31. Bowling A, Ebrahim S. Measuring patients' preferences for treatment and perceptions of risk. *Qual Health Care* 2001; 10(Suppl 1): i2-8.
32. Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, et al. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. *JAMA* 2000; 284(10): 1290-6.
33. McColl A, Smith H, White P, Field J. General practitioners' perceptions of the route to evidence based medicine: a questionnaire survey. *BMJ* 1998; 316(7128): 361-5.
34. Emanuel EJ, Emanuel LL. Four models of the physician-patient relationship. *JAMA* 1992; 267(16): 2221-6.
35. Stewart MA. Effective physician-patient communication and health outcomes: a review. *CMAJ* 1995; 152(9): 1423-33.
36. Swift JK, Callahan JL. The impact of client treatment preferences on outcome: A meta-analysis. *J Clin Psychol* 2009; 65(4): 368-81.
37. Siminoff LA. Incorporating patient and family preferences into evidence-based medicine. *BMC Med Inform Decis Mak* 2013; 13(Suppl 3): S6.

## An Overview of Statistical and Clinical Concepts: The Use and Interpretation of Confidence Interval

Mahdi Sepidarkish<sup>1</sup>, Zahra Mohammadi-Pirouz<sup>2</sup>

### Review Article

#### Abstract

Application and interpretation of statistical significance of association are the basic and necessary principle in medical research. Traditionally, hypothesis testing and reporting p-values have been extensively utilized to quantify the statistical significance of observed results. The majority of published research that involves statistical inference seems to make exclusive use of hypothesis testing, and summarize their findings only to statistical significance. Most importantly, p-value fail to provide two crucial pieces of information: (1) the magnitude of an effect of interest, and (2) the precision of the estimate of the magnitude of that effect. Currently, most of the reputable journals no longer accept statistical significance alone; and there has been increasing attention focused on the Effect size index, confidence limits along with clinical significant. However, there is no single guideline for reporting statistical and clinical significance, and there are inconsistencies between journals. The aim of this paper is to provide a correct and integrated instruction for reporting statistical and clinical significance in medical sciences employing an estimation approach (reporting confidence interval).

**Keywords:** Confidence interval; Data interpretation; Effect size; Estimation; Statistics

**Citation:** Sepidarkish M, Mohammadi-Pirouz Z. **An Overview of Statistical and Clinical Concepts: The Use and Interpretation of Confidence Interval.** J Isfahan Med Sch 2023; 41(731): 696-707.

1- Assistant Professor, Department of Biostatistics and Epidemiology, School of Public Health, Babol University of Medical Sciences, Babol, Iran

2- MSc Student, Department of Biostatistics and Epidemiology, School of Public Health, Babol University of Medical Sciences, Babol, Iran

**Corresponding Author:** Mahdi Sepidarkish, Assistant Professor, Department of Biostatistics and Epidemiology, School of Public Health, Babol University of Medical Sciences, Babol, Iran; Email: mahdi.sepidarkish@gmail.com