

## پژوهش در جراحی

# راهنمای یک جراح در استنتاج‌های آماری

ترجمه و تلخیص: دکتر سیدعباس میرمالک\* و گروه مترجمین\*\*

متون پزشکی که به شکل مقاله ارائه می‌شوند، اغلب برای اثبات از پشتوانه آماری استفاده می‌کنند. تفسیر آمارهای ارائه شده و نتایج حاصل از آن مستلزم آشنایی اندک و درک مفاهیم روش شناختی و چالش‌های آن است. مفاهیمی مانند؛ گروه‌های مورد مطالعه، تصادفی بودن کارآزمایی‌ها، مفهوم ارزشی  $P$  (P-value)، فاصله اطمینان، توان و ... راهنمای قضاوت در مورد هر مقاله است. در صورت ناآشنایی با این مفاهیم ممکن است برای مقاله‌ای کم ارزش اعتبار زیادی قائل شویم و یا نوشتاری معتبر را فاقد ارزش تلقی کنیم.

### مقدمه

از آخرین دهه قرن بیستم جراحان خود را در حال تصمیم‌گیری براساس تجزیه و تحلیل‌های آماری می‌یابند. این تجزیه و تحلیل‌ها که بطور مرتب در مجلات تخصصی و سخنرانی‌های علمی ارائه می‌شوند، توسط دولت‌ها و شرکت‌های بیمه نیز برای تدوین اسناد مربوط به خط مشی مراقبت‌های بهداشتی مورد استفاده قرار می‌گیرند. وسایل ارتباط جمعی غالباً ممکن است این اطلاعات را بدون در نظر گرفتن چالش‌های جدی همراه با روش‌های آماری مورد استفاده در دستیابی به نتایج انتشار یافته، منتشر کنند. استفاده‌کنندگان معمولاً تصور می‌کنند آزمون‌های آماری بکار گرفته شده در مقالات و نشریات بسیار دقیق، مناسب و کافی هستند. اما بطوری که تخمین زده شده است، حدود دو سوم مطالعات منتشر شده در بهترین نشریات پزشکی دارای اشتباهات آماری و نتایج غیرموجه می‌باشند. برای ارزیابی نقادانه این اطلاعات، داشتن درکی کلی از روش‌های آمار حیاتی و محدودیت‌های آنها لازم است. علاوه بر درک مفاهیم روش شناختی و چالش‌های آنها، جراحان نیاز به تفسیر نتایج در مفهوم بالینی آن دارند. برخی از کارشناسان احساس می‌کنند که ارقامی غیرضروری در تأکید بر گزارش اطلاعات خام تا آزمون فرضیه وجود دارد، که موجب افزایش برای دانستن شاخص آماری و مسائل بالینی می‌شوند. برای مثال برخی مطالعات، نتایج بالینی را با استفاده از نقطه‌های پایان حساس اما فاقد شاخص بالینی به اثبات رسیده، گزارش می‌کنند. امروزه کارآزمایی بالینی تصادفی آینده‌نگر در حال تبدیل شدن به روش استاندارد طلایی برای ارزیابی اثر بخشی مداخلات تشخیصی و درمانی است. این گونه تصور می‌شود که این کارآزمایی بهترین شواهد مربوط به روابط علت و معلولی را در اختیار می‌گذارد. این کارآزمایی پرهزینه‌ترین کارآزمایی‌ها از نظر اجرائی می‌باشد.

نویسنده پاسخگو: دکتر سیدعباس میرمالک

تلفن: ۸۸۷۸۷۵۶۱

Email: SAM@Mirmalek.net

\* استادیار گروه جراحی عمومی، دانشگاه علوم پزشکی آزاد اسلامی، واحد تهران

\*\* دکتر لیلا پرورش، دکتر شقایق تهرانی، دکتر پوریا حسینی، دکتر پانته رضائیان، دکتر مریم سعیدیان

دکتر مروا طهماسبی‌راد، دکتر علی غلامرضانژاد، دکتر مهدی کلاتری، دکتر الهام کنی، دکتر امیر تیمور مرعشی،

دکتر امید میرمطلبی، دکتر علی ناظمیان

تاریخ وصول: ۱۳۹۰/۰۱/۱۵

## راهنمایی

مراقب نوشته‌هایی که سعی در اثبات روابط علت و معلولی بر پایه مطالعات گذشته‌نگر و مقطعی دارند، باشید. این روابط به بهترین شکل در متن یک کارآزمایی تصادفی بالینی به اثبات می‌رسد.

روش‌شناسی کارآزمایی از ۳۰ سال پیش که اولین کارآزمایی‌های بالینی تصادفی آینده‌نگر منتشر گردیدند، به شکل زیربنایی تکامل یافته است. قدرت این رویکرد، در شباهت بین گروه‌های درمان و شاهد حاصل از توزیع تصادفی بیماران نهفته است. مطالعات مشاهده‌ای تاریخی در مقایسه با کارآزمایی‌های تصادفی به میزان بالاتری از شاخص آماری دست می‌یابند. این مسئله معمولاً قابل استناد به این واقعیت است که گروه‌های شاهد در این مطالعات ضعیف‌تر عمل می‌کنند و علت این امر وابسته به دشواری در حذف تورش درون بخشی در مواردی است که توزیع تصادفی بیماران، اعمال نشده باشد. ممکن است تغییرات داده نشده نیز در گروه‌های تاریخی در برخی عرصه‌ها وجود داشته باشند. در عین حال انجام کارآزمایی بالینی به منظور پاسخ به یک سؤال جراحی بنا به دلایل بسیار، همواره میسر نیست، به عنوان مثال، شرایط مورد نظر ممکن است نادر باشند، به این ترتیب تکمیل یک کارآزمایی میزان غیرقابل قبولی از زمان را بخود اختصاص می‌دهد. از سوی دیگر زمانی که هزینه کارآزمایی در مقابل اهمیت سؤال مورد نظر در کفه سنگین‌تری قرار گیرد، به نحوی که هزینه سرسام‌آور آن مانع اجرای آن شود. در انتها عدم ارائه درمان به یک گروه شاهد آن هم در صورتی که بیماری مورد نظر به طور یکسان کشنده باشد، اخلاقی به نظر نمی‌رسد. این مقاله به بحث در مورد قابلیت قیاس گروه‌های مورد مطالعه، پیش‌آگهی، طبقه‌بندی، انتخاب تصادفی، موضوع ابعاد نمونه و *P-values* می‌پردازد. همچنین به تفاوت بین کارآزمایی‌های مثبت و منفی و تفاوت بین شاخص آماری و بالینی پرداخته می‌شود. مقایسه آماری دو شیوه درمانی تنها زمانی دارای مفهوم است که بتواند تفاوت مهم بالینی را شناسایی کند.

## بررسی گروه‌های مورد مطالعه

پاراگراف اول بخش نتیجه‌گیری یک مقاله منتشره در نشریات، معمولاً گروه مورد مطالعه را در مفهوم ویژگی‌های جمعیت‌شناختی و عوامل پیش‌آگهی دهنده توضیح می‌دهد. این توضیح خواننده را قادر می‌سازد تا در مورد شباهت کافی گروه‌ها نسبت به یکدیگر تحقیق نماید. شباهت بین گروه‌های مورد مطالعه، امری مهم است، زیرا آزمون‌های مربوط به شاخص آماری بر اساس این فرض که گروه‌های مورد مطالعه به طور تصادفی از درون جمعیت مورد نظر انتخاب شده‌اند، مورد پیش‌بینی قرار می‌گیرند. البته این مسئله تنها در کارآزمایی‌های بالینی تصادفی دقیق می‌باشد. این بدان معنا نیست که نتایج آزمون‌های مربوط به شاخص آماری که به روش دیگر به دست آمده‌اند، دقیق نیستند، بلکه این نتایج باید با احتیاط بیشتری تفسیر شوند. سطح شاخص گزارش شده و قابلیت تعمیم نتایج به کل جمعیت، تحت تأثیر میزان شکست در توزیع تصادفی بیماران در شاخه‌های مختلف درمانی می‌باشد.

در کارآزمایی‌های تصادفی، لازم به اثبات شباهت گروه‌های شاهد و مورد با یکدیگر با استفاده از آزمون مربوط به تفاوت شاخص آماری بین آنان نیست. گزارش تعداد مطلق بیماران با هر یک از متغیرهای وابسته به آنان به صورت جدول کافی خواهد بود. در کارآزمایی‌های غیرتصادفی، بررسی عدم توازن بین گروه‌های مورد مطالعه به سبب تورشی که ممکن است در انتخاب یک بیمار ویژه برای یک شاخه درمانی در مقابل شاخه درمانی دیگر، مداخله داده شود، بسیار مهم است. اگر عدم توازن بین گروه‌ها با توجه به متغیر ویژه‌ای وجود داشته باشد، یک آزمون شاخص آماری جهت تعیین توان این عدم توازن در تأثیر بر روی نتایج باید انجام شود. افزون بر این نگاهی به گذشته برای بررسی چگونگی گردآوری بیماران برای مطالعه ممکن است نشان دهد که تورش انتخاب می‌تواند نتایج را تحت تأثیر قرار دهد.

اگر جراح در مورد عدم شرکت دادن کلیه بیماران واجد شرایط در مطالعه تردید داشته باشد، تورش در انتخاب ممکن است به راستی وارد عمل شود. در یک مطالعه واقعاً تصادفی نویسندگان باید، نه تنها تعداد بیمارانی که در نهایت به طور تصادفی انتخاب شده‌اند، بلکه همچنین تعداد بیماران واجد شرایطی را که در مطالعه شرکت داده نشده‌اند را به همراه علت کوتاهی در انجام این کار ارائه دهند. برای مثال ممکن است فرد بیمار در مورد انجام مطالعه آگاه نشده باشد. ممکن است جراح احساس کرده باشد که یک بیمار با خطر بیشتر یا کمتر مواجه است یا این که به دلایلی برای مطالعه مزبور غیرقابل قبول است یا این که برای مشارکت در مطالعه فراموش شده باشد. ممکن است پزشکی که بیمار به وی ارجاع شده است تقاضای رویکردی ویژه برای وی کرده باشد.

خواننده باید براساس اطلاعات داده شده قضاوت کند و معین کند که تورش انتخاب ممکن است عمل کرده باشد یا خیر و اگر چنین تورش وجود داشته تا چه اندازه قابلیت تعمیم نتایج به بیماران را تحت تأثیر قرار داده است. سه سؤال وجود دارد که خواننده محتاط باید برای حصول اطمینان از درستی اطلاعات خلاصه شده، قادر باشد به آنها پاسخ دهد: (۱) آیا همه اعداد با هم جمع شده‌اند، یعنی آیا همه بیماران با توجه به تک تک پارامترهای گزارش شده محاسبه شده‌اند یا خیر؟ (۲) آیا مشاهدات به درستی محاسبه شده‌اند یا خیر؟ (۳) آیا سطح پایه متغیرها بین گروه شاهد و درمان یکسان است یا خیر؟

باید طبقه‌بندی متغیرهای مربوط به پیش‌آگهی بیماری انجام شود، زیرا بسیار با مفهوم‌تر است که بیماران قبل از انتخاب تصادفی رده‌بندی شوند تا این که تجزیه و تحلیل بر روی زیر گروه‌ها در انتها انجام شود. در مثال پیشین محققان تفاوتی در پاسخ بین زیر گروه‌های تعریف شده به وسیله نسبت را پیش‌بینی کرده و تلاش نمودند تا اطمینان حاصل شود که تعداد برابری از بیماران با این نسبت به صورت تصادفی در هر شاخه درمانی قرار گرفته‌اند، در حالی که در نمونه اخیر، محققان تفاوتی در میزان پاسخ‌ها یافتند و ترجیح دادند که پس از این اطلاعات را براساس زیر گروه‌ها بررسی نمایند. به عنوان مثال برای پروفیلاکسی آنتی‌بیوتیک در بیماران جراحی، فرد باید اختلاف در میزان وقوع عفونت زخم بین افراد دیابتی در مقایسه با افراد غیردیابتی را پیش‌بینی کند، برای حصول اطمینان از وجود تعداد برابر افراد دیابتی در هر گروه، وجود دیابت قبل از انتخاب تصادفی را طبقه‌بندی کند، از سوی دیگر تجزیه و تحلیل زیر گروه‌ها هم می‌تواند آموزنده و هم به صورت بالقوه گمراه‌کننده باشد. قدرت استنتاج در تأثیر درمانی بین زیر گروه‌ها با توجه به تفاوت مورد نظر در تأثیر درمان در بین افراد زیر گروه‌ها وابسته به عوامل متعدد می‌باشد. اگر بزرگی اختلاف، تأثیری بر روی تصمیم‌گیری بالینی نداشته باشد، تجزیه و تحلیل زیر گروه مفید نیست. افزون بر این اجرای آزمون‌های مجزای شاخص بر روی زیر گروه‌های مختلف احتمال دستیابی به تفاوت شاخص آماری با تکیه بر شانس را افزایش می‌دهد و بنابراین ممکن نیست شاهد مستقیمی، مبنی بر این که عوامل تعیین‌کننده پیش‌آگهی بر تفاوت مشاهده شده تأثیر بگذارند، به دست دهد. در این مثال آزمونی برای بررسی تأثیر متقابل می‌تواند انجام شود و روشی معتبر برای بررسی ارتباط بین متغیر و پاسخ درمانی خواهد بود. اگر نویسندگان تأثیر درمانی متفاوتی را پیش‌بینی کردند، تجزیه و تحلیل زیر گروه‌ها قابل اعتمادتر از یافتن تأثیر پس از انجام آزمون می‌باشد، زیرا ممکن است در این راه متهم به مطرح کردن مجدد اطلاعات شوند. یافتن تأثیر درمانی متفاوت در مطالعه مجزا بر میزان اعتماد آن می‌افزاید. در نهایت این تأثیر درمانی متفاوت که مفهومی زیست‌شناختی را ایجاد می‌کند، حتی موجه نیز خواهد بود.

## راهنمایی

مراقب باشید تا اطلاعاتی را که مورد بررسی چندین تجزیه و تحلیل زیر گروهی بوده‌اند، مورد تفسیر بیش از حد قرار ندهید. زیر گروه‌ها ممکن است دارای ابعاد متفاوتی باشند و در نتیجه مقادیر احتمال مربوط به تجزیه و تحلیل هر زیر گروه باید براساس اندازه همان گروه محاسبه شود.

اثبات این که آیا کلیه بیماران شرکت داده شده در مطالعه، در قسمت نتیجه‌گیری به حساب آمده‌اند یا خیر؟ دارای اهمیت است. اگر تعداد موارد حذف اختیاری و حذف اجباری در یک شاخه درمانی به شکلی نامتناسب بالا باشد، باید آزمون مربوط به شاخص اجرا شود تا تعیین گردد که آیا این تعداد بین گروه‌های درمان و شاهد عدم تساوی ایجاد کرده است یا خیر؟ این امر از آن جهت اهمیت دارد که عدم تساوی بین گروه‌ها می‌تواند بر روی اعتبار مطالعه تأثیر داشته باشد. به طور مثال در مقایسه بین دو دارو برای درمان فشار خون یک دارو ممکن است موجب بروز حوادث ناشی از عوارض جانبی نامطلوب که در حد متناسبی بالا است، شود که در نتیجه منجر به حذف اختیاری تعداد بیشتری از بیماران شاخه درمانی شده و در نهایت عدم تعادل بین گروه درمان و شاهد را ایجاد نماید. در عین حال باید اطمینان حاصل شود که تعداد مشاهدات انجام شده در مورد هر یک از افراد مورد مطالعه با هم برابر است. به عنوان مثال اگر محقق در مورد تأثیر دو داروی ضد فشار خون یک ساعت پس از تجویز بخواهد گزارش دهد اما در گزارش خود در مقابل ۱۰۰ مورد مشاهده در گروه درمان تنها ۲۰ مورد مشاهده را در گروه شاهد ذکر نماید، خواننده ممکن است در مورد وجود تورش گزارش به شک بیفتد. تورش در گزارش می‌تواند گروه درمان یا گروه شاهد را تحت تأثیر قرار دهد.

بیماران به دلایل متعدد از مطالعه خارج شده یا کنار گذاشته می‌شوند (حذف اختیاری و حذف اجباری). برای مثال آنها ممکن است (۱) از عوارض جانبی نامطلوب رنج برده یا به علت عوارض ناشی از درمان، از مطالعه خارج شوند، (۲) بهبودی یافته و تمایل به ادامه پیگیری نداشته باشند، (۳) فوت نمایند، (۴) تغییر مکان داده و قادر به ادامه پیگیری نباشد. در گزارش در مورد پیش‌آگهی، این که سرنوشته هر بیمار ثبت نام شده در مطالعه ذکر شود و هر یک از آنان در تجزیه و تحلیل نتایج گنجانده شوند، دارای اهمیت می‌باشد. اگر در مطالعه‌ای همه بیماران به حساب نیامده باشند، تأثیر آن را گاهی اوقات می‌توان با محاسبه مجدد آماری و با فرض بر این که همه این بیماران حداقل نتایج مطلوب را تجربه کرده‌اند، بررسی نمود. اگر نتیجه با در نظر گرفتن این محاسبه تغییر نکند پزشکان می‌توانند اینگونه تصور کنند که نتیجه به دست آمده، نتیجه‌ای استوار است، در غیر این صورت اهمیت بالینی آن باید به زیر سؤال برود.

## تجزیه و تحلیل نتایج P-value

هنگامی که P-value به مقالات جراحان پیوست، پزشکان آموختند که  $P < 0.05$  به معنای آن است که تفاوت در هر آن چه که سنجیده شده است، از نظر آماری شاخص است. معنای این سخن چیست؟

بیش از هر چیز باید گفت آمارگران ناگزیر از انتشار P-value حقیقی هستند (مثلاً  $P = 0.04$  به جای  $P < 0.05$ ) همانگونه که ناچارند ارزش آزمون آماری که P-value از آن طریق به دست آمده است را بیان کنند. برای بیشتر مطالعات، جداول آماری دو دامنه‌ای باید مورد استفاده قرار گیرند. همان‌گونه که قبلاً گفته شد این امور محافظه‌کارانه‌تر بوده و نتیجه، شاخص آماری مشکل‌تر بدست می‌آید. با استفاده از جداول دو دامنه‌ای محقق مجاز است در مورد امکان وقوع تأثیر مثبت یا منفی تحقیق نماید. هنگامی که مقدار  $P = 0.05$  می‌باشد، بدین معنا است که یک احتمال پنج درصدی برای تصادفی بودن تفاوت مشاهده شده بین دو گروه شاهد و درمان با در نظر گرفتن هر آن چه که مورد مقایسه قرار گرفته است، وجود دارد. P-values اطلاعات اندکی در مورد ابعاد تفاوت بین گروه‌ها در اختیار قرار می‌دهند. هنگامی که  $P < 0.001$  باشد به این معنا نیست که نتیجه به دست آمده دارای اهمیت زیادی است، بلکه بیشتر بیانگر بزرگ بودن ابعاد نمونه مورد استفاده می‌باشد و از سوی دیگر نشان می‌دهد که تصادفی بودن، توجه مناسبی برای تفاوت مشاهده شده نمی‌باشد. در هر مطالعه‌ای نویسنده باید میزان تأثیر درمانی را که از نظر بالینی معنی‌دار در نظر گرفته می‌شود، از پیش تخصیص دهد.

دلایل متعددی را که موجب می‌شود یک تفاوت شاخص آماری (برای مثال  $P = 0.05$ ) احتمالاً به عنوان یک تفاوت مرتبط بالینی مورد تفسیر قرار نگیرد در زیر آمده است:

۱- ممکن است جمعیتی از بیماران که مداخلات درمانی بر روی آنان انجام شده نمایانگر بیمارانی که درمان کرده‌اید، نباشند. به عنوان مثال ممکن است بیماران از گروه‌های مختلف سنی، نژادی بوده یا دارای بیماری‌های متفاوت همراه باشند و این اطلاعات باید از مقاله مربوطه گردآوری شوند تا قابلیت اجرای بالینی نتیجه بدست آمده برای بیماران تعیین شود.

۲- ممکن است وضعیت هزینه مداخلات نسبت به اهمیت اختلاف مشاهده شده در نتیجه، حالت بازدارنده داشته باشد، هر چند که نتیجه از نظر آماری شاخص باشد. برای مثال استفاده از یک رژیم جدید شیمی درمانی ممکن است طول عمر یک زن مبتلا به سرطان پستان را ۴ هفته بیشتر نسبت به رژیم متداول افزایش دهد، اما آیا ۴ هفته از نظر بالینی معنی‌دار است یا خیر؟ آیا عوارض جانبی رژیم اخیر کیفیت زندگی در بیمار را نسبت به رژیم مورد استفاده فعلی برای مدت زمان بیشتری کاهش داده و در نتیجه منافع واقعی زندگی ۲۸ روزه را در بیمار کاهش می‌دهد؟ آیا ۲۸ روز برای بیماری که از سرطان پستان رنج می‌برد، از نظر بالینی مدت زمان مهمی است؟ هزینه مداخله درمانی چقدر است؟ باید تصمیم بگیرید که آیا رژیم درمانی جدید در محدودیت‌های عملکرد بالینی شما قابل اجرا و عملی است یا خیر؟

۳- ممکن است خطر مداخله در جمعیت بیماران شما به طور قابل توجهی بالا بوده و شما تمایلی برای پذیرفتن حتی یک احتمال پنج درصدی تفاوت در نتیجه را صرفاً با اتکاء به شانس نداشته باشید. برای مثال براساس یک کارآزمایی بالینی ممکن است قطع و آناستوموز اولیه در بیماران مبتلا به سرطان انسدادی کولون نسبت به رویکرد دیگر در این بیماری به صورت قطع کولونوستومی ثانویه ارجحیت پیدا کند. در جمعیت بیماران شما ممکن است درصد بالاتری از مبتلایان به

دیابت، بیماران تحت درمان با استروئید و بیماران سالخورده نسبت به گروه کارآزمایی وجود داشته باشد و در نتیجه، رویکرد جراحی در گروه شما نتواند نتیجه یکسانی را به دست دهد. از سوی دیگر چنانچه مداخله مورد بحث خطر یا زبانی برای گروه بیماران شما در بر نداشته باشد، ممکن است مایل به قبول احتمال بالاتر تفاوت که با اتکاء بر شانس به وقوع پیوسته (مثلاً ده درصد) باشید.

## راهنمایی

به نویسنده‌ای که ارزش دقیق آزمون آماری انجام شده را به همراه  $P$ -value مربوطه گزارش می‌کند، می‌توانید اعتماد بیشتری داشته باشید.

هنگامی که  $P$ -value بزرگ‌تر از  $0/05$  یا NS گزارش شود، تصمیم‌گیری در مورد ارتباط بالینی این اطلاعات مشکل‌تر خواهد بود. اختلافات کوچک در تأثیرات درمانی می‌تواند برای یک متخصص بالینی دارای اهمیت باشد هر چند که از نظر آماری شاخص نباشند. گزارش به صورت  $P=NS$  خواننده را از یک سلسله اطلاعات بالقوه مهم محروم می‌کند، در حالی که  $P=0/07$  در مقابل  $P=0/070$  ممکن است دارای ارتباط بالینی باشد. افزایش ابعاد نمونه اندکی بزرگ‌تر در مثال قبلی می‌تواند یک نتیجه مهم آماری به دست دهد در حالی که افزایش ابعاد نمونه در مثال اخیر بعید به نظر می‌رسد که بتواند تفسیر نتیجه را تغییر دهد. بسیاری از خوانندگان و محققان به غلط نتیجه‌گیری می‌کنند که دو درمان یا آزمون زمانی که ارزش  $P > 0/05$  باشد، معادلند. تنها چیزی که در واقع می‌توان از این نتیجه به دست آورد این است که تفاوتی بین گروه‌ها وجود ندارد که مطالعه موفق به شناسایی آن نشده است. عدم موفقیت در شناسایی تفاوت شاخص آماری ممکن است به معنای عدم وجود تفاوت در طبابت بالینی یا عدم وجود ارتباط بالینی نباشد. سه توضیح ممکن برای این مسئله وجود دارد: (۱) تعداد مواردی که قرار است مقایسه شوند آنقدر کم هستند که موجب به دست آمدن  $P$ -value بزرگ‌تر از  $0/05$  شده‌اند، تنها به دلیل این که ابعاد نمونه کوچک‌تر از آن بوده است که بتواند تفاوت موجود را شناسایی کند، در حالیکه تفاوت وجود داشته است. (۲) ممکن است روش مورد استفاده در اندازه‌گیری اثر درمان، بیش از آن فاقد حساسیت بوده که بتواند تفاوتی را که وجود داشته، شناسایی کند. (۳) یک تفاوت شاخص ممکن است در واقع وجود داشته باشد، اما مطالعه به طور اتفاقی موفق به شناسایی آن نشده باشد. در بازنگری ۷۱ کارآزمایی منفی منتشر شده در تقریباً همه فرضیه‌های صفر پذیرفته شده به طور غلط به گونه‌ای تفسیر شده بودند که نشان دهند هیچ‌گونه تأثیر درمانی وجود نداشته است، در حالی که تفاوت‌هایی به بزرگی ۲۵٪ وجود داشت. علیرغم این که این اختلافات ارتباط بالینی قابل ملاحظه‌ای را دارا بوده‌اند لیکن نتوانستند به حد شاخص آماری برسند.

## مطالعات منفی و فاصله اطمینان

یک روش کار برای ارزیابی ارزش یک مطالعه منفی (مطالعه‌ای که تفاوت شاخص آماری) را بین گروه‌های شاهد و درمان نشان نداده است)، محاسبه ۹۵ درصدی فاصله‌های اطمینان حول محور تفاوت مشاهده شده می‌باشد. فاصله‌های اطمینان منعکس‌کننده طیف ارزش‌هایی است که برای جمعیت بیمارانی که نمونه از آن گرفته شده است، قابل قبول‌اند. این ارزش‌ها کامل‌کننده  $P$ -values می‌باشند و به همراه هم باید در مطالعات منفی گزارش شوند. فاصله‌های اطمینان در مورد ابعاد تفاوت مشاهده شده اطلاعاتی ارائه می‌کنند. اگر فاصله باریک بوده و شامل عدد صفر باشد این تصور را که تفاوت مشاهده شده واقعاً از لحاظ آماری شاخص نیست، تقویت می‌کند. اغلب فاصله‌های اطمینان برای کمک در امر پیش‌بینی و مقایسه نسبت‌ها و حوادث وابسته به زمان بکار می‌روند. برای مثال در مطالعه تصادفی من (J.O) بر روی بیمارانی که تحت عمل توراکتومی به صورت انتخابی قرار می‌گیرند، تعداد عفونت ناشی از زخم با استفاده از یک دوز آنتی‌بیوتیک قبل از عمل و شش دوز آنتی‌بیوتیک بعد از عمل مورد مقایسه قرار می‌گیرد، فاصله اطمینان ۹۵٪ از  $-0/008$  تا  $0/048$  بود. این مطلب نشانگر آن بود که شانس بهتر بودن تجویز آنتی‌بیوتیک به صورت تک‌دوزی نسبت به تجویز آن به صورت شش دوزی بیشتر از  $0/48$  نبوده و شانس بهتر بودن تجویز آنتی‌بیوتیک به صورت شش دوزی نسبت به تجویز آن به صورت تک دوزی بیشتر از  $0/8$  نیست. این حقیقت که فاصله اطمینان بدست آمده بسیار کم بوده و شامل عدد صفر می‌باشد، بیانگر آن است که این امر غیرمحمتمل است که وجود تفاوت شاخص بالینی در میزان عفونت‌های زخم انطباق دو رویکرد با یکدیگر باشد. من پیش‌بینی کردم که نتیجه منفی خواهد بود، یعنی این که موارد عفونت زخم در گروهی که یک

دوز دارو دریافت کرده بود با گروهی که شش دوز دریافت کرده بوده، مشابه خواهد بود. با پیش‌بینی عدم وجود تفاوت، حجم نمونه مورد نیاز برای محاسبه فاصله اطمینان ۹۵٪ را برای اختلاف در نسبت عفونت‌های زخم تخمین زدیم. بدین‌گونه با محاسبه این مسئله که وجود مزیت مهم بالینی برای بیماران در استفاده از شش دوز آنتی‌بیوتیک در مقایسه با یک دوز آن غیرمحمتمل است، احساس آرامش کردم.

مکان دقیق و وسعت فاصله اطمینان معیار خوبی برای تشخیص این مطلب است که حقیقت امر کجاست و کفایت ابعاد حجم نمونه را مشخص می‌کند. فواصل اطمینان برای استفاده در داده‌های توصیفی (مثلاً وزن بیماران) که باید در آنان از میانگین‌ها و انحراف استاندارد شوند، مناسب نمی‌باشند. ارتباط یک مطالعه منفی بستگی به بزرگی تفاوتی دارد که آن مطالعه بتواند شناسایی کند و آن را توان (Power) مطالعه می‌نامند.

## راهنمایی

در یک مطالعه منفی در پی فاصله‌های اطمینان باشید تا به شما در تعیین این که ابعاد نمونه آنقدر بزرگ بوده است که بتواند تفاوتی را در صورت وجود شناسایی نماید کمک کند و این موضوع زمانی اتفاق می‌افتد که توان مطالعه کافی باشد. وسعت فاصله وابسته به خطای استاندارد و در نتیجه وابسته به هر دو مورد ابعاد نمونه و انحراف استاندارد دو اندازه‌گیری نمونه می‌باشد. یک فاصله وسیع باید خواننده را از احتمال کوچکی مفرط ابعاد نمونه آگاه سازد و این یافته‌ای است که به دفعات در نوشته‌ها با آن مواجه می‌شویم. با احتمال کمتر یک فاصله اطمینان وسیع ممکن است بیانگر تنوع طبیعی باشد. سطح اطمینان گزارش شده هم (۹۰٪ در مقابل ۹۵٪) در اندازه فاصله تأثیر دارد. یک فاصله اطمینان ۹۰ درصدی همواره باریک‌تر از فاصله اطمینان ۹۵ درصدی است، زیرا محقق شانس بیشتری (۱۰٪) دارد که حقیقت خارج از محدودیت‌های محاسبه قرار گرفته باشد. فواصل اطمینان احتمالاً در موقعیت‌هایی که ابعاد نمونه نه خیلی بزرگ و نه خیلی کوچک است بیش از سایر موارد ارزشمند است.

## راهنمایی

فواصل اطمینان، راه مناسبی برای بیان طیف تفاوت‌های مورد انتظار در نتیجه، بین گروه‌های شاهد و درمان است. اگر فاصله شامل صفر باشد خواننده می‌داند که امکان آن وجود دارد که تفاوت مشاهده شده در نتیجه، تنها بر اتکای شانس بوده باشد. اگر فاصله باریک باشد خواننده می‌تواند مطمئن شود که ابعاد نمونه کافی بوده است. اکنون تعدادی از مجلات تمایل به استفاده از فاصله اطمینان به جای گزارش خطاهای استاندارد دارند، زیرا روش آماری اخیر شامل اطلاعات اضافی است.

## راهنمایی

مشکل مشابهی ممکن است با محاسبه فواصل اطمینان چندگانه به وجود آید همان‌گونه که در محاسبه  $P$ -value های متعدد هم ممکن است ایجاد شود با هر ۹۵٪ فاصله گزارش شده یک شانس ۵ درصدی وجود دارد که حقیقت در خارج از محدوده وجود داشته باشد.

## توان

توان یک مطالعه به توانایی آن مطالعه در شناسایی تفاوت بین گروه‌ها، در صورت موجود بودن، مربوط می‌شود. این مورد هم به ابعاد نمونه مرتبط است. سطوح توان معمولاً در ۸۰٪ قرار داده می‌شود. در یک مطالعه منفی برای مثال سطح قدرت ۸۰٪ به این معناست که یک احتمال ۸۰ درصدی برای عدم وجود اختلاف بین گروه‌های درمان و شاهد موجود است و این بدان معناست که نتیجه منفی درست است. به سخن دیگر نویسنده مایل است قبول کند که یک امکان ۲۰ درصدی وجود دارد که نتیجه مثبت به سبب عدم کفایت ابعاد نمونه از حیطة شناسایی آنان خارج شده باشد.

بدون داشتن آگاهی نسبت به توان مطالعه، تفسیر در مورد شاخص نبودن نتیجه از نظر آماری اگر غیرممکن نباشد دشوار است. در بازنگری ۸۴ کارآزمایی درمانی چاپ شده در شش نشریه جراحی عمومی از جولای ۱۹۸۱ تا ژوئن ۱۹۸۲ فقط ۵٪ از نویسندگان در مورد توان مطالعه خود صحبت کرده بودند. رید و همکارانش (Reed et al) اعلام کردند که ۸۶٪ مقالات چاپ شده در شش مجله پزشکی که نتایج منفی را گزارش دادند. ممکن است نتایج کوچکی را از قلم انداخته باشند ۶۰٪ ممکن است نتایج متوسط را از دست داده باشند و ۴۰٪ ممکن است تأثیرات بزرگ را نادیده گرفته باشند. ارتباط بالینی یک مطالعه منفی معمولاً وابسته به اندازه تفاوتی است که مطالعه می‌تواند بین گروه‌ها شناسایی کند، به سخنی دیگر وابسته به توان آن است. به‌طور کلی در تفسیر ارتباط بالینی فواصل اطمینان بیش از محاسبه توان مطالعه مورد استفاده است، زیرا در محاسبه توان، نتایج بدست آمده به حساب نمی‌آیند.

## میزان بقاء

### روش‌های جدول زندگی

میزان بقایی که به درستی محاسبه شده باشد، بهترین شاخص آماری در دسترس برای اندازه‌گیری کارآئی (برای مثال درمان سرطان در مقابل درمانی دیگر) است.

در یک کارآزمایی بالینی میانگین طول بقاء ممکن است برای هر گروه مورد مطالعه محاسبه شود. مفهوم داشتن آن به وسیله طول مدت پیگیری محدود می‌شود. خوانندگان ممکن است نتوانند اطمینان یابند که آیا پیش‌آگهی یا بقاء بین گروه‌های مورد مطالعه متفاوت است یا خیر؟ مگر این که همه بیماران در زمان محاسبات در گذشته باشند.

به جای آن ممکن است میزان مرگ و میر محاسبه شود. این درصد بیماران فوت شده در پایان مطالعه است. در ضمن این میزان بستگی به طول مطالعه نیز دارد اگر طول مدت پیگیری بیش از حد کوتاه باشد ممکن است موارد کافی مرگ و میر در هر شاخه از مطالعه برای دادن امکان تصمیم‌گیری به خوانندگان در خصوص منافع و خطرهای درمان مورد مطالعه، داده نشده باشد.

ممکن است گونه دیگری از میزان بقاء محاسبه شود و آن درصد بیماران زنده در هر شاخه از مطالعه در انتهای یک دوره زمانی خاص می‌باشد. میزان اخیر، اطلاعات بالقوه ارزشمندی را هدر می‌دهد، زیرا به بیمارانی که برای مدت کوتاه‌تری پیگیری شده‌اند اجازه مشارکت در میزان داده نمی‌شود.

یک میزان بقاء تعدیل شده بازتاب نسبتی از بیماران در گروهی است که به علت بیماری مورد بحث در مطالعه فوت نکرده‌اند. این میزان، بیماران فوت شده به دلایل دیگر را نادیده می‌گیرد (و در نتیجه میزان بقاء را تعدیل می‌کند) استفاده از این میزان ممکن است در زمان حصول یقین به بالا بودن نامناسب میزان مرگ و میر در یک گروه بنا بر علت‌های غیرمرتبط با بیماری مورد بحث نسبت به سایر موارد مرگ و میر مناسب باشد. وقوع این امر در درون یک کارآزمایی تصادفی غیر محتمل است.

## راهنمایی

مراقب اطلاعات نویسنده‌ای که میزان بقاء تعدیل شده را گزارش می‌دهد باشید، از خود بپرسید چه چیزی تعدیل شده است و آیا این تعدیل مناسب بوده است یا خیر؟

میزان بقای نسبی بازتاب نسبت بیماران فوت شده در یک گروه در مقایسه با میزان قابل انتظار در گروهی از جمعیت عمومی که از نظر نژاد، جنسیت، سن و دوره زمانی انجام مشاهده مشابهند، است. این روش تنها در زمانی که اطلاعات مربوط به علت مرگ کامل نیست، مورد استفاده قرار می‌گیرد.

از سوی دیگر روش‌های آماری یا جدول زندگی (Actuarial or Life-Table) وسیله بکارگیری اطلاعات حاصل از پیگیری را که تا روز بررسی گردآوری شده‌اند، فراهم می‌کند.

تجزیه و تحلیل جدول زندگی روشی متداول برای جمع‌بندی اطلاعات مربوط به بقاء در کارآزمایی‌های بالینی است. همچنین در گزارش نتایج حاصل از تحقیقات طولی با هدف تعیین مثلاً پیش‌آگهی یا نتیجه تحقیق در مورد یک بیماری مزمن مورد استفاده قرار می‌گیرد. مزیت اصلی این روش نسبت به اطلاعات مربوط به گزارش بقای ۵ ساله این است که همه اطلاعات جمع‌آوری شده مربوط به بقاء در بیماران ثبت‌نام شده در یک تحقیق و از جمله کسانی را که ۵ سال مورد نظر را به اتمام نرسانده باشند، مورد استفاده قرار می‌دهد. این روش استاندارد شده نیست. بر نویسندگان لازم است که در مورد چگونگی هدایت بیمارانشان توضیح دهند.

این عبارت «زمان برابر صفر است» که به روشنی تعریف شده است، باید مورد اشاره قرار گیرد (به عنوان مثال تاریخ تشخیص تاریخ اولین درمان یا تاریخ در بیمارستان). افزودن بر این یک نقطه پایان تعریف شده هم باید ذکر شود (مثلاً مرگ). نویسندگان برای اجتناب از نتایجی که به طور بالقوه با تورش همراهند، باید سعی کنند که تعداد بیمارانی را که از پیگیری حذف می‌شوند به حداقل برسانند و چگونگی انجام این کار را گزارش دهند.

تجزیه و تحلیل جدول زندگی این حقیقت را که بیماران ثبت نام شده در مطالعات در زمان‌های مختلف برای مدت‌های متفاوتی از زمان، در مطالعه مشارکت دارند و ممکن است زنده باشند، بیمارانی که فوت می‌کنند، کسانی که از ادامه شرکت در مطالعه منصرف می‌شوند یا از پیگیری حذف می‌شوند را در انتهای دوره مطالعه بحساب می‌آورد. جدول‌های زندگی نسبت به بیماران زنده در انتهای فاصله‌های زمانی مختلف را در مقایسه با بیماران زنده در ابتدای آن فاصله‌های زمانی به کمیت در می‌آورد. به طور کلی بیمارانی که کمتر از یک دوره زمانی کامل زنده می‌مانند، بیشتر مواقع به گونه‌ای در نظر گرفته می‌شوند که نصف یک دوره زمانی زنده بوده‌اند.

مزیت استفاده از تمامی بیماران به این ترتیب است که (۱) اگر تنها تعداد اندکی از بیماران مبتلا به یک بیماری خاص وجود داشته باشند، کلیه اطلاعات بالقوه در جدول زندگی گنجانده می‌شوند. (۲) اگر میزان مرگ و میر در چند سال اول قابل توجه باشد و متعاقباً کارآزمایی متوقف شود، بیمارانی که برای فاصله کوتاه‌تر زمانی مشارکت داشته‌اند، اطلاعاتی مرتبط به اندازه کسانی که در تمام جدول زندگی مشارکت داشته‌اند، ارائه می‌دهند. (۳) اطلاعات مربوط به بقاء به طور کامل تری توضیح داده می‌شوند و درصد کمتری از بیماران برای حذف از پیگیری در نظر گرفته می‌شوند. (۴) اگر تجزیه و تحلیل زیر گروه انجام شود، قابلیت اعتماد میزان بقاء با استفاده از همه اطلاعات در دسترس مربوط به بیماران بنابر دلیلی مشابه با مورد (۲) افزایش می‌یابد.

به عنوان مثال در کارآزمایی‌های مربوط به سرطان میزان مرگ و میر همانند میزان عود بیماری اغلب در طی دو تا سه سال اول بعد از عمل جراحی بالاتر است. گرچه در گزارش بقای ۵ ساله یا فوت دیررس که به عنوان درصد گزارش می‌شود، بیماری که بیش از ۲ سال زنده مانده است چیز خاصی را بیان نمی‌کند، در حالی که در جدول زندگی همین بیمار در فهرست‌بندی داده‌های ۳ سال اول شرکت خواهد داد. گزارش داده‌های مربوط به بقای ۵ ساله مرگ و میر، توزیع بیمار - رویکردها در طول زمان را به عنوان درصدی ساده از قلم می‌اندازد، اطلاعاتی که به طور بالقوه ارزشمندند، تنها به این دلیل که ممکن است طرح پیگیری را تحت تأثیر قرار دهد. بنابراین استفاده از اطلاعات مربوط به بقای هر بیمار اشتباهات استاندارد ناشی از داده‌های مربوط به بقاء را کاهش می‌دهد.

## راهنمایی

مواظب مطالعه‌ای که میزان‌های بقاء را هنگامی ارائه می‌دهد که پیگیری بیمار تا روز ارزیابی از نظر زمانی غیر کافی است، باشید. ممکن است مهمی وجود داشته باشند که در میزان‌ها گنجانده نشده باشند. اطلاعاتی که باید در روش جدول زندگی گنجانده شوند. کلیه بیمارانی که در جدول زندگی مشارکت می‌کنند، باید در روزهای اولیه شروع مطالعه طبقه‌بندی شوند به صورت (۱) زنده (۲) فوت شده (۳) حذف اختیاری یا (۴) حذف اجباری.

بیمارانی که از مطالعه کناره‌گیری می‌کنند (حذف اختیاری) معمولاً به عنوان کسانی در نظر گرفته می‌شوند که دقیقاً تا لحظه قبل از خروج به طور کامل در جدول زندگی شرکت داشته‌اند و به عنوان بیمارانی به حساب می‌آیند که یک دوم طول دوره زمانی را که طی آن خارج شده‌اند، در مطالعه شرکت داشته‌اند. نویسندگان باید تعداد بیمارانی را که از مطالعه کناره‌گیری کرده‌اند ثبت کند و دلیل یا دلایل این حذف اختیاری را در گزارش خود ارائه دهد.

از طریق تعدادی از روش‌ها می‌توان بر مشکل بیمارانی که از پیگیری کنار گذاشته شده‌اند، (حذف اجباری) غلبه کرد. محافظه‌کارانه‌ترین رویکردها این است که آنان را پس از آخرین پیگیری فوت شده تلقی کرد. گرچه این تصور در اغلب موارد در تناقض با واقعیت قرار می‌گیرد. تجارب ثبت شده مربوط به تحقیقات میدانی بر روی موارد از دست رفته به بهبودی برخی از بیماران منجر شده و نشان می‌دهد که این بیماران اغلب برای سال‌ها پس از آخرین تماس با آنان به زندگی خود ادامه داده‌اند. روشی که معمولاً انتخاب می‌شود این است که بیمارانی که از پیگیری کنار گذاشته شده‌اند، از تجربه بقاء همسانی نسبت به موارد باقی مانده در مطالعه، برخوردار بوده‌اند. این رویکرد ممکن است میزان بقاء را بیش از آنچه که باید باشد تخمین بزند. روش



دیگر در مورد برخورد با بیماران حذف شده از پیگیری، حذف کلیه این بیماران از جدول است. بیشتر اپیدمیولوژیست‌ها بر این مسئله که روش اخیر، اطلاعات در دسترس و بالقوه سودمندی را از دور خارج می‌کند، اتفاق نظر دارند. هر روشی که پیاده شود نویسنده باید تعداد بیمارانی که از پیگیری کنار گذاشته شده‌اند، چگونگی تعیین وضعیت آنان و چگونگی هدایت آنان در تحلیل داده‌ها را، بیان کند.

## راهنمایی

هرچند در هنگام تدوین جدول زندگی با بیمارانی که از پیگیری کنار گذاشته شده‌اند (حذف اجباری) و کسانی که خود کناره‌گیری کرده‌اند (حذف اختیاری) به شکلی یکسان برخورد می‌شود، اما تمایز قائل شده بین این دو گروه مهم است. اگر برای مثال تعداد زیادی از بیماران در مطالعه‌ای خاص از پیگیری کنار گذاشته شده باشند، ممکن است به واسطه تورش، نتیجه‌گیری کلی به نتیجه‌ای محدود تبدیل شود.

نویسندگان یک مطالعه ممکن است در جهت دستیابی به هدف دست به کنار گذاشتن برخی از بیماران بزنند. هر یک از بیماران که وضعیت آنها در طی دوره مشاهده تغییر کند یا کسی که دیگر حائز شرایط شمول در جدول نباشند، ممکن است در جهت رسیدن به هدف از پیگیری کنار گذاشته شوند. انجام این کار ممکن است نتیجه را تحت تأثیر تورش قرار دهد. علت کنارگذاری چنین بیمارانی باید بیان شود.

روش‌های جدول زندگی افزودن بر بقاء می‌توانند برای توضیح داده‌های مربوط به نتایج نیز به کار گرفته شوند. تجزیه و تحلیل رویدادهای غیرکنشده به وسیله روش‌های آماری هم مناسب است.

هنگام تفسیر جدول زندگی باید به تعداد نقاط داده‌های مشارکت‌کننده در منحنی در نقاط مختلف در طول محور Xها توجه شود. این نقاط بیانگر تعداد بیمارانی خواهد بود که با افزایش زمان ورود به منحنی به آن پیوسته‌اند. اگر این تعداد انتشار نیافته باشند، باید همچنان که منحنی از نقطه صفر دور می‌شود، ویژگی کمتری برای آن قائل شد. تعداد بیماران شرکت‌کنندگان در منحنی معمولاً در طول زمان کاهش می‌یابد، در نتیجه این امر منحنی را برای تخمین حقیقت غیرقابل اعتمادتر می‌گرداند. هنگامی که کمتر از ۱۰ نفر در یک نقطه در منحنی بقاء شرکت کرده باشند، معمولاً احساس می‌شود که باید به منحنی مورد بحث خاتمه داد. منازعات بسیاری در مورد مناسب بودن روش تجزیه و تحلیل جدول زندگی وجود داشته است که آیا منحنی‌ها باید در تمامیت خود مقایسه شوند یا مقایسه در نقاط انتخاب شده در طول زمان هم مناسب خواهد بود. روش دوم تجزیه و تحلیل وقتی که متخصص بالینی در مورد انتخاب‌های درمان با افراد بیمار بحث می‌کند، مناسب است، هر چند که ممکن است در زمان تعمیم نتایج از نمونه موری مطالعه به جمعیت عمومی قابل اجرا نباشد. روش تجزیه و تحلیل بقای کاپلان میبر (Kaplan-Meier) مشابه روش جدول زندگی می‌باشد. این روش محاسبه نسبت بیماران زنده در هر نقطه از زمان که مرگی اتفاق می‌افتد را به جای بقاء در مقطع زمانی حال انجام می‌دهد. این روش یک منحنی تدریجی را ایجاد می‌کند در حالی که روش جدول زندگی به یک خط منجر می‌شود.

## تجزیه و تحلیل تصمیم

تجزیه و تحلیل تصمیم، روشی برای مبادرت به طرح یک سؤال بهداشتی است وقتی که یک کارآزمایی بالینی غیرعملی بوده یا انجام آن با توجه به اهمیت سؤال مورد نظر پرهزینه باشد، این روش می‌تواند در ارتقاء اثبات یا حمایت از تصمیم‌گیری مبتنی بر احساس در زمانی که مشکل بالینی نادر بوده و در نتیجه سؤال آن با روش‌های معمول قابل پاسخ‌گویی نیست. مورد استفاده قرار گیرد. این روش به صورت یک روش متداول مورد استفاده طراحان خط مشی بهداشتی و نویسندگان کتاب‌های مرجع در آمده است. این روش تکنیکی بر پایه استدلال کمی و مدل‌های وابسته به احتمالات می‌باشد. روش تجزیه و تحلیل تصمیم شامل ایجاد و گسترش تدابیر می‌باشد و تأثیر عدم قطعیت بر روی نتایج این تدابیر را به وسیله تغییر تخمین‌ها (تجزیه و تحلیل حساسیت) می‌سنجد و یک نرم‌افزار کامپیوتری را مورد استفاده قرار می‌دهد که یک الگوریتم را که شاخه‌هایش تدابیر مختلف درمانی را توصیف می‌کنند، پدید می‌آورد. خطرها و منافع هر یک از این شاخه‌ها، زیر شاخه‌هایی را پدید می‌آورد و به نقطه پایان هر یک از زیر شاخه‌ها یک کاربردپذیری متصل بود. کاربردپذیری به صورت اولویت بیماران برای موقعیت‌های مختلف سلامتی می‌باشد و می‌تواند با استفاده از هر یک از تکنیک‌های یاد شده محاسبه شود.

در سال ۱۹۹۰ یکی از نویسندگان (J.O) از تجزیه و تحلیل تصمیم در مقایسه سه تدبیر مربوط به اداره بیمار تحت عمل جراحی ازوفازوگاسترکتومی به علت کارسینوم قسمت دیستال مری ناحیه کاردیا در معده استفاده کرد. سؤال این بود آیا برای همه، برخی یا هیچ یک از بیماران تحت عمل ازوفازوگاسترکتومی با ازوفازوگاسترکتومی برای درمان کارسینوم باید یک عمل درناژ در زمان عمل اول انجام شود یا خیر؟ مطالعات اندکی که به این سؤال پرداخته بودند، تا آن زمان دارای این معایب بودند تعداد کم بیماران، مورد مطالعه زمان‌های کوتاه پیگیری بالینی بیمار، استفاده از روش‌های مختلف قطع مری یا استفاده از اعمال مختلف درناژ. هدف تجزیه و تحلیل، کاهش نسبت بیمارانی بود که دارای مشکل عوارض چه به عنوان نتیجه انسداد راه خروج معده و چه به عنوان نتیجه مشکل در انجام درناژ بودند. الگوریتم سه شاخه داشت: (۱) برای همه بیماران درن گذاشته شده و خطرات مربوط به عوارض آن پذیرفته شود (۲) برای هیچ بیماری درن گذاشته نشود و خطر انسداد راه خروج معده پذیرفته شود (۳) در مورد هر بیمار خطر انسداد به طور جداگانه بررسی شده و در بیماران دارای میزان بالای خطر عمل برقرار کردن درناژ انجام شود. کلیه خطرات با استفاده از مطالعات در دسترس تخمین زده شد. سرانجام تصمیم گرفته شد که برای همه بیماران با خطر انسداد معده بیشتر از ۱۰٪، مادامی که عمل جراحی درناژ دارای اثر بخشی ۹۵٪ باشد، باید درناژ برقرار نمود، علاوه بر این تجزیه و تحلیل نشان داد که اگر آزمون برای طبقه‌بندی بیماران به گروه‌های دارای خطر پائین و میزان خطر بالای انسداد راه خروج شکل می‌یافت، اختصاصی بودن آن ۱۰٪ و حساسیت آن ۸۰٪ می‌شد. احتمال این که آزمون به خوبی این آزمون می‌توانست شکل بگیرد، اندک بود. این یافته‌ها از دورنمای خط مشی سلامتی نشان می‌داد که سرمایه‌گذاری در تحقیق برای ایجاد یک چنین آزمونی ممکن است نامعقول باشد.

تجزیه و تحلیل تصمیم می‌تواند هم در مسائل کلی بالینی و هم در تصمیم‌هایی که شامل بیماران منحصر به فرد می‌شود همانند مثال بالا بکار گرفته شود. این روش مناسب است هر چند تا جایی که وضع بالینی را تغییر دهد. بنابراین خواننده باید درکی پایه‌ای از مفاهیم زمینه‌ای ایجاد و توسعه تجزیه و تحلیل داشته باشد تا بتواند به نتایج آن اطمینان کند.

همبستگی می‌تواند به عنوان جایگزینی برای آزمون فرضیه‌ها بکار گرفته شود. هم چنین قدرت ارتباط خطی یا همراهی دو متغیر را توضیح می‌دهد. ضریب همبستگی بین ۱- و ۱+ قرار دارد. همبستگی اغلب برای توضیح ارتباط بین عوامل خطر و یک بیماری بکار می‌رود. یک ضریب همبستگی بین ۱- و صفر بیانگر یک ارتباط معکوس بین یک عامل و یک بیماری است (برای مثال استفاده بسیار کم از الکل و بیماری عروق کرونری)، در حالی که تغییر ضریب بین صفر و ۱+ نشان‌دهنده تغییر یک عامل در مسیری مشابه به عنوان خطر ایجاد بیماری است (مثلاً تعداد پاکت‌های سیگار در سال که بیمار مصرف کرده است و خطر ایجاد کارسینوم ریه).

## راهنمایی

یک همبستگی کوچک بین دو متغیر می‌تواند به این دلایل بستگی داشته باشد، (۱) همراهی اندک خطی بین دو متغیر یا (۲) اشتباهات فاحش در اندازه‌گیری متغیرها.

نوع ضریب همبستگی که در یک مقاله ویژه محاسبه می‌شود بستگی به نوع اطلاعاتی دارد که جمع‌آوری می‌شود.  $r$  (Pearson Product-Moment  $r$ ) برای به کمیت درآوردن قدرت ارتباط خطی بین متغیرهای متوالی (متناوب یا نسبی) از اطلاعات توزیع شده به صورت عادی استفاده می‌شود. Spearman Rank Order  $r$  (Spearman  $r$ ) برای داده‌های که نه از نوع متناوب هستند و نه از نوع نسبت و به صورت عادی توزیع نشده‌اند یا ارتباط خطی ندارند، بکار برده می‌شود. همبستگی قدرت ارتباط بین یک عامل خطر و یک بیماری و همچنین مسیر ارتباط آن را به کمیت تبدیل می‌کند. هر چند علت را بیان نمی‌کند. برای تعیین این که آیا ارتباط بین دو متغیر از نظر آماری شاخص است یا خیر؟ آزمون فرضیه استاندارد می‌تواند انجام شود. پرسشی که با انجام یک آزمون شاخص آماری باید به آن پاسخ داد این است که آیا  $r$  به طور شاخص از صفر به علت اتکاء بر شانس و یا به علت صفر بودن ضریب همبستگی در جمعیت حقیقی متفاوت است یا خیر؟

## راهنمایی

معمولاً احساس می‌شود که ضریب همبستگی باید با سطح شاخص‌های مربوطه گزارش شود. یک بار دیگر سؤال ارتباط بالینی به بصیرت خواننده واگذار می‌شود. همبستگی در واقع به گروه بزرگ‌تری از تکنیک‌های آماری با عنوان رگرسیون تعلق دارد. رگرسیون می‌تواند برای توضیح ارتباط بین دو یا چند متغیر بکار رود. یک خط یا منحنی را برای نقاط زوج داده متناسب می‌سازد. استفاده اصلی آن در توضیح ارتباط بین دو متغیر غیرخطی یا بین متغیرهای چندگانه می‌باشد. رگرسیون هم در مورد قدرت و هم در مورد بزرگی ارتباط بین متغیرها توضیح می‌دهد. تکنیک‌های رگرسیون اغلب در نوشته‌های جراحی برای توضیح مدل‌های پیش‌بینی کننده درمانی و تشخیصی بکار برده شده است. فرض‌های اصلی یک رگرسیون خطی ساده شامل موارد زیر می‌شود (۱) ارزش‌های متغیر غیر وابسته (X) به وسیله محقق تعیین می‌شود (۲) متغیر غیر وابسته (Y) باید بدون اشتباهات تجربی اندازه‌گیری شود (۳) برای هر متغیر وابسته زیر مجموعه‌ای از متغیرهای Y وجود دارد که به صورت عادی توزیع شده‌اند (۴) واریانس‌های متغیرهای Y باید همگن باشند (۵) متغیرهای Y که به صورت خطی توزیع شده‌اند و (۶) ارزش‌های Y مستقل از یکدیگرند.

بسیاری از تکنیک‌های رگرسیون توضیح داده شده‌اند که همه آنها شامل دستکاری متغیرهای چندگانه به طور همزمان هستند تا مشخص شود کدامیک نتیجه مورد نظر را بهتر پیش‌بینی می‌کند. این تکنیک‌ها شامل رگرسیون خطی منطقی، رگرسیون خطی لگاریتمی، رگرسیون چندگانه گام به گام و تحلیل با قابلیت تفکیک می‌باشد.

آزمون شاخص برای رگرسیون خطی از نظر مفهومی مشابه ANOVA می‌باشد. یک نسبت F برای یک P-value ایجاد می‌کند که می‌تواند با استفاده از جدول مناسبی یافته شود آزمون شاخص به این پرسش پاسخ می‌دهد که آیا شیب خط رگرسیون از نظر آماری متفاوت از صفر است یا خیر؟

## نتیجه‌گیری

در این مقاله سعی بر آن بوده است که آزمون‌های متداول آماری مورد استفاده در جراحی مورد بازنگری قرار گرفته شده و قدرتشان به همراه چالش‌های مربوط به استفاده و تفسیر آنان بررسی گردد و سعی بر آن بوده که این گفتار بتواند تفاوت بین شاخص بودن آماری و بالینی را نیز بیان کند. افزایش درک چگونگی استفاده از آمار در مجموعه داده‌ها و چگونگی تفسیر نتایج آزمون‌های آماری، کیفیت مطالعات جراحی را ارتقاء داده و از تعداد مطالعاتی که نتایج خود را بدون تکیه بر داده‌های ارائه شده گزارش می‌دهند، می‌کاهد.