

آزمون های تشخیصی و غربال گری

در خلال جنگ جهانی دوم و متعاقب اختراع رادار مسئله مطرح شده برای ارتش انگلیس کارآیی این وسیله در تشخیص درست وجود هواپیمای دشمن بر آسمان انگلستان بود و در این مورد دو سؤال اساسی مطرح بود : ۱- زمانی که رادار وجود هواپیمای دشمن را اعلام می کند چقدر احتمال دارد که هواپیمایی در آسمان دیده شود؟

۲- زمانی که رادار هیچگونه علامتی از وجود هواپیمای دشمن را نشان نمی دهد چقدر احتمال دارد که هواپیمایی در آسمان دیده نشود؟

برای پاسخ به سئوالات فوق روشی ابداع شد که بعدها این روش توسط علمای علم آمار ، اپیدمیولوژی و پزشکی کاربرد گسترده ای در تحقیقات پزشکی پیدا کرد تا بر اساس آن روش های گوناگون غربال گری یا تشخیصی از نظر کارآیی به محک گذاشته شود. اما چون در علم پزشکی کارآیی هر آزمایشی تحت تاثیر عوامل مختلف از جمله مرحله بیماری و گستره آن در جمعیت مورد مطالعه قرار می گیرد استفاده از این روش با مشکلاتی روبرو می شود. مثلا تست توبرکولین هفته ها بعد از مواجهه فرد با باسیل سل مثبت می شود یا ماه ها بعد از مواجهه فرد با HIV آنتی بادی در خون او قابل اندازه گیری خواهد بود. در هیپرپاراتیروئیدیسم سطح کلسیم سرم اغلب بالا می رود ولی چنین افزایشی از زمانی به زمان دیگر متفاوت است و بنابراین زمانی که سطح کلسیم خون بالا نیست نتیجه منفی کاذب بدست می آید و بر عکس زمانی که سطح کلسیم خون در فرد غیر مبتلا به هیپرپاراتیروئیدیسم (و مبتلا به مثلا کانسر ، سارکوئیدوز ، میلوم مولتیپل یا ...) بالا باشد نتیجه برای هیپرپاراتیروئیدیسم مثبت کاذب خواهد بود. با توجه به مطالب فوق بود که اصل زیر در بکار گیری این روش مورد موافقت همگان قرار گرفت:

جمعیتی که در آن آزمایش غربال گر یا تشخیصی مورد مطالعه قرار می گیرد

باید شاخص های مشابهی با جمعیتی که قرار است آزمایش بر روی آنان مورد

استفاده قرار گیرد داشته باشد. بر این اساس بررسی یک آزمایش بر روی مثلا کودکان یا مردان

نمی تواند در مورد بزرگسالان یا زنان کاربرد داشته باشد.

برای تعیین کارآیی یک آزمایش نیاز به یک معیار طلایی (Gold standard) است که بدون در نظر گرفتن وضعیت بیماری (مرحله و گستره آن) در افراد مورد مطالعه بتواند بیمار بودن یا نبودن آنان را تشخیص دهد. اغلب چنین منبع خارجی حقیقت (External source of truth) یک آزمایش دیگری است که دقیق تر بوده و اغلب تهاجمی تر می باشد (مثل کاتتریزاسیون قلبی یا بیوپسی بافتی).

با توجه به جدول صفحه بعد در مقایسه یک آزمایش با آزمایشی که بعنوان معیار طلایی بکار گرفته می شود ۹ شاخص بدست خواهد آمد :

| TRUE DISEASE STATUS | | | | |
|---------------------|----------|------------|--------------|----------------|
| | | Diseased | Non-diseased | Total |
| TEST RESULT | Positive | A | B | A+B |
| | Negative | C | D | C+D |
| | Total | A+C | B+D | A+B+C+D |

$$\frac{A}{(A+C)} = \text{Sensitivity}$$

$$\frac{D}{(B+D)} = \text{Specificity}$$

$$\frac{(A+C)}{(A+B+C+D)} = \text{Prevalence}$$

$$\frac{B}{(B+D)} = \text{False - positive - error - rate}$$

$$\frac{C}{(A+C)} = \text{False - negative - error - rate}$$

$$\frac{A}{(A+B)} = \text{Positive - predictive - value}$$

$$\frac{D}{(C+D)} = \text{Negative - predictive - value}$$

$$\frac{\frac{A}{(A+C)}}{\frac{B}{(B+D)}} = \frac{\frac{A}{B}}{\frac{(A+C)}{(B+D)}} = \text{Likelihood - ratio - positive (LR+)}$$

$$\frac{\frac{C}{(A+C)}}{\frac{D}{(B+D)}} = \frac{\frac{C}{D}}{\frac{(A+C)}{(B+D)}} = \text{Likelihood - ratio - negative (LR-)}$$

- ۱- حساسیت (Sensitivity): توانایی یک آزمایش در تشخیص بیماری در افراد بیمار
- ۲- اختصاصیت (Specificity): توانایی یک آزمایش در تشخیص صحیح افراد بدون بیماری
- ۳- میزان خطای مثبت کاذب (false-positive error rate or alpha error rate or Type 1 rate): نشان دهنده میزانی است که آزمایش مورد مطالعه می تواند افراد غیر بیمار را در طیف افراد بیمار قرار دهد. مثلا در بررسی کارایی اندازه گیری سطح کلسیم سرم درمقایسه با معیار طلایی برای تشخیص هیپرپاراتیروئیدیسم اگر FPER پنج درصد محاسبه شود به این معنی است که اندازه گیری سطح کلسیم در ۵٪ موارد می تواند افراد غیر مبتلا به هیپرپاراتیروئیدیسم را بیمار تشخیص دهد. هرچقدر میزان فوق زیاد باشد نیاز به استفاده از آزمایش دیگری که بیماری را رد کند بیشتر می شود.
- ۴- میزان خطای منفی کاذب (False-negative error rate or Beta error rate or Type 2 rate): نشان دهنده میزانی است که آزمایش مورد مطالعه می تواند افراد بیمار را در طیف افراد سالم

قرار دهد. مثلا اگر در مقایسه اندازه گیری سطح کلسیم با یک معیار طلایی برای تشخیص هیپرپاراتیروئیدیسم FNER چهل درصد محاسبه شود به معنی این است که ۴۰٪ افراد دارای بیماری با این آزمایش می توانند غیر بیمار تلقی شوند. هر چقدر میزان فوق زیاد باشد نیاز به آزمایش دیگری که تأیید کننده بیماری باشد بیشتر می شود.

۵- میزان شیوع بیماری (Prevalence) : که از تقسیم افراد واقعا بیمار به کل افراد مورد مطالعه حاصل می شود.

۶- ارزش پیشگویی کنندگی (ارزش اخباری) مثبت (Positive predictive value) : نشان دهنده این است که وقتی آزمایش فردی را بیمار اعلام می کند تا چه اندازه ی تشخیص صحیح بوده و او بیمار است.

۷- ارزش پیشگویی کنندگی (ارزش اخباری) منفی (Negative predictive value) : نشان دهنده این است که وقتی آزمایش فردی را غیر بیمار نشان داد تا چه اندازه ای احتمال دارد که او غیر بیمار باشد. دو شاخص فوق برای پزشکان از جذابیت خاصی برخوردار است ولی باید خاطر نشان شود که این دو شاخص دارای محدودیت های قابل توجه هستند. به عبارتی فقط در صورت نداشتن هیچ میزانی از خطای مثبت یا منفی کاذب در یک آزمایش ، که خود امر تقریبا محالی است ، می توان اطمینان ۱۰۰٪ به ارزش پیشگویی کنندگی مثبت یا منفی آن آزمایش داشت و از طرفی دیگر این دو شاخص شدیداً تحت تاثیر شیوع بیماری مورد مطالعه قرار دارند. با مثال زیر موضوع مشخص تر می شود:

الف : در بررسی کارایی یک آزمایش برای تشخیص بیماری دارای شیوع یک درصدی (شیوع نسبتاً نادر) در جامعه A و بر روی ۱۰۰۰۰ نفر یافته های زیر حاصل شده است:

| TRUE DISEASE STATUS | | | | |
|---------------------|----------|------|----------|-------|
| Test Result | | Sick | Not sick | Total |
| | Positive | 99 | 495 | 594 |
| | Negative | 1 | 9405 | 9406 |
| | Total | 100 | 9900 | 10000 |

Positive predictive value = $99/594 = 17\%$

ب : در بررسی کارایی همان آزمایش برای تشخیص همان بیماری دارای شیوع ۵٪ در جامعه B یافته های زیر حاصل می شود:

| TRUE DISEASE STATUS | | | | |
|---------------------|----------|------|----------|-------|
| Test Result | | Sick | Not sick | Total |
| | Positive | 495 | 475 | 970 |
| | Negative | 5 | 9025 | 9030 |
| | Total | 500 | 9500 | 10000 |

Positive predictive value = $495/970 = 51\%$

از مطالب فوق چنین بر می آید که اگر پزشکان در پی بررسی یک بیماری نادر باشند (در مطب برای بررسی یک بیمار یا در جامعه برای غربالگری مردم از نظر یک بیماری) باید خود را برای مواجهه با مثبت

کاذب بودن یافته های آزمایش مورد مطالعه آماده نموده و برای افرادی که دارای نتیجه مثبت هستند آزمایش تأیید کننده دیگری بکار گیرند. با این حال و علیرغم محدودیت فوق بکار گیری آزمایش های غربالگر با توجه به اینکه تعداد افراد یکه نیاز به آزمایش تأیید کننده خواهند داشت نسبت به کل جمعیت کم خواهد بود در جای خودش مفید می تواند باشد.

۸- درستنمایی مثبت یا نسبت امر محتمل برای آزمایش مثبت ($\text{likelihood ratio positive} = \text{LR}+$) با توجه به اجزای این نسبت یعنی حساسیت آزمایش (صورت کسر) و میزان خطای مثبت کاذب (منجر کسر) که هیچکدام به شیوع بیماری وابسته نیستند بنابراین $\text{LR}+$ نیز تحت تاثیر شیوع بیماری مورد مطالعه نخواهد بود و این مسئله چنین شاخصی را برای پزشکان مطلوب تر از دیگر شاخص ها می نمایاند. معنی $\text{LR}+$ عبارت است از شانس وجود آزمایش مثبت در افراد مبتلا به بیماری (حساسیت) به وجود آزمایش مثبت در افراد بدون ابتلا به بیماری (میزان خطای مثبت کاذب) . هرچقدر عدد حاصل از این نسبت بیشتر باشد نشان دهنده خوب بودن آزمایش مورد مطالعه است و برای اینکه آزمایشی یکی از بهترین ها باشد لازم است $\text{LR}+$ آن خیلی بیشتر از عدد یک باشد!

۹- درستنمایی منفی یا نسبت امر محتمل برای آزمایش منفی ($\text{Likelihood ratio negative} = \text{LR}-$) عبارت است از شانس وجود آزمایش منفی در افراد بدون بیماری (اختصاصیت) به شانس وجود آزمایش منفی در افراد دارای بیماری (میزان خطای منفی کاذب) . هرچقدر عدد حاصل از چنین نسبتی به صفر نزدیکتر باشد نشان از مناسب بودن آزمایش مورد مطالعه است. نهایتاً اگر $\text{LR}+$ یک آزمایش بزرگتر و $\text{LR}-$ آن کوچکتر باشد احتمالاً می توان آن آزمایش را برای بیماری مورد نظر آزمایش خوبی تلقی کرد! در بکار گیری یک آزمایش جدید ، با حساسیت و اختصاصیت معلوم و قابل قبول ، وجود نتایج مثبت کاذب (بیمار تلقی کردن افرادی که بیمار نیستند) از دل مشغولی های عمده محسوب می شود. برای کاستن از احتمال وارد شدن افراد سالم به گروه بیماران ناچار از بکار گیری یک آزمایش تأیید کننده دیگری ، با حساسیت و اختصاصیت بالاتر از آزمایش اول ، خواهیم بود. حال سؤال این است که آیا با بکارگیری این دو آزمایش می توان با اطمینان بالاتری از ورود افراد بدون بیماری به گروه بیماران جلوگیری کرد یا نه ؟ با مثال زیر موضوع را پی می گیریم :

الف : فرض کنید در جامعه ای شیوع بیماری A پنج در صد است و آزمایشی با حساسیت ۷۰٪ و اختصاصیت ۸۰٪ در ۱۰۰۰۰ نفر از جامعه فوق با یک آزمایشی بعنوان معیار طلایی مورد مقایسه قرار گرفته است. با توجه به جدول زیر از ۵۰۰ بیمار واقعی آزمایش ۳۵۰ نفر را مبتلا و از ۹۵۰۰ نفر سالم آزمایش ۷۶۰۰ نفر را بدون بیماری تشخیص می دهد. با این آزمایش ۱۹۰۰ نفری که واقعا سالم هستند بیمار بحساب می آیند.

| TRUE DISEASE STATUS | | | | |
|---------------------|----------|------------|-------------|-------------|
| Test Result | | Sick | Not sick | Total |
| | Positive | <u>350</u> | <u>1900</u> | <u>2250</u> |
| | Negative | 150 | 7600 | 7750 |
| | Total | 500 | 9500 | 10000 |

ب : حال آزمایش دیگری که دارای حساسیت ۹۰٪ و اختصاصیت ۹۰٪ در صدی است در همان جامعه و برای همان بیماری با همان معیار طلایی بکار می گیریم . در این صورت از ۵۰۰ بیمار واقعی ۴۵۰

نفر مبتلا تشخیص داده می شوند (۵۰ نفر مثبت کتذب) و آزمایش قادر خواهد بود از ۹۵۰۰ فرد سالم ۹۴۵۰ نفر را بدون بیماری تشخیص دهد .

| TRUE DISEASE STATUS | | | | |
|---------------------|----------|------------|------------------|------------|
| Test Result | | Sick | Not sick | Total |
| | Positive | 450 | <u>50</u> | 500 |
| | Negative | 50 | 9450 | 9500 |
| | Total | 500 | 9500 | 10000 |

ج: حال اگر بخواهیم فقط افرادی را که در آزمایش اول دارای نتیجه مثبت بودند ($350 + 1900 = 2250$) با آزمایش دوم ، که آزمایشی با حساسیت و اختصاصیت بالایی است ، مورد بررسی قرار دهیم تا وضعیت آنان از نظر داشتن یا نداشتن بیماری روشن شود یافته های زیر بدست خواهد آمد:

| TRUE DISEASE STATUS | | | | |
|---------------------|----------|-------------------|--------------------|--------------------|
| Test Result | | Sick | Not sick | Total |
| | Positive | 315 | <u>190</u> | 505 |
| | Negative | 35 | 1710 | 1745 |
| | Total | <u>350</u> | <u>1900</u> | <u>2250</u> |

$$\text{Net sensitivity} = 315:500 = 63\%$$

$$\text{Net specificity} = (7600+1710): 9500 = 98\%$$

با تقسیم تعداد افراد واقعا بیماری که آزمایش دوم نیز آنان را بیمار تشخیص داده (مثبت واقعی در آزمایش دوم=۳۱۵ نفر) به کل افرادی که واقعا بیمار هستند (۵۰۰ نفر) حساسیت مطلق (Net sensitivity) برای دو آزمایش فوق بدست می آید (۳۱۵ تقسیم بر ۵۰۰ برابر می شود با ۶۳٪) و با تقسیم مجموع افراد واقعا سالم در آزمایش اول (۷۶۰۰ نفر) و آزمایش دوم (۱۷۱۰ نفر) به کل افراد واقعا سالم (۹۵۰۰ نفر) اختصاصیت مطلق (Net specificity) برای دو آزمایش (۹۸٪) حاصل می شود که بر اساس آن فقط ۱۰ نفر دارای نتیجه مثبت کاذب خواهند بود. با توجه به اعداد بدست آمده مشاهده می شود که با بکارگیری دو آزمایش فوق و علیرغم کم شدن حساسیت مطلق ، اختصاصیت دو آزمایش در مجموع به اندازه ای افزایش یافته است که با اطمینان بسیار بالاتری نسبت به هر کدام از دو آزمایش مورد مطالعه (۹۸٪ یا ۱۰ نفر در مقابل ۸۰٪ یا ۱۹۰۰ نفر و ۹۰٪ یا ۵۰ نفر) می توان از ورود افراد سالم به گروه بیماران جلوگیری کرد.

Receiver Operating Characteristic Curve *ROC curve*

برای آزمایش‌هایی که متغیرهای کمی پیوسته مثل کلسیم سرم، قند خون یا فشار خون را اندازه‌گیری می‌کنند انتخاب نقطه برش (Cut-off point) مناسب اغلب دشوار است. اگر نقطه برش در سطح بالایی قرار داده شود با موارد مثبت کاذب کم و منفی کاذب بیشتری مواجه خواهیم شد و برعکس. هم‌چنین با توجه به اینکه چنین متغیرهایی از فردی به فرد دیگر، صرف نظر از بیمار یا سالم بودنشان، در نوسان است همپوشانی (overlap) مقادیر در افراد سالم و بیمار دیده شده و مسئله ساز خواهد بود.

به منظور تعیین نقطه برش مناسب از منحنی ROC استفاده می‌شود. برای اینکار دانستن دو شاخص حساسیت (Sensitivity) و میزان خطای مثبت کاذب (False-positive error rate) برای هر نقطه برش انتخاب شده لازم است. حساسیت یک آزمایش از تقسیم موارد مثبت واقعی (True positive) به مجموع موارد مثبت واقعی و منفی کاذب بدست می‌آید. میزان خطای مثبت کاذب نیز از تقسیم موارد مثبت کاذب به مجموع مثبت کاذب و منفی واقعی حاصل می‌شود. این میزان را با کم کردن اختصاصیت از عدد یک (1 - Specificity) نیز می‌توان بدست آورد. مثال زیر برای تعیین نقطه برش مناسب از طریق منحنی ROC کمک‌کننده خواهد بود:

۱۰۰۰ نفر از یک جمعیتی مورد بررسی قرار می‌گیرند تا مناسب‌ترین نقطه برش فشار خون در آن جمعیت تعیین شود بطوریکه بتوان با بکارگیری آن افراد دارای پرفشاری خون و افراد دارای فشارخون طبیعی را در آن جمعیت مشخص کرد. به این منظور نیاز به یک معیار طلایی برای تشخیص پرفشاری خون است. در این مثال چنین معیاری می‌تواند یک آزمایش تهاجمی باشد. حال با انتخاب نقاط برش گوناگون حساسیت و میزان خطای مثبت کاذب برای هر نقطه را محاسبه می‌کنیم. در این مثال دو شاخص فوق را برای فشار خون‌های ۱۲۰، ۱۳۰، ۱۴۰، ۱۵۰ و ۱۶۰ میلی‌متر جیوه محاسبه می‌کنیم (انتخاب چنین اعدادی کاملاً به بیماری مورد مطالعه وابسته است):

| TRUE DISEASE STATUS | | | | |
|---------------------------|----------|--------------|--------|-------|
| Test Result (120mm/Hg) | | Hypertensive | Normal | Total |
| | Positive | 380 | 360 | 740 |
| | Negative | 20 | 240 | 260 |
| | Total | 400 | 600 | 1000 |

Sensitivity=95% FPER=60%

| TRUE DISEASE STATUS | | | | |
|---------------------------|----------|--------------|--------|-------|
| Test Result (130mm/Hg) | | Hypertensive | Normal | Total |
| | Positive | 360 | 240 | 600 |
| | Negative | 40 | 360 | 400 |
| | Total | 400 | 600 | 1000 |

Sensitivity=90% FPER=40%

| TRUE DISEASE STATUS | | | | |
|---------------------|----------|--------------|--------|-------|
| Test Result | | Hypertensive | Normal | Total |
| (140mm/Hg) | Positive | 300 | 120 | 420 |
| | Negative | 100 | 480 | 580 |
| | Total | 400 | 600 | 1000 |

Sensitivity=75% FPER=20%

| TRUE DISEASE STATUS | | | | |
|---------------------|----------|--------------|--------|-------|
| Test Result | | Hypertensive | Normal | Total |
| (150mm/Hg) | Positive | 220 | 60 | 280 |
| | Negative | 180 | 540 | 720 |
| | Total | 400 | 600 | 1000 |

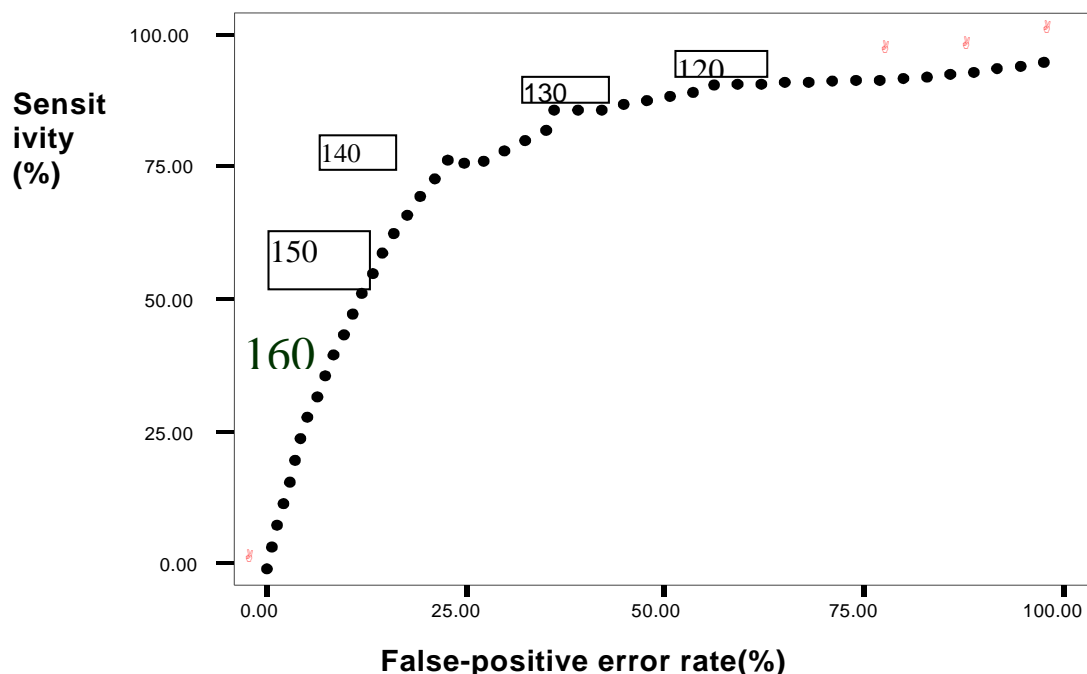
Sensitivity=55% FPER=10%

| TRUE DISEASE STATUS | | | | |
|---------------------|----------|--------------|--------|-------|
| Test Result | | Hypertensive | Normal | Total |
| (160mm/Hg) | Positive | 160 | 30 | 190 |
| | Negative | 240 | 570 | 810 |
| | Total | 400 | 600 | 1000 |

Sensitivity=40% FPER=5%

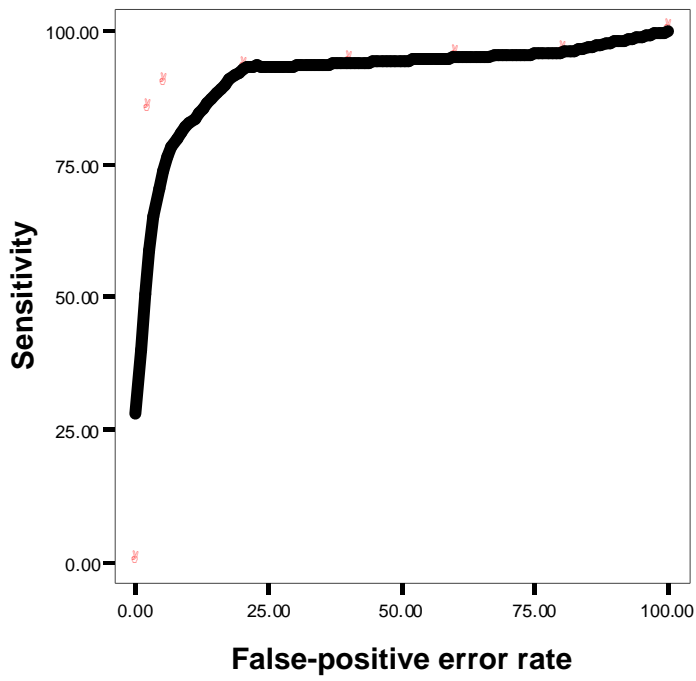
حال اعداد بدست آمده از حساسیت آزمایش در نقاط برش مختلف را در محور Y ها (محور عمودی) و اعداد مربوط به میزان های مثبت کاذب را در محور X ها (محور افقی) قرار می دهیم (چون طبق تعریف LR+ یک آزمایش از تقسیم حساسیت به میزان خطای مثبت کاذب آن حاصل می شود پس می توان منحنی ROC را نمودار LR+ نیز تلقی کرد). اما قبل از این کار باید نکته ای را مورد نظر قرار داد و آن اینکه اگر نقطه برش بر روی فشارخون صفر میلی متر جیوه قرار داده شود تمام افراد مورد مطالعه دارای پرفشاری خون تلقی شده و نیز تمام افراد دارای پرفشاری خون نیز تشخیص داده می شوند (حساسیت آزمایش برابر 100%) و از طرفی تمام افرادی که فشارخون طبیعی دارند نیز جزء افراد دارای پرفشاری خون محسوب می شوند (میزان خطای مثبت کاذب برابر 100%) چنین نقطه ای در محور فوق در بالاترین قسمت طرف راست (100%-100%) قرار می گیرد. اگر نقطه برش را بر روی فشارخون بسیار بالا (مثلا 500 میلی متر جیوه) قرار دهیم هیچ فرد دارای پرفشاری خون تشخیص داده نمی شوند (حساسیت آزمایش برابر

صفر) و از طرفی هیچ فرد دارای فشارخون طبیعی نیز بیمار تلقی نخواهد شد (میزان خطای مثبت کاذب برابر صفر) چنین نقطه ای در محور فوق در پائین ترین قسمت طرف چپ (0%-0%) قرار می گیرد.



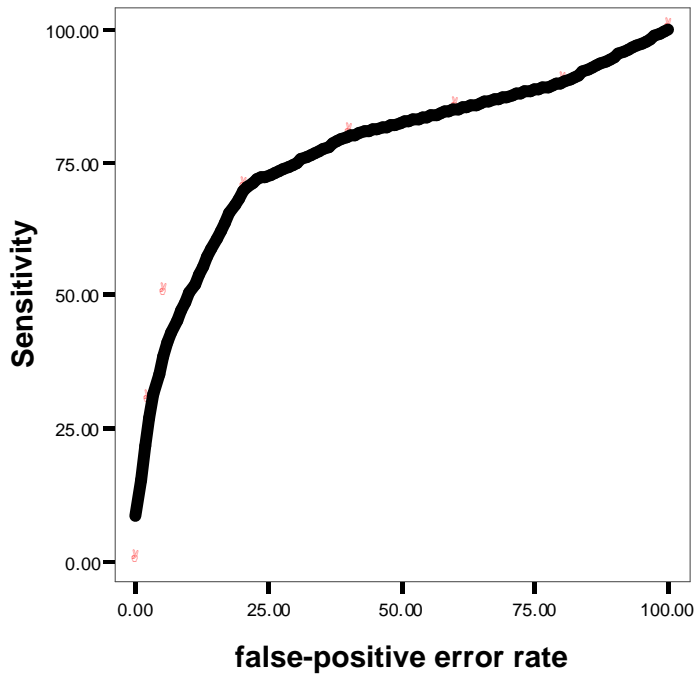
از آنجائیکه مناسب ترین نقطه برش نقطه ای است که بیشترین بیماران را تشخیص داده (دارای بیشترین حساسیت باشد) و کمترین حد از افراد دارای وضعیت طبیعی را بیمار تلقی نماید (کمترین میزان خطای مثبت کاذب داشته باشد) و چنین نقطه ایده آلی بندرت در آزمایش های مورد مطالعه بدست می آید (به مطلب پائین توجه شود) لذا با توجه به اهمیت بیماری مورد مطالعه نقطه ای که بطور نسبی دارای شرایط فوق است را انتخاب می کنیم. پس بر روی منحنی نقطه ای بعنوان نقطه برش فشار خون انتخاب می شود که به حد بالا و راست منحنی (نقطه دارای حساسیت ۱۰۰٪ و میزان خطای مثبت کاذب ۰٪) نزدیک تر باشد. چنین نقطه ای ۱۲۰ یا ۱۳۰ میلی متر جیوه خواهد بود.

منحنی ایده آل ROC برای یک آزمایش منحنی است که از پائین ترین نقطه محور Y ها بطور عمودی تا بالاترین نقطه آن دامه یافته و سپس بطور افقی و به موازات محور X ها امتداد یابد (نمودار شماره ۱). اگر حساسیت یک آزمایش برابر میزان خطای مثبت کاذب آن باشد منحنی از پائین ترین نقطه محور ها (نقطه صفر) شروع و بطور مستقیم تا بالاترین نقطه آن (نقطه تلاقی ۱۰۰٪ ها) ادامه پیدا خواهد کرد (نمودار شماره ۴). اگر آزمایشی چنین منحنی بدست دهد استفاده از آن آزمایش برای تشخیص بیماری مناسب نخواهد بود. منحنی ROC برای اغلب آزمایش های پزشکی بین دو منحنی فوق قرار می گیرند {نمودار شماره ۲ (نمودار مناسب) و نمودار شماره ۳ (نمودار نسبتا مناسب)}



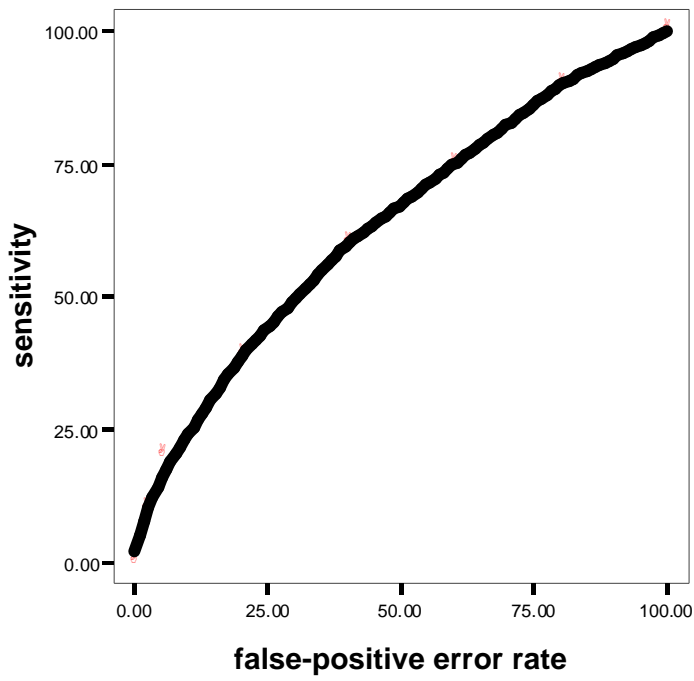
LLR Smoother

نمودار ۱ (نمودار عالی)



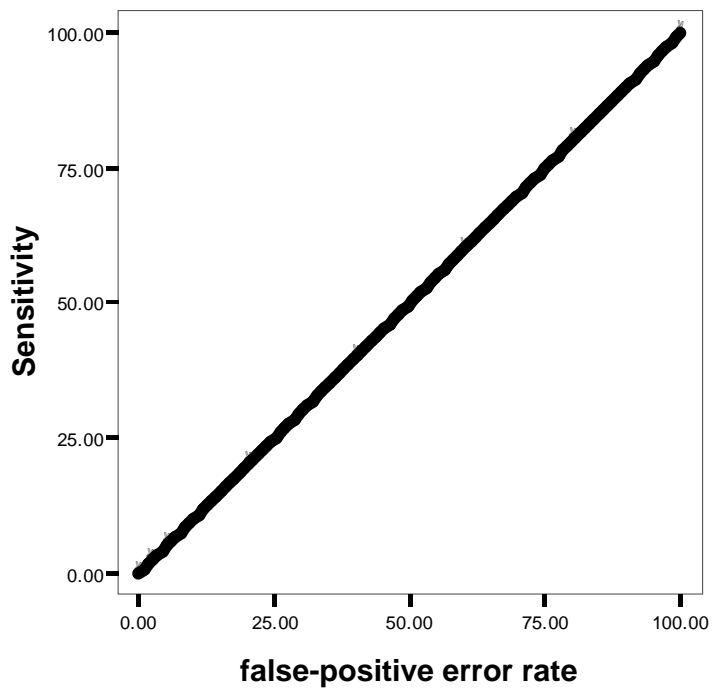
LLR Smoother

نمودار ۲ (نمودار مناسب)



LLR Smoother

نمودار ۳ (نمودار نسبتاً مناسب)



LLR Smoother

نمودار ۴ (نمودار نامناسب)

اندازه گیری درجه توافق

در تحقیقات پزشکی تعیین تکرار پذیری اندازه گیری ها و ارزیابی درجه نزدیکی مقادیر اندازه گیری شده اهمیت خاصی دارد. اگر کمیتی توسط یک نفر دوبار اندازه گیری شود و مقادیر حاصل مقایسه گردد شاخصی بنام اعتبار شخصی (Intrarater reliability) بدست می آید و اگر کمیتی توسط دو یا چند نفر اندازه گیری شده و مورد مقایسه قرار گیرد شاخص تغییرات بین گروهی (Interrater reliability) بدست خواهد آمد. در این نوشته نحوه اندازه گیری شاخص های فوق مورد بحث قرار می گیرد. اگر نتیجه معاینه جداگانه تعداد مشخصی از بیماران توسط دو پزشک را در جدول زیر خلاصه نمائیم :

| OBSERVER NO.1 | | | | |
|---------------|----------|----------|----------|---------|
| OBSERVER NO.2 | | Positive | Negative | Total |
| | Positive | A | B | A+B |
| | Negative | C | D | C+D |
| | Total | A+C | B+D | A+B+C+D |

می توان درصد مجموع A و D را بعنوان درصد توافق کلی و درصد مجموع B و C را بعنوان درصد عدم توافق کلی اعلام نمود اما در چنین حالتی مسایل زیر مطرح می شود :

۱- در صورتیکه درصد مجموع توافق مورد نظر قرار گیرد نمی توان به توافق مشاهده گرها در مورد تک تک بیماران پی برد.

۲- چون در اعلام درصد مجموع عدم توافق ، عدم توافق در اعلام نتیجه مثبت یا منفی پخش شده است و ممکن است یکی از مشاهده گرها بطور مستمر نتایج مثبت بیشتری نسبت به مشاهده گر دیگری اعلام نماید پس با چنین شاخصی نمی توان نحوه بروز عدم توافق را مشخص کرد.

۳- چون در اعلام نظرات موافق شانس و تصادف می تواند نقش داشته باشد درصد توافق کلی نمی تواند تعیین کند که در کدام قسمت از توافق (توافق در اعلام موارد مثبت یا منفی) چنین اتفاقی رخ می دهد. برای حل مسایل فوق از آماره کاپا (Kappa Test Ratio) استفاده می شود که با مثال زیر آنرا پی می گیریم :

فرض کنید که دو پزشک ۱۰۰ بیمار را جداگانه برای جستجوی سوفل قلبی مورد معاینه قرار داده و نتایج زیر حاصل شده است :

| OBSERVER NO.1 | | | | |
|---------------|----------------|----------------|---------------|-------|
| OBSERVER NO.2 | | Murmur Present | Murmur absent | Total |
| | Murmur present | 30 | 7 | 37 |
| | Murmur absent | 3 | 60 | 63 |
| | Total | 33 | 67 | 100 |

برابر مشاهدات فوق یافته های زیر بدست می آید:

۱- توافق مشاهده شده : ۳۰ بعلاوه ۶۰ مساوی ۹۰

۲- حداکثر توافق ممکن : ۱۰۰

۳- توافق مورد انتظار تصادفی (توافق که اگر شانس رخ می نمود اتفاق می افتاد) : چنین توافقی برای خانه های A (۳۰ نفر) و D (۶۰ نفر) و طبق روشی که برای آماره Chi-square بکار برده می شد (جمع سطر ضرب در جمع ستون تقسیم بر تعداد کل) قابل محاسبه است و بر این اساس توافق مورد انتظار تصادفی برای خانه A (۳۰ نفر) برابر می شود با ۳۳ ضرب در ۳۷ تقسیم بر ۱۰۰ یعنی ۱۲/۲ و برای خانه D (۶۰ نفر) برابر می شود با ۶۷ ضرب در ۶۳ تقسیم بر ۱۰۰ یعنی ۴۲/۲ .

۴- توافق کلی مورد انتظار تصادفی : از مجموع توافق مورد انتظار تصادفی دو خانه A و D حاصل می شود:
 $۱۲/۲ + ۴۲/۲ = ۵۴/۴$

۵- حال اگر سهم تصادف (شانس) را از توافق مشاهده شده (بند ۱) و حداکثر توافق ممکن (بند ۲) کم کنیم عملاً سهم شانس را حذف خواهیم کرد و با تقسیم دو عدد حاصل عدد کاپا , که یک نسبت است , بدست خواهد آمد:

$$۹۰ - ۵۴/۴ = ۳۵/۴$$

$$۱۰۰ - ۵۴/۴ = ۴۵/۴$$

$$کاپا = \frac{۳۵/۴}{۴۵/۴} = ۰/۷۸ = ۷۸\%$$

گستره عدد کاپا از منهای ۱ تا باضافه ۱ است و عدد ۱- نشان دهنده حداکثر عدم توافق و عدد ۱+ نشان دهنده حداکثر توافق است. عدد کاپا معمولاً با درصد نشان داده می شود و توافق تا ۲۰٪ غیر قابل اعتنا (قابل صرف نظر کردن) , از ۲۰٪ تا ۴۰٪ حداقل توافق , از ۴۰٪ تا ۶۰٪ توافق نسبتاً مناسب , از ۶۰٪ تا ۸۰٪ توافق مناسب و توافق از ۸۰٪ با بالا توافق عالی محسوب می شود. عدد کاپای حاصل از مثال فوق , ۷۸٪ , به این معنی است که با حذف شانس , بین دو پزشک مورد مطالعه توافق عالی برای تشخیص سופل قلبی وجود دارد!

محاسبه عدد کاپا بصورت کلی خود در زیر نشان داده شده است :

| OBSERVER NO.1 | | | | |
|---------------|----------|----------|----------|---------|
| | | Positive | Negative | Total |
| OBSERVER NO.2 | Positive | A | B | A+B |
| | Negative | C | D | C+D |
| | Total | A+C | B+D | A+B+C+D |

A+B=observed agreement (Ao)

A+B+C+D=maximum possible agreement (N)

(A+D) : (A+B+C+D) = overall percent agreement

{(A+B)(A+C)} : (A+B+C+D) = cell A agreement expected by chance(Ac)

{(C+D)(B+D)} : (A+B+C+D) = cell D agreement expected by chance(Dc)

Ac + Dc = total agreement expected by chance (Tc)

(Ao-Tc) : (N-Tc) = Kappa