

ارزیابی مطالعات درستی آزمون تشخیصی

دکتر علیرضا خاتمی

مرکز آموزش و پژوهش بیماری‌های پوست و جدام، دانشگاه علوم پزشکی تهران، ایران

نویسنده‌ی مسئول:

دکتر علیرضا خاتمی

تهران، خیابان طالقانی غربی، شماره‌ی ۴۱۵، کدپستی: ۱۴۱۶۶۱۳۶۷۵، پست الکترونیک: akhatami@tums.ac.ir

تعارض منافع: اعلام نشده است.

در سال‌های اخیر، پیشرفت فراوان شاخه‌های مختلف فناوری سبب شده تا استفاده از آزمون‌ها و روش‌های تشخیصی در رشته‌های مختلف پزشکی بالینی به گونه‌ای چشمگیر افزایش یابد. این امر افزایش قابل توجه انتشار مقالات مرتبط با استفاده از این آزمون‌ها در تشخیص و پیشگیری از بیماری‌ها را به دنبال داشته است و رشته‌ی تخصصی بیماری‌های پوست نیز از این امر مستثنی نیست. با توجه به آن که رویکرد مبتنی بر شواهد (Evidence-based) بر لزوم استفاده‌ی صحیح از بهترین شواهد حاصل از پژوهش‌های بالینی در اخذ مناسب‌ترین تصمیم برای تشخیص یا درمان بیماری‌ها با در نظر گرفتن خواسته‌های بیمار تأکید دارد، اهمیت توانایی متخصصان پوست در ارزیابی و استفاده از مقالاتی که نتایج پژوهش‌های بالینی در حیطه‌ی درستی آزمون‌های تشخیصی را گزارش می‌کنند مشخص می‌شود. به این منظور در این مبحث، پس از مروری اجمالی بر برخی از مفاهیم مورد استفاده در این گروه از مطالعات، اصول ارزیابی نقادانه مطالعات تشخیصی بر اساس رویکرد مبتنی بر شواهد ارائه می‌شود. به طور کلی راهبردهایی برای تعیین روایی (Validity)، کاربرد (Applicability) و ارزیابی کیفیت مطالعه مورد بحث قرار می‌گیرد.

کلیدواژه‌ها: درستی تشخیص، آزمون‌های تشخیصی، پزشکی مبتنی بر شواهد

دریافت مقاله: ۸۹/۱۰/۱ پذیرش مقاله: ۸۹/۱۱/۴

پوست و زیبایی؛ بهار ۱۳۹۰، دوره‌ی ۲ (۱): ۴۷-۵۹

مقدمه

بیماری‌های پوست نیز از این امر مستثنی نیست و از سوی دیگر بر اساس رویکرد مبتنی بر شواهد که بر لزوم استفاده‌ی صحیح از بهترین شواهد حاصل از پژوهش‌های بالینی در اخذ مناسب‌ترین تصمیم برای تشخیص یا درمان بیماری‌ها با در نظر گرفتن خواسته‌های بیمار تأکید دارد، اهمیت توانایی متخصصان پوست در ارزیابی و استفاده از مقالاتی که نتایج پژوهش‌های بالینی را گزارش می‌کنند مشخص می‌شود. به این منظور در این مبحث، پس از مروری اجمالی بر برخی از مفاهیم مورد استفاده در این گروه از مطالعات، اصول ارزیابی نقادانه مطالعات تشخیصی بر اساس رویکرد مبتنی بر شواهد (Evidence-based) ارائه می‌شود. به طور کلی راهبردهایی برای تعیین روایی (Validity)، کاربرد (Applicability) و ارزیابی کیفیت مطالعه مورد بحث قرار می‌گیرد.

آزمون‌های تشخیصی (diagnostic tests) گروهی از مداخلات هستند که به منظور افتراق افرادی که دارای وضعیتی خاص هستند، مانند یک بیماری، از کسانی که آن وضعیت را ندارند، به کار می‌روند. هدف نهایی به‌کاربردن یک آزمون تشخیصی شناسایی افرادی می‌باشد که از یک مداخله‌ی درمانی یا پیشگیری سود می‌برند (شکل ۱)^{۱،۲}.

با توجه به پیشرفت فراوان شاخه‌های مختلف فناوری در سال‌های اخیر، استفاده از آزمون‌ها و روش‌های تشخیصی در رشته‌های مختلف پزشکی بالینی به گونه‌ای چشمگیر افزایش یافته است. این امر سبب شده تا مقالات متعددی در خصوص استفاده از این آزمون‌ها در تشخیص و پیشگیری از بیماری‌ها منتشر شوند. با توجه به این نکته که رشته‌ی تخصصی

جدول ۱: جدول ۲×۲ نشان دهنده وضعیت بیماری و نتیجه یک تست تشخیصی با پیامد دوتایی.

مجموع	بیماری		نتیجه تست
	مثبت	منفی	
a+b	a (مثبت حقیقی)	b (مثبت کاذب)	مثبت
c+d	c (منفی کاذب)	d (منفی حقیقی)	منفی
a+b+c+d	a+c	b+d	مجموع

دارای تستی با نتیجه‌ی مثبت (غیر طبیعی) باشد.

- ویژگی (Specificity): عبارت است از احتمال آن که فردی که واقعاً بیمار نیست (سالم است)، آزمایشی با نتیجه‌ی منفی (طبیعی) داشته باشد.
- ارزش اخباری مثبت (Positive predictive value [PPV]): عبارت است از احتمال آن که فردی که دارای تستی با نتیجه‌ی مثبت (غیرطبیعی) باشد، واقعاً به بیماری مدنظر مبتلا باشد.
- ارزش اخباری منفی (Negative predictive value [NPV]): عبارت است از احتمال آن که فردی که آزمایش با نتیجه‌ی منفی (طبیعی) دارد، واقعاً به بیماری مدنظر مبتلا نباشد (سالم باشد).
- شانس (Odds): عبارت است از احتمال رخ دادن یک واقعه تقسیم بر احتمال رخ ندادن آن واقعه.
- احتمال بیماری پیش از انجام تست (Pre-test probability of disease): عبارت است از درست‌نمایی (Likelihood) تخمین زده شده در این خصوص که بیمار پیش از آن که با تست تشخیصی مورد ارزیابی قرار گیرد به بیماری مدنظر مبتلا باشد.
- شانس بیماری پیش از انجام تست (Pre-test odds of disease): عبارت است از احتمال تخمین‌زده‌شده در این خصوص که بیمار پیش از آن که با تست تشخیصی مورد ارزیابی قرار گیرد به بیماری مدنظر مبتلا باشد تقسیم بر احتمال آن که بیمار مورد بحث، بیمار نباشد.

• تست مرجع (Reference test): عبارت از بهترین روشی است که بودن یا نبودن وضعیت مورد نظر را تعیین کند. این روش با عنوان مرجع استاندارد (معیار) (Reference standard) یا معیار طلایی (Gold standard) خوانده می‌شود.

• تست مورد ارزیابی (Under-evaluation test) یا ایندکس (Index): همان‌طور که از نام آن مشخص است، به تستی اطلاق می‌شود که خصوصیات آن تحت ارزیابی قرار دارد.

• درستی (Accuracy): درستی یک تست تشخیصی عبارت است از میزان همخوانی (Agreement) بین اطلاعات به دست آمده از تست ایندکس و تست مرجع. راه‌های مختلفی برای بیان درستی یک تست تشخیصی وجود دارد.

لازم به ذکر است که برای ساده‌کردن مطلب، تنها به موارد پایه که در ساده‌ترین نوع تست‌های تشخیصی که وضعیت بیماری مورد نظر و نتیجه‌ی تست هر دو به‌صورت متغیرهای دوتایی (Dichotomous) توصیف شده‌اند، اشاره می‌شود (جدول ۱).

- مثبت حقیقی (True-positive): عبارت است از نتیجه‌ی مثبت تست در یک فرد هنگامی که آن فرد حقیقتاً به بیماری مورد نظر مبتلا باشد.
- مثبت کاذب (False-positive): به وضعیتی اطلاق می‌شود که نتیجه‌ی تست مثبت (غیرطبیعی) باشد در حالی که در حقیقت بیماری وجود نداشته باشد.
- منفی کاذب (False-negative): به وضعیتی اطلاق می‌شود که نتیجه‌ی تست منفی (طبیعی) باشد در حالی که در حقیقت بیماری وجود داشته باشد.
- منفی حقیقی (True-negative): عبارت است از نتیجه‌ی منفی تست در یک فرد هنگامی که آن فرد حقیقتاً به بیماری مورد نظر مبتلا نباشد.
- حساسیت (Sensitivity): عبارت است از احتمال آن که فردی که واقعاً به بیماری مبتلا است

$[LR+]$ به نسبت درست‌نمایی برای تستی با نتیجه‌ی منفی $[LR-]$.
 برای کمک به درک بهتر مطلب، در ابتدا نحوه‌ی محاسبه‌ی حساسیت، ویژگی، و ویژگی، PPV و NPV بر اساس اطلاعات موجود در جدول ۱ ارایه شده و سپس رابطه‌ی بین حساسیت، ویژگی و $LR+$ و $LR-$ بیان می‌شود.

$$\text{حساسیت} = \frac{a}{a+c} = \frac{\text{مثبت - حقیقی}}{\text{مثبت - حقیقی} + \text{منفی - حقیقی}}$$

$$\text{ویژگی} = \frac{d}{b+d} = \frac{\text{منفی - حقیقی}}{\text{منفی - حقیقی} + \text{مثبت - حقیقی}}$$

$$\text{ارزش اخباری مثبت (PPV)} = \frac{a}{a+b} = \frac{\text{مثبت - حقیقی}}{\text{مثبت - حقیقی} + \text{کاذب + مثبت - حقیقی}}$$

$$\text{ارزش اخباری منفی (NPV)} = \frac{d}{b+d} = \frac{\text{منفی - حقیقی}}{\text{منفی - حقیقی} + \text{کاذب + منفی - حقیقی}}$$

معمولاً، مقدار عددی محاسبه‌شده برای حساسیت، ویژگی، PPV و NPV در عدد ۱۰۰ ضرب می‌شود و حاصل آن به صورت درصد (%) بیان می‌شود.

$$\frac{\text{حساسیت}}{\text{ویژگی} - ۱} = \text{نسبت درست‌نمایی برای تست با نتیجه‌ی مثبت (LR+)}$$

$$\frac{\text{حساسیت} - ۱}{\text{ویژگی}} = \text{نسبت درست‌نمایی برای تست با نتیجه‌ی منفی (LR-)}$$

$$\frac{LR+}{LR-} = \text{نسبت شانس تشخیصی (DOR) (Diagnostic odds ratio)}$$

از جمله روش‌های ساده محاسبه‌ی احتمال بیماری پس از انجام تست بر اساس احتمال بیماری پیش از انجام تست در صورتی که نسبت بزرگ‌نمایی آن تست نیز مشخص باشد، استفاده از Fagan plot یا نوموگرام Bayes می‌باشد^۲. برای استفاده از این نوموگرام (شکل ۲) باید خط مستقیمی از نقطه‌ای که احتمال بیماری پیش از انجام تست را نشان می‌دهد به گونه‌ای ترسیم کرد که پس از قطع نقطه‌ای که روی خط نشان‌دهنده‌ی

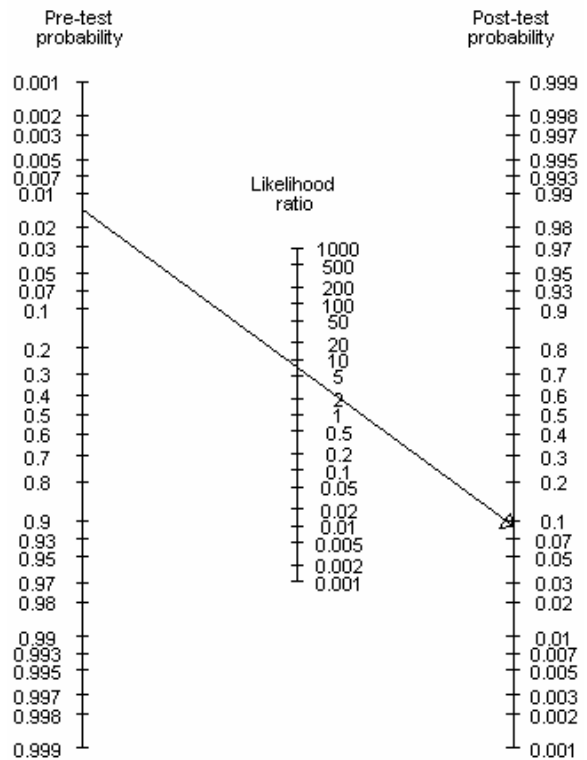
- احتمال بیماری پس از انجام تست (Post-test probability of disease): عبارت است از درست‌نمایی تخمین‌زده‌شده در این خصوص که بیمار پس از آن که با تست تشخیصی مورد ارزیابی قرار گیرد به بیماری مدنظر مبتلا باشد.
- شانس بیماری پس از انجام تست (Post-test odds of disease): عبارت است از احتمال تخمین‌زده‌شده در این خصوص که بیمار پس از آن که با تست تشخیصی مورد ارزیابی قرار گیرد به بیماری مدنظر مبتلا باشد تقسیم بر احتمال آن که بیمار مورد بحث، بیمار نباشد.
- نسبت درست‌نمایی (LR) (Likelihood ratio): عبارت است از حاصل تقسیم احتمال یک نتیجه‌ی مشخص از یک تست در بیماری که به بیماری مدنظر مبتلا می‌باشد بر احتمال همان نتیجه‌ی مشخص از آن تست در بیماری که به بیماری مدنظر مبتلا نمی‌باشد.
- نسبت درست‌نمایی برای تستی با نتیجه‌ی مثبت (Likelihood ratio for a positive test result ($LR+$): عبارت است از حاصل تقسیم احتمال یک نتیجه‌ی مثبت از یک تست در بیماری که به بیماری مدنظر مبتلا می‌باشد بر احتمال یک نتیجه‌ی مثبت از آن تست در بیماری که به بیماری مدنظر مبتلا نمی‌باشد.
- نسبت درست‌نمایی برای تستی با نتیجه‌ی منفی (Likelihood ratio for a negative test result ($LR-$): عبارت است از حاصل تقسیم احتمال یک نتیجه‌ی منفی از یک تست در بیماری که به بیماری مدنظر مبتلا می‌باشد بر احتمال یک نتیجه‌ی منفی از آن تست در بیماری که به بیماری مدنظر مبتلا نمی‌باشد.
- نسبت شانس تشخیصی (Diagnostic odds ratio [DOR]): عبارت است از حاصل تقسیم نسبت درست‌نمایی برای تستی با نتیجه‌ی مثبت

در بالین، لازم است وضعیت حقیقی بیمار (بیمار بودن یا بیمار نبودن) مشخص باشد. در وضعیت معمول که از ابتلاء یک فرد به بیماری مورد نظر اطلاعی در دست نیست، استفاده‌ی بالینی از این دو شاخص محدود می‌شود. حساسیت و ویژگی از خصوصیات ذاتی تست بوده و با شیوع بیماری در جمعیت مورد مطالعه ارتباطی ندارند.

۲. ارزش اخباری مثبت و منفی: ارزش اخباری یک تست نه تنها منعکس‌کننده‌ی حساسیت و ویژگی آن تست، بلکه تحت‌تأثیر شیوع بیماری مورد نظر در جمعیت مورد مطالعه نیز می‌باشد. وابستگی ارزش اخباری به شیوع بیماری مورد نظر از جمله محدودیت‌های استفاده از این شاخص می‌باشد.

۳. با توجه به محدودیت‌های بیان شده در خصوص استفاده از شاخص‌هایی چون حساسیت، ویژگی و ارزش اخباری مثبت و منفی برای ارزیابی درستی یک تست تشخیصی، به مزایای استفاده از دیگر شاخص درستی یک تست به نام "نسبت درست نمایی" (LR) اشاره‌ی مختصری می‌شود.

بر اساس تعریفی که در قسمت تعاریف و مفاهیم از "تست" ارائه شده است، مشخص می‌شود که در عمل، هدف از درخواست یک تست تشخیصی، افزایش اطلاعات درخواست‌کننده در مورد فردی است که تست برای او درخواست شده است. به عبارت دیگر وقتی یک تست تشخیصی ارزش درخواست‌کردن دارد که نتیجه‌ی آن احتمال وجود یا عدم وجود بیماری مورد نظر را در فرد مورد بررسی تغییر دهد. بدیهی است هرچقدر احتمال بیماری مورد نظر پس از انجام یک تست تشخیصی تغییر بیشتری کرده باشد، تست مذکور به تشخیص بیماری کمک بیشتری کرده است. اهمیت LR در آن است که با استفاده از این شاخص می‌توان با اطلاع از شیوع بیماری مورد نظر در جمعیت هدف (احتمال



شکل ۲: Fagan plot یا نمودار Bayes.

نسبت درست‌نمایی تست مورد نظر می‌باشد، خط احتمال بیماری پس از انجام تست را قطع کند.^۲

شاخص‌های درستی تست تشخیصی: محدودیت‌ها و مزایا

پس از مروری اجمالی بر اصطلاحاتی که به‌طور معمول در مطالعه‌های تشخیصی مورد استفاده قرار می‌گیرند، به ذکر محدودیت‌ها و مزایای خصوصیات برخی از شاخص‌هایی که برای بیان درستی یک تست تشخیصی به کار می‌روند، می‌پردازیم.^{۱۰-۶}

۱. حساسیت و ویژگی: این دو شاخص به ترتیب نسبت افراد دارای تست با نتیجه‌ی مثبت در افراد بیمار و نسبت افراد دارای تست با نتیجه‌ی منفی در افرادی که به آن بیماری مبتلا نیستند را نشان می‌دهند. همان‌طور که از تعریف ارائه‌شده برای هر یک از این دو شاخص آشکار است، برای استفاده‌ی مستقیم از این دو شاخص

حساسیت‌ها (محور عمودی) برحسب یک منهای ویژگی‌های (محور افقی) یک تست تشخیصی که دارای نقاط cut-off متعدد و به تبع آن حساسیت‌ها و ویژگی‌های متعددی برای تشخیص یک وضعیت خاص (مثلاً یک بیماری) می‌باشد، به کار می‌رود. در اینجا سطح زیر منحنی (Area Under the Curve [AUC]) نشان‌دهنده‌ی عملکرد (Performance) کلی تست می‌باشد و بزرگترین اندازه‌ی آن برابر ۱ (تست ایده‌آل) است. AUC برابر ۰/۵ نشانگر تستی است که به افتراق موارد دارای وضعیت مورد نظر و موارد فاقد آن کمک نمی‌کند. هرچه قدر AUC یک تست به عدد ۱ نزدیکتر باشد عملکرد تشخیصی کلی آن آزمایش بهتر است. به‌منظور پایبندی به سادگی مطلب از پرداختن بیشتر به این شاخص خودداری شده و به علاقه‌مندان به کسب اطلاعات بیشتر در این ارتباط پیشنهاد می‌شود به منبع ارایه‌شده مراجعه کنند.

محاسبه شاخص‌های درستی یک تست تشخیصی: یک مثال

در قسمتی از یک مطالعه تشخیصی، مقایسه‌ای بین توان تشخیص (Diagnostic performance) پزشکان عمومی براساس یافته‌های بالینی با روش تشخیصی استاندارد این بیماری که بررسی نتیجه‌ی کشت مواد برداشت شده با روش خراش از سطح ضایعه‌های مشکوک به درماتوفیتوز پس از ۴ هفته می‌باشد، انجام شد.^۷ در این مطالعه، پزشکان عمومی بر اساس یافته‌های بالینی در ۲۱ مورد از ۲۶ نفری که در آنها نتیجه‌ی کشت (تست استاندارد) مثبت بود، درماتوفیتوز را به‌درستی تشخیص دادند. هم‌چنین، پزشکان در ۵۵ مورد از ۱۲۲ موردی که در آنها نتیجه‌ی کشت منفی بود به‌درستی عدم وجود درماتوفیت را تشخیص دادند. در خصوص توان تشخیص بالینی پزشکان عمومی شاخص‌های زیر را محاسبه کنید:

وجود بیماری در فرد مورد بررسی پیش از انجام تست، احتمال وجود بیماری را در فرد مورد بررسی پس از انجام تست محاسبه کرد.

از دیگر مزایای استفاده از LR می‌توان به موارد زیر اشاره کرد:

- ◇ سهولت استفاده از LR: شانس بیماری پس از انجام تست با ضرب شانس بیماری پیش از انجام تست در LR به آسانی به دست می‌آید.
- ◇ در صورتی که مقدار عددی LR+ از ۱۰ بزرگتر یا مقدار عددی LR- از ۰/۱ کوچک‌تر باشد، به تشخیص بیماری (در صورت مثبت بودن جواب تست) یا رد بیماری (در صورت منفی بودن جواب تست) کمک بسیاری می‌کند.
- ◇ در صورتی که از چند تست برای تشخیص بیماری در فرد مورد بررسی استفاده شود، با ضرب LR آنها در یکدیگر و در شانس بیماری پیش از انجام تست می‌توان شانس بیماری پس از انجام تست را به دست آورد.
- ◇ استفاده از LR در تست‌هایی که نتیجه‌ی آنها دوتایی نیست امکان‌پذیر است.
- ◇ LR تحت تأثیر شیوع بیماری قرار نمی‌گیرد.

۴. نسبت شانس تشخیصی (DOR): می‌تواند

به‌صورت نسبت شانس موارد دارای نتیجه‌ی تست مثبت در گروه دارای بیماری نسبت به نسبت شانس موارد دارای تست مثبت در گروه فاقد بیماری یا نسبت شانس بیماری در افراد دارای تست مثبت نسبت به نسبت شانس بیماری در افراد دارای نتیجه‌ی تست منفی در نظر گرفته شود. برای اطلاع بیشتر از نقاط مزایا و محدودیت‌های این شاخص پیشنهاد می‌شود به مقاله‌ی Glas و همکاران مراجعه شود.^۵

۵. منحنی خاصه‌ی محرکه‌ی گیرنده (Receiver Operating Characteristics [ROC] curve) حاصل از ترسیم منحنی پراکنش (Scatter plot)

جدول ۲: سلسله مراتب شواهد بر اساس قوت مطالعات تشخیصی.

سطح شاهد	نوع مطالعه
1a	مرور نظاممند مطالعات همگون سطح 1b
1b	مطالعه‌ی هم‌گروهی روایی‌ده (Validating cohort study) با استاندارد طلایی خوب
1c	SPIns* و SNouts** مطلق
2a	مرور نظاممند مطالعات همگون سطح 2b
2b	مطالعه‌ی هم‌گروهی اکتشافی (Exploratory)
2c	پژوهش‌های پیامد (Outcome researches)
3a	مرور نظاممند مطالعات همگون سطح 3b
3b	مطالعاتی که در آنها ورود و ارزیابی موارد به‌دنبال هم نیست (Nonconsecutive) یا استاندارد طلایی به‌طور ثابت استفاده نشده است.
4	سری موارد (Case series) یا مطالعات با استاندارد طلایی ضعیف یا وابسته
5	نظر خبره بدون آنکه با ارزیابی نقادانه همراه باشد

* ویژگی (زیاد)، مواردی را که نتیجه‌ی تست مثبت دارند به‌درستی تشخیص می‌دهد. ** حساسیت (زیاد)، مواردی را که نتیجه‌ی تست منفی دارند به‌درستی از تشخیص کنار می‌گذارد.

۴. به‌کارگیری شواهد برای بیمار

۵. ارزیابی گام‌های قبلی

سلسله مراتب شواهد در مطالعات تشخیصی

سلسله مراتب شواهد حاصل از پژوهش در ارتباط با مطالعات تشخیصی از ابعاد مختلفی از جمله مدل اثربخشی (Efficacy model)، میزان آزادی مطالعه از سوگرایی (Bias-freeness) و قوت شواهد (Strength of evidence) تقسیم‌بندی می‌شوند (جدول ۲-۴) ۱۵-۱۳.

نکته‌های اساسی در ارزیابی نقادانه مطالعه‌های تشخیصی

سه نکته در ارزیابی نقادانه مطالعه‌های تشخیصی حائز اهمیت هستند ۱۶ و ۹.

- روایی (اعتبار [validity]) نتایج مطالعه
- نتایج مطالعه
- آیا نتایج مطالعه بر مراقبت از بیمار تأثیر دارد؟

جدول ۳: سلسله مراتب شواهد مطالعات تشخیصی بر اساس مدل اثربخشی.

سطح	اثربخشی مورد ارزیابی	مثال
۱	فنی (Technical)	کیفیت تصویری که توسط یک سامانه‌ی درماتوسکوپی دیجیتال ایجاد می‌شود.
۲	درستی تشخیصی (Diagnostic accuracy)	حساسیت و ویژگی
۳	تفکر تشخیصی (Diagnostic thinking)	نسبت درست‌نمایی مثبت یا منفی
۴	درمانی (Therapeutic)	تصمیماتی که بر اساس نتایج آزمون برای درمان/تغییر درمان بیماران گرفته می‌شود.
۵	پیامد [مدنظر] بیمار (Patient outcome)	کیفیت زندگی (Quality of life)
۶	اجتماعی (Societal)	هزینه - اثربخشی (Cost-effectiveness)

۱. موارد مثبت و منفی حقیقی
۲. موارد مثبت و منفی کاذب
۳. حساسیت و ویژگی
۴. نسبت درست‌نمایی تست با نتیجه‌ی مثبت و تست با نتیجه‌ی منفی
۵. نسبت شانس تشخیصی
۶. درستی (Accuracy) تشخیص پزشکان عمومی چند درصد است؟

رویکرد مبتنی بر شواهد به مطالعات تشخیصی

پایه‌های رویکرد مبتنی بر شواهد به مطالعات تشخیصی با آنچه در مورد این رویکرد به سایر مطالعات مطرح است تفاوتی ندارد. این پایه‌ها (گام‌ها) عبارت هستند از ۱۱ و ۱۲:

۱. طرح یک سؤال بالینی دارای ساختار مناسب و قابل پاسخ‌گویی
۲. جستجوی نظاممند برای یافتن بهترین شواهد حاضر
۳. ارزیابی نقادانه‌ی شواهد یافت‌شده

جدول ۴: سلسله مراتب شواهد مطالعات درستی تشخیصی بر اساس آزادی از سوگرایی.

سطح شاهد	نوع مطالعه
I	مرور نظام‌مند مطالعات سطح II
II	یک مطالعه‌ی درستی تشخیصی: که با استفاده از یک استاندارد طلایی روا مقایسه‌ای مستقل و بدون آگاهی [منظور از نتیجه‌ی هریک از دو تست است] در میان افرادی که به صورت پشت سرهم و با تظاهر بالینی تعریف شده‌ای داخل مطالعه می‌شوند، انجام دهد
III-1	یک مطالعه‌ی درستی تشخیصی: که با استفاده از یک استاندارد طلایی روا (Valid) مقایسه‌ای مستقل و بدون آگاهی (Blinded) [منظور از نتیجه‌ی هر یک از دو تست است] در میان افرادی که غیر پشت سرهم و با تظاهر بالینی تعریف شده‌ای داخل مطالعه می‌شوند، انجام دهد.
III-2	مقایسه‌ای که با استاندارد طلایی که فاقد معیارهای سطوح II و III-1 هستند.
III-3	مطالعات تشخیصی مورد - شاهدهی
IV	مطالعات دستاورد تشخیص (Diagnostic yield) (بدون استاندارد طلایی)

در قسمت بعدی هر یک از موارد فوق به اختصار توضیح داده می‌شود.

روایی نتایج مطالعه

برای ارزیابی روایی نتایج یک مطالعه لازم است به ارزیابی وجود سوگرایی (Bias) در آن مطالعه پرداخت. امکان رخداد سوگرایی‌های متعددی در یک مطالعه‌ی تشخیصی وجود دارد. در اینجا به ذکر مهم‌ترین این سوگرایی‌ها می‌پردازیم.^{۸-۱۰}

- یکی از موارد مهم در ارزیابی یک تست تشخیصی توجه به این نکته است که آیا تست مورد بررسی در حضور یک عدم اطمینان از تشخیص (Diagnostic uncertainty) انجام شده است یا خیر. بدیهی است که اگر عدم اطمینان از تشخیص وجود نداشته باشد (یعنی تشخیص صحیح قبل

از انجام تست مشخص باشد)، انجام تست و نتیجه‌ی آن تغییری در احتمال رسیدن به تشخیص صحیح ایجاد نمی‌کند. به بیانی دیگر اگر تست مورد ارزیابی تنها در بیمارانی که تشخیص صحیح در آنها دانسته شده باشد استفاده شود باعث می‌شود که احتمال موارد مثبت کاذب در افراد فاقد بیماری مورد نظر و موارد منفی کاذب در مبتلایان کمتر از مواردی که تست مورد نظر در بیمارانی که در آنها عدم اطمینان از تشخیص وجود داشته باشد به کار رود.

یک سؤال: به نظر شما استفاده از یک تست در گروهی از مبتلایان به یک بیماری که عدم اطمینان از تشخیص صحیح حداقل باشد، چه تأثیری بر روی حساسیت و ویژگی تست مورد اشاره می‌گذارد؟

- مورد دیگری که در ارتباط با گروهی که تست تشخیصی بر روی آنها انجام شده است باید مدنظر قرار گیرد آن است که تست بر روی گروهی از بیماران که با آنچه در عمل در بالین وجود دارد انجام شده باشد. در صورتی که طیف بیماران مورد ارزیابی با یک تست تشخیصی با آنچه در عمل وجود دارد یکسان نباشد می‌تواند نوعی از سوگرایی به‌نام سوگرایی طیف (Spectrum bias) را ایجاد کند. توجه به شدت بیماری افراد مورد بررسی با تست، امکان افتراق بیماری مد نظر با تست مورد بررسی از تشخیص‌های افتراقی آن و سطح ارجاع محل انجام تست از مواردی هستند که به ارزیابی مطالعه از نظر وجود این سوگرایی کمک می‌کنند. وجود سوگرایی در انتخاب بیماران می‌تواند توان تشخیص یک تست را تا سه برابر بیشتر از واقعیت نشان دهد.^۹
- درستی یک تست در تشخیص یک وضعیت هنگامی امکان دارد که نتایج حاصل از انجام آن

آنها متغیرهای Indicator و پیامد (Outcome) (در اینجا دو تست ایندکس و مرجع) را همزمان بررسی می کنند، این نوع سوگیری حداقل است.

- نوعی دیگر از سوگرایی که با نام Incorporation bias خوانده می شود، هنگامی رخ می دهد که تست ایندکس جزئی از استاندارد طلایی باشد. در حقیقت اطلاع از این امر که آیا انجام یا تفسیر نتایج تست مرجع یا ایندکس تحت تأثیر اطلاع از نتایج تست دیگر بوده است، حائز اهمیت فراوان است.
- از دیگر مواردی که توجه به آن در ارزیابی روایی یک مطالعه حائز اهمیت است توجه به این مطلب است که آیا تست های ایندکس و مرجع براساس دستورالعمل مدون و به طور یکسانی در طول مطالعه انجام شده اند. در حقیقت، آیا روش انجام هر تست با جزئیات کافی که بر اساس آن بتوان آن تست را انجام داد، توصیف شده است؟ به غیر از موارد ذکر شده، انواع دیگری از سوگرایی در مطالعه های تشخیصی مطرح هستند که در منابع مرتبط به خوبی مورد بحث قرار گرفته اند.^۱

نتایج مطالعه

هدف از درخواست یک تست تشخیصی افزایش احتمال تشخیص درست در مقایسه با آن احتمال قبل از اطلاع از نتیجه ی تست درخواست شده است. شکل ۱ به تبیین جایگاه درست درخواست یک تست تشخیصی کمک می کند.

در صورت روایی نتایج یک مطالعه ی درستی یک تست تشخیصی امکان آن وجود دارد که با توجه به نتایج آن مطالعه در مورد قابلیت تست مورد بررسی را افتراق میان افراد مبتلا به وضعیت مورد نظر و افرادی که به آن وضعیت مبتلا نیستند اظهار نظر کرد. برای این کار لازم است به ارزیابی شاخص های درستی (توان

با واقعیت مقایسه شود. در مبحث مطالعه های تشخیصی، نتایج حاصل از انجام تست "استاندارد طلایی" به عنوان واقعیت در نظر گرفته می شود. در صورتی که تستی با "استاندارد طلایی" مقایسه نشده باشد امکان ارزیابی شاخص های درستی تشخیص حاصل از انجام آن ممکن نخواهد بود. از سوی دیگر، برای حصول نتایج معتبر (Valid)، لازم است این مقایسه با رعایت برخی راهبردها (Strategies) انجام گیرد. یکی از نکاتی که توجه به آن ضروری است این است که آیا تمامی مواردی که با تست ایندکس ارزیابی شده اند با "استاندارد طلایی" نیز مورد بررسی قرار گرفته اند. از جمله راهبردهای اصلی که در تعیین روایی نتایج یک مطالعه ی تشخیصی مهم است می توان به کورسازی (Blinding) و ارزیابی مستقل (Independent) نتایج تست اشاره کرد. منظور از ارزیابی مستقل کور آن است که کسی که نتایج تست ایندکس را ارزیابی می کند از نتایج تست استاندارد طلایی آگاه نباشد و کسی که نتایج تست استاندارد طلایی را ارزیابی می کند نیز از نتایج تست ایندکس آگاه نباشد. این راهکار باعث به حداقل رساندن نوعی از سوگرایی با نام سوگرایی ناشی از ارزیابی (Work-up [review] bias) می شود. برخی از مطالعه ها نشان داده اند وجود این سوگرایی می تواند تا ۳۰٪ نسبت شانس تشخیصی یک تست را بیشتر از میزان حقیقی آن تخمین بزند.^۲

- در صورتی که مطالعه ای که در آن تست های مرجع و ایندکس بر روی بیماران انجام می شوند، در طول زمان انجام گیرد (Longitudinal باشد)، ممکن است منجر به پدید آمدن نوعی سوگرایی موسوم به سوگرایی ناشی از پیشرفت بیماری (Disease progression bias) شود. در مطالعه های مقطعی (Cross-sectional) که در

- نسبت سود (benefit) و زیان (harm)^{۱۱}، و
- موارد حائز اهمیت برای بیمار (Patient preferences)^{۱۱}

ارتقاء گزارش مطالعه‌های تشخیصی و STARD

تا حدود یک دهه‌ی قبل روش شناختی (Methodology) و نحوه‌ی گزارش بخش عمده‌ای از مطالعه‌های درستی تشخیص کیفیت پایینی داشت. این امر باعث شد تا همان‌گونه که برای ارتقاء نحوه‌ی انجام و کیفیت گزارش مطالعه‌های کارآزمایی بالینی تصادفی (Randomized controlled trials [RCTs]) دستورالعمل CONSORT (در حال حاضر CONSORT 2010 Statement) تکوین یافت^{۱۸}، دستورالعمل STARD که مخفف Reporting of Diagnostic Accuracy است، در سال ۱۹۹۹ تدوین شود تا به ارتقاء کیفیت گزارش مطالعه‌های درستی تشخیص کمک کند. تاکنون مقالات متعددی در مورد STARD و نحوه‌ی استفاده از آن در نشریات معتبر پزشکی منتشر شده‌اند که مطالعه‌ی آنها به علاقمندان کسب اطلاعات بیشتر در این مورد توصیه می‌شود^{۱۹-۲۱}.

ابزارهای ارزیابی کیفیت مطالعه‌های تشخیصی

ابزارهای مختلفی برای ارزیابی نظام‌مند کیفیت مطالعه‌هایی که به بررسی درستی تست‌های تشخیصی می‌پردازند وجود دارند. از جمله‌ی آنها می‌توان به QUADAS مخفف Quality Assessment of Diagnostic Accuracy Studies^{۲۲} و ابزار ارزیابی مطالعه‌های تشخیصی از ابزارهای ارزیابی نقادانه CASP مخفف Critical Appraisal Skills Programme استفاده کرد^{۲۳}. به منظور آشنایی بیشتر با ابزار ارزیابی مطالعه‌های تشخیصی CASP توصیه می‌شود این ابزار را که به‌طور رایگان در دسترس است بارگیری (Download) و موارد مطرح شده در آن را مطالعه

تشخیصی) تست مورد ارزیابی در مطالعه توجه شود. این شاخص‌ها پیش از این توصیف شده‌اند. به نظر شما حداقل شاخص‌هایی که لازم است اطلاعات مربوط به آنها در یک مقاله در خصوص درستی یک تست تشخیصی ذکر شده باشد، کدام‌ها هستند؟ با توجه به مطالب ذکر شده در خصوص $LR+$ و $LR-$ آیا استفاده از این شاخص‌ها را برای پی بردن به نتایج یک مطالعه در مورد درستی یک تست تشخیصی مناسب می‌دانید؟ پاسخ خود را توضیح دهید.

لازم به ذکر است که دقت (Precision) نتایج یک مطالعه‌ی تشخیصی به حجم نمونه‌ی مطالعه و عوامل تعیین‌کننده‌ی آن بستگی دارد و باید مورد توجه قرار گیرد. به‌عنوان مثال می‌توان به فاصله‌ی اطمینان (Confidence interval) که برای ویژگی یا حساسیت یک تست تشخیصی محاسبه شده است توجه کرد.^{۱۷}

آیا نتایج مطالعه بر مراقبت از بیمار تأثیر دارد؟

در ارتباط با این نکته توجه به موارد زیر اهمیت دارند^۸:

- به‌خاطر داشتن این مطلب که آیا انجام تست به تشخیص درست بیماری کمک می‌کند،
- پدید آمدن تغییر در اداره‌ی بیمار (Management) بر اساس نتیجه‌ی تست (شکل ۱ را به یاد داشته باشید)،
- وجود شباهت کافی بین بیمار ما با بیمارانی که در مطالعه‌ای که می‌خواهیم از نتایج آن استفاده کنیم، تست مورد نظر بر روی آنها انجام شده است (شباهت از نظر شیوع بیماری، خصوصیات مردم شناختی [Demographic] و خصوصیات بالینی بیماران از جمله شدت بیماری)،
- قابل انجام بودن تست مورد مطالعه در شرایطی که ما در آن به طبابت می‌پردازیم،
- در دسترس بودن تست مورد مطالعه،

روش‌های تشخیصی بررسی شده در این مطالعه بدانند. برای پاسخ به سؤالات ایشان، این مطالعه را با استفاده از ابزار ارزیابی مطالعه‌های تشخیصی CASP به‌طور نقادانه ارزیابی کنید.

با توجه به اهمیت ویژه‌ی مرورهای نظام‌مند در پزشکی مبتنی بر شواهد، پیشنهاد می‌شود خوانندگان گرامی به نمونه‌هایی از مقالات مرور نظام‌مند در ارتباط با مطالعات درستی تشخیص در درماتولوژی به مواردی مانند مرور نظام‌مند منتشرشده توسط Brennkmeijer و همکاران که در سال ۲۰۰۸ در نشریه‌ی درماتولوژی بریتانیا منتشر شده است، مراجعه کنند.^{۲۶}

گفتنی است، نگارنده به اتفاق دکتر Gorouhi این مبحث را در ویراست دوم کتاب درماتولوژی مبتنی بر شواهد تألیف Maibach and Gorouhi که انتظار می‌رود در اکتبر سال ۲۰۱۱ میلادی به بازار عرضه شود به‌طور کامل‌تری مورد بحث قرار داده‌اند.^{۲۷}

کنید.^{۲۴} در صورتی که علاقمند به کسب تجربه در مورد کاربرد این ابزار هستید از این ابزار در هنگام مطالعه و ارزیابی مطالعه‌های تشخیصی استفاده کنید.

یک تمرین: یکی از بستگان شما که پزشک عمومی و ۴۰ ساله‌ای است، با علایم بالینی مطرح‌کننده‌ی عفونت قارچی ناخن شست پای راست مراجعه کرده است. ایشان مقاله‌ی دکتر کریم‌زادگان‌نیا و همکاران^{۲۵} با عنوان "ارزیابی روش‌های مختلف تشخیصی در عفونت‌های قارچی ناخن" که در شماره‌ی ۲ هفتمین جلد فصل‌نامه‌ی بیماری‌های پوست منتشر شده و در آن به مقایسه روش‌های مختلف تشخیصی در عفونت‌های قارچی ناخن پرداخته شده را خوانده‌اند (متن کامل این مقاله از قسمت بایگانی (Archive) وب سایت فصلنامه‌ی بیماری‌های پوست به آدرس www.iranjd.ir به‌طور رایگان در دسترس است) و مایل هستند نظر شما را در مورد درستی تشخیصی

References

1. Dwamena B. Evidence-based radiology: Step 3- diagnostic systematic review and meta-analysis (critical appraisal). *Semin Roentgenol* 2009; 44: 170-9.
2. de Sousa MR, Ribeiro AL. Systematic review and meta-analysis of diagnostic and prognostic studies: a tutorial. *Arq Bras Cardiol* 2009; 3: 229-38.
3. Khatami A. Evaluation of diagnostic studies. In: Firooz A, Nassiri-Kashani M, editors. *Evidence-based dermatology*. Tehran, Mehr-e-Novin Publications; 2008: 9-25 (In Persian).
4. Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. *Med Edu* 2007; 41: 1140-5.
5. Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; 56: 1129-35.
6. Diagnostic testing. In: Greenberg RS, Daniels SR, Flanders WD, et al, editors. *Medical Epidemiology*. New York: Mc Graw-Hill; 2005: 92-106.
7. Last JM. *A dictionary of epidemiology*. 4th Ed. New York: Oxford University Press; 2000.
8. Whiting P, Rutjes AWS, Dinnes J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004; 8: 1-234.
9. Bhandar M, Guyatt GH. How to appraise a diagnostic test? *World J Surg* 2005; 29: 561-6.
10. Jaeschke R, Guyatt G, Lijmer J. Diagnostic testing. In: Guyatt G, Rennie D, editors. *Users' guides to the medical literature. A manual for the evidence-based clinical practice*. AMA Press 2002: 121-40.

11. Sackett DL, Straus SE, Richardson WS, et al. Evidence-Based Medicine: How to Practice and Teach EBM. London: Churchill Livingstone; 2000.
12. Firooz A, Khatami A. A review on evidence-based medicine. Iranian Journal of Dermatology 2004; 7: 101-11.
13. Kelly AM. Evidence-based radiology: Step 1-Ask. Semin Roentgenol. 2009; 44: 140-6.
14. Oxford Center for Evidence Based Medicine. Levels of Evidence. Available at: <http://www.cebm.net/index.aspx?o=1025>, Accessed: April 15, 2010.
15. Merlin T, Weston A, Tooher R. Extending an evidence hierarchy to include topics other than treatment: revising the Australian "levels of evidence". BMC Medical Research Methodology. 2009; 9:34.
16. McKibbin A. Diagnosis and screening. In: Maibach HI, Bashir SJ, McKibbin A. Evidence-based dermatology. 1st Ed., Hamilton: BC Decker Inc.; 2002: 31-40.
17. Altman DG, Bland JM. Statistical notes: Diagnostic studies 1: Sensitivity and specificity. BMJ 1994; 308: 1552-3.
18. CONSORT. Available at: <http://www.consort-statement.org>. Last accessed: December 15, 2010.
19. Bossuyt PM, Reitsma JB. The STARD initiative. Lancet 2003; 361-71.
20. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. Ann Intern Med 2003; 138:W1-W12.
21. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 2003; 326:41-4.
22. Whiting P, Rutjes AWS, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Research Method 2003; 3: 25.
23. Critical appraisal tools. Available at: http://www.phru.nhs.uk/casp/critical_appraisal_tools.htm. Last accessed: December 15, 2010.
24. Critical Appraisal Skills Programme (CASP). 12 questions to help you make sense of a diagnostic test study. Available at: http://www.phru.nhs.uk/Doc_Links/Diagnostic%20Tests%2012%20Questions.pdf. Accessed: December 15, 2010.
25. Karimzadegan-nia M, Mir Amin Mohammadi A, Firooz A, Shidfar MR. Evaluation of different diagnostic methods in onychomycosis. Iranian Journal of Dermatology 2004; 7: 95-100.
26. Breninkmeijer EEA, Schram ME, Leeflong MMG, et al. Diagnostic criteria for atopic dermatitis: a systematic review. Br J Dermatol 2008; 158: 754-65.
27. Khatami A, Gorouhi F. Studies and systematic reviews on diagnostic test accuracy. In: Maibach and Gorouhi, editors. Evidence-based dermatology. 2nd Ed., Hamilton: BC Decker Inc. (In press).

Evaluation of diagnostic test accuracy studies

Alireza Khatami, MD, MSPH

Center for Research and Training in Skin Diseases and Leprosy, Tehran University of Medical Sciences, Tehran, Iran.

In the recent years, advances in different fields of technology have been resulted in a dramatic increase in the diagnostic methods and tests in different branches of clinical medicine with a consequent increase in the number of scientific publications related to the application of those tests and methods in the diagnosis or prevention of diseases. Dermatology is not an exception. Since, evidence-based medicine emphasizes on the appropriate use of the current best clinical research evidence in the diagnosis or treatment of each individual patient acknowledging his/her values, the capability of dermatologists in appraisal and application of the publications, which reports the results of diagnostic test accuracy studies cannot be ignored. In this article, after a brief review on basic issues in diagnostic test accuracy studies, principles of critical appraisal of these studies, from an evidence-based viewpoint, will be discussed. General strategies for evaluation of validity, applicability and quality assessment of a diagnostic test study will be provided.

Keywords: diagnostic accuracy, diagnostic tests, evidence-based medicine

Received: Dec 22, 2010 Accepted: Jan 24, 2011

Dermatology and Cosmetic 2011; 2 (1): 47-59

Corresponding Author:

Alireza Khatami, MD, MSPH

No. 415, Taleqani Avenue, Tehran,
14166-13675, Iran.
Email: akhatami@tums.ac.ir

Conflict of interest: None to declare