التیام، دوره ۱۰، شماره ۲: ۲۳-۱۳، ۱۴۰۲ Eltiam 10 (2): 13-23, 2024

التیام
شاپا الکترونیکی:۲۷۸۳۳۲۹۱
eltiam.ivsa@yahoo.com
http://eltiamjournal.ir/

مقاله مروری

ارزیابی آزمون‌های تشخیصی و غربالگری در دامپزشکی

محمدآراد زندیه[۱]، فاطمه شیخیان[۲]، حمید شریفی[۳]، حسام‌الدین اکبرین[۴*]

۱- دانشجوی PhD اپیدمیولوژی، دانشکده دامپزشکی، دانشگاه تهران

۲- دانشجوی دکتری عمومی دامپزشکی، دانشکده دامپزشکی دانشگاه تهران

۳- استاد اپیدمیولوژی، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان

۴- استادیار بخش اپیدمیولوژی و بیماری‌های مشترک، گروه بهداشت و کنترل مواد غذایی، دانشکده دامپزشکی، دانشگاه تهران

*akbarein@ut.ac.ir

تاریخ دریافت: ۱۴۰۲/۹/۲۴ تاریخ پذیرش:۱۴۰۲/۱۰/۱۴

https://doi.org/10.61186/eltiamj.10.2.2

کپی‌رایت © مجله التیام: دسترسی آزاد؛ کپی‌برداری، توزیع و نشر برای استفاده کامل با ذکر منبع آزاد است، © نویسندگان. ناشر: انجمن جراحی دامپزشکی ایران.

چکیده

آزمون‌های غربالگری نوع خاصی از آزمون‌های تشخیصی هستند که در جمعیت به‌ظاهر سالم انجام می‌شود. هدف از انجام آزمون‌های تشخیصی تشخیص درست بیماران، تفریق حیوانات بیمار از حیوانات سالم، تشخیص موارد و کنترل‌ها و تشخیص موارد طبیعی و غیر طبیعی است. آزمون‌های غربالگری باید ساده، ارزان بودن، سریع و معتبر باشند. آزمون‌های تشخیصی و غربالگری به‌طور عمده در پایش بیماری‌ها مورد استفاده قرار می‌گیرند، از این‌رو آشنایی با اصطلاحات تخصصی مرتبط با ارزیابی این آزمون‌ها از جمله حساسیت، ویژگی، ارزش‌های پیشگویی مثبت و منفی، صحت و دقت و همچنین مفاهیمی مانند استاندارد طلایی به‌دلیل این‌که اغلب به‌جای یکدیگر استفاده می‌شوند و یا اشتباه تفسیر می‌شوند، اهمیت ویژه‌ای دارد. هم‌چنین با توجه به این‌که آزمون‌های استاندارد طلایی اغلب گران‌تر و زمان‌برتر هستند، از آزمون‌های تشخیصی ناقص استفاده می‌شود که در این شرایط با محاسبه‌ی حساسیت و ویژگی، محاسبه‌ی شیوع واقعی و ظاهری که توسط آزمون تشخیصی به‌دست آمده است؛ قابل محاسبه می‌شود، اما در صورتی که آزمون استاندارد طلایی موجود نباشد، ارزیابی آزمون‌ها از روش‌های دیگر از جمله آزمون شاخص کاپا استفاده می‌شود. یافته‌های تفسیر این آزمون‌ها رویکردی جامع و مبتنی بر شواهد را به کلینیسین‌ها و خبرگان ارائه می‌دهد که در نهایت به پایش دقیق‌تر، جامع‌تر، ارزان‌تر و سریع‌تر منجر می‌شود. در این مقاله مروری با بررسی کامل و جامع روی مفاهیم مرسوم در ارزیابی

آزمون‌های تشخیصی به‌همراه حل مثال‌هایی کاربردی به بسط و ارائه‌ی جامعی از این مفاهیم پرداخته خواهد شد. هم‌چنین آخرین مطالعه‌های اصیلی که در زمینه‌ی ارزیابی آزمون‌های تشخیصی انجام شده‌اند، نیز در این مقاله مرور خواهد شد.

واژه‌های کلیدی: آزمون تشخیصی، آزمون غربالگری، حساسیت، ویژگی، استاندارد طلایی

مقدمه

آزمون‌های غربالگری (Screening tests) کاربست یک آزمون تشخیصی در جمعیت به‌ظاهر سالم به منظور مشخص نمودن عفونت یا بیماری تحت بالینی است. هدف از انجام آزمون‌های غربالگری، پیدا کردن بیماری به هر شکلی (تحت بالینی و بالینی) در جمعیت به ظاهر سالم است؛ یعنی جمعیتی که ظاهراً هیچ نوعی بیماری از خود نشان نمی‌دهند. آزمون‌های غربالگری جزء پیشگیری سطح دوم طبقه‌بندی می‌شود و هدف پیدا کردن بیماری به هر شکلی در جمعیت به‌ظاهر سالم است تا بیماری در همان مراحل اولیه خودش را نشان دهد که از مثال‌های آن می‌توان استفاده از آزمون کالیفرنیایی ورم پستان (California Mastitis Test; CMT) برای تشخیص ابتدایی ورم پستان در گاوها، آزمون توبرکولین برای تشخیص سل در جمعیت گاوهای به ظاهر سالم، آزمون مالئین برای تشخیص مشمشه در اسب، یا آزمون یونین برای تشخص بیماری یونه، آزمون آگلوتیناسیون میکروسکوپیک (Microscopic Agglutination Test; MAT) برای تشخیص لپتوسپیروز و رزبنگال برای تشخیص بروسلوز را نام برد.

آزمون‌های غربالگری نوع خاصی از آزمون‌های تشخیصی هستند که در جمعیت به‌ظاهر سالم انجام می‌شود. هدف از انجام آزمون‌های تشخیصی تشخیص درست بیماران، تفریق حیوانات بیمار از حیوانات سالم، تشخیص موارد و کنترل‌ها و تشخیص موارد طبیعی و غیر طبیعی است. آزمون‌های غربالگری باید ویژگی‌های زیر را داشته باشند:

ساده باشند و نیاز به آموزش خاص و پیچیده‌ای نداشته باشند و افراد با کم‌ترین سطح دانش بتوانند این آزمون را انجام دهند. مثلاً نیاز به میکروسکوپ الکترونی نداشته باشد.

آزمون‌ها ارزان باشند، چون در جمعیت زیادی قرار است استفاده شوند.

سریع باشند، مثل آزمون توبرکولین برای تشخیص سل در گاو که ۷۲ ساعت زمان بین آزمون و قرائت نیاز دارد.

معتبر (valid) باشند.

ارزیابی اعتبار آزمون‌های تشخیصی

همواره دامپزشکان بالینی بر اساس آزمون‌های تشخیصی نسبت به تشخیص بیماری‌ها و شکلیت بیماران اقدام می‌کنند، در حالت ایده‌آل داده‌های تشخیصی حیوانات سالم و بیمار را به‌طور کامل از یکدیگر مجزا می‌کنند؛ به‌طوری‌که هیچ هم‌پوشانی با یکدیگر نداشته باشند، اما در واقعیت به‌علت نتایج مثبت و منفی کاذب، حیوانات سالم و بیمار با یکدیگر هم‌پوشانی دارند و دامپزشکان با قطعیت دست به داوری و اقدام نمی‌زنند. برای ارزیابی اعتبار یک آزمون غربالگری یا تشخیصی، نیاز به مقایسه آن با یک آزمون استاندارد طلایی (Gold Standard) وجود دارد. آزمون استاندارد طلایی باید یافته‌های قطعی و غیر قابل تغییر داشته باشد به بیان دیگر استاندارد طلایی روش قطعی تشخیص بیماری است که آزمونی است که بدون خطا بوده و حرف نهایی را در تشخیص می‌زند، مانند کشت یک باکتری در محیط اختصاصی خود. به‌عنوان مثال برای تشخیص بروسلوز ابتدا آزمون رزبنگال انجام می‌شود، اگر مثبت شد، سپس آزمون رایت و در نهایت یافته‌های مثبت آزمون رایت، توسط آزمون ۲-مرکاپتو اتانول (2ME) ارزیابی می‌شوند و اگر یافته این آزمون هم مثبت شد، قطعاً پاسخ مثبت است. مثلاً برای اندازه‌گیری طول یک ماژیک و یک مداد از یک وسیله‌ی اندازه‌گیری استاندارد که حداقل مساوی یا بزرگ‌تر از طول ماژیک یا خودکار باشد و به‌درستی مدرج شده باشد، باید استفاده شود. این خودکار یا ماژیک آزمون‌های غربالگری یا تشخیصی هستند

و وسیله استاندارد، همان آزمون استاندارد طلایی (Gold standard). در نتیجه آزمون استاندارد بررسی می‌کند که آزمون‌های غربالگری معتبر هستند یا خیر. آزمون‌های استاندارد طلایی هیچ‌وقت خطا یا یافته کاذب ندارند. علاوه بر کشت در محیط اختصاصی، یافته‌های پاتوگنومونیک هیستوپاتولوژی (مانند اجسام نگری در شاخ آمون یا هیپوکمپ مغز برای تشخیص هاری یا اجسام بولینجر در تشخیص آبله) نیز استاندارد طلایی محسوب می‌شوند، اما حال چرا اغلب از آزمون‌های ناقص (Imperfect tests) استفاده می‌شود؟ به این دلیل که اغلب آزمون‌های استاندارد طلایی نسبت به دیگر آزمون‌های تشخیصی گران‌تر و زمان‌برتر یا شرایط انجام آن وجود ندارد، هم‌چنین در موارد دیگری هم برای تشخیص بیماری استاندارد طلایی وجود ندارد. در این شرایط مفاهیمی همچون حساسیت، ویژگی، مثبت و منفی کاذب مطرح می‌شود (۳-۱).

محاسبه‌ی حساسیت و ویژگی راهی است برای کمی کردن توانایی آزمون تشخیصی که در یک جدول کلاسیک مقایسه آزمون غربالگری و تشخیصی با یک آزمون استاندارد طلایی، چهار خانه به وجود می‌آید (جدول شماره ۱):

اگر آزمون استاندارد طلایی مواردی که آزمون غربالگری مثبت تشخیص داد را تأیید کند؛ به آن مثبت واقعی (True Positive; TP) گفته می‌شود.

اگر آزمون استاندارد طلایی مواردی که آزمون غربالگری مثبت تشخیص داد را تأیید نکند؛ به آن مثبت کاذب (False Positive; FP) گفته می‌شود.

اگر آزمون استاندارد طلایی مواردی که آزمون غربالگری منفی تشخیص داد را تأیید نکند؛ به آن منفی کاذب (False Negative; FN) گفته می‌شود.

اگر آزمون استاندارد طلایی مواردی که آزمون غربالگری منفی تشخیص داد را تأیید کند؛ به آن منفی واقعی (True Negative; TN) گفته می‌شود.

نوع آزمون		آزمون استاندارد طلایی		جمع
		مثبت	منفی	
آزمون غربالگری/ تشخیصی	مثبت	مثبت واقعی (TP)	مثبت کاذب (FP)	TP+FP
	منفی	منفی کاذب (FN)	منفی واقعی (TN)	TN+FN
جمع		TP+FN	TN+FP	TP+FP+FN+TN

جدول ۱- یافته‌های آزمون‌های غربالگری/ تشخیصی و آزمون استاندارد طلایی

بر اساس جدول شماره (۱)، شاخص‌های حساسیت، ویژگی، شیوع واقعی، شیوع ظاهری و ارزش‌های پیشگویی مثبت و منفی قابل محاسبه هستند.

حساسیت (Sensitivity): قدرت یک آزمون تشخیصی در تشخیص حیوانات واقعاً بیمار است؛ یا به‌عبارتی چه تعدادی از حیوانات بیمار به‌وسیله آزمون غربالگری مشخص می‌شوند یا نسبت حیوانات واقعاً بیمار جامعه غربالگری که به‌وسیله آزمون غربالگری بیمار تشخیص داده شده‌اند که از روی فرمول شماره (۱) محاسبه می‌شود:

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$ فرمول (۱)

ویژگی (Specificity): قدرت یک آزمون در تشخیص درست حیوانات سالم یا چه تعدادی از حیوانات سالم به‌عنوان طبیعی (غیر بیمار) طبقه‌بندی می‌شوند. نسبت حیوانات واقعاً سالم که به‌وسیله آزمون غربالگری سالم تشخیص داده شده‌اند که از روی فرمول شماره (۲) محاسبه می‌شود:

$$\text{Specificity} = \frac{TN}{TN+FP}$$ فرمول (۲)

شیوع ظاهری (Appeared Prevalence; AP):

یعنی موارد مثبتی که توسط آزمون غربالگری تأیید می‌شود که از روی فرمول شماره (۳) محاسبه می‌شود:

$$AP = \frac{TP+FP}{TP+FP+FN+TN}$$ فرمول (۳)

شـیوع واقعی (Real Prevalence; RP): یعنی موارد مثبتی که توسط آزمون استاندارد طلایی طلایی (حساسیت و ویژگی آزمون استاندارد طلایی ۱۰۰ درصد است) تأیید می‌شود که از روی فرمول شماره (۴) محاسبه می‌شود:

$$RP = \frac{TP+FN}{TP+FP+FN+TN}$$ فرمول (۴)

رابطه بین شیوع واقعی و شیوع ظاهری بر اساس فرمول شماره (۵) محاسبه می‌شود:

$$RP = \frac{AP+SP-1}{Se+SP-1}$$ فرمول (۵)

اهمیت حساسیت و ویژگی

حساسیت و ویژگی یک آزمون بیان‌گر درستی آن آزمون است، اما یک دامپزشک آیا می‌تواند صرفاً بر اساس همین مفاهیم اقدام کند؟ آیا صرفاً بر اساس یافته‌های مثبت و منفی یک آزمون تشخیصی ناکافی می‌تواند نتیجه‌گیری کند؟ اهمیت این دو شاخص در این است که موارد بیمار و سالم را مشخص می‌کند. حال در این‌جا دو سؤال مطرح می‌شود که «احتمال بیمار بودن حیوانی که نتیجه آزمون غربالگری‌اش مثبت است؛ چه قدر است؟»و «احتمال سالم بودن حیوانی که نتیجه آزمون غربالگری‌اش منفی است؛ چه قدر است؟» دو شاخص حساسیت و ویژگی نمی‌توانند به این دو سؤال پاسخ بدهند و باید از شاخص‌های دیگری به نام ارزش‌های پیشگویی مثبت و پیشگویی منفی استفاده شوند.

پیش از بررسی ارزش پیشگویی باید با مفهوم دیگری به نام احتمال پیشین (Pre Test) آشنا شد. پیش از انجام آزمایش بر اساس شرح حال و و تجربه‌ی قبلی احتمال داشتن آن بیماری چقدر است؟ مثلاً احتمال ابتلا به یک بیماری ۲۰ درصد است. حال احتمالاتی که پس از انجام آزمون تشخیصی به دست می‌آید به آن احتمال پسین (Post test) گفته می‌شود که خود دو جزء ارزش اخباری مثبت و ارزش اخباری منفی دارد (۱).

ارزش پیشگویی (اخباری) مثبت (Positive Predictive Value; PPV): نسبت بیماران واقعی در بین حیواناتی که نتیجه آزمون غربالگری آن‌ها مثبت بوده است یا اگر نتیجه آزمون حیوانی مثبت شود؛ احتمال بیمار بودن آن حیوان چقدر است؟ یا احتمال مبتلا بودن حیوانی که نتیجه آزمون وی مثبت است. بنا به تعریف احتمال (تعداد حالت‌های مطلوب تقسیم بر حالت‌های ممکن) شاخص ارزش پیشگویی مثبت از روی فرمول شماره (۶) محاسبه می‌شود:

$$PPV = \frac{TP}{TP+FP}$$ فرمول (۶)

ارزش پیشگویی (اخباری) منفی (Negative Predictive Value; NPV): نسبت حیوانات سالم در بین حیواناتی که نتیجه آزمون غربالگری آن‌ها منفی بوده است یا اگر نتیجه آزمون حیوانی منفی شود، احتمال سالم بودن آن حیوان چقدر است یا احتمال مبتلا نبودن (سالم بودن) حیوانی که آزمون وی منفی است؛ چقدر است که از روی فرمول شماره (۷) محاسبه می‌شود:

$$NPV = \frac{TN}{TN+FN}$$ فرمول (۷)

بر مبنای فرمول‌های ۶-۱، جدول توافقی زیر یافته‌های آزمون تشخیصی الایزا را درمقابل استاندارد طلایی تشخیص بیماری یون (کشت مدفوع) در ۴۰۴ رأس گاو از یک دامداری نمایش می‌دهد که شیوع، حساسیت، ویژگی و ارزش های اخباری مثبت و منفی به صورت زیر محاسبه می‌شوند (۴).

		FECAL CULTURE Positive	FECAL CULTURE Negative	
ELISA	Positive (0.35)	102	40	142
ELISA	Negative (< 0.35)	38	224	262
		140	264	

جدول ۲- مقایسه ی آزمون الایزا با کشت مدفوع (۴)

$$\text{Specificity}=\frac{224}{224+40}=\frac{224}{264}=0.85$$

$$\text{Sensitivity}=\frac{102}{102+38}=\frac{102}{140}=0.72$$

$$\text{RP}=\frac{102+38}{102+40+38+224}=\frac{140}{404}=0.35$$

$$\text{AP}=\frac{102+40}{102+40+38+224}=\frac{142}{404}=0.35$$

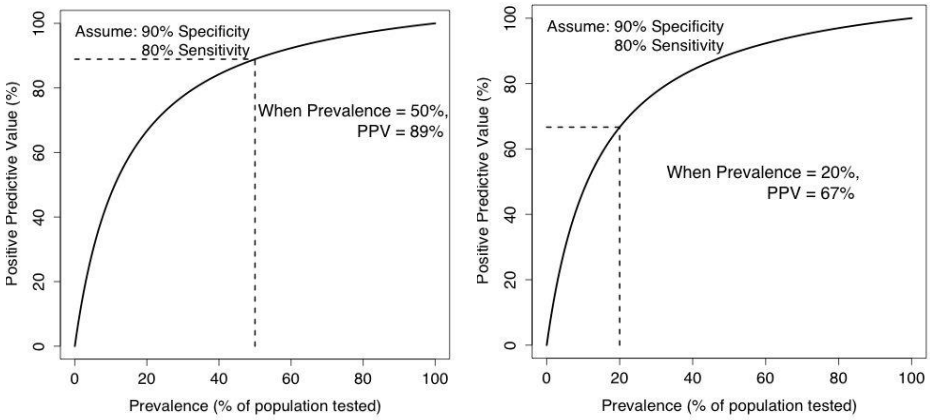
$$\text{NPV}=\frac{224}{224+38}=\frac{224}{262}=0.85$$

$$\text{PPV}=\frac{102}{102+40}=\frac{102}{142}=0.72$$

پرواضح است که چنان‌چه حساسیت و ویژگی از عدد ۱۰۰ کم شود، به‌ترتیب درصد منفی کاذب و درصد مثبت کاذب به دست می‌آید که در مثال یاد شده به ترتیب ۲۸ و ۱۵ درصد است.

عوامل مؤثر بر ارزش پیشگویی (اخباری): شیوع بیماری و ویژگی و حساسیت روی ارزش اخباری اثر می‌گذارند. شیوع در تعیین ارزش اخباری مهم تر است. چون در عمل نوسان حساسیت و ویژگی به ندرت بیش از ۲ برابر است، در حالی که شیوع بیماری می‌تواند دامنه نوسانات شیوع وسیع‌تر باشد.

رابطه شیوع و ارزش اخباری مثبت: بسته به شیوع ارزش پیشگویی مثبت تغییر می‌کند؛ هرچه شیوع بیشتر باشد، ارزش پیشگویی آزمون بیشتر است.



تصویر ۱– مثالی از رابطه ی ارزش اخباری مثبت با شیوع (۳)

اگر در جمعیت فرضی شیوع یک بیماری ۵۰ درصد و حساسیت و ویژگی با هم برابر و ۹۹ درصد باشد؛ ارزش‌های پیشگویی مثبت و منفی هر دو ۰/۹۹ خواهند بود (روش حل در پایین درج شده است).

$$\text{PPV}=\frac{0.50\times0.99}{0.5\times0.99+(1-0.99)(1-0.5)}=0.99$$

$$\text{NPV}=\frac{0.99\times(1-0.5)}{(1-0.99)\times0.5+0.99\times(1-0.5)}=0.99$$

و اگر شیوع بیماری به یک درصد برسد، با فرض ثابت ماندن حساسیت و ویژگی، ارزش‌های پیشگویی مثبت و منفی به ترتیب ۰/۵ و ۰/۹۹ خواهند بود (روش حل در پایین درج شده است).

$$PPV=\frac{0.01\times0.99}{0.01\times0.99+(1-0.99)\times(1-0.01)}=0.5$$

$$NPV=\frac{0.99\times(1-0.01)}{(1-0.99)\times0.01+0.99\times(1-0.01)}=0.99$$

با توجه به مثال بالا شیوع بیماری شدیداً بر PPV اثرگذار است.

محاسبه ارزش‌های پیشگویی مثبت و منفی زمانی که جدول توافقی ۲*۲ در دسترس نباشد: هنگامی که جدول توافقی در دسترس نباشد، نمی‌توان از روی فرمول‌های شماره (۶) و (۷) ارزش‌های پیشگویی مثبت و منفی را محاسبه کرد و باید از فرمول‌های شماره (۸) و (۹) محاسبه شوند:

فرمول (۸)

$$PPV=\frac{\text{شیوع}\times\text{حساسیت}}{(\text{شیوع}\times\text{حساسیت})+(\text{شیوع}-1)(\text{ویژگی}-1)}$$

فرمول (۹)

$$NPV=\frac{(\text{شیوع}-1)\times\text{ویژگی}}{(\text{شیوع}\times(\text{حساسیت}-1)+(\text{شیوع}-1)\text{ویژگی}}$$

بر اساس آمارنامه رسمی وزارت جهاد کشاورزی در خصوص عملیات مبارزه با سل گاوی در سال ۱۴۰۰، نسبت راکتور مثبت به تست ۰/۰۰۰۹ برآورد شده است (۵). اگر حساسیت و ویژگی آزمون توبرکولین بر اساس مطالعه خاوری و همکاران (۶) به‌ترتیب ۹۸ و ۷۲ درصد باشد، ارزش‌های پیشگویی مثبت و منفی به‌ترتیب ۳۱ درصد و ۹۹/۹۹ درصد به‌دست می‌آیند. (نسبت راکتور به تست معادل شیوع در نظر گرفته می‌شود)

$$PPV=\frac{0.98\times0.0009}{0.98\times0.0009+(1-0.72)(1-0.0009)}=31\%$$

یعنی احتمال ابتلا به سل در گاوهایی که به کشتارگاه اعزام می‌شوند، ۳۱ درصد است.

$$NPV=\frac{0.72\times(1-0.0009)}{(1-0.9991)\times0.0009+0.72\ (1-0.0009)}=99.99\%$$

یعنی ۹۹/۹۹ درصد گاوهایی که منفی هستند؛ واقعاً منفی هستند؛ یعنی NPV عدد خوبی است، اما PPV که احتمال مثبت بودن است؛ عدد پایینی (۰/۳۱) است.

نتایج آزمون‌ها در داده‌های پیوسته: زمانی که داده‌ها اعدادی باشند که پیوستگی دارند، باید برای نتیجه آزمون یک نقطه مرزی (cut off point)تعیین کرد که از چه نقطه‌ای به بعد سالم و چه نقطه‌ای به بعد بیمار است.

نقطه برش یا حد مرزی (cut off point): هرجایی که قرار بگیرد، سمت راست آن جمعیت بیمار و سمت چپ آن جمعیت سالم است. جمعیت مثبت کاذب زیر مجموعه مثبت واقعی است. جمعیت منفی کاذب زیر مجموعه منفی واقعی است.

بر اساس مطالعه سخا و همکاران آزمون BHB (بتا هیدروکسی بوتیریک اسید) برای تشخیص کتوز در گاوها استفاده می‌شود که در واقع BHB≥۱/۵ را بیمار تشخیص می‌دهند و ۱/۵ BHB کرا سالم یا غیر بیمار تشخیص می‌دهند و سیستم آزمون طلایی مانند جدول ۲×۲ گفته شده است (۷).

	Disease +	Disease –
Test + 'BHB ≥ 1.5'	True Positive	False Positive
Test – 'BHB < 1.5'	False Negative	True Negative

جدول ۳ – جدول طراحی شده با توجه به مطالعه‌ی سخا و همکاران (۷)

در BHB نقطه‌ی مرزی ۱/۵ است. اگر نقطه‌ی برش به سمت راست برده شود، نتیجه‌ی آزمون تغییر می‌کند؛ یعنی اگر نقطه‌ی مرزی از ۱/۵ به ۱/۷ تغییر داده شود، به همین نسبت حساسیت کاهش و ویژگی افزایش می‌یابد.

برآورد شیوع بر پایه آزمون‌های ناقص

آزمون ناقص (Imperfect test)، آزمونی است که حساسیت و ویژگی آن ۱۰۰ درصد نیست. حساسیت و ویژگی آزمون استاندارد طلایی ۱۰۰ درصد و بقیه آزمون‌ها، ناقص تلقی

می‌شوند. چنان‌چه شیوع بیماری با استفاده از آزمونی با حساسیت و ویژگی ۸۰ درصد برابر ۲۳ درصد به‌دست آید .شیوع واقعی این بیماری بر اساس فرمول شماره (۵) عبارت است از:

$$RP = \frac{AP + SP - 1}{Se + SP - 1}$$

$$RP = \frac{0.23 + 0.8 - 1}{0.8 + 0.8 - 1} = 0.05$$

درستی یا صحت (Accuracy): درستی یک آزمون وابسته به توانایی آن در ارائه اندازه واقعی و صحیح در یک آزمون است . ممکن است مطالعه‌ای که خطای سیستماتیک کوچکی داشته باشد (مثلاً دستگاه کالیبر هنباشد) در نتیجه دارای درستی بالایی شود. درستی تحت تأثیر تعداد نمونه قرار نمی‌گیرد. درستی یک آزمون وابسته به دو مشخصه اعتبار یا روایی (Validity) و پایایی (Reliability) وابسته است. درستی شاخص قابل استنادی نیست. اعتبار، توانایی یک تست در تعیین حیوانات بیمار و سالم است که دارای دو جزء حساسیت و ویژگی است. پایایی در برگیرنده قابلیت تکرار (Repeatability) و دقت (Precision) است. قابلیت تکرار، توانایی یک آزمون در ارایه نتایج ثابت (مشابه و نه یکسان) در تکرار آزمایش است. ارائه نتایج ثابت روی یک نمونه در یک آزمایشگاه نشان دهنده کالیبر بودن دستگاه و عملکرد درست آن است. در ذیل فرمول صحت درج شده است.

صحت (احتمال صحیح بودن نتیجه) = (TN + TP) / Total

تکرارپذیری (Reproducibility) به‌معنای ارائه نتایج ثابت روی یک نمونه در آزمایشگاه‌های مختلف است. دقت (Accuracy) میزان انحراف نتایج با مقدار واقعی است و درجه انحراف یک مجموعه آزمون‌ها حول وحوش یک اندازه مرکزی (Standard Error; SE) گفته می‌شود.

استراتژی‌های تشخیص

آزمون‌های چندتایی یا آزمون‌های پی در پی (Serial testing): درشرایط بالینی که ارزیابی سریع بیماران لازم نیست یا وقتی که آزمون‌ها گران یا پر مخاطره باشند. معمولاً از آزمون‌های پی در پی استفاده می‌شود مثل تست بروسلوز برای مبارزه با این بیماری در ایران. ابتدا تست رزبنگال انجام می‌شود سپس روی رزبنگال مثبت‌ها تست رایت انجام می‌شود ودر نهایت روی رایت مثبت‌ها، آزمون 2ME انجام می‌شود. ارزیابی هر آزمون با توجه به یافته‌های آزمون قبلی است. این آزمون‌ها معمولاً وقت‌گیرتر هستند. در این آزمون‌ها ویژگی و ارزش اخباری افزایش می‌یابد، اما حساسیت و ارزش اخباری منفی کاهش می‌یابند. نتایج مثبت دال بر بیماری است، اما خطر نادیده گرفته شدن بیماری افزایش می‌یابد.

فرمول‌های ۱۰ و ۱۱، محاسبه حساسیت و ویژگی در آزمون‌های پی در پی A و B:

$$Se = Se_A \times Se_B$$

$$SP = 1 - (1 - SP_A) \times (1 - SP_B)$$

آزمون‌های موازی (Parallel tests): وقتی که ارزیابی سریع لازم باشد؛ از این روش استفاده می‌شود؛ مثلاً برای بیماری‌یابی سریع بروسلوز هر دامی که علامت بالینی بروسلوز داشته باشد یا سقط داشته باشد یا تست سرولوژی‌اش مثبت باشد، «بیمار» تلقی می‌شود و مانند تست پی در پی لازم نیست تأیید مرحله‌ای داشته باشد و اگر مثلاً از ۳ آزمون، یک آزمون هم مثبت شود؛ کفایت می‌کند. در آزمون‌های موازی، حساسیت و ارزش اخباری منفی افزایش می‌یابند، اما ویژگی و ارزش اخباری مثبت کاهش می‌یابد و در غربالگری با احتمال کم‌تری بیماری از نظر دور می‌ماند، اما احتمال تشخیص مثبت کاذب افزایش می‌یابد. در انگلستان برای ریشه‌کنی بروسلوز از این روش استفاده می‌شود.

فرمول‌های ۱۲ و ۱۳، محاسبه حساسیت و ویژگی در آزمون‌های موازی A و B:

$$Sp = Sp_A \times Sp_B$$

$$Se = 1 - (1 - Se_A) \times (1 - Se_B)$$

تکرار آزمون (Repeat testing, Herd retest):

نوع تغییر یافته آزمون آزمون پی درپی است که حیوانات با جواب منفی باید با همان آزمون در فاصله زمانی معینی دوباره آزمون شوند. اگر شیوع بیماری در جمعیت کاهش یابد افزایش ویژگی تست مورد استفاده مهم است. مثلاً حیوان اگر مشکوک به سل باشد؛ ۷۰ روز بعد دوباره آن را تست می‌کنند. در این حالت درصد مثبت کاذب افزایش می‌یابد؛ به‌ویژه هنگامی که حیوانات بیمار به کشتارگاه فرستاده می‌شوند. تکرار آزمون نوعی برنامه

تست وکشتار است که طراحی شده است تا بیماری را تا حد خاصی کنترل حذف و ریشه کن کند به مانند برنامه کنترل سل، بروسلوز و مشمشه.

ارزیابی آزمون‌ها وقتی که معیار طلایی وجود ندارد

برخی باکتری‌ها دیر رشد هستند (مانند مایکوباکتریوم‌ها) و برخی از آن‌ها سخت رشد هستند (مانند لپتوسپیراها). برای این باکتری‌ها و موارد دیگر استفاده از آزمون استاندارد طلایی (کشت) بسیار زمان‌بر و سخت است و به دلیل لزوم آغاز زودهنگام مداخله درمانی، نمی‌توان منتظر پاسخ کشت ماند. در این موارد از توافق بین آزمون‌ها استفاده می‌شود. در توافق بین آزمون‌ها شاخص کاپا (ضریب هم‌خوانی، ضریب توافق کاپا) کمک‌کننده است.

برای محاسبه‌ی آماره کاپا از فرمول شماره (۱۴) استفاده می‌شود، که در این فرمول OP بیانگر نسبت مشاهده شده و EP بیان‌گر نسبت مورد انتظار است. n تعداد کل موارد است که با مجموع a+b+c+d در این فرمول است (۳).

$$kappa = \frac{OP-EP}{1-EP}$$ فرمول (۱۴)

$$OP = a + d/n$$

$$EP = \left(\frac{a+b}{n}\right) * \left(\frac{a+c}{n}\right) * \left(\frac{b+d}{n}\right) * \left(\frac{c+d}{n}\right)$$

برای تفسیر آزمون کاپا بدین شرح اقدام می‌شود؛ اگر نتیجه‌ی آزمون به ترتیب ۱، بیش از ۰/۸، بین ۰/۸ تا ۰/۶، بین ۰/۶ تا ۰/۴، بین ۰/۴ تا ۰/۲ و زیر ۰/۲ به‌ترتیب به عنوان توافق کامل، توافق عالی، توافق خوب، توافق متوسط، توافق ضعیف و عدم توافق گزارش می‌شود. البته لازم به ذکر است که اگر کاپا بیش از ۰/۶ باشد ،هردو تست خوب هستند و به‌جای یکدیگر می‌توان آن‌ها را به‌کار برد. همان‌طور که در دو جدول زیر دیده می‌شود، در جدول بالایی مقایسه‌ی موارد مثبت و منفی تست شیر و خون دیده می‌شود و در جدول پایینی کاپا ۶۷/ است که نشان دهنده‌ی توافق خوب است. هیچ‌کدام از تست شیر و تست شاخص خون استاندارد طلایی نیستند، اما به علت کاپا که توافق خوبی را نشان می‌دهد این دو تست را می‌توان به جای یکدیگر به کار برد (۸).

شیر / خون	مثبت	منفی	کل
منفی	۲۵	۰	۲۵
مثبت	۳	۴	۷
کل	۲۸	۴	۳۲

جدول ۴– مقایسه‌ی موارد مثبت و منفی تست شیر و تست شاخص خون برای پی بردن به حساسیت و ویژگی (۸)

	Kappa	Asymp. Std. Error	Approx. T[b]	. p-Value
ضریب توافق	۰.۶۷۶	۰.۱۶۹	۴.۰۴۱	۰.۰۰۰۵
کل	۳۲			

جدول ۵- ضریب توافق تست شیر و تست خون (۸)

در مطالعه‌ی جمشیدی و همکاران روی مقایسه‌ی ارزش روش‌های مختلف تشخیصی در تعیین آلودگی در گربه‌ها به هلیکوباکترهای معدی، ضریب کاپا بین این روش‌ها ۰/۵۵ به دست آمد که نشان دهنده‌ی توافق نسبی بین روش‌های مختلف است (۹).

مروری بر برخی مطالعه‌های ارزیابی آزمون‌های تشخیصی

در مطالعه‌ی وکیلی و همکاران، حساسیت و ویژگی آزمون‌های الایزا در تشخیص بروسلوز اندازه‌گیری شد که به‌ترتیب حساسیت و ویژگی آزمون الایزا برای IgG به‌ترتیب ۹۳/۷ و ۷۰/۶ درصد و آزمون IgM به‌ترتیب ۱۲/۵ و ۱۰۰ درصد ارزیابی شد (۱۰).

در مطالعه‌ای دیگر که روی تیلریوز بدخیم گوسفند انجام گرفت، حساسیت و ویژگی آزمایش پادتن درخشان با روش غیر مستقیم ارزیابی شد. در این مطالعه حساسیت و ویژگی به‌ترتیب ۹۴/۵۷ و ۷۱/۷۰ درصد برآورد شد (۱۱).

ارزیابی تست ممانعت از تشکیل rosette برای تشخیص زودهنگام آبستنی در تلیسه‌های شیری توسط وجگانی و همکاران انجام شد که در این ارزیابی، حساسیت ۸۵ درصد و ویژگی ۲۶ درصد گزارش شد (۱۲).

برای تعیین میزان حساسیت و ویژگی تست PCR در مقایسه با نتایج کشت در تشخیص عفونت با M. Tuberculosis نادری و همکاران روی نمونه‌های بیمار در شهرستان زاهدان مطالعه کردند. در این مطالعه حساسیت و ویژگی تست PCR به‌ترتیب ۸۶/۹۱ و ۸۷/۹۶ درصد گزارش شدد. همچنین ارزش اخباری مثبت و منفی نیز به‌ترتیب ۵۳/۹۷ و ۸۵/۸۹ درصد گزارش شد (۱۳).

برای تعیین ارزش حساسیت و ارزش اخباری صدای زنگی در اختلالات حفره‌ی بطنی در گاو، نوروزیان و همکاران در مطالعه‌ی خود، در طرف چپ جدار حفره بطنی وجود صدای زنگی را دارای اعتبار بالایی در تشخیص دانستند و تغییر محل شیردان به طرف چپ بیانگر آن بود که ارزش اخباری آن ۸۱/۵۶ درصد به‌دست آمد. هم‌چنین ارزش اخباری در طرف راست جدار حفره‌ی بطنی اختلالات توپوگرافیکی شیردان، اختلالات سکوم و اختلالات روده‌های کوچک به‌ترتیب ۸۵/۹۴، ۸۷/۵ و ۸۵/۷۰ درصد گزارش شد (۱۴).

در مطالعه‌ی قاسم‌زاده نوا و همکاران حساسیت و ویژگی CMT برای تشخیص ورم پستان تحت بالینی در گاوهای شیری در مقایسه با کشت به‌ترتیب ۸۴/۱ و ۶۱/۲ درصد و حساسیت و ویژگی آن در مقایسه با SCC به ترتیب ۹۷/۵ و ۷۹/۶ درصد گزارش شد. همچنین در مطالعه‌ی دیگری برای بررسی حساسیت و ویژگی دستگاه MAS-D-TEC، حساسیت و ویژگی این دستگاه برای شناسایی ورم پستان تحت بالینی در گاوهای شیری به‌ترتیب ۱۰۰ و ۴۳/۳ درصد گزارش شد (۱۵).

در مطالعه‌ی اسلامی و همکاران، دقت روش‌های مختلف کشتارگاهی با روش متداول کشتارگاهی برای تشخیص سیستی‌سرکوس بویس در گاوهای ذبح شده بررسی شد که در بررسی آماری، سه روش با روش متداول کشتارگاهی اخلاف معنی‌داری وجود داشت (۱۶).

تعارض منافع

بین نویسندگان تعارض در منافع گزارش نشده‌است.

منابع

۱. حق‌دوست ع‌ا، ایرانمنش الهام. ۱۳۸۷. ارزیابی دقت آزمون‌های تشخیصی. مجله دانشگاه علوم پزشکی کرمان، دوره شانزده، شماره ۱، ص ۱۰۵-۹۳.

۲. بشردوست ن. ارزیابی آزمون‌های تشخیصی، در: ملک‌افضلی ح، مجدزاده س‌ر، فتوحی ا، توکلی س. ۱۳۸۳. روش‌شناسی‌پژوهش‌های کاربردی، انتشارات دانشگاه علوم پزشکی تهران، چاپ اول، ص. ۲۲۹-۴۱

3. Thrusfield M. Veterinary epidemiology: 4th edition, John Wiley & Sons; 2018.

4. Smith RD. Veterinary clinical epidemiology, 3rd eition: CRC press; 1995.

۵. معاونت آمار مرکز فناوری اطلاعات و ارتباطات وزارت جهاد کشاورزی (۱۴۰۱). آمارنامه کشاورزی سال ۱۴۰۰، جلد دوم، مرکز فناوری اطلاعات و ارتباطات معاونت برنامه‌ریزی و اقتصادی، وزارت جهاد کشاورزی ص ۱۱۹.

٦. خاوری خراسانی ا. ۱۳۷۸. پژوهشی در سل گاوی در ایران با تکیه بر روش‌های مختلف تشخیصی در جهت ارزیابی اعتبار آزمون توبرکولین و ارزشیابی اقتصادی برنامه‌های مبارزه با سل گاوی، انتشارات سازمان دامپزشکی کشور.

۷. سخا م، شریفی ح، طاهری ا، صافی ش. تعیین میزان فراوانی کتوز تحت درمانگاهی در گاوداری‌های شیری شهرستان کرمان با استفاده از روش اندازه‌گیری بتاهیدروکسی بوتیرات سرم. مجله تحقیقات دامپزشکی، ۱۳۸۳؛ ۵۹(۳): ۲۴۹.

۸. باستانی ب، شقایق ع، موسی‌خانی ف. ارزیابی حساسیت و ویژگی آزمایش نواری BHBA شیر در مقایسه با آزمایش BHBA خون جهت تشخیص کتوز تحت بالینی در گاوهای شیری. پژوهش‌های بالینی دامپزشکی ۱۳۹۲؛۴(۴)،پیاپی ۱۶):۲۳۵-۲۲۷.
20.1001.1.20088159.1392.4.4.1.7

۹. جمشیدی ش، اختردانش ب، محمدی م، ساسانی ف، بکایی س، زهرایی‌صالحی ت و همکاران. مقایسه ارزش روش‌های مختلف تشخیصی در تعیین آلودگی گربه‌ها به هلیکوباکترهای معدی. مجله تحقیقات دامپزشکی، ۱۳۸۶؛ ۶۲(۲): ۱۳۱.

۱۰. وکیلی ز، مؤمن‌هروی م، شریف ع، معصومی م. حساسیت و ویژگی آزمون الایزا در تشخیص بیماری بروسلوز. مجله پزشکی کوثر. ۱۳۸۹، دوره ۱۵، شماره ۲، ص ۹۸-۹۵.

۱۱. خاکی ز، رهبری ص، نوروزیان ا. تعیین حساسیت و ویژگی آزمایش پادتن درخشان با روش غیر مستقیم در تیلریوز بدخیم گوسفند. مجله تحقیقات دامپزشکی. ، دوره ۵۳، شماره ۳و۴، ص ۳۰-۲۷.

۱۲. وجگانی م، قراگوزلو ف، وجگانی م، میرترابی س‌و. تعیین حساسیت و ویژگی تست ممانعت از تشکیل Rosette برای تشخیص زود هنگام آبستنی در تلیسه‌های شیری. مجله تحقیقات دامپزشکی. ۱۳۸۱؛۵۷(۱):۴-۱.

۱۳. نادری م، شریفی‌مود ب، کوهپایه حر، ناصرپورفریور ت، احمدی من. تعیین میزان حساسیت و ویژگی تست PCR در مقایسه با نتایج کشت در تشخیص عفونت با M. Tuberculosis. مجله دانشگاه علوم پزشکی گیلان. ۱۳۸۳؛۱۳(۵۲):۵۵-۵۹.

۱۴. نوروزیان ا، کلانتری‌خاندانی ع، نادعلیان مق، دهقان مم. تعیین ارزش حساسیت و پیشگوئی صدای زنگی در اختلالات حفره بطنی گاو. پژوهش و سازندگی. ۱۳۷۹؛۱۳(۱ (پی آیند ۴۶)):۱۳۸-۱۳۲.

۱۵. اسلامی ع، مشگی ب، باهنر ع، موسوی س‌ع. مطالعه دقت روش‌های مختلف جهت تشخیص سیستی سرکوس بوویس در گاوهای ذبح شده. مجله تحقیقات دامپزشکی (دانشگاه تهران) ۱۳۸۴؛۶۰(۲):۱۸۰-۱۷۷.

Abstract in English

Evaluation of Diagnostic and Screening Test in Veterinary Medicine

Mohammad Arad Zandieh[1], Fateme Sheikhian[2], Hamid Sharifi[3], Hesameddin Akbarein[4]

1- PhD candidate of Epidemiology, Faculty of Veterinary Medicine, University of Tehran, Tehran, Iran

2- Undergraduate Student, Faculty of Veterinary Medicine, University of Tehran, Tehran, Iran

3- Professor of Epidemiology, Institute for Future Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

4- Assistant professor of Epidemiology, Division of Epidemiology and Zoonoses, Department of Food Hygiene& Quality Control, Faculty of Veterinary Medicine, University of Tehran, Tehran, Iran

* akbarein@ut.ac.ir

Screening tests are a special type of diagnostic tests that are performed in an apparently healthy population. The purpose of performing diagnostic tests is to correctly diagnose patients, distinguish affected animals from healthy animals, distinguish between cases and controls, and distinguish between normal and abnormal cases. Screening tests should be simple, cheap, rapid and valid. Diagnostic and screening tests are mainly used in the monitoring of diseases. Related terms to the evaluation of these tests, including sensitivity, specificity, positive and negative predictive value, accuracy and precision, as well as concepts such as the golden standard, because they are often used interchangeably or misinterpreted, it is especially important to learn them. Also, due to the fact that gold standard tests are often more expensive and time-consuming, incomplete diagnostic tests are used, which can be calculated by calculating the sensitivity and specificity of the actual and apparent prevalence obtained by the diagnostic test. However, if the golden standard test is not available, other methods are used to evaluate the tests, including the Kappa index test. The interpretation results of these tests provide a comprehensive and evidence-based approach to clinicians and experts, which ultimately leads to more accurate, comprehensive, cheaper and faster monitoring. In this review article, with a complete and comprehensive review of the conventional concepts in the evaluation of diagnostic tests, along with the solution of practical examples, we will expand and provide a comprehensive presentation of these concepts. Also, the latest original studies that have been done in the field of evaluation of diagnostic tests will also be reviewed in this article.

Keywords: Diagnostic test, screening test, sensitivity, specificity, gold standard