

Odabrane teme iz biostatistike

Lessons in biostatistics

Odabir i tumačenje dijagnostičkih pretraga Selecting and interpreting diagnostic tests

Marc A. Raslich¹, Ronald J. Markert², Shahan A. Stutes¹

¹Klinika za medicinu i pedijatriju, Sveučilište „Wright State“, Medicinski fakultet „Boonshoft“, Dayton, Ohio, SAD

¹Departments of Medicine and Pediatrics, Wright State University Boonshoft School of Medicine, Dayton, Ohio, USA

²Klinika za internu medicinu, Sveučilište „Wright State“, Medicinski fakultet „Boonshoft“, Dayton, Ohio, SAD

²Department of Internal Medicine, Wright State University Boonshoft School of Medicine, USA

Sažetak

Dijagnostički testovi uključuju sve od simptoma bolesti opisanih u bolesnikovoj anamnezi i uočenih fizičkim pregledom do sofisticiranih laboratorijskih pretraga i slikovnih tehnika široke upotrebe u današnjoj praksi. Dijagnostičkim testom uspoređuje se klinička informacija dobivena manje invazivnim i/ili jeftinijim testom s takozvanim „zlatnim standardom“.

U ovom članku najprije definiramo osnovne odrednice dijagnostičkog testiranja, primjerice, specifičnost i osjetljivost testa te pozitivnu i negativnu prediktivnu vrijednost kao mjeru učinkovitosti testa. Također raspravljamo o korisnosti ovih odrednica u odabiru testa i tumačenju dobivenih rezultata. Isto tako, objašnjavamo i Bayesov pristup dijagnostičkom ispitivanju kroz raspravu o vjerojatnosti neke dijagnoze prije i poslije učinjenog testa te o pozitivnom i negativnom omjeru vjerojatnosti.

Potom tumačimo sva pitanja koja se tiču sažimanja mjera učinkovitosti testa, s osobitim naglaskom na površinu ispod ROC krivulje, dijagnostičku točnost i dijagnostički omjer izgleda. Također se pojašnjava i učinak prevalencije na dijagnostičku točnost i prediktivne vrijednosti. Na kraju, predstavljamo dijagnostički omjer izgleda kao mjeru učinkovitosti testa koja udružuje osjetljivost i specifičnost, ali ne ovisi o prevalenciji.

Prije nego što liječnik zatraži neki dijagnostički test, mora procijeniti njegovu moguću dobrobit i rizike te kako će rezultat testa utjecati na skrb o bolesniku. Osnovni preduvjet za postavljanje dijagnoze temeljene na dokazima je dobro poznavanje i ispravno tumačenje svih dostupnih mjera diskriminacijske učinkovitosti testa.

Ključne riječi: medicina temeljena na dokazima, dijagnostički test, vjerojatnost, osjetljivost i specifičnost, kliničko odlučivanje

Abstract

Diagnostic tests range from the signs and symptoms obtained from the patient's history and physical examination to the sophisticated laboratory and imaging tests widely used in medical practice today. The typical diagnostic test compares clinical information gathered in a less invasive and/or less costly manner to the so-called gold standard.

First, we define the basic components of diagnostic testing – i.e., the test characteristics of sensitivity and specificity and the test performance measures of positive predictive value and negative predictive value. We discuss the traditional use of these components in selecting and interpreting diagnostic test results. We then explain the Bayesian model for diagnostic testing through a discussion of pre-test probability and post-test probability and positive and negative likelihood ratios.

We discuss the issue of integrating indicators of a test's characteristics and performance, highlighting the area under the ROC curve, diagnostic accuracy and the diagnostic odds ratio. The impact of prevalence on accuracy and predictive values is clarified. And lastly, the diagnostic odds ratio is presented as a measure of test performance that combines sensitivity and specificity but is independent of test prevalence.

Before ordering a diagnostic test, the clinician must evaluate the potential benefits and risks of the test and how the results will alter patient management. Familiarity with the available measurement options for evaluating a test and how to interpret the results are the first steps toward making evidence-based diagnostic decisions.

Key Words: evidence-based medicine, diagnostic tests, probability, sensitivity and specificity, decision making

Pristiglo: 15. listopada 2007.

Prihvaćeno: 5. studenog 2007.

Received: October 15, 2007

Accepted: November 5, 2007

Uvod

Dijagnostička pretraga predstavlja jedan od temelja medicinskog obrazovanja i kliničkog rada. S korijenima u

Introduction

The diagnostic test is one of the cornerstones of medical education and clinical practice. With roots in ancient Gree-

starij Grčkoj i skorašnjim povijesnim značenjem u radu Williama Oslera, liječnici su u vezi s pretragama odavno shvatili da moraju postati vješti u odabiru primjerenih metoda za određivanje oboljenja ili stanja bolesnika. Dijagnostičke pretrage obuhvaćaju znakove i simptome zabilježene u anamnezi te otkrivene fizikalnim pregledom do složenih laboratorijskih i slikovnih pretraga koji su danas u medicinskoj praksi u širokoj uporabi. Liječnici početno često odluče ne koristiti se tzv. „zlatnim standardom“ kao pomoći u postavljanju dijagnoze, te izabiru radije manje skup, manje invazivan i/ili manje neprijatan pristup. Primjerice, elektrokardiogram (EKG) se nakon vježbanja može prvi koristiti kao dijagnostička pretraga kod 60-godišnjeg bolesnika s bolovima u prsima, umjesto određenijeg (no i manje sigurnog, a skupljeg) koronarnog angiograma.

Ciljevi dijagnostičkih ispitivanja uključuju otkrivanje ili isključenje bolesti, predstavljaju pomoć u obradi bolesnika, procjeni prognoze, praćenju kliničkog tijeka, te mjerenju općeg zdravlja ili sposobnosti (1).

U ovom broju časopisa *Biochemia Medica* predstavljamo nekoliko obilježja pretraga na koje se kliničari oslanjaju kod odabira i tumačenja dijagnostičkih pretraga, a ujedno ističemo nekoliko tradicijskih pristupa te se osvrćemo na nekoliko karakteristika pretraga koje su opisane kao više „utemeljene na dokazima“ (2). U našem ćemo sljedećem članku opisati kako kritički ocijeniti medicinski članak koji izvještava o dijagnostičkoj pretrazi (3-7). Izložiti ćemo značenje metodologije istraživanja, pružiti pregled nedavno podrobno opisanih normi za nalaze, te razjasniti kako su dijagnostički rezultati ugrađeni u kliničko odlučivanje.

Dijagnostički slučaj

Gosp. Brown, star 74 godine, dolazi u Hitnu službu žaleći se na akutnu zaduhu. Ne navodi nikakav bol u prsima ili nesvjesticu. Njegov nalaz EKG ukazuje na fibrilaciju pretkljetki. U njegovoj povijesti bolesti značajan je podatak jedino o hipertenziji. Bolesnik ne navodi nikakvu bubrežnu bolest. Klinički gledano, vjerujemo da njegov dolazak opravdava razmatranje sistoličke kardijalne disfunkcije uz kongestivno zatajenje srca (engl. *congestive heart failure*, CHF). Naše središnje kliničko pitanje u takvoj situaciji glasi: Kakva je u bolesnika s fibrilacijom pretkljetki i akutnom zaduhom točnost B tipa natriuretskog peptida (engl. *B-type natriuretic peptide*, BNP) u dijagnosticiranju kongestivnog srčanog zatajenja?

Odabir dijagnostičke pretrage: obilježja pretrage

Osjetljivost i specifičnost

Prije nego je medicina utemeljena na dokazima potaknula Bayesov pristup dijagnostičkom ispitivanju, kliničari su

ce and more recent historical prominence in the work of William Osler, medical and surgical practitioners have long realized that they must become skilled at choosing appropriate methods for determining the patient's disease or condition. Diagnostic tests range from the signs and symptoms obtained from the history and physical examination to the sophisticated laboratory and imaging tests widely used in medical practice today. Often physicians initially choose not to use the so-called “gold standard” to assist in making a diagnosis, preferring a less costly, less invasive, and/or less painful approach. For example, an exercise electrocardiogram (EKG) might first be used as a diagnostic test with a 60-year old patient with chest pain rather than the more definitive (as well as less safe and more costly) coronary angiogram. Objectives of diagnostic investigations include detecting or excluding disease; contributing to management; assessing prognosis; monitoring clinical course; and measuring general health or fitness (1).

In this issue of *Biochemia Medica* we present several test characteristics which clinicians rely on when selecting and interpreting diagnostic tests – highlighting several traditional approaches as well as reviewing a few test characteristics which have been described as more “evidence-based” (2). In our next article we will outline how to critically appraise a medical article reporting on a diagnostic test (3-7). We will describe the importance of study methodology, review recently detailed reporting standards and illustrate how diagnostic results are incorporated into clinical decision-making.

Diagnostic case

Mr. Brown, a 74-year-old male, comes to the Emergency Department with the complaint of acute dyspnea. He reports no chest pain and no syncope. His EKG identifies atrial fibrillation. His past medical history is significant only for hypertension. He reports no renal disease. Clinically, we feel his presentation warrants consideration of systolic cardiac dysfunction with congestive heart failure (CHF). Our focused clinical question in this situation becomes: In patients with atrial fibrillation and acute dyspnea, what is the accuracy of B-type natriuretic peptide (BNP) for diagnosing congestive heart failure?

Selecting a diagnostic test: test characteristics

Sensitivity and specificity

Before the Bayesian approach to diagnostic testing was encouraged by Evidence Based Medicine (EBM), clinicians

se oslanjali isključivo na uobičajena obilježja pretraga / model 2 x 2 učinkovitosti pretrage (Slika 1.). Kao pomoć u postavljanju dijagnoze liječnik bi odabrao pretragu s odgovarajućim karakteristikama, tj. osjetljivošću i specifičnošću. Osjetljivost je sposobnost dijagnostičke pretrage da osobe s određenom bolešću razvrsta kao pozitivne (iznad ili ispod točno određene granične vrijednosti, npr. visoka glukoza nakon gladovanja (iznad granične vrijednosti) u dijagnosticiranju šećerne bolesti, ali nizak feritin (ispod granične vrijednosti) kod dijagnoze anemije zbog pomanjkanja željeza). Osjetljivost opisuje vjerojatnost da će neka osoba s ciljanim poremećajem na ispitivanju biti pozitivna; specifičnost je sposobnost dijagnostičke pretrage da razvrsta osobe bez oboljenja kao negativne (tj. ispod/iznad granične vrijednosti, ovisno o pretrazi). Specifičnost opisuje vjerojatnost da će pojedinac bez ciljanog poremećaja po ispitivanju biti negativan. Ako se BNP ≥ 100 pg/mL koristi kao granična vrijednost kojom se bolesnici razvrstavaju kao pozitivni, Knudsen i sur. (8) su utvrdili da osjetljivost BNP iznosi 95% (95%-tni interval pouzdanosti [engl. *confidence interval*, CI] = 91-97%), te specifičnost 40% (95%-tni CI = 28-52%) za dijagnosticiranje kongestivnoga srčanog zatajenja (CHF) u bolesnika s fibrilacijom atriya (engl. *atrial fibrillation*, AF) i zaduhom. S obzirom na sliku 1, osjetljivost se izračuna kao broj stvarno pozitivnih (TP, engl. *true positives*) podijeljen s ukupnim brojem bolesnika s oboljenjem (stvarno pozitivni plus lažno negativni [FN, engl. *false negatives*]). Stoga: $207 / 219 = 0,945$ ili 95%. Drugim riječima, devetnaest od 20 bolesnika (95%) s CHF koji se prijave u Hitnu službu s AF i zaduhom bilo je pozitivno kad se BNP ≥ 100 pg/mL primijenila kao granična vrijednost. Specifičnost se dobiva dijeljenjem stvarno negativnih (TN, engl. *true negatives*) s ukupnim brojem bolesnika bez oboljenja (lažno pozitivni [FP, engl. *false positives*] plus stvarno negativni). Stoga: $29 / 73 = 0,397$ ili 40%. Ispitanici u studiji koji su imali i AF i zaduhu bez kongestivnoga srčanog zatajenja bili su negativni (BNP < 100 pg/mL) samo u 40% slučajeva.

relied exclusively on the traditional test characteristics/ test performance 2 x 2 model (Figure 1).

To aid in making a diagnosis, the physician would choose a test with adequate test characteristics – i.e., sensitivity and *specificity*. Sensitivity is the ability of a diagnostic test to classify those with the disease as *positive* (above or below a specified cutoff – e.g., high fasting glucose (above the cutoff) in diagnosing diabetes mellitus but low ferritin (below the cutoff) in diagnosing iron deficiency anemia). Sensitivity describes the probability that a person with the target disorder will test positive; specificity is the ability of a diagnostic test to classify those without the disease as *negative* (i.e., below/above the cutoff, depending on the test). Specificity describes the probability that an individual without the target disorder will test negative. When a BNP ≥ 100 pg/mL is used as the cutoff to classify patients as positive, Knudsen et al. (8) found that BNP had a sensitivity of 95% (95% confidence interval = 91% to 97%) and a specificity of 40% (95% CI = 28% to 52%) for diagnosing congestive heart failure (CHF) in patients with atrial fibrillation (AF) and dyspnea. Referring to Figure 1, sensitivity is calculated as the true positives (TP) divided by the total number of patients with the disorder (true positives plus false negatives [FN]). Thus: $207 / 219 = 0.945$ or 95%. In other words, nineteen of 20 patients (95%) with CHF presenting to the Emergency Department with AF and dyspnea were positive when a BNP ≥ 100 pg/mL is used as the cutoff value. Specificity is derived from dividing the true negatives (TN) by the total number of patients without the disorder (false positives [FP] plus true negatives). Thus: $29 / 73 = 0.397$ or 40%. The study participants with both AF and dyspnea *without* congestive heart failure were negative (BNP < 100 pg/mL) only 40% of the time.

Traditional 2x2 table definition

	Disorder present	Disorder absent
Result positive	TP	FP
Result negative	FN	TN

BNP as diagnostic test for CHF

	CHF	No CHF
BNP ≥ 100 pg/mg	207	44
BNP < 100 pg/mg	12	29

SLIKA 1. BNP kao dijagnostička pretraga za kongestivno zatajenje srca: uobičajeni model 2 x 2

FIGURE 1. BNP as a diagnostic test for congestive heart failure: traditional 2 x 2 model

Tumačenje rezultata pretrage: učinkovitost pretrage

Prediktivne vrijednosti

Kliničar se kod odluke o pretrazi koja će mu pripomoći u dijagnosticiranju obično vodi obilježjima dostupnih pretraga, naime njihovom osjetljivošću i specifičnošću.

Nakon što je, međutim, pretraga obavljena, liječnika zanima učinkovitost pretrage. U tu svrhu koristimo prediktivne vrijednosti pretrage. Pozitivna prediktivna vrijednost (PPV) kliničaru otkriva koliki postotak osoba s pozitivnim nalazom ima neku bolest. Negativna prediktivna vrijednost (NPV) otkriva koliki postotak osoba s negativnim rezultatom nema bolest. Vraćajući se na naš primjer BNP-a i kongestivnoga srčanog zatajenja (8), PPV predstavlja bolesnike TP podijeljene s TP i FP (tj. $207 / 251 = 0,824$ ili 82%), dok NPV predstavlja ispitanike TN podijeljene s TN i FN (tj. $29 / 41 = 0,707$ ili 71%). Pozitivna prediktivna vrijednost od 82% ukazuje da su bolesnici navedeni u ovoj studiji, koji su se prijavili u Hitnu službu sa zaduhom i AF, imali kongestivno zatajenje srca kao uzrok simptoma u 4 od 5 slučajeva kad im je BNP bio isti ili veći od 100 pg/mL. Negativna prediktivna vrijednost od 71% govori nam da 7 od 10 bolesnika s AF i zaduhom koji su bili negativni nakon pretrage (tj. BNP < 100 pg/mL) nisu imali CHF.

Omjeri vjerojatnosti

Jedno od ključnih pitanja koje se razmatra kod tumačenja rezultata dijagnostičkih pretraga je: „Razlikuju li se rezultati pretraga u oboljelih bolesnika od rezultata dobivenih kod normalnih pojedinaca? (9). Pri postavljanju dijagnoze kliničari se koriste dijagnostičkim pretragama kako bi preinačili svoju početnu sumnju na bolest i učinili određenu dijagnozu što vjerojatnijom. Osjetljivost ili specifičnost određene pretrage mogu biti poznate, no kako tu informaciju spojiti s kliničkom intuicijom, ili provjeriti vjerojatnost prije pretrage da bi se odredila vjerojatnost bolesti koja je već prisutna, ili vjerojatnost poslije pretrage? Matematičko pomagalo u toj integraciji je omjer vjerojatnosti. Jednostavnim riječima, omjerima vjerojatnosti izražava se veličina kojom je vjerojatnost bolesti u određenom bolesniku promijenjena rezultatom pretrage (10). Omjerima vjerojatnosti uspoređuju se dvije vjerojatnosti: vjerojatnost da će osoba s određenom bolešću imati određeni rezultat pretrage u odnosu na vjerojatnost da će osoba bez takve bolesti imati isti rezultat pretrage.

Pozitivan omjer vjerojatnosti izračuna se na sljedeći način:

LR za pozitivnu pretragu = osjetljivost/(1-specifičnost)

Negativan omjer vjerojatnosti:

LR za negativan rezultat pretrage =
(1-osjetljivost)/ specifičnost.

Interpreting test results: Test performance

Predictive values

In choosing a test to help make a diagnosis, the clinician is typically guided by the characteristics of the available tests, namely, their sensitivities and specificities. However, after the test is administered, the physician is interested in test performance. For this objective we can use the test's predictive values. Positive predictive value (PPV) tells the clinician what percent of those with a positive finding has the disease. Negative predictive value (NPV) reveals what percent of those with a negative result do not have the disease. Turning to our BNP and congestive heart failure example (8), PPV is the TPs divided by the TPs and FPs (i.e., $207 / 251 = 0.824$ or 82%) while NPV is the TNs divided by TNs and FNs (i.e., $29 / 41 = 0.707$ or 71%). A positive predictive value of 82% tells us that a patient coming to the Emergency Department with dyspnea and who has AF in this study had congestive heart failure as the cause for his/her symptoms in 4 of 5 cases when their BNP is equal to or greater than 100 pg/mL. A negative predictive value of 71% tells us that 7 of 10 of such AF and dyspnea patients who are negative on the diagnostic test (i.e., BNP < 100 pg/mL) turned out to not have CHF.

Likelihood ratios

One of the core questions we consider when interpreting diagnostic test results is “Do test results in affected patients differ from those in normal individuals?” (9) In the process of diagnosis, clinicians use diagnostic tests to modify their initial clinical suspicion of disease to make a particular diagnosis more or less likely. The sensitivity or specificity of a particular test may be known, but how does one integrate that information with their clinical intuition, or pretest probability, to determine likelihood of a disease being present, or post-test probability? The mathematical tool to aid in this integration is the likelihood ratio. Simply stated likelihood ratios express the magnitude by which the probability of disease in a specific patient is modified by the result of a test (10). Likelihood ratios compare two probabilities: the probability that a person with a particular disease will have a specified test result versus the probability that a person *without* that disease will have the same test result.

The positive likelihood ratio is calculated as follows:

LR for a positive test = sensitivity/(1-specificity).

The negative likelihood ratio as follows:

LR for a negative test result =
(1-sensitivity)/specificity.

The LR value can range from zero to infinity. Values less than 1 correspond to a decrease in the post-test proba-

Raspon vrijednosti LR je od ničice do beskonačnosti. Vrijednosti manje od 1 odgovaraju padu vjerojatnosti bolesti poslije pretrage, dok vrijednosti veće od 1 upućuju na porast vjerojatnosti bolesti poslije pretrage. Vrijednost LR koja iznosi 1 naznačuje da za pretragu postoji ista vjerojatnost da bude pozitivna ili negativna kod ispitanika koji imaju bolest i onih koji je nemaju, te stoga nije korisna.

Kod primjene LR-a početna klinička sumnja ili vjerojatnost bolesti prije pretrage mora se najprije prevesti u izgled prije pretrage pomoću sljedeće formule: izgledi = vjerojatnost/(1-vjerojatnost). Izgledi prije pretrage množe se s LR, a dobiveni umnožak predstavlja izgled poslije pretrage. Da bi se izgledi poslije pretrage preveli nazad u vjerojatnost (format s kojim je upoznata većina kliničara), treba se koristiti sljedećom formulom: vjerojatnost = izgledi/(1+izgledi). U našem slučaju, prevalencija CHF u spomenutoj studiji bila je 75% (219 od ukupno 292 bolesnika)(8). Kao što je ranije ukazano, BNP sa 100 pg/mL kao graničnom vrijednošću ima osjetljivost od 95% i specifičnost od 40%. Pozitivan LR bi trebao biti $0,95/(1-0,4) = 1,58$, a negativan LR = $(1 - 0,95)/0,4 = 0,125$. Primjenom 75% kao naše vjerojatnosti prije pretrage, tu prevalenciju prevodimo u izgled na sljedeći način: $0,75/(1-0,75) = 3:1$. Ako je BNP = 300 (pozitivan od ≥ 100), upotrebljavamo naše izgled prije pretrage i LR (+) za izračun izgleda poslije pretrage kao $3:1 \times 1,58 = 4,74:1$, te to prevodimo u vjerojatnost poslije pretrage kao $4,74/(1+4,74) = 83\%$. U tom slučaju LR(+) od 1,58 je povećao našu vjerojatnost prije pretrage sa 75% na samo 83%, što možda i nije istaknut porast. Slijedeći isti postupak za negativan BNP, vjerojatnost poslije pretrage je određena kao 27% (pokušajte to izračunati i provjeriti naš rezultat!). Pad sa 75% na 27% je zasigurno od veće važnosti nego za LR(+). Slika 2. se može koristiti kao vodič kliničarima za tumačenje različitih vrijednosti LR u kliničkim studijama.

Isto tako, nomogram se može koristiti da bi se odbacila potreba za izračunima ili procjenama (Slika 3). Koristi se ucrtavanjem linije od vjerojatnosti prije pretrage kroz LR (pozitivan, a zatim negativan) radi utvrđivanja vjerojatnosti nakon pretrage.

Positive LR	Magnitude of change
1	No difference
2	Small
5	Moderate
10	Large

SLIKA 2. Omjeri vjerojatnosti i očekivane razlike u vjerojatnosti prije i nakon pretrage

bility of disease while values greater than 1 indicate an increase in the post-test probability of disease. A LR value of 1 indicates the test has the same likelihood of being positive or negative in those with disease as in those without disease and is not helpful.

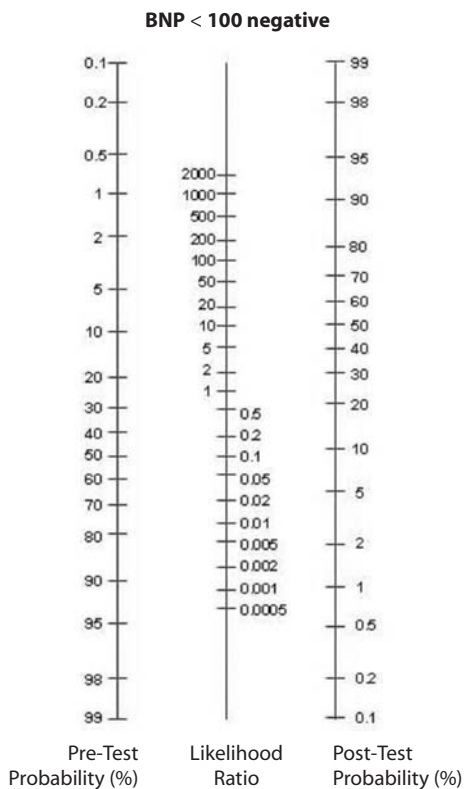
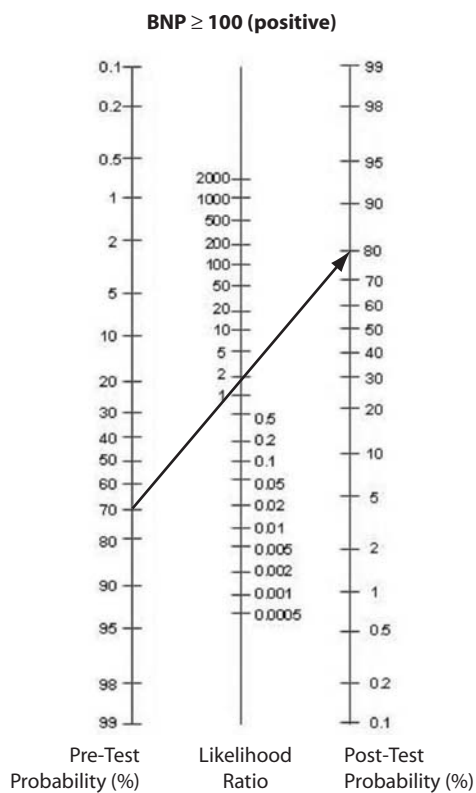
To use LR's the initial clinical suspicion or pre-test probability of disease must first be converted to pretest odds with the following formula: $\text{odds} = \text{probability}/(1 - \text{probability})$. The pre-test odds are multiplied by the LR and the product yielded is the post-test odds. To convert the post-test odds back to probability (the format most clinicians are familiar with) use the following formula: $\text{probability} = \text{odds}/(1+\text{odds})$.

In our case, the prevalence of CHF in their study was 75% (219 out of 292 total patients) (8). As previously shown, BNP with 100pg/ml as a cut off has a sensitivity of 95% and a specificity of 40%. The positive LR would then be $0.95 / (1 - 0.4) = 1.58$. And, the negative LR = $(1 - 0.95)/0.4 = 0.125$. Using 75% as our pre-test probability, we then convert this prevalence to odds as follows: $0.75/(1-0.75) = 3:1$. If the BNP = 300 ("positive" since ≥ 100), we use our pre-test odds and LR (+) to calculate our post-test odds as $3:1 \times 1.58 = 4.74:1$ and convert this to post-test probability as $4.74/(1 + 4.74) = 83\%$. In this case, the LR(+) of 1.58 increased our pre-test probability from 75% to only 83%, perhaps not a notable increase. Following the same procedure for a negative BNP, the post-test probability is determined to be 27% (try the math and double-check our work!). The drop from 75% to 27% was certainly of greater magnitude than for the LR(+). Figure 2 can be used to guide clinicians on how to interpret different values for LR's from clinical studies.

Alternatively, a nomogram can be used to eliminate the need for calculations or estimations (Figure 3). It is used by drawing a line from the pre-test probability through the LR (positive, then negative) to identify the post-test probability.

Negative LR	Magnitude of change
1	No difference
0.5	Small
0.2	Moderate
0.1	Large

FIGURE 2. Likelihood ratios and expected pre-test to post-test probability difference



SLIKA 3. Faganov nomogram – Vjerojatnosti prije i poslije pretrage

FIGURE 3. Fagan's Nomogram – Pre-test to Post-test probabilities.

Integracija pokazatelja obilježja i učinkovitosti pretraga

U eri prakse zasnovane na dokazima liječnicima su nužni visokokvalitetni podatci kojima će poduprijeti odluke o tome hoće li ili ne koristiti dijagnostičku pretragu u točno određenoj kliničkoj situaciji, koju će pretragu izabrati, te kako uporabiti rezultate kao pomoć u donošenju odluke. Kao što je već ranije naglašeno, obilježja pretraga (osjetljivost i specifičnost, LR(+) i LR(-)) koriste se za *odabir* dijagnostičke pretrage, dok se mjere učinkovitosti pretrage (pozitivne i negativne prediktivne vrijednosti; vjerojatnost nakon pretrage) koriste u *tumačenju* bolesnikovih rezultata pretrage. Međutim, s obzirom da ne postoje jednostavna pravila kombiniranja pokazatelja obilježja i učinkovitosti pretraga, kliničari, istraživači i nastavnici su istražili korisne pristupe za integraciju pokazatelja dijagnostičkih pretraga kao pomoć u donošenju mišljenja o bolesničkoj skrbi. U nastavku ćemo dati pregled nekih od pristupa integraciji pojedinačnih pokazatelja koji se pojavljuju kod dijagnostičkog ispitivanja.

Krivulje ROC

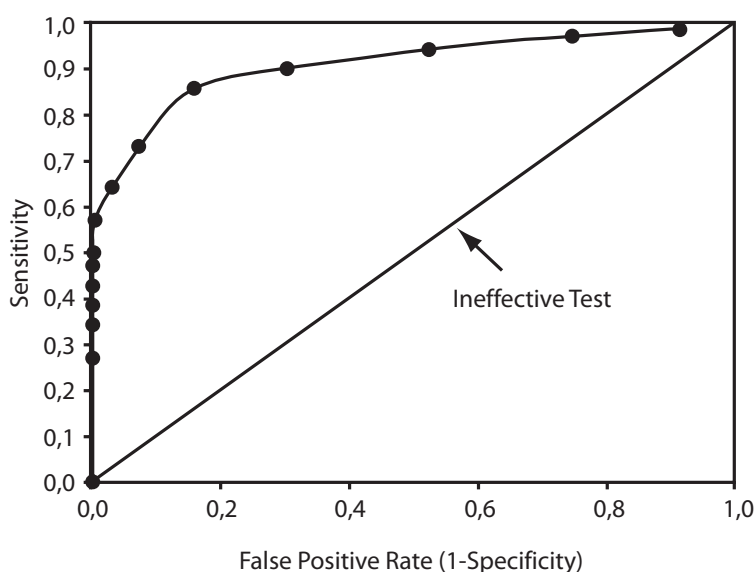
Krivuljama ROC (engl. *Receiver Operator Characteristic*, ROC) mogu se opisati promjene u osjetljivosti i specifičnosti za pojedinačnu pretragu s brojnim točkama granične vrijednosti (poput mjerenja BNP u našem primjeru), te usporediti učinkovitost različitih pretraga koje se koriste kod dijagnoze iste bolesti. Krivulje ROC su dijagrami stvarnog pozitivnog postotka (tj. osjetljivosti) na okomitoj ili osi y u odnosu na lažan pozitivan postotak (1-specifičnost) na vodoravnoj ili osi x za višestruke točke granične vrijednosti dijagnostičke pretrage. Slika 4. prikazuje krivulju

Integrating indicators of test characteristics and performance

In an era of evidence-based practice, physicians need high quality data to support decisions about whether or not to use a diagnostic test in a specific clinical situation, which test to choose, and how to use the results to aid in decision-making. As emphasized earlier, test characteristics [sensitivity and specificity; LR (+) and LR (-)] are used for selecting a diagnostic test while measures of test performance (positive and negative predictive values; post-test probability) are utilized in interpreting a patient's test result. However, since there are no simple aggregation rules for combining indicators of test characteristics and performance, clinicians, researchers, and educators have sought helpful approaches to integrating diagnostic test indices to aid in making judgments in patient care. We will now review some of the approaches to integrating the individual indices that emerge from diagnostic testing.

ROC Curves

Receiver Operator Characteristic (ROC) curves can describe the changes in sensitivity and specificity for a single test with numerous cutoff points (like the BNP measurement in our example) as well as compare the performance of different tests used to diagnose the same disease. ROC curves are plots of the true positive rate (i.e., sensitivity) on the vertical or y axis versus the false positive rate (1-specificity) on the horizontal or x axis for the multiple cutoff points of a diagnostic test. Figure 4 illustrates an ROC curve. We can use ROC curves to choose the optimal



SLIKA 4. Krivulja ROC

FIGURE 4. Receiver Operator Characteristic Curve

ROC. Krivulje ROC mogu se upotrijebiti za odabir optimalne granične vrijednosti koja minimalno smanjuje postotak lažno pozitivnih i lažno negativnih rezultata. Treba primijetiti da je na krivulji ROC optimalna točka smještena na položaju na krivulji koja je najbliža gornjem lijevom uglu dijagrama.

Nadalje, ukupna učinkovitost pretrage ili točnost (da upotrijebimo tehnički izraz) može se odrediti mjerenjem područja ispod krivulje ROC (vrijednost koju daju svi poznati kompleti statističke programske podrške). ROC loših ili beskorisnih pretraga nalazi se u području ispod krivulje blizu 0,50, a dijagram se čini dijagonalnim kad je stvaran pozitivan postotak jednak lažnom pozitivnom postotku (vidjeti dijagonalnu crtu na slici 4.). Kod odlične pretrage krivulja ROC se strmo penje i prolazi blizu gornjeg lijevo ugla; kod savršene pretrage područje ispod krivulje je 1,0, a osjetljivost i specifičnost iznose 100%.

Ovisno o ishodu i populaciji bolesnika, krivulje ROC mogu se u laboratorijskoj medicini primjenjivati dijagnostički, za praćenje ili za prognozu bolesti. Primjerice, za praćenje učinkovitosti dvaju postupaka ili strategija koje se koriste kod pretrage, područja ispod krivulja mogu se ispitati kako bi se otkrila razlika: pretraga ili strategija sa značajno većim područjem ispod krivulje ROC ima bolju učinkovitost.

Dijagnostička točnost

Točnost je vjerojatnost da će pojedinac biti ispravno razvrstan temeljem neke pretrage; tj. zbroj stvarno pozitivnih i stvarno negativnih podijeljen s ukupnim brojem ispitanih osoba. U našem slučaju s BNP-om, izračun bi glasilo: $(207 + 29) / 292 = 0,81$. Drugim riječima, 85% bolesnika u toj studiji bili su razvrstani adekvatno u skupine oboljelih ili zdravih s graničnom vrijednošću BNP od 100 pg/ml.

Točnost predstavlja odmjereni prosjek osjetljivosti i specifičnosti, pri čemu se osjetljivost mjeri prevalencijom ishoda u ispitivanoj populaciji, a specifičnost komplementom prevalencije. Na žalost, točnost će se samo u rijetkim primjerima približiti osjetljivosti i specifičnosti, tj. kad su osjetljivost i specifičnost istovjetne ili gotovo istovjetne ili kad se prevalencija bolesti približava 50%. Unatoč intuitivnoj dopadljivosti točnosti kao jedinstvene mjere valjanosti pretrage, njena ovisnost o prevalenciji čini je manje vrijednom od pažljivog i uravnoteženog razmatranja pojedinačnih mjera učinkovitosti i obilježja pretrage.

Učinci prevalencije bolesti

Omjeri vjerojatnosti, područje ispod krivulja ROC te dijagnostička točnost predstavljaju oba obilježja pretrage, tj. osjetljivost i specifičnost. Pozitivni i negativni omjeri vjerojatnosti te područje ispod krivulje ROC izračunavaju se isključivo pomoću osjetljivosti i specifičnosti. Stoga se ti pokazatelji ne mijenjaju s prevalencijom bolesti. Suprotno tome, točnost varira s prevalencijom bolesti (11). Taj

cutoff that minimizes the percent of false positive and false negative results. Note that on the ROC curve the optimal point is located at the curve location closest to the upper-left hand corner of the plot.

Further, overall test performance, or accuracy (to use the technical term), can be determined by measuring the area under the ROC curve (a value reported with all the familiar statistical software packages). Poor to useless tests have ROC under the curve areas close to 0.50, and the plot appears as a diagonal where the true positive rate equals the false positive rate (see the diagonal line in Figure 4). For an excellent test, the ROC curve rises steeply and passes close to the upper left-hand corner; the perfect test has an area under the curve of 1.0, and both the sensitivity and specificity are 100%.

Depending upon the outcome and patient population, ROC curves can be used for diagnostic, monitoring, and prognostic applications in laboratory medicine. For example, to compare test performance of two test procedures or strategies, the areas under the curves can be examined for a difference; the test or strategy with the significantly greater area under the ROC curve has the better performance.

Diagnostic accuracy

Accuracy is the probability that an individual will be correctly classified by a test; that is, the sum of the true positives plus true negatives divided by the total number of individuals tested. For our BNP case, our calculation would be: $(207 + 29) / 292 = 0.81$. In other words, 81% of the patients in this study were classified appropriately in their diseased or non-diseased groups with the BNP cutoff value of 100 pg/mL.

Accuracy represents the weighted average of sensitivity and specificity, where sensitivity is weighted by the prevalence of the outcome in the study population and specificity is weighted by the complement of the prevalence. Unfortunately, only in rare instances will accuracy closely approximate both sensitivity and specificity, such as when sensitivity and specificity are equal or nearly equal to each other, or, when disease prevalence is close to 50%. Despite the intuitive appeal of accuracy as a single measure of test validity, its dependence on prevalence renders it inferior to the careful and balanced consideration of individual measures of test characteristics and performance.

Effects of disease prevalence

Likelihood ratios, area under ROC curves, and diagnostic accuracy incorporate both test characteristics – i.e., sensitivity and specificity. The positive and negative likelihood ratios and the area under ROC curve are calculated exclusively with sensitivity and specificity. Thus, these indices do not vary with disease prevalence. In contrast, accuracy does vary with

učinak prevalencije bolesti također je prisutan kod tumačenja prediktivnih vrijednosti neke pretrage: kako prevalencija opada, tako opada i pozitivna prediktivna vrijednost, dok negativna prediktivna vrijednost raste.

Dijagnostički omjer vjerojatnosti

Kao mjera učinkovitosti pretrage, dijagnostički omjer vjerojatnosti (engl. *diagnostic odds ratio*, DOR) kombinira snagu osjetljivosti i specifičnosti kao neovisnih pokazatelja prevalencije s korisnošću točnosti kao jedinstvenog pokazatelja. Omjer vjerojatnosti je poznat statistički podatak u epidemiologiji kojim se izražava jačina povezanosti između izloženosti i bolesti. Kao takav, može se također primijeniti za izražavanje jačine povezanosti između rezultata pretrage i bolesti. Dijagnostički omjer vjerojatnosti neke pretrage jest omjer izgleda za pozitivnost na bolest u odnosu na izgleda za pozitivnost u neoboljelih osoba. Vrijednost DOR je u rasponu od 0 do beskonačnosti; više vrijednosti ukazuju na bolju razlikovnost pretrage.

$$\text{DOR} = (\text{TP}/\text{FN})/(\text{FP}/\text{TN}) = (\text{osjetljivost}/(1 - \text{osjetljivost})/((1 - \text{specifičnost})/\text{specifičnost})$$

i

$$\text{DOR} = (\text{TP}/\text{FP})/(\text{FN}/\text{TN}) = (\text{PPV}/(1 - \text{PPV})/((1 - \text{NPV})/\text{NPV}) = \text{LR+}/\text{LR-}$$

U našem bi slučaju izračun glasio: $\text{DOR} = (0,95/0,05)/(0,6/0,4) = 12,67$.

S obzirom da dijagnostičku učinkovitost izražava kao jedinstven izraz, dijagnostički omjer vjerojatnosti je osobito koristan kod usporedbe učinkovitosti paralelnih pretraga. Takva usporedba, uz primjenu uparenih pokazatelja kao što su osjetljivost i specifičnost, može biti komplicirana posebice ako jedna pretraga ne nadmaši drugu u oba pokazatelja (13). Zbog toga se izvještavanje o omjeru dijagnostičke učinkovitosti treba smatrati poželjnijom metodom za izvješćivanje o učinkovitosti pretrage.

Zaključci

Kliničar u praksi razvija niz mogućih dijagnoza ili uzroka za zapaženu bolest te svakoj dodjeljuje određenu vjerojatnost. Takve vjerojatnosti prije pretraga su istovjetne prevalenciji bolesti u bolesnika s istim obilježjima, znakovima, simptomima, anamnezom i ukupnom kliničkom slikom. Nakon procjene vjerojatnosti prije pretraga tražimo ispitivanja od službi kao što su radiologija, patologija i/ili laboratorijska medicina. Kao što smo pokazali, kliničar može procijeniti vjerojatnost poslije pretraga koristeći se vjerojatnošću prije pretraga i temeljnim znanjem o obilježjima učinkovitosti pretrage. Zatim moramo odlučiti o dijagnozi, procijeniti vjerojatnost mogućih ishoda, te odrediti opcije liječenja i raspoređivanja u kontekstu anamneze pojedinog bolesnika, ostalih pretraga te osobnih prio-

ritetima (11). This effect of disease prevalence is also present when interpreting a test's predictive values - as the prevalence falls, so too does the positive predictive value, while the negative predictive value rises.

Diagnostic odds ratio

As a measure of test performance the diagnostic odds ratio (DOR) combines the strengths of sensitivity and specificity, as prevalence independent indicators, with the advantage of accuracy as a single indicator. The odds ratio is a familiar statistic in epidemiology expressing the strength of association between exposure and disease. As such, it also can be applied to express the strength of association between test result and disease.

The diagnostic odds ratio of a test is the ratio of the odds of positivity in disease relative to the odds of positivity in the nondiseased. The value of a DOR ranges from 0 to infinity, with higher values indicating better discriminatory test performance.

$$\text{DOR} = (\text{TP}/\text{FN})/(\text{FP}/\text{TN}) = (\text{sensitivity}/(1 - \text{sensitivity})/((1 - \text{specificity})/\text{specificity})$$

and

$$\text{DOR} = (\text{TP}/\text{FP})/(\text{FN}/\text{TN}) = (\text{PPV}/(1 - \text{PPV})/((1 - \text{NPV})/\text{NPV}) = \text{LR+}/\text{LR-}$$

For our article the calculation would be: $\text{DOR} = (0,95/0,05)/(0,6/0,4) = 12,67$.

Since it expresses diagnostic performance as a single term, the diagnostic odds ratio is particularly useful when comparing the performance of competing tests. Comparing competing tests using paired indicators, such as sensitivity and specificity, can be complicated especially if one test does not outperform the other on both indicators (13). For this reason, reporting the diagnostic odds ratio should be considered a preferred method for reporting test performance.

Conclusions

In practice, the clinician develops a list of possible diagnoses or causes for the observed illness and assigns a probability to each. These pre-test probabilities are equivalent to the prevalence of the disease in patients having the same characteristics, signs, symptoms, history, and overall clinical picture. After estimating the pre-test probability we request testing from services such as radiology, pathology, and/or laboratory medicine. As we have demonstrated, the clinician can estimate the post-test probability using the pre-test probability and background knowledge of the test's performance characteristics. We must then decide the diagnosis, estimate likelihood of potential outcomes, and determine the disposition and treatment options in the context of the individual patient's medical

riteta. Ako pretraga nema potencijal da utječe na odluke vezane za dijagnozu, prognozu, raspoređivanje ili liječenje, tada se takva pretraga ne bi uopće trebala zatražiti. Imajući na umu takav klinički proces, prikazali smo pregled osnovnih obilježja pretrage koja predstavljaju važne čimbenike kod razmatranja odabira i tumačenja dijagnostičkih pretraga. Svaki pokazatelj ima svoje vlastite prednosti i nedostatke u kliničkoj primjeni. Na primjer, globalnim se mjerama točnosti pretrage procjenjuje samo sposobnost pretrage za razlikovanje osoba sa i bez oboljenja. Te mjere uključuju područje ispod krivulje ROC te omjer dijagnostičke vjerojatnosti. Takvi rezultati mogu biti dovoljni za odluke vezane za zdravstvenu politiku - primjerice, kod odlučivanja je li neka nova pretraga općenito bolja od postojeće za ciljano oboljenje. Nijedna, međutim, od tih mjera razlikovne snage pretrage nije osobito korisna za određivanje vjerojatnosti bolesti kod pojedinačnih bolesnika. S druge strane prediktivnom se vrijednošću, osjetljivošću, specifičnošću i omjerima vjerojatnosti ili izravno procjenjuje vjerojatnost bolesti u osoba ili se mogu koristiti za procjenu vjerojatnosti bolesti u pojedinaca primjenom Bayesovog teorema (14).

U cjelini, dijagnostičke pretrage su uglavnom vrijedne kao dopunska informacija kliničkoj procjeni, osobito kod srednje vjerojatnosti bolesti prije pretrage. Kliničari često zaboravljaju tu temeljnu postavku i zahtijevaju pretrage da bi "dijagnosticirali" ili "isključili" oboljenja, a ne da bi potvrdili ili otklonili kliničku pretpostavku.

Napretkom dijagnostičkih modaliteta skrb o bolesniku može se pojačati samo ako je vođena pažljivim kliničkim prosuđivanjem. Prije traženja dijagnostičke pretrage nužno je razmotriti ne samo vrijednost pretrage za prihvaćanje ili odbacivanje dijagnoze; kliničar mora procijeniti moguće koristi i rizike pretrage te način na koji će rezultati promijeniti obradu bolesnika (10). Upoznavanje s različitim dostupnim mogućnostima mjerenja kod pretraga te načinima kako tumačiti rezultate predstavlja prvi korak prema donošenju dijagnostičkih odluka zasnovanih na dokazima. U našem ćemo se sljedećem članku nadovezati na te vještine te pokazati kako tu informaciju povezati sa skrbi za pojedinačnog bolesnika. Također ćemo dati pregled važne dijagnostičke istraživačke metodologije te normi za nalaze pretraga.

Sažetak slučaja

Kod našeg bolesnika, koji se prijavio s akutnom zaduhom i fibrilacijom pretklijetke, razmotrili smo dijagnozu kongestivnog srčanog zatajenja. Koristeći se medicinskim pristupom kliničkom odlučivanju zasnovanom na dokazima, poslužili smo se objavljenim dijagnostičkim istraživanjima da bismo procijenili B-tip natriuretskog peptida (BNP) u CHF; - rezultati su sažeto prikazani na slici 5.

history, other tests, and personal preferences. If the test has no potential for impacting decisions regarding diagnosis, prognosis, disposition, or treatment, then the test should not have been requested in the first place.

We have reviewed the basic features of a test that are important factors in considering selection and interpretation of diagnostic tests with this clinical process in mind.

Each index has its own advantages and disadvantages in clinical application. For example, global measures of test accuracy assess only the ability of the test to discriminate between people with and without a disease. These include the area under the ROC curve and the diagnostic odds ratio. Such results may suffice for health policy decisions – for example, to decide whether a new test is generally better than an existing test for the target condition. However, neither of these measures of a test's discriminatory power is especially useful for determining the probability of disease in individual patients. On the other hand, the predictive value, sensitivity, specificity, and likelihood ratios either directly estimate probability of disease in individuals or can be used to estimate probability of disease in individuals using Bayes' theorem (14).

As a whole, diagnostic tests are mostly valuable as complementary information to clinical assessment, particularly when the pre-test probability of a disease is intermediate. Clinicians often forget this basic tenet and order tests to 'diagnose' or 'to rule out' conditions rather than to corroborate or challenge a clinical hypothesis.

Advances in diagnostic modalities may enhance patient care only if guided by careful clinical judgment. Before ordering a diagnostic test, one must consider more than the value of the test for accepting or rejecting a diagnosis; the clinician must evaluate the potential benefit and risks of the test and how the results will alter patient management (10). Becoming familiar with the various test measurement options that are available to us and how to interpret these results is the first step toward making evidence-based diagnostic decisions. In the next article we will build on these skills, demonstrating how to link this information to the care of an individual patient. Important diagnostic research methodology and reporting standards will also be reviewed.

Case Summary

For our patient that presented with acute dyspnea and atrial fibrillation we considered the diagnosis of congestive heart failure. Following the evidence-based medicine approach to clinical decision making, we referred to published diagnostic research to assess the accuracy of B-type natriuretic peptide (BNP) in CHF – the results are summarized in figure 5.

TEST CHARACTERISTIC	FORMULA	RESULT
Sensitivity (Sens)	$TP/(TP + FN)$	95%
Specificity (Spec)	$TN/(TN + FP)$	40%
Positive Predictive Value	$TP/(TP + FP)$	82%
Negative Predictive Value	$TN/(TN + FN)$	71%
Positive Likelihood Ratio	$Sens/(1 - Spec)$	1.58
Negative Likelihood Ratio	$(1 - Sens)/Spec$	0.12
Diagnostic accuracy	$(TP + TN)/All\ patients$	81%
Diagnostic Odds Ratio	Pos LR / Neg LR	12.67

SLIKA 5. Dijagnostička učinkovitost BNP - sažetak obilježja pretrage

FIGURE 5. BNP diagnostic performance - test characteristic summary

Adresa za dopisivanje:

Marc A. Raslich
 Combined Internal Medicine & Pediatrics Residency Program
 Wright State University Boonshoft School of Medicine
 PO Box 927
 Dayton, OH 45401-0927, USA
 e-pošta: marc.raslich@wright.edu
 tel: +937 224 3078
 faks: +937 224 3112

Corresponding author:

Marc A. Raslich
 Combined Internal Medicine & Pediatrics Residency Program
 Wright State University Boonshoft School of Medicine
 PO Box 927
 Dayton, OH 45401-0927, USA
 e-mail: marc.raslich@wright.edu
 phone: +937 224 3078
 fax: +937 224 3112

Literatura/References

1. Knottnerus JA, van Weel C, Muris JWM. Evidence base of clinical diagnosis: Evaluation of diagnostic procedures. *BMJ*. 2002;324:477-80.
2. Hawkins RC. The Evidence Based Medicine approach to diagnostic testing: practicalities and limitations. *Clin Biochem Rev*. 2005;26:7-18.
3. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. Are the results of the study valid? *JAMA*. 1994;271:389-91.
4. Jaeschke R, Gordon H, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. what are the results and will they help me in caring for my patients? *JAMA*. 1994;271:703-7.
5. Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller M. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120:667-76.
6. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. for the STARD group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin Chem Lab Med*. 2003;41:68-73.
7. Straus SE, Richardson WS, Glasziou P, Haynes RB. Evidence-Based Medicine: How to Practice and Teach EBM. 3rd ed. Edinburgh, UK: Churchill-Livingstone; 2005.
8. Knudsen CW, Omland T, Clopton P, Westheim A, Wu AH, Duc P, et al. Impact of atrial fibrillation on the diagnostic performance of B-type natriuretic peptide concentration in dyspneic patients: an analysis from the Breathing Not Properly Multinational Study. *J Am Coll Cardiol*. 2005;46:838-44.
9. Sackett DL, Haynes RB. Evidence Base of Clinical Diagnosis: The architecture of diagnostic research. *BMJ*. 2002;324:539-41.
10. Halkin A, Reichman J, Schwaber M, Paltiel O, Brezis M. Likelihood ratios: Getting diagnostic testing into perspective. *QJM*. 1998;91:247-58.
11. Alberg AJ, Park JW, Hager BW, Brock MV, Diener-West M. The Use of "Overall Accuracy" to Evaluate the Validity of Screening or Diagnostic Tests. *J Gen Intern Med*. 2004;19:460-5.
12. Haynes RB. Department of Clinical Epidemiology and Biostatistics, McMaster University. Interpretation of diagnostic data: 2. How to do it with a simple table (Part A). *CMAJ*. 1983;129:559-65.
13. Christenson RH. Evidence-based laboratory Medicine – a Guide for Critical Evaluation of in vitro Laboratory testing. *Ann Clin Biochem*. 2007;44:111-30.
14. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Evidence base of clinical diagnosis: Designing studies to ensure that estimates of test accuracy are transferable. *BMJ*. 2002;324:669-71.