

Analýza dat v neurologii

XXVI. Hodnocení diagnostických testů – věrohodnostní poměr a diagnostický poměr šancí

V minulém díle seriálu jsme otevřeli novou kapitolu zaměřenou na hodnocení diagnostických testů. Představili jsme nové ukazatele správnosti a diagnostické síly testů, které produkují binární data (výstupy typu ano/ne; zdravý/nemocný apod.). Zároveň jsme doložili, že žádný jednotlivý ukazatel nepřináší zcela vyčerpávající informaci o kvalitě testu, a je tedy vysoce doporučeno je kombinovat. Krátce připomeneme hlavní ukazatele a jejich význam:

1. Sensitivita: schopnost testu správně diagnostikovat nemocného člověka. Vyjadřuje pravděpodobnost, že test bude pozitivní u skutečně nemocných lidí. Počítáme ji jako podíl počtu po-

zitivních výsledků testu u nemocných a celkového počtu nemocných lidí.

2. Specificita: schopnost testu negativně vyhodnotit zdravé pacienty. Vyjadřuje pravděpodobnost, že test bude negativní u zdravých lidí. Počítáme jako podíl počtu negativních výsledků testu u zdravých jedinců a celkového počtu zdravých lidí.

3. Pozitivní prediktivní hodnota: podíl správně pozitivních výsledků ze všech pozitivních výsledků testu. Vyjadřuje pravděpodobnost, že člověk je skutečně nemocný při pozitivním testu.

4. Negativní prediktivní hodnota: podíl správně negativních výsledků ze

L. Dušek, T. Pavlík, J. Jarkovský, J. Koptíková

Institut biostatistiky a analýz, Masarykova univerzita, Brno



doc. RNDr. Ladislav Dušek, Dr.
Institut biostatistiky a analýz
Masarykova univerzita, Brno
e-mail: dusek@cba.muni.cz

všech negativních výsledků testu. Vyjadřuje pravděpodobnost, že člověk je zdravý při negativním testu.

Popis problému: Je ověřován diagnostický test, na základě jehož výsledku se má lékař rozhodnout, zda je testovaný člověk zdravý, nebo nemocný. Negativní výsledek testu ukazuje na zdravého člověka, pozitivní výsledek naopak na nemocného. Výsledek testu je konfrontován se skutečností.

		Skutečnost		
		nemocný	zdravý	Celkem
Diagnostický test	pozitivní	48	36	84
	negativní	12	174	186
	Celkem	60	210	270

Pro vyjádření užitečnosti diagnostického testu můžeme využít tzv. věrohodnostní poměr (likelihood ratio, LR).

Pro pozitivní výstup testu vyjadřuje LR podíl pravděpodobnosti, že nemocný pacient je testem diagnostikován jako pozitivní proti pravděpodobnosti, že zdravý pacient je chybně diagnostikován jako pozitivní

$$LR+ = \text{senzitivita} / (1 - \text{specificita})$$

Pro negativní výsledek testu vyjadřuje LR podíl pravděpodobnosti, že nemocný pacient je testem chybně diagnostikován jako negativní proti pravděpodobnosti, že zdravý pacient je diagnostikován jako negativní

$$LR- = (1 - \text{senzitivita}) / \text{specificita}$$

Po dosažení do vztahu počítáme věrohodnostní poměr pro výsledky v příkladu:

$$LR+ = 0,800 / (1 - 0,829) = 4,67 \quad LR- = (1 - 0,800) / 0,829 = 0,24$$

Při standardním značení tabulky četností a–d (viz níže) lze výpočet věrohodnostního poměru vyjádřit následujícími vztahy:

$$LR+ = \frac{a}{b} \cdot \frac{a+c}{b+d} \quad LR- = \frac{c}{d} \cdot \frac{a+c}{b+d}$$

Skutečnost – stav pacienta

Diagnostický test	Skutečnost – stav pacienta	
	nemocný	zdravý
pozitivní	správná pozitivita SP	falešná pozitivita FP
negativní	falešná negativita FN	správná negativita SN

Skutečnost – stav pacienta

Diagnostický test	Skutečnost – stav pacienta	
	nemocný	zdravý
pozitivní	a	b
negativní	c	d

Vedle pravděpodobnostní interpretace lze na LR nahlížet také pomocí šancí. Vyjádření $a / (a + c)$ ve výpočtu $LR+$ totiž interpretačně odpovídá šanci a ku c , zatímco $b / (b + d)$ interpretačně odpovídá šanci b ku d . Ve výpočtu $LR-$ naopak vyjádření $c / (a + c)$ poměrem šancí interpretačně odpovídá šanci c ku a , zatímco $d / (b + d)$ interpretačně odpovídá šanci d ku b .

Příklad 1. Výpočet věrohodnostního poměru pro hodnocení diagnostického testu.

Popis: Tabulky srovnávají tři různé testy pro diagnostiku onemocnění, vždy při stejném počtu nemocných a zdravých pacientů.

Diagnostický test	Skutečnost		
	nemocný	zdravý	Celkem
pozitivní	54	12	66
negativní	6	198	204
Celkem	60	210	270

1. Počítáme hodnoty senzitivity a specifity spolu s 95% intervaly spolehlivosti

Senzitivita: 90,0 % (82,4 %; 97,6 %)
 Specifita: 94,3 % (91,1 %; 97,4 %)

2. Pro vyjádření užitečnosti testu počítáme věrohodnostní poměr (LR) a diagnostický poměr šancí (DOR)

$LR+ = 0,900 / (1 - 0,943) = 15,75$
 $LR- = (1 - 0,900) / 0,943 = 0,11$
 $DOR = 15,75 / 0,11 = 148,5$

Senzitivita i specifita nabývají vysokých hodnot. Rovněž věrohodnostní poměr indikuje, že diagnostický test je dobrý prediktor přítomnosti onemocnění.

Diagnostický test	Skutečnost		
	nemocný	zdravý	Celkem
pozitivní	35	85	120
negativní	25	125	150
Celkem	60	210	270

1. Počítáme hodnoty senzitivity a specifity spolu s 95% intervaly spolehlivosti

Senzitivita: 58,3 % (45,9 %; 70,8 %)
 Specifita: 59,5 % (52,9 %; 66,2 %)

2. Pro vyjádření užitečnosti testu počítáme věrohodnostní poměr (LR) a diagnostický poměr šancí (DOR)

$LR+ = 0,583 / (1 - 0,595) = 1,44$
 $LR- = (1 - 0,583) / 0,595 = 0,70$
 $DOR = 1,4 / 0,7 = 2,06$

Senzitivita i specifita nabývají nízkých hodnot. Nízká hodnota LR+ a vysoká hodnota LR- potvrzují, že diagnostický test je špatný prediktor přítomnosti onemocnění.

Diagnostický test	Skutečnost		
	nemocný	zdravý	Celkem
pozitivní	34	20	54
negativní	26	190	216
Celkem	60	210	270

1. Počítáme hodnoty senzitivity a specifity spolu s 95% intervaly spolehlivosti

Senzitivita: 56,7 % (44,1 %; 69,2 %)
 Specifita: 90,5 % (86,5 %; 94,4 %)

2. Pro vyjádření užitečnosti testu počítáme věrohodnostní poměr (LR) a diagnostický poměr šancí (DOR)

$LR+ = 0,567 / (1 - 0,905) = 5,95$
 $LR- = (1 - 0,567) / 0,905 = 0,48$
 $DOR = 6,0 / 0,5 = 12,42$

Diagnostický test má nízkou senzitivitu, nicméně vysokou specifitu. Test by byl využitelný za určitých podmínek, např. v situaci, kdy je nežádoucí falešná pozitivita.

Příklad 2. Hodnocení výsledků různě citlivých diagnostických testů a výpočet věrohodnostního poměru.

V tomto díle rozšíříme výše uvedenou sadu ukazatelů o tzv. **věrohodnostní poměr („likelihood ratio“, LR)**. Věrohodnostní poměr je podílem pravděpodobnosti výskytu jevu 1 za platnosti určité podmínky (jev 2 nastal) a pravděpodobnosti výskytu jevu 1, pokud daná podmínka neplatí (jev 2 nenastal). Z diagnostického hlediska tak hodnotíme věrohodnost, že výskyt sledovaného klinického znaku (symptomu) je vázán na výskyt určitého onemocnění. V matematické statistice věrohodnostní poměr obecně znamená podíl dvou pravděpodobností, které jsou výsledkem dvou různých modelů.

Pro hodnocení diagnostických testů můžeme věrohodnostní poměr definovat následovně:

$$LR = \frac{[\text{pravděpodobnost, že test dosáhne daného výsledku u nemocných pacientů}]}{[\text{pravděpodobnost, že test dosáhne daného výsledku u zdravých osob}]}$$

V našem hodnocení může diagnostický test nabídnout pouze dva typy výstupů, pozitivní a negativní. Rozlišujeme tedy LR+ (LR pro pozitivní test) a LR- (LR pro negativní test):

- Pro pozitivní výstup testu vyjadřuje LR podíl pravděpodobnosti, že nemocný pacient je testem diagnostikován jako pozitivní, a pravděpodobnosti, že zdravý pacient je chybně diagnostikován jako pozitivní.

$$LR+ = \text{senzitivita} / (1 - \text{specifita})$$

- Pro negativní výsledek testu LR vyjadřuje podíl pravděpodobnosti, že nemocný pacient je testem chybně diagnostikován jako negativní, a pravděpodobnosti, že zdravý pacient je diagnostikován jako negativní.

$$LR- = (1 - \text{senzitivita}) / \text{specifita}$$

Věrohodnostní poměr značený jako LR+ tedy vyjadřuje schopnost pozitivního výsledku testu odlišit nemocného a zdravého jedince. Naopak věrohodnostní poměr značený jako LR- vyjadřuje obrácenou situaci, kdy test mylně označí nemocného jedince za zdravého (viz příklad 1). Z definice tedy vyplývá, že u kvalitního testu je LR+ žádoucí co nejvyšší a naopak LR- co nejnižší. V literatuře lze najít zdůvodnění, že hodnoty LR+ > 10 a naopak hodnoty LR- < 0,1 umožňují věrohodné určení nebo vyloučení nemoci na základě daného testu (McGee, 2002). Vysoká hodnota LR+ jasně ukazuje na test, který poskytuje vysokou jistotu v určení nemoci, i přesto však není 100% zárukou, že osoba s pozitivním testem musí být vždy nemocná.

Výpočet věrohodnostního poměru pro různé výsledky diagnostického testu přibližují příklady 1 a 2. I věrohodnostní poměr je odhadem, a lze jej tedy doplnit výpočtem intervalu spolehlivosti, jak dokládá

Popis problému: Máme diagnostický test, u kterého chceme vyjádřit, jak dobře odlišuje pacienty se sledovanou nemocí od zdravých osob.

		Skutečnost		
		nemocný	zdravý	Celkem
Diagnostický test	pozitivní	43	31	74
	negativní	7	169	176
	Celkem	50	200	250

Používáme následující značení:

		Skutečnost	
		nemocný	zdravý
Diagnostický test	pozitivní	a	b
	negativní	c	d

Poznámka:

- ✓ Výpočet lze provést i pro kategoriální znaky, nejen pro dva výstupy diagnostického testu, např. pro kategorie hodnot laboratorního vyšetření.
- ✓ Věrohodnostní poměr – angl. likelihood ratio (LR)

Sackett DL. Interpretation of diagnostic data: 5. How to do it with simple maths. Can Med Assoc J 1983; 129(9): 947–954.
 Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample-size estimation for diagnostic-test studies. J Clin Epidemiol 1991; 44(8): 763–770.

Věrohodnostní poměr vyjadřuje kvalitu diagnostického testu:

- Počítejme **věrohodnostní poměr pozitivního testu**, který porovnává relativní podíl pacientů a zdravých osob, kteří mají **pozitivní diagnostický test**

$$LR+ = \frac{a / (a + c)}{b / (b + d)} = \frac{\text{senzitivita}}{1 - \text{specifická}}$$

- Dosažením do vztahu v bodě 1 dostaneme $LR+ = \frac{43 / (43 + 7)}{31 / (31 + 169)} = 5,55$

- Pro věrohodnostní poměr pozitivního testu počítáme **95% interval spolehlivosti**

$$\exp\left[\ln \frac{\text{senzitivita}}{1 - \text{specifická}} \pm 1,96 \sqrt{\frac{1 - \text{senzitivita}}{a} + \frac{\text{specifická}}{b}}\right]$$

- Dosažením do vztahu v bodě 3 dostaneme

$$\exp\left[\ln \frac{43 / (43 + 7)}{31 / (31 + 169)} \pm 1,96 \sqrt{\frac{7 / (43 + 7)}{43} + \frac{169 / (31 + 169)}{31}}\right] \rightarrow (3,94; 7,81)$$

- Počítejme **věrohodnostní poměr negativního testu**, který porovnává relativní podíl pacientů a zdravých osob, kteří mají **negativní diagnostický test**

$$LR- = \frac{c / (a + c)}{d / (b + d)} = \frac{1 - \text{senzitivita}}{\text{specifická}}$$

- Dosažením do vztahu v bodě 5 dostaneme $LR- = \frac{7 / (43 + 7)}{169 / (31 + 169)} = 0,17$

- Pro věrohodnostní poměr negativního testu počítáme **95% interval spolehlivosti**

$$\exp\left[\ln \frac{1 - \text{senzitivita}}{\text{specifická}} \pm 1,96 \sqrt{\frac{\text{senzitivita}}{c} + \frac{1 - \text{specifická}}{d}}\right]$$

- Dosažením do vztahu v bodě 7 dostaneme

$$\exp\left[\ln \frac{7 / (43 + 7)}{169 / (31 + 169)} \pm 1,96 \sqrt{\frac{43 / (43 + 7)}{7} + \frac{31 / (31 + 169)}{169}}\right] \rightarrow (0,08; 0,33)$$

Příklad 3. Odhad věrohodnostního poměru a jeho intervalu spolehlivosti.

příklad 3. Ačkoli jsou vztahy v příkladu 3 na první pohled poněkud složité, lze do nich jednoduše dosadit hodnoty z tabulky četností a interval spolehlivosti tak vypočítat.

Věrohodnostní poměr je často používán pro srovnávání výstupů různých testů, přičemž je souhrnně kalkulován tzv. **diagnostický poměr šancí („Diagnostic Odds Ratio“, DOR)**. DOR vyjadřuje šanci, že test bude pozitivní u nemocného člověka násobenou šancí, že test bude negativní u zdravého člověka.

$$DOR = LR+ / LR-$$

Hodnota DOR může růst od 0 do nekonečna, přičemž vyšší hodnoty znamenají lepší schopnost testu rozlišovat mezi nemocnými a zdravými jedinci, a tedy obecně větší schopnost dosahovat správného výsledku. Hodnoty menší nebo rovny 1 ukazují na test se slabou diagnostickou silou. Logicky požadujeme, aby

diagnostický poměr šancí byl co nejvyšší a zejména hodnoty 1 a nižší nejsou pro klinickou praxi přijatelné.

Pojem **šance (odds)** je blízký pojmu věrohodnostní poměr a vztahuje se k pravděpodobnosti, že nějaký jev nastane. Ačkoli v českém jazyce možná slovo šance nezní příliš exaktně, jde o jasně definovaný pojem, který se i v medicíně velmi často používá. Šanci, že nějaký jev nastane, vyjadřujeme i v běžné řeči formou poměru, například šance výskytu jevu A je 1 ku 4. To znamená, že jev A nastává s pravděpodobností $P(A) = 1 / 5 = 0,20$. Šanci nastání jevu A označujeme $O(A)$ (odds). Pro náhodný jev A potom obecně platí:

$$O(A) = \frac{P(A)}{1 - P(A)}$$

Výše uvedený vztah také někdy čteme jako „šanci ve prospěch A“. Ze vztahu je

patrné, že šance je podíl pravděpodobnosti nastání jevu A „ku“ pravděpodobnosti nenastání jevu A. Pokud pravděpodobnosti přímo odhadujeme z relativních četností jevů, např. z kontingenční tabulky, pak můžeme do vztahu dosadit přímo četnosti nastání jevu A a četnosti nenastání jevu A:

$$O(A) = \frac{\text{četnost jevu A}}{\text{četnost nenastání jevu A}}$$

Šance dvou jevů lze dávat do poměru, hovoříme o poměru šancí („odds ratio“). Ten může nabývat nejnižší hodnoty 0, v maximu ale nijak ohraničen není.

Vedle věrohodnostního poměru a diagnostického poměru šancí lze v literatuře najít i další ukazatele kvality diagnostických testů, které ale povětšinou vycházejí z odhadu senzitivity a specifické. Zde se pouze krátce zmíníme o dvou ukaza-

Tab. 1. Ukázky výpočtu různých ukazatelů správnosti diagnostických testů¹.

Diagnostický test	Značení četnosti				Skutečnost		Celkem
	pozitivní	negativní	Celkem	nemocný	zdravý	a + b c + d n	
				a	b		
	c	d	a + c	b + d			
Výsledek testu	Senzitivita (%)	Specifická (%)	PH+² (%)	PH-² (%)	LR+³	LR-³	DOR⁴
a b c d							
1 6 9 2	10,0 (0,3; 44,5)	25,0 (3,2; 65,1)	14,3 (0,4; 57,9)	18,2 (2,3; 51,8)	0,13 (0,02; 0,89)	3,60 (1,07; 12,17)	0,04
2 5 8 3	20,0 (2,5; 55,6)	37,5 (8,5; 75,5)	28,6 (3,7; 71,0)	27,3 (6,1; 61,0)	0,32 (0,08; 1,24)	2,13 (0,83; 5,50)	0,15
3 4 7 4	30,0 (6,7; 65,3)	50,0 (15,7; 84,3)	42,9 (9,9; 81,6)	36,4 (10,9; 69,2)	0,60 (0,19; 1,94)	1,40 (0,63; 3,13)	0,43
4 3 6 5	40,0 (12,2; 73,8)	62,5 (24,5; 91,5)	57,1 (18,4; 90,1)	45,5 (16,8; 76,6)	1,07 (0,33; 3,45)	0,96 (0,46; 2,01)	1,11
5 2 5 6	50,0 (18,7; 81,3)	75,0 (34,9; 96,8)	71,4 (29,1; 96,3)	54,6 (23,4; 83,3)	2,00 (0,52; 7,72)	0,67 (0,32; 1,39)	3,00
6 1 4 7	60,0 (26,2; 87,8)	87,5 (47,4; 99,7)	85,7 (42,1; 99,6)	63,6 (30,8; 89,1)	4,80 (0,72; 32,16)	0,46 (0,20; 1,02)	10,50
12 2 1 26	92,3 (64,0; 99,8)	92,9 (76,5; 99,1)	85,7 (57,2; 98,2)	96,3 (81,0; 99,9)	12,92 (3,37; 49,59)	0,08 (0,01; 0,55)	156

¹jednotlivé charakteristiky jsou uváděny jako hodnota s 95% intervalem spolehlivosti; ²prediktivní hodnota pozitivního, resp. negativního testu; ³věrohodnostní poměr (LR, likelihood ratio) pozitivního, resp. negativního testu; ⁴diagnostický poměr šancí (DOR, diagnostic odds ratio) jako poměr LR+ a LR-

telích, jejichž výpočet i interpretace jsou velmi jednoduché:

- Dosažená jistota („gain in certainty“) kalkulovaná jako prostý součet specifity a senzitivity (Connell a Koepsell, 1985). Součet může nabývat hodnot od 0 do 2, přičemž hodnoty blízké 1 indikují test, jehož výsledky odpovídají náhodnému tipování. Sčítání hodnot senzitivity a specifity je vhodné pro srovnávání různých testů, zvláště pokud byly testovány při různé prevalenci dané choroby.
- Jinou, podstatně starší mírou je tzv. Youdenův index (Youden, 1950), značený „J“ = senzitivita + specifita - 1. Tento výpočet lze alternativně vyjádřit jako 1 - (falešná pozitivita + falešná negativita)

Výpočty věrohodnostního poměru, diagnostického poměru šancí a dalších ukazatelů diagnostické hodnoty testu shrnuje pro různé datové soubory tab. 1.

Věrohodnostní poměr, jakož i další zmíněné míry nabízejí alternativní vyjád-

ření hodnoty diagnostických testů k hodnocení senzitivity a specifity. Je samozřejmě na autorovi, zda využije všechny tyto míry souběžně či nikoli. Čtenáře jistě napadne zcela legitimní otázka, není-li ukazatelů kvality testu příliš mnoho a jestli skutečně přispívají k interpretaci. Pravdou je, že žádný ukazatel nedá svou hodnotou 100% záruku, že při pozitivním testu je pacient skutečně nemocný. Výsledek je ovlivňován řadou faktorů, mimo jiné i velikostí testovacích souborů dat a způsobem jejich získání. I tyto faktory jsou důvodem vývoje různých ukazatelů, které nabízejí různé interpretace hodnoty testů. V literatuře nalezneme i prospektivní randomizované studie, jež řeší, zda jsou lékaři v praxi schopni správně využívat ukazatele kvality diagnostických testů pro svá rozhodování. Například práce Puhana et al z roku 2005 takto srovnávala senzitivitu a specifitu s věrohodnostním poměrem a došla k závěru, že jsou ve svém vlivu na rozhodování lékařů srovnatelné. Nicméně věrohodnostní poměr je stále poměrně málo využíván při praktické validaci diagnostických

testů (Steurer et al, 2002). Obdobně i diagnostický poměr šancí by si zasloužil častější aplikaci především pro svou jasnou interpretaci (Ingelfinger et al, 1987).

Tento díl seriálu jsme pojali více i jako rešerši relevantní literatury a výklad dokládáme větším počtem citací vědeckých prací. Udělali jsme to záměrně, abychom upozornili na překvapivý fakt, jak je svým způsobem velmi staré a jednoduché téma v současné literatuře živé. Důvodem je z velké části medicína sama. S nástupem personalizované medicíny přichází na trh velké množství nových biomarkerů a s nimi také roste heterogenita a cena diagnostických testů. Srovnávání různých testů je tedy velmi aktuální téma řešené z hlediska diagnostické síly i nákladové efektivity (Cornell et al, 2008; Hayen et al, 2010). Jde o trend, který bude s rozvojem genomiky a farmakogenomiky jistě pokračovat. Je stimulačím vidět, že i v dnešním přetechizovaném a složitém světě lze publikovat významnou vědeckou práci prostou diskuzí nad součtem dvou pravděpodobností ©.

Literatura

Connell FA, Koepsell TD. Measures of gain in certainty from a diagnostic test. *Am J Epidemiol* 1985; 121(5): 744–753.

Cornell J, Mulrow CD, Localio AR. Diagnostic test accuracy and clinical decision making. *Ann Intern Med* 2008; 149 (12): 904–906.

Hayen A, Macaskill P, Irwig L, Bossuyt P. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. *J Clin Epide-*

miol 2010; DOI:10.1016/j.jclinepi.2009.08.024. PMID 20079607.

Ingelfinge, JA, Mosteller F, Thibodeau LA, Ware JH. *Biostatistics in Clinical Medicine*. 2nd ed. New York: Macmillan 1987.

McGee S. Simplifying likelihood ratios. *J Gen Intern Med* 2002; 17(8): 646–649.

Puhan MA, Steurer J, Bachmann LM, ter Riet G. A randomized trial of ways to describe test accuracy: the effect on physicians' post-test probability estimates. *Ann Intern Med* 2005; 143(3): 184–189.

Sackett DL. Interpretation of diagnostic data: 5. How to do it with simple maths. *Can Med Assoc J* 1983; 129(9): 947–954.

Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991; 44(8): 763–770.

Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002; 324(7341): 824–826.

Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; 3(1): 32–35.

ERRATUM

V prvním čísle časopisu Česká a slovenská neurologie a neurochirurgie jsme omylem neuvedli jména recenzentů minimonografie *Monitoring pacientů s těžkým poraněním mozku* (autor prof. MUDr. Martin Smrčka, CSc.).

Tento článek recenzovali:

prof. MUDr. Roman Gál, Ph.D.
MUDr. Aleš Hejčl, Ph.D.
MUDr. Martin Filaun, CSc.

Všem recenzentům se omlouváme.