proLékaře.cz | 24.11.2026

Analýza dat v neurologii

XXX. Validita klinických testů v širším kontextu

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

Předcházející díly seriálu č. XXV–XXVIII jsme věnovali hodnocení různých diagnostických testů, u kterých jsme rozlišovali dva zásadní typy výstupů. Testy s binárním výstupem typu ano/ne vyžadují jiný analytický scénář než testy poskytující na výstupu spojitou, kvantitativní proměnnou. Takto pojaté třídění klinických testů je ale v jistém smyslu omezující, neboť pracuje pouze s jednorozměrnými diagnostickými testy, tedy s nástroji vedoucími k jediné proměnné, jejíž hodnoty mají co nejlépe odlišit nemocné a zdravé jedince. Problematika klinických šetření je však mnohem širší a zahrnuje i testy směřující k detailní typologii již nemocných jedinců, dále celou řadu skríningových testů, které mají vyloučit přítomnost jistého rizika, nebo dotazníková šetření, která na výstupu nabízejí více než jednu hodnotu a jsou složena z dílčích položek. Příkladem takových šetření může být hodnocení kvality života nebo kognitivních funkcí či posuzování stupně invalidity pacienta. Díl XXIX seriálu se zabýval hodnocením spolehlivosti těchto položkových testů.

V tomto díle problematiku hodnocení klinických testů zakončíme komplexním přehledem ukazatelů spolehlivosti a validity, které jsou aplikovatelné pro všechny typy klinických testů. Většina dosud probíraných ukazatelů kvality testů měřila především diagnostickou validitu (senzitivita, specificita, prediktivní hodnota, věrohodnostní poměr). Kromě tohoto hodnocení se v odborné klinické literatuře setkáváme s řadou dalších indikátorů spolehlivosti a validity, které stojí za to terminologicky přiblížit. S většinou těchto ukazatelů se lékaři setkávají, kdykoli je do praxe zaváděn nový laboratorní test, klinické prediktivní skóre nebo parametr hodnotící výsledky léčby. A naopak, týmy, které aspirují na vytvoření nového klinického testu nebo hodnotícího skóre, jsou nuceny číselně doložit komplexní validitu navrhovaného nástroje.

Tab. 1 přináší ucelený přehled skupin ukazatelů, které jsou při komplexním hodnocení určitého testu využitelné a měly by být vyžadovány především při vývoji testů nových. Tabulka obsahuje některé nové pojmy, jež jsme dosud nevysvětlili. To učiníme v následujícím textu. Především jde o **ukazatele validity kli-**

Tab. 1. Přehled hlavních kategorií ukazatelů kvality klinických testů.

Kategorie – ukazatel	Komentář
Diagnostická validita	Vyjadřuje schopnost testu rozlišit nemocné a zdravé jedince podle přítomnosti určitých jevů nebo měřením spojité proměnné. Nejčastěji používanými indikátory jsou senzitivita a specificita, věrohodnostní poměr a prediktivní hodnota testu. U testů se spojitými, kvantitativními výstupy používáme analýzu ROC křivek k stanovení hraničních bodů pro odlišení daných skupin subjektů.
Spolehlivost	Vyjadřuje stabilitu (konzistenci) výstupů testu při opakovaném měření za stejných podmínek, kdy se vlastnosti měřeného subjektu nemění. Hodnocení spolehlivosti souvisí s analýzou reprodukovatelnosti a opakovatelnosti testu. Existují různé metody měření spolehlivosti: test-retest přístup využívající opakované měření v čase, paralelní měření ekvivalentních forem testu, hodnocení shody více hodnotitelů (*inter-observer variability*). U složitých testů (typicky u dotazníkových šetření) je používána metoda odděleného hodnocení dílčích položek (*split-half design*) anebo hodnocení vnitřní konzistence (*reliability*) na základě Cronbachova koeficientu alfa.
Validita	Validita je komplexní ukazatel, který vyjadřuje schopnost testu měřit skutečně charakteristiku, kterou měřit chceme (tzv. konstrukt). Objektivní metody hodnocení jsou založeny na sledování shody výstupů testu s jinými proměnnými, které rovněž vyjadřují sledovanou charakteristiku či vlastnost subjektu: konstruktová validita a kriteriální validita. Méně objektivní metody dokládají validitu obsahu testu rozborem dostupných informací a literatury: obsahová validita, zjevná validita.
Praktická využitelnost	Využitelnost testu (*utility*) představuje významný rozměr hodnocení, které musí odpovídat situaci a kontextu nasazení testu. Jiná kritéria jsou aplikovatelná u skríningových testů a jiná pro testy diferencující symptomy již prokázané choroby. Do této kategorie patří dostupnost testu (včetně ceny), realizovatelnost měření v praxi, akceptovatelnost pro pacienta.

proLékaře.cz | 24.11.2026

nických testů a šetření (*validity*). Validitou rozumíme schopnost testu měřit jev (koncept), který skutečně chceme měřit. Jinými slovy, validita vyjadřuje, do jaké míry test skutečně měří to, co deklaruje, že měří. Validita testu je základní předpoklad toho, že z výstupů testu odvodíme správná rozhodnutí o vlastnostech hodnocených subjektů. Z těchto důvodů jsou také ve vztahu k validitě testů používány pojmy, jako je smysluplnost nebo využitelnost závěrů.

Validitu testů nesmíme zaměňovat za jejich spolehlivost (*reliability*); jde o odlišné atributy kvality. Spolehlivost vyjadřuje schopnost testu poskytovat stejné (konzistentní) výstupy při opakovaném měření za stejných podmínek, kdy se vlastnosti měřeného objektu (subjektu) nemění (viz díl XXIX seriálu). Hodnocení spolehlivosti je tedy zaměřeno na vlastní test (nástroj), zatímco hodnocení validity se zaměřuje spíše na obsah a interpretaci výstupů testu a na odvozená rozhodnutí. Spolehlivost testu je ovšem jednou z nutných podmínek validity.

Hodnocení validity testů je bohužel složitější než hodnocení spolehlivosti (viz díl XXIX seriálu). Máme-li objektivně posoudit, zda výstup testu skutečně sleduje charakteristiky, které chceme měřit, musíme nutně využít i jiné ukazatele dané charakteristiky (přímé i nepřímé ukazatele, nejlépe již nějak zavedené či standardizované). V této souvislosti je v literatuře často používán pojem **konstrukt** (*construct*). Konstrukt lze jednoduše vysvětlit jako pojem nebo záměr, který je předmětem výzkumu, experimentu či modelu. Složitější konstrukty mohou vznikat kombinováním jednoduchých pojmů, mezi kterými mohou být popsány vztahy v rámci dané teorie či konceptu. Podstatné je, že konstrukty (koncepty) používáme na teoretické, modelové úrovni; na úrovni empirické (při vlastním měření) hovoříme o sledovaných veličinách (měřených proměnných).

Příkladem konstruktu na teoretické úrovni může být např. deprese pacienta či jeho inteligence, které v praxi měříme pomocí konkrétních proměnných (klinická skóre, výstupy testů apod.). Příkladem komplexního, vícerozměrného konstruktu může být následující požadavek na hodnocení kvality života v klinické studii zahrnující pacienty po náročné operaci: „hodnocení kvality života komplexně pokryje všechny podstatné složky fyzického a duševního zdraví, s důrazem na vývoj v období rekonvalescence pacienta". V praxi takový konstrukt měříme standardizovaným dotazníkem kvality života, kde je každá komponenta zdravotního stavu pokryta příslušnými otázkami a vyjádřena číselným skóre.

Konstruktem bývá velmi často popisná charakteristika, která je teoreticky (koncepčně) definována, ale není v praxi přímo měřitelná nebo ji nelze postihnout jedinou měřenou proměnnou. Jako příklad uveďme konstrukty jako hyperaktivita dítěte, sociální inteligence či odolnost vůči stresu. Více klinickým příkladem konstruktu může být pokročilost nebo rizikovost nemoci, kterou vyjadřujeme pomocí různých prediktivních faktorů a skóre.

Pojem konstrukt jsme uvedli proto, že je základem velmi významného způsobu hodnocení kvality klinických testů. Hovoříme o tzv. **konstruktové validitě** (*construct validity*) testu. Tento ukazatel by měl dokládat, že výstup testu skutečně odráží (měří, vyjadřuje) záměr, proč dané měření provádíme. Je přitom lhostejné, zda jde o měření výzkumné nebo o hodnocení pacientů v klinické praxi. V obou případech bychom měli mít záměr (teoretický konstrukt) měření jednoznačně daný. Je patrné, že měření a vyjadřování konstruktové validity testů je složité, nicméně bývá recenzenty odborných časopisů často vyžadováno. Konstruktovou validitu nejlépe doložíme korelací mezi výstupem testu (měřený indikátor) a jinými proměnnými, které vyjadřují stejnou věc. Hodnotíme-li např. klinickým skóre pohyblivost pacientů, pak by lépe hodnocení pacienti měli skutečně lépe chodit nebo běhat. Při hodnocení rozlišujeme **konvergentní a diskriminační konstruktovou validitu**:

- **Konvergentní validita** (*convergent validity*) dokládá, že výstupy hodnoceného testu jsou ve vztahu (korelaci) k ostatním ukazatelům měřeného stavu či charakteristiky. Z toho usuzujeme, že hodnocený test nebo klinické šetření je ve vztahu k sledované charakteristice.
- **Diskriminační validita** (*discriminant validity*) naopak vyjadřuje situaci, kdy u testu a jeho výstupů vztah k nějaké charakteristice nebo jevu neočekáváme a v souladu s tím není vztah u provedených měření pozorován.

Konstruktová validita je objektivně měřitelná, neboť pracuje s více ukazateli sledovaného jevu a vzájemné vztahy těchto ukazatelů číselně vyjadřuje. V literatuře tedy bývá preferována více než jiné formy měření validity (Messick, 1995). Nejpřesvědčivější doklad validity předložíme, jsme-li schopni vyjádřit validitu daného testu jako jasně definovaný konstrukt, který má svou oporu v dalších stejně zaměřených ukazatelích.

Další objektivní formou vyjadřování validity je **kriteriální validita**, která dokládá shodu výstupu testu s již zavedeným (ověřeným, standardizovaným) ukazatelem měřené charakteristiky. Tato ověřená proměnná bývá někdy nazývána „kriteriální standard" nebo „zlatý standard". S již validovanou kriteriální proměnnou pracujeme buď souběžně v daném čase (souběžná validita, *concurrent validity evidence*) nebo je k dispozici až po určité době (prediktivní validita, *predictive validity evidence*).

Za méně objektivní formy hodnocení validity testů je považována **validita obsahová** (*content validity*). Její podstatou je doklad, že měření nebo výstup testu skutečně reprezentuje sledovanou charakteristiku nebo jev. Obsahovou validitu nedokládáme analýzou shody více ukazatelů, jako je tomu u validity konstruktové či kriteriální; její ověření je založeno na věcném rozboru struktury testu, jeho výstupů a opírá se např. o znalost určité teorie nebo o odbornou literaturu. Zvláštní formou je potom tzv. **zjevná validita** (*face validity*), kde je obsah a výstup testu validován míněním expertů. Tento doklad validity však bývá v literatuře kritizován jako neobjektivní a nedostatečný, především pro závazné klinické testy. Je také nutno zdůraznit, že obsahová a zjevná validita nejsou hodnotitelné statisticky, neboť jsou založeny na práci s externími informacemi a důkazy. Příkladem může být rozbor obsahu testu hodnotícího akutní toxicitu po léčbě. Obsahový rozbor teoreticky hodnotí, zda komponenty šetření pokrývají všechny relevantní a možné symptomy popsané v literatuře.

Hodnocení validity je samozřejmě triviální, někdy až zbytečné, pokud měříme přesně danou a kvantifikovatelnou hodnotu, jako je např. povrch těla, výbavnost či intenzitu signálů při elektrofyziologickém měření apod. V takovém případě je obsah výstupu měření jasně definován

proLékaře.cz | 24.11.2026

a hodnocení testu se zaměří spíše na spolehlivost a diagnostickou hodnotu. Pokud ale měříme komplexní charakteristiky, které integrují řadu faktorů (např. inteligenci, osobnostní charakteristiky, kvalitu života apod.), má správná interpretace výstupů testu klíčový význam. Ověření, zda používaný nástroj skutečně měří sledovanou charakteristiku (konstrukt), je zásadní. I proto koncept hodnocení validity zdomácněl především v psychologii, psychiatrii nebo sociologii, kde se pro jeho hodnocení používají i velmi sofistikované modely (např. model MIMIC Multiple Indicators – MultIple Causes). Rozsah tohoto textu neumožňuje detailní výklad podstaty těchto nástrojů, zájemci mohou nalézt velmi inspirativní vysvětlení např. v práci Kreidl (2005). Velká část testovacích nástrojů vyvíjených sociology či psychology našla ovšem významné uplatnění i v medicíně (např. Bjorner a Kristensen 1999; Mast et al 2002).

Hodnocení spolehlivosti i validity klinických testů úzce souvisí s jejich využitelností v praxi. Nespolehlivý test s výstupem, který má nadto diskutabilní interpretaci, by jistě nikdo nedoporučil jako klinický standard. Z těchto důvodů se při hodnocení klinických testů často objevují i pojmy vázané k praktické realizovatelnosti daného měření, jako vhodnost, reprezentativnost, dostupnost v běžné praxi (často spojována s cenou) a také akceptovatelnost pro pacienta. K ucelené teorii hodnocení testů existuje velmi rozsáhlá odborná literatura, jejíž základy položil již v 50. a 60. letech minulého století profesor L. J. Cronbach, kterého jsme představili v minulém díle jako autora koeficientu alfa pro měření položkové spolehlivosti testů.

Doufejme, že se nám podařilo výše uvedené pojmy dostatečně odlišit a že čtenářům nesplývají. V žádném případě nejde o hru se slovíčky. Ukazatele shrnuté v tab. 1 mají striktní definice a odpovídající metody měření. To jistě potvrdí každý odborník, který se pokusil navrhnout nový klinický test nebo hodnotící skóre; všechny kvalitní odborné časopisy v takovém případě vyžadují doložení komplexního panelu indikátorů diagnostické validity, spolehlivosti i validity výstupů testu.

Literatura

Kreidl M. Zhodnocení vlivu práce výzkumných agentur na konstruktovou validitu škál Sociologický časopis/Czech Sociological Review 2005; 41(1): 103–123.
Bjorner JB, Kristensen TS. Multi-Item Scales for Measuring Global Self-Rated Health. Investigation of Construct Validity Using Structural Equations Models. Research on Aging 1999; 21 (3): 417–439.
Mast BT, MacNeill SE, Lichtenberg PA. A MIMIC Model Approach to Research in Geriatric Neuropsychology: the Case of Vascular Dementia. Aging Neuropsychology and Cognition 2002; 9 (1): 21–37.
Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist 1995; 50(9): 741–749.
Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluation of patient-based outcome measures for use in clinical trials. Health Technology Assessment 1998; 2(14): 1–74.



Na sjezd je přihlášeno celkem 326 sdělení, z toho 280 přednášek v 55 odborných sekcích (z toho 21 topických symposií a 7 satelitních symposií) a posterová sekce s 46 příspěvky.

Bohatý doprovodný program zahrnuje Slavnostní zahájení v činoherní scéně MDB s muzikálem My fair lady (ze Zelňáku), Společenský večer s Ivou Bittovou v pavilonu A2 a posezení ve Vinárně u Elišky nebo v Pivovarské pivnici.

Těšíme se brzy na viděnou v Brně!

ČESKÁ NEUROLOGICKÁ SPOLEČNOST ČLS JEP

NEUROLOGICKÁ KLINIKA LF MU a FN BRNO

POŘADATEL SJEZDU
- Česká neurologická společnost ČLS JEP
- Česká společnost pro klinickou neurofyziologii ČLS JEP

SPOLUPOŘADATELÉ
- Slovenská neurologická spoločnosť SLS
- Slovenská spoločnosť pre klinickú neurofyziologii SLS
- Společnost biomedicínského inženýrství a lékařské informatiky ČLS JEP
- Neurologická klinika LF MU a FN Brno
- NeuroBrno, a.s.

ORGANIZAČNÍ SEKRETARIÁT SJEZDU
GUARANT International, spol. s r.o., Opletalova 22, 110 00 Praha 1
Tel.: +420 284 001 444, fax: +420 284 001 448, e-mail: csns2011@guarant.cz

www.csns2011.cz