

Sledování skeletálních změn radiologickými technikami

CLAUS-C. GLUER¹

ABSTRAKT

Dlouhodobá citlivost techniky, tj. její schopnost sledovat skeletální změny, je ovlivněna dvěma parametry: dlouhodobou chybou přesnosti (PE_{lt}) a poměrem odezvy specifickým pro skupinu subjektů (tj. ročního poměru změny). Obojí je třeba vzít do úvahy, aby se předešlo desinterpretaci měřených změn. Je navržena nová koncepce dlouhodobého měření na pomoc klinickému rozhodování, založená na třech typech měřítek: kritéria pro zjišťování změn—"nejmenší významná změna" (LSC) je nejmenší změna, která se má považovat za statisticky významnou, ale pro některé klinické otázky může být dostačující menší rozmezí, "rozmezí trendu hodnocení" (TAM - trend assessment margin); intervaly kontrolního sledování—na kontrolní vyšetření musí být pacienti zváni přibližně v intervalu určeném (podle populace) "intervalem monitorování" (MTI - monitoring time interval) nebo o jednu třetinu času dříve, podle "intervalu trendu hodnocení" (TAI - trend hodnocení interval), v závislosti na tom, zda může být rozhodnutí založeno na LSC nebo TAM; a standardní chyba přesnosti (stdPE)—čím menší je stdPE, tím citlivější je technika sledování skeletálních změn. Spolu tato tři měřítka poskytují dobrou charakterizační schopnost techniky sledovat skeletální změny. Ve srovnání s předchozími koncepcemi činí navrhovaná standardizace podle poměru odezvy místo měření rozptylu nebo míry odezvy chybu stdPE podstatně méně závislou na skupině subjektů. Umožňuje srovnání stdPE napříč různými studiemi a mohla by nahradit zavádějící koncepci vyjadřování přesnosti jako koeficientu variace. Aplikace této koncepce by měla usnadnit interpretaci měřených skeletálních změn. (J Bone Miner Res 1999;14:1952-1962)

¹ Arbeitsgruppe Medizinische Physik, Klinik für Diagnostische Radiologie, Universitätsklinikum an der Christian-Albrechts-Universität zu Kiel, Kiel, Německo.

ÚVOD

Pro vyhodnocení progresu nemoci, odezvy na léčení a odhad rizika fraktury je důležité interpretovat změřené změny hustoty minerálu v kosti (BMD) a jiné skeletální parametry uvážlivým způsobem. Densitometrie kosti je přesná metoda, ale vinou omezení techniky změřené výsledky pouze přibližně vyjadřují skutečné změny. Dlouhodobá citlivost techniky, definovaná jako schopnost sledovat změny ve skeletálním statutu,⁽¹⁾ je limitována nepřesností techniky. Aby se umožnilo srovnání nepřesností technik specifikovaných v různých jednotkách, chyba přesnosti se často uvádí jako procento, vypočítané např. jako koeficient variace (CV) opakovaných měření. Avšak je známo, že zdánlivá srovnatelnost procentních jednotek může být zavádějící a proto byly navrženy různé způsoby standardizace chyb přesnosti⁽²⁻⁸⁾: dělení chyby přesnosti variací populace, rozsah 10-90 %, normální pokles za rok v závislosti na věku atd. Do této chvíle nebyla žádná z těchto metod standardizace důkladně prozkoumána ve srovnání s jinými přístupy, natož aby byla jednomyslně přijata. Chybí standardizace metod, jak standardizovat přesnost. K posouzení výhod a omezení soupeřících přístupů je třeba definovat problémy a ukázat na cíle standardizace: Jaký smysl má standardizace chyb přesnosti?

Když hodnotíme dlouhodobé změny v čase, je často třeba se při klinickém rozhodování věnovat těmto třem otázkám:

- Interpretace měřených změn: jsou kalkulované změny významné a klinicky relevantní?— Náhodné kolísání se někdy mylně považuje za skutečnou změnu.
- Plánování kontrolní návštěvy pro určení míry změny: Jaký časový interval je požadován pro přesné vyhodnocení odezvy na léčení nebo progresu nemoci?—Kontrolní měření provedené příliš brzo neumožňuje posoudit významnost změřených změn.
- Srovnání technik: Která technika se nejlépe hodí ke zjištění změn přesně a rychle?— Používá se matoucí množství nedostatečných metod pro standardizaci chyb přesnosti.

Abychom motivovali pro navrhovanou koncepci, podáváme několik vysvětlení týkajících se obtížnosti odpovědi na třetí otázku. Jednoduchá odpověď, že pro monitorování změn v čase se nejlépe hodí technika s nejlepší reprodukovatelností nebo nejmenší chybou přesnosti, je chybná. Především je zřejmé, že chyba přesnosti různých parametrů měření se nemůže přímo srovnávat, je-li specifikovaná v absolutních jednotkách (např. v g/cm² vs. mg/cm³ vs. dB/MFz vs. m/s). Vyjádření chyby přesnosti procentuálně je celkem oblíbené, ale neřeší tento problém. Zcela naopak: tím, že implikuje, že jsou nyní výsledky ihned srovnatelné, může být tato častá procentní jednotka velmi zavádějící; stačí změnit definici parametru (např. přidáním kompenzace odchylky) a může být dosažena jakákoli úroveň procentní přesnosti. Například hodnoty původně přímo kalkulovaného útlumu širokopásmového ultrazvuku (BUA) se typicky pohybují v rozsahu mezi 30 dB/MHz a 80 dB/MHz a chyba přesnosti, např. 2 dB/MHz by přinesla procentní chybu přesnosti 2,5 % (2/80 x 100) až 6,7 % (2/30 x 100). U některých kvantitativních ultrazvukových přístrojů (QUS) podléhají původní hodnoty kompenzaci odchylky. Vezmeme-li pro jednoduchost odchylku 50, výsledný rozsah bude 80-130 dB/MHz. Tato jednoduchá manipulace by snížila chybu přesnosti na rozsah 1,5 % (2/130 x 100) až 2,5 % (2/80 x 100), aniž by došlo ke skutečnému zlepšení dlouhodobé citlivosti. Tyto příklady také ilustrují druhý problém, že se procentní chyba přesnosti zdá být lepší (tj. nižší) u zdravých subjektů jednoduše kvůli většímu jmenovateli (např. 2,5 % vs. 6,7 %). To opět neodráží žádnou skutečnou diferenci v dlouhodobé citlivosti.

V hlubším kontextu je však třeba uznat, že chyba přesnosti sama o sobě vypovídá málo o schopnosti techniky sledovat změny: pro určení dlouhodobé citlivosti je třeba vzít do úvahy také citlivost monitorovacích parametrů. Dobrým příkladem pro demonstraci, co se stane, když je toto hledisko přehlíženo, jsou parametry QUS. Změny v rychlosti zvuku (SOS) jsou typicky

v rozsahu pouze několika metrů za sekundu za rok z asi 1500-2000 m/s, což reflektuje reakci menší než 1 % za rok; když porovnáme změny v množství BUA o několik decibelů na megahertz za rok z asi 50-100 dB/MHz, odráží to odezvu ve výši několika procentních bodů. Není překvapením, že procento chyby přesnosti pro SOS je typicky menší než pro BUA nejméně o jeden řád velikosti. To však nezbytně nepředstavuje ekvivalentní výhodu ve schopnosti SOS sledovat změny, protože nebyla vzata do úvahy nižší citlivost SOS. Tak se zdá, že dělení procentních chyb nějakým měřítkem citlivosti pro dosažení dlouhodobé citlivosti, nazývané standardizovanou chybou přesnosti, je užitečné a byly navrženy různé metody. Takže jaký je problém s tímto typem standardizace?

Jedním problémem je, že různá měřítka citlivosti vedou k rozdílně standardizované chybě přesnosti, z nichž většina je stále podstatně zkreslená, např. ovlivněna kohortou. Vezmeme si jako typický příklad jednu z běžných metod standardizace chyby přesnosti, tj. dělení standardní deviací hodnot skupiny subjektů. V případě úzce definované skupiny subjektů (např. mladí normální) bude výsledná "standardizovaná chyba přesnosti" větší než chyba kalkulovaná pro smíšenou skupinu zdravých a osteoporotických jedinců, i když technika měla identickou přesnost (vyjádřenou v absolutních jednotkách) pro zdravé i osteoporotické subjekty. Velikosti vzorků pro studie přesnosti jsou typicky celkem malé. Výběr subjektů proto může snadno vnést podstatné zkreslení. Skutečně můžeme snadno "zlepšit" standardizovanou přesnost jednoduše přidáním několika extrémnějších případů (velmi zdravý nebo velmi osteoporotický) ke skupině subjektů. Dokud omezujeme srovnání přesnosti technik na měření, která byla všechna získána ze stejné skupiny subjektů, je tento typ standardizace užitečný. Ale jakmile je tato "standardizovaná" přesnost považována za univerzální charakteristiku, která popisuje schopnost techniky sledovat změny v jakékoli skupině subjektů, je zde prostor pro desinterpretace. Je jasné, že standardizované chyby přesnosti kalkulované z různých skupin subjektů se nemohou přímo srovnávat; nebyly ve skutečnosti vůbec standardizované smysluplným způsobem.

Co je třeba udělat pro vyřešení tohoto problému? Můžeme vyjmenovat čtyři požadavky pro užitečné způsoby standardizace: měřítko musí reflektovat jak nepřesnost, tak citlivost; měřítko musí umožnit přímé srovnání výkonů techniky testované v různých studiích; měřítko musí mít intuitivní klinický význam; a měřítko musí být co nejméně citlivé k odchýlkám ve výběru subjektů.

Citlivost se různí pro různá pohlaví, věkové skupiny a terapie; může se lišit i reprodukovatelnost. Takto je třeba zkoumat tyto rozdílné kohorty odděleně, aby se zjistila příslušná hladina standardizované přesnosti. Důsledkem je, že standardizovaná přesnost jako měřítko, které reflektuje reprodukovatelnost i citlivost, již nemůže být vyjádřena jediným číslem.

A nakonec, kvalita odhadu standardizované přesnosti bude záviset na typu struktury analyzované studie. Jako obvykle jsou výsledky odvozené z dlouhodobých studií preferovány před údaji napříč sekcemi.

Budeme-li mít na paměti tato varování, měla by tato navrhovaná koncepce usnadnit objektivní hodnocení schopnosti techniky sledovat dlouhodobé změny.

MATERIÁLY A METODY

Interpretace měřených změn:

Zavedení kritérií změny

Pro klinické rozhodování je důležité vědět, jaká velikost měřených změn je požadovaná pro jistotu, že pacient skutečně ztratil kostní hmotu. Jinými slovy, která změna je statisticky významná, když vezmeme do úvahy limity výkonnosti přístroje? Jak již dříve ukázali jiní autoři,⁽⁹⁾ pro měření dvou bodů v čase mohou být za kritérium skutečné změny (se

spolehlivostí 95 %) považovány pouze změny přesahující 2,8 krát chybu přesnosti techniky. Odpovídající kritérium změny bylo nazváno "nejmenší významná změna" (LSC):

$$LSC = 2,8 \times PE_{It}$$

kde PE_{It} je největší chyba přesnosti použité techniky.

Avšak kliničtí lékaři musí vybalancovat přání dosáhnout statistické jistoty s potřebou pacienta dostat léčení co nejrychleji, je-li pro něj platná indikace. Aby se nezdržovala potencionálně důležitá medikace, musí se klinický lékař spokojit s úrovní spolehlivosti < 95 %.⁽¹⁰⁾ Skutečně je představitelné, že kolísající limity spolehlivosti mohou být v různých klinických situacích vhodné. Například, když identifikujeme někoho, kdo skutečně reagoval na terapii v situaci, kdy se reakce očekává, pak může být požadovaná spolehlivost poněkud menší. Avšak v situaci, kdy se zvažuje změna průběhu terapie, může klinický lékař požadovat 95 % spolehlivost, aby změnil zákrok. Statisticky mohou být definovány intervaly pro každou úroveň spolehlivosti. Abychom se vyhnuli nadbytku různých úrovní spolehlivosti, navrhuje zavést jedno další, méně striktní kritérium změny, "rozmezí trendu hodnocení" (TAM). Je-li dáno

$$TAM = 1,8 \times PE_{It}$$

může to být považováno za kritérium skutečné změny při hladině spolehlivosti 80 % pro dvoustranné testy nebo 90 % pro jednostranné testy. Slovo "trend" by mělo implikovat, že je třeba splnit méně přísné požadavky, než pro test s významností při 95 % úrovni spolehlivosti.

Obě kritéria změny, LSC i TAM, se musí kalkulovat s použitím dlouhodobé, nikoli krátkodobé chyby přesnosti specifikované pro měření in vivo ve srovnatelné skupině subjektů.

Plánování kontrolních návštěv: Zavedení intervalů kontrolního sledování

Po stanovení počátečního statutu parametru (např. BMD) je třeba určit míru změny tohoto parametru, aby se mohla vyhodnotit progresa nemoci nebo odezva na léčení. Jaký časový interval mezi počátečním a následným měřením je dostatečný pro přesné a platné hodnocení? Když budeme hledat odpověď na tuto otázku, budeme čelit dilematu, zda se máme rozhodnout pro rychlou odpověď při časně kontrolní návštěvě spojené s větší statistickou nejistotou při odhadování skutečné změny ze změřené změny, nebo pro solidnější odpověď při pozdější návštěvě s rizikem podstatné ztráty kostní hmoty a fraktur v mezidobí. Proto, podobně jako v předešlém oddíle, se mohou definovat dva různé intervaly kontrolního sledování.

"Časový interval sledování" pro hodnocení progresa nemoci (MTI_p) je odhad časového období, po kterém polovina pacientů s normální ztrátou kostní hmoty vykáže změřenou změnu překračující kritérium změny LSC. To je dáno vzorcem:

$$\begin{aligned} MTI_p &= LSC / \text{medián odezvy} \\ &= 2,8 \times PE_{It} / \text{medián změny za rok} \end{aligned}$$

Podobně "interval trendu hodnocení" (TAI) je odhad (kratšího) sledovacího období, po kterém polovina pacientů s normálním úbytkem kostní hmoty bude demonstrovat změny překračující kritérium TAM. Je dán vzorcem:

$$\begin{aligned} TAI_p &= TAM / \text{medián odezvy} \\ &= 1,8 \times PE_{It} / \text{medián změny za rok} \end{aligned}$$

Například pro techniku s dlouhodobou chybou přesnosti $PE_{It} = 1,5 \%$ a pacienta, u kterého se dá očekávat roční změna 1 %, bude TAI_p a MTI_p 2,7 let a 4,2 let. Pro subjekt s rychlejší očekávanou roční ztrátou 3 % bude TAI_p a MTI_p 0,9 roku a 1,4 roku.

Při plánování měření pacienta pro vyhodnocení odezvy na léčení lze použít podobnou strategii. V předchozím oddíle byla stanovena kritéria, podle kterých lze usuzovat, že pacient pozitivně reagoval na léčení: jestliže změřená změna byla větší, než normální úbytek před léčbou, nejméně o TAM nebo LSC. V kterém okamžiku lze očekávat, že se tak stane u většiny léčených pacientů? MTI_T a TAI_T pro léčbu se mohou definovat jako:

$$\begin{aligned} MTI_T &= LSC/\text{medián odezvy na léčení} \\ &= 2,8 \times PE_i/\text{medián zlepšení vs. placebo za rok} \end{aligned}$$

$$\begin{aligned} TAI_T &= TAM/\text{medián odezvy na léčení} \\ &= 1,8 \times PE_{it}/\text{medián zlepšení vs. placebo za rok} \end{aligned}$$

Index "T" znamená léčení, ale měl by se specifikovat podle zkoumané léčby. Jestliže se například očekává, že estrogen zlepšit kost o 3 % za rok (medián), zatímco neléčení jedinci ztratí kostní hmotu průměrným tempem -1 % za rok, medián léčebného účinku bude 4 % za rok a pro techniku s 1,5 % chybou přesnosti bude doporučený TAI_{estr} a MTI_{estr} 8,1 a 12,6 měsíců. Po těchto obdobích bude střední zisk BMD 2,02 % a 3,15 %, což představuje hodnoty, při kterých můžeme mít 80 % nebo 95 % jistotu, že subjekt skutečně ztrácí kostní hmotu menším než normálním tempem (2,02 % = -0,68 % +2,7 % a 3,15 % = -1,05 % - 4,2 %).

Srovnání technik: Zavedení redefinované chyby přesnosti

Jak MTI , tak TAI definované v předchozím oddíle jsou vhodné pro měřítko charakterizující schopnost techniky sledovat skeletální změny: čím kratší je MTI a TAI , tím lepší je dlouhodobá citlivost. Alternativně se může dlouhodobá citlivost vyjádřit chybou přesnosti, jestliže je upravena pro rozdíly v citlivosti. Taková standardizační procedura by umožnila setrvat u známé koncepce vyjádření chyby přesnosti v procentech (místo roků pro TAI a MTI). To usnadňuje interpretaci, protože to je druh měřítka, na která je většina výzkumníků a klinických lékařů zvyklá.

Standardizace lze dosáhnout opravou chyby přesnosti techniky A zkoumané podle poměru odezvy (rr); rr se udává jako poměr míry odezvy referenční techniky R dělený mírou odezvy techniky A:

$$rr(AvsR) = \text{míra odezvy (R)}/\text{míra odezvy (A)}$$

"Standardizovaná chyba přesnosti," sPE_{it} techniky A, která byla standardizovaná relativně k referenční technice R, je dána vzorcem:

$$\begin{aligned} sPE_{it}(AvsR) &= PE_{it}(A) \times rr(AvsR) \\ &= \text{míra odezvy (R)}/\text{míra odezvy (A)} \end{aligned}$$

Jakmile se standardizuje tímto způsobem, může se standardizovaná chyba přesnosti nyní přímo srovnávat s chybou přesnosti referenční techniky R.

Tato metoda standardizace transformuje chybu přesnosti techniky A do stupnice referenční techniky R. Vynásobením rr se standardizace chyby přesnosti stává skutečně srovnatelnou napříč technikami. Všechny chyby přesnosti technik A, B, C..., které byly standardizované tímto způsobem, se nyní mohou přímo srovnávat navzájem a také s chybou přesnosti referenční techniky (která se podle definice rovná standardizované chybě přesnosti, protože je standardizovaná na sama sebe).

Například jestliže přístroj QUS má (dlouhodobou) chybu přesnosti pro SOS 0,3 % a chceme ho srovnat s hlášenou výkonností BUA tohoto přístroje, v tomto případě nastavenou na 1,5 %, budeme standardizovat jeden nebo jiný parametr versus druhý parametr. Označme (arbitrárně) BUA jako referenční techniku. Chyba přesnosti SOS bude standardizovaná vynásobením rr BUA versus SOS. Jestliže se například zjistí, že se rovná 5 (tj. roční změna

BUA je pětikrát větší než u SOS), bude standardizovaná chyba přesnosti SOS 1,5 %, tj. rovna chybě přesnosti BUA. Oba přístroje budou mít v tomto příkladu stejnou dlouhodobou citlivost.

Kdybychom místo toho stanovili SOS jako referenční techniku, chyba přesnosti BUA by se musela standardizovat. V tomto případě r je 0,2 a standardizovaná chyba přesnosti BUA by se rovnala $1,5 \times 0,2 = 0,3$ %, tj. opět by byla stejná jako chyba přesnosti SOS. Nehledě na to, která technika byla vybrána jako referenční technika, výsledek "rovná standardizovaná přesnost" zůstává týž. Avšak stupnice standardizované chyby přesnosti závisí na volbě referenční techniky. V prvním případě jsme kalkulovali standardizovanou chybu přesnosti 1,5 % pro obě techniky, zatímco kdybychom přehodili referenční techniku, byla by standardizovaná chyba přesnosti 0,3 %.

V důsledku se musíme dohodnout na výběru univerzální referenční techniky, aby byly techniky skutečně srovnatelné napříč studii. Navrhujeme používat jako referenční techniku posterior-anteriorní rentgenové měřiče absorpce s dvojí energií pro měření lumbální páteře (DXA_{sp}), protože jsou v dlouhodobých studiích nejpoužívanější. Abychom jasně pojmenovali tuto volbu, navrhujeme nazývat standardizovanou chybu přesnosti techniky A, která byla standardizovaná versus DXA_{sp} , "standardní chybou přesnosti," $stdPE_{it}(A)$:

$$stdPE_k(A) = sPE_{,t}(A vs DXA_{sp})$$

Používání standardní chyby přesnosti by se mělo preferovat před standardizovanou chybou přesnosti, kdykoli je to možné, tj. kdykoli výzkumná technika a DXA_{sp} se mohou měřit u stejných subjektů. Míra odezvy výzkumné techniky a DXA_{sp} se musí získat na stejných subjektech.

Jestliže techniky A a R mají různé jednotky (např. m/s a g/cm²), je třeba vyjádřit chybu přesnosti i míru odezvy procentuálně. Jestliže techniky A a R mají stejné jednotky, může být standardizovaná přesnost alternativně také hodnocena v absolutních jednotkách, ale pak chyba přesnosti i míra odezvy se musí vyjádřit konzistentně v absolutních jednotkách. Aby byly všechny rovnice co neuniverzálněji aplikovatelné, jsou všechny chyby přesnosti vyjádřeny procentuálně v celém zbytku tohoto rukopisu.

Poměry odezvy budou pravděpodobně méně závislé na skupině subjektů než míry odezvy (část zkreslení kohorty se anuluje). Přesto může pohlaví a etnická skupina, zdravotní stav (zdravý, osteopenický, osteoporotický atd.), a—pokud lze aplikovat—typ a dávkování léčby, mít určitý dopad a proto je třeba je specifikovat. Standardní chyba přesnosti se tak může lišit a klasifikace dlouhodobé citlivosti může záviset na kohortě. Proto se musí prozkoumat tyto scénáře, dříve než se bude moci vynést obecné prohlášení o klasifikační technice:

- Dlouhodobá citlivost pro zjišťování normálních procesů stárnutí;
- Dlouhodobá citlivost pro zjišťování progresu nemoci u osteoporotických jedinců; a
- Standardizovaná dlouhodobá chyba přesnosti pro zjišťování změny vyvolané léčbou, která je specifická pro léčení a proto je třeba specifikovat druh léčby a dávkování.

Aplikace koncepcí

Abychom ilustrovali jejich užitečnost, použili jsme odvozené koncepce na data z literatury. Krátkodobá chyba přesnosti (protože dlouhodobá chyba přesnosti není dosud pro QUS stanovena) a typické míry odezvy byly sebrány pro dva parametry DXA a dva QUS. Protože účelem této přednášky není srovnávat techniky, ale představit koncepce, udávaná čísla je třeba brát pouze jako příklady aplikace koncepce, nikoli jako hodnocení dlouhodobé citlivosti těchto čtyř parametrů.

TABULKA 1. HYPOTETICKÝ PŘÍKLAD PRO ODVOZENÉ KONCEPCE

<i>Technika</i>	<i>Míra odezvy (% p.a.)</i>	<i>Chyba přesnosti (%)</i>		<i>Kritéria změny (%)</i>		<i>Doba sledování (roky)</i>		<i>Standardní chyba přesnosti (%)</i>
		<i>PE_{st}</i>	<i>TAM</i>	<i>LSC</i>	<i>TAI</i>	<i>MTI</i>	<i>stdPE_{st}</i>	
BMD _{spine}	0,9	0,7	1,3	2	1,4	2,2	0,7	
BMD _{femtot}	0,6	0,7	1,3	2	2,2	3,3	1,1	
SOS _{calc}	0,07	0,16	0,29	0,45	4,1	6,4	2,1	
BUA _{calc}	0,3	1,2	2,2	3,4	7,3	11,3	3,6	

Skeletální parametry zahrnují hustotu minerálu v kosti (BVID) měřenou posterior-anteriorním rentgenovým měřením s dvojitou energií (DXA) v lumbální páteři (BMD_{spine}), DXA celého proximálního femuru (BMD_{femtot}), rychlosti zvuku (SOS_{calc}) a útlumu širokopásmového ultrazvuku (BUA_{calc}) patní kosti. Přes velké rozdíly v neopravené krátkodobé chybě přesnosti (PE_{st}) a míry odezvy, mohou být napříč technikami srovnávány parametry reflektující dlouhodobou citlivost, such interval trendu hodnocení (TAI), časový interval sledování (MTI) a standardní krátkodobá chyba přesnosti (stdPE_{st}). Kritéria změn jako rozpětí trendu hodnocení (TAM) a nejmenší významná změna (LSC) poskytují prahové hodnoty pro hodnocení, zda došlo k významným změnám na úrovni 95 % a 80 % spolehlivosti (dvoustranné testy).

VÝSLEDKY

Navrhované koncepce byly aplikovány na hypotetické výsledky pro dva DXA přístupy (BMD posterior-anteriorní DXA lumbální páteře, BMD_{spine} a DXA celého proximálního femuru, BMD_{femtot} a dvou přístupů k QUS (SOS a BUA patní kosti) prezentované v tabulce 1.

DISKUSE

Potřeba nových koncepcí

"Hledání rozdílu může být pro současný výzkum jako hledání kamene mudrců pro alchymii nebo Svatý grál pro rytíře z okouzlujících legend – nepostižitelné, a příliš často iluzorní."⁽¹¹⁾ Tucet let poté, co Robert Heaney upozornil na tuto záležitost, zůstává jeho hodnocení z větší míry pravdivé. Mezitím přibyly důležité příspěvky, ale v klinické praxi přetrvává dodnes značný zmatek ohledně interpretace měřených změn a technik srovnávání výsledků. S rostoucím používáním ultrazvuku se tyto problémy ještě posílily, protože přesnost QUS a densitometrie kosti se nedá snadno porovnat kvůli různým jednotkám a klamným představám o procentuálním vyjádření. Při řešení těchto problémů byla vyvinuta nová koncepce na pomoc klinickým lékařům při rozhodování o sledování a léčení jednotlivých pacientů. Výzkumníci mohou těžit z toho, že získali nástroj pro objektivnější způsoby srovnávání dlouhodobé citlivosti techniky. Koncepční centra kolem tří otázek byla popsána v úvodním oddíle. Toto jsou otázky a komponenty navrhované koncepce jako řešení:

- Interpretace měřených změn: jsou kalkulované změny významné a klinicky relevantní? Navrhovaná odpověď: Ano, jestliže překračují kritéria změny LSC nebo TAM.
- Plánování kontrolní návštěvy pro určení míry změn: Jaký časový interval je třeba pro přesné vyhodnocení odezvy na léčení nebo progresu nemoci? Navrhovaná odpověď: znovu vyšetřete pacienta po intervalu kontrolního sledování MTI nebo TAI.
- Srovnání technik: Která technika se nejlépe hodí pro zjištění změny přesně a rychle? Navrhovaná odpověď: technika s nejnižší standardní chybou přesnosti (stdPE).

Také čtyři požadavky na užitečné způsoby standardizace vyjmenované v úvodu jsou převážně splněny. Odchytky při výběru subjektů jsou stále problémem pro MTI, ale pouze proto, že je míněn jako specifický pro populace s různými mírami změny. Pro stdPE je tento problém minimální, pokud míry odezvy použité pro kalkulaci rr byly získány od stejných jedinců pro obě techniky. Jestliže tomu tak není, je třeba porovnávat podobné populace s opatrností. StdPE se nejlépe hodí pro přímé srovnání různých technik, i napříč různými studii.

Všechny tři parametry mají poměrně intuitivní význam. Změna menší než TAM nemůže být interpretována jako klinicky relevantní; změna menší než LSC není statisticky prokázána změna. Sledovací interval kratší než TAI nebo MTI přinese ve většině případů nedostatečné změny. StdPE může být snadno interpretována, protože stupnice je jednoduchá a známá: výsledek stdPE 1-1,5 % je možno považovat za celkem dobrý. To je podobné úrovni přesnosti hlášené v mnoha studiích pro DXA_{sp}, která je známá většině výzkumníků.

Odvozené koncepce nejsou omezeny na radiografické diagnostické přístupy. Kritéria změn, intervaly kontrolního sledování a standardní chyba přesnosti by také mohly být kalkulovány například pro markéry kostního metabolismu. Obrovská diference v míře odezvy a chybě přesnosti pro markéry versus radiografické parametry není překážkou, protože se vyruší při kalkulaci intervalu kontrolního sledování nebo standardní chyby přesnosti. Proto standardní chyba přesnosti markéru kostní resorpce se do budoucna může udávat přímo s odpovídajícími výsledky pro radiografické parametry.

Plánování kontrolních návštěv

Interpretace MTI (nebo TAI) jako měřítka dlouhodobé citlivosti je intuitivní a jednoduché: představuje čas pro sledování, požadovaný k otestování, zda došlo ke klinicky relevantní změně. Čím kratší je MTI, tím citlivější je technika.

Přesto je třeba uvést několik varování. Zaprvé, není jediný MTI (nebo TAI) pro každou techniku. Velikost parametru bude zřejmě rozdílná pro studie progresu nemoci (a opět mezi normálním a rychlým úbytkem) a odezvy na léčení (zde to může také záviset na typu zkoumané léčby). Když sledujeme druhou záležitost, můžeme si také všimnout, že odezva na léčení je zcela variabilní i pro zavedenou účinnou medikaci jako estrogen.^(12 13) Podle definice bude polovina pacientů vykazovat odezvu, která je menší než medián odezvy, and jako důsledek bude jejich měřené zlepšení během MTI_T (nebo TAI_T) menší než LSC (nebo TAM).

Pacienti, kteří nedosáhli úrovně změny očekávané po MTI_T (nebo TAI_T) mohou přesto těžit z léčby, i když na poněkud nižší úrovni. Jak interpretujeme takovou "negativní" nedostatečnou odezvu? Jak zjistíme skutečné nerespondéry? Dokud je pacientova měřená změna "lepší" než úbytek očekávaný bez léčení, bude pacient spíše těžit z léčby než ne. Avšak statistická nejistota bude nepříjemně vysoká. V závislosti na zdravotním stavu pacienta můžeme přesto brát vzestupnou tendenci jako povzbuzující a naplánovat další kontrolní návštěvu za dvojnásobek MTI_T (nebo TAI_T). V takovém okamžiku se dá očekávat, že i pacienti s pouhou polovinou mediánu míry odezvy (ve srovnání mezi léčenými a neléčenými pacienty) vykážou změnu, která překročí LSC (nebo TAM). Podle publikovaných studií to splní ~60 % pacientů na estrogeneru^(12,14) a ~80 % pacientů na alendronátu,⁽¹⁵⁾ za předpokladu normální distribuce odezvy. Další omezení kritérií změny se zdají klinicky sporné, nejen kvůli tomu, že je odezva menší, ale protože intervaly kontrolního sledování požadované pro testování citlivosti by se staly nepříjemně dlouhými.

Alternativně lze také obhájit plánování kontrolních návštěv v kratších časových intervalech než TAI nebo MTI za účelem identifikace pacientů, kteří nadále utrácejí kostní hmotu rychle. Úbytek kostí překračující TAM nebo LSC by představoval vhodné kritérium testu.

(Re-)definovaná standardizovaná chyba přesnosti

V přílohách je uvedeno množství vyvinutých různých definicí standardní chyby přesnosti. Aby se předešlo zmatku, měl by se používat termín $stdPE_L$ pouze tehdy, byla-li standardní chyba přesnosti získána skutečně z dlouhodobých dat. $stdPE_H$ se preferuje pro jiné přístupy. Kdyby normativní data všech výrobců byla stejně dobré kvality, mohla by být $stdPE_N$ dobrým odhadem dlouhodobé citlivosti pro zjištění změn stárnutím. Avšak je známo, že nedávno byly hlášeny rozdíly pro DXA,⁽¹⁶⁾ a opět se lze setkat s nesouladem u nově zavedených přístrojů a metod. Proto by se měl používat tento typ standardizace s opatrností.

Ve srovnání s předchozími navrhovanými přístupy nabízejí nové definice standardní (a standardizované) chyby přesnosti zde prezentované výhodu snadné interpretace (všech parametrů), vhodnost pro srovnání jakýchkoli dvou technik (standardizovaná chyba přesnosti), srovnatelnosti napříč různými studiemi (standardní chyba přesnosti), minimální odchylkou kohorty (opravy spíše pomocí rr než mírou odezvy) a aplikovatelnost pro radiografické i biochemické přístupy.

Tyto výhody budou prodiskutovány a poté budou nastíněna omezení definic dříve navrhovaných jinými autory.

Proč zavádět dvě koncepce standardní a standardizované chyby přesnosti?

Výhoda koncepce standardizované chyby přesnosti tkví v tom, že se může okamžitě použít pro srovnání chyby přesnosti jakýchkoli dvou technik, za předpokladu, že neopravená chyba přesnosti a míra odezvy je známa pro obě techniky. To umožní srovnávání v různých výzkumných situacích, zatímco koncepce standardní chyby přesnosti požaduje od výzkumníků, aby zjistili chybu přesnosti i míru odezvy DXA_S ve své populaci, která nemusí být vždy proveditelná. Avšak shodneme-li se na společném referenčním standardu—což je požadavek standardní chyby přesnosti—přineseme dobře definované robustní měřítko pro srovnání dlouhodobé citlivosti technik i napříč různými studii.

Usnadnění hodnocení standardní chyby přesnosti pro velký počet technik by usnadnila publikace samotných *rr*. Takové údaje by poskytly výzkumníkům metodologii pro určení *stdPE* pro novou techniku, i kdyby se ve středisku nemohlo provést přímé srovnání s posterior-anteriorním rentgenovým měřičem s dvojitou energií (PA-DXA) lumbální páteře. Bylo by pouze třeba porovnat novou techniku s referenční technikou, pro kterou je *rr* versus PA-DXA páteře již dostupné z literatury.

Proč používat BMD páteře z PA-DXA jako referenční standard?

Předchozí přístupy k standardizaci nedokázaly dosáhnout cíle standardizace, protože výsledek stále velmi závisel na skupině subjektů. Navrhovaná koncepce snižuje dopad tohoto zdroje chyby. Přesto je třeba vzít do úvahy jiné formy zkreslení. BMD páteře z PA-DXA je podstatně ovlivněna degenerativními změnami. Subjekty ovlivněné degenerativními změnami se musí vyloučit z kalkulace standardní chyby přesnosti, zvláště když byly hodnoceny z údajů napříč sekcemi. Volba DXA_{sp} jako referenční techniky pro kalkulaci standardizované chyby přesnosti neznamená, že DXA_{sp} je technika s nejlepší dlouhodobou citlivostí; pouze je považována za nejlepší referenční standard.

*Proč opravovat *sPE* a *stdPE* spíše s použitím poměrů odezvy než míry odezvy?*

Opravování chyby přesnosti dělením mírou odezvy přináší dobré měřítko dlouhodobé citlivosti, ale toto měřítko je velmi citlivé na vzorek studované populace (zkreslení kohorty). Tento přístup byl použit k definování MTI (nebo TAI), protože v kontextu odhadování kontrolních sledovacích časů je dopad populace kriticky důležitý. Pro generické srovnávání technik se preferuje robustnější měřítko jako je *stdPE*. Dokud různé techniky měří podobné hledisko kosti, jejich míry odezvy se částečně korelují. Proto je eliminována podstatná část dopadu studijní populace, když se místo míry odezvy použije *rr*. Navíc ponechává násobení *RR*, které je jednotné pro referenční techniky, procentní chybu přesnosti v rozsahu hodnot (typicky 1-5 %), které výzkumníci a kliničtí lékaři znají, což zvyšuje pravděpodobnost přijetí a usnadnění interpretace. To neplatí pro většinu definic standardizované chyby přesnosti, navrhovaných dříve.

Proč upravovat pro roční míry změny a nikoli pro měřítko variability mezi subjekty?

Pro kalkulaci míry odezvy pro standardizovanou chybu přesnosti bylo navrhované měřítko "roční míry změny". Je-li standardizovaná chyba přesnosti zamýšlena k použití jako měřítko dlouhodobé citlivosti, zdá se logické, že odezva by se měla definovat jako změna v čase. To by měl být nejvíce intuitivní přístup ke kvantifikaci schopnosti technika sledovat dlouhodobé změny. Pro dlouhodobá studie je to stejně volba, která je nasnadě, ale pro odhady

dlouhodobé citlivosti napříč sekcemi lze uvažovat jiné měřítko citlivosti. Zde však byla stejně zvolena standardizace podle roční míry změny, aby (krátkodobá, napříč sekcemi) definice stdPE_{st} byla co nejpodobnější standardizaci (dlouhodobé) stdPE_{lt} (viz Příloha 2). Navíc se můžeme všimnout, že veškerá měřítka rozptylu nebo dynamického rozsahu zahrnují chybovou složku, způsobenou chybami přesnosti (a výstižnosti). Proto u dvou technik se srovnatelně věrnou citlivostí technika s větší chybou přesnosti ukáže větší zdánlivou citlivost. V důsledku toho jsou odhady stdPE založeny na podhodnocení rozdílů v měřítku rozptylu v dlouhodobé citlivosti mezi technikami. Techniky s horší přesností budou demonstrovat uměle zvětšený dynamický rozsah a následně jejich kalkulovaná standardizovaná chyba přesnosti bude vypadat lépe, než jaké ve skutečnosti je (zkreslení přesnosti). To lze opravit, čili odstranit chybu přesnosti z rozptylu měřítka dvoucestnou hnízdovou analýzou variací.

Proč musí být parametry dlouhodobé citlivosti založeny spíše na dlouhodobých údajích přesnosti než na krátkodobých?

Hodnocení skeletálních změn prostřednictvím radiologických technik, jako je densitometrie kosti nebo OUS, obvykle vyžaduje časové intervaly mezi kontrolními měřeními v délce jednoho roku nebo delší. Proto musí být reprodukovatelnost technik založena na dlouhodobé chybě přesnosti, která je obvykle větší než krátkodobá chyba přesnosti.⁽¹⁷⁾ Existují další zdroje chyb (např. dlouhodobá stabilita zařízení, variabilita tělesné teploty pro měření SOS atd.), které se mohou zjistit pouze z dlouhodobých dat.⁽¹⁸⁾ Chyba přesnosti odvozená z opakovaných krátkodobých měření pouze přibližně vyjadřuje skutečné chyby reprodukovatelnosti. Přesto může být jejich výpočet užitečný, zejména můžeme-li předpokládat, že podíl krátkodobé a dlouhodobé chyby přesnosti (tj. poměr chyb přesnosti) zkoumané techniky a referenční techniky bude podobný. Pak nebude klasifikace citlivosti technik ovlivněna (ale absolutní velikosti dlouhodobé citlivosti budou nadhodnoceny).

Předchozí koncepce

Dříve publikované koncepce standardizace měly všechny problémy, o kterých jsme se zmiňovali. Miller et al. zavedl standardizovaný CV založený na normalizaci dynamického rozsahu daného 90 % percentilem,⁽³⁾ a Greenspan et al. použil podobný přístup, ale standardizovaný s 95 % percentilem.⁽⁸⁾ Obě měřítka rozptylu populace závisí na kritériích pro výběr subjektů a tak jsou ovlivněna značným zkreslením kohorty a přesnosti. Langton navrhol koncepci ZSD, tj. standardní deviaci Z skóre, které je bráno jako měřítko standardizované přesnosti.⁽⁵⁾ Zde jsou problémy se zkreslením vzorků méně závažné, protože variace populace použitá ke kalkulaci Z skóre je obvykle získána z velkých měřených populací, ze kterých se odvozují normativní data. Avšak současná diskuse o platnosti a srovnatelnosti normativních dat poskytnutých výrobcem vnáší do problému několik otázek. Ovšem mnohem důležitější než to, že je dobrým měřítkem citlivosti, je fakt, že větší variace populace by mohla být také způsobena problémy techniky (zkreslení přesnosti) a růzností subjektů, která není vztažena k osteoporose (zkreslení výstižnosti). Skutečně technika s velkým poklesem souvisejícím s věkem relativně k své variaci populace spíše umožní sledování skeletálních změn než technika, která—v extrému—nevykáže žádné změny ve spojitosti s věkem, i když tato druhá technika má stejně velikou nebo dokonce větší variaci populace. Variace populace se nezdá být spolehlivým měřítkem citlivost v čase a pro charakterizování diagnostické citlivosti může být vhodnější ZSD.

Jako další přístup navrhl Blumsohn et al. index individuality⁽⁴⁾, který je ovlivněn značným zkreslením ze vzorků, protože v sobě zahrnuje měřítko variability mezi subjekty. Problémy jsou podobné těm, kterých jsme si všimli u Quan a Shih u jiného měřítka standardizované přesnosti, mezitřídního CV.⁽¹⁹⁾ Obě tato měřítka se možné lépe hodí

k hodnocení diagnostické citlivosti. Machado a kolegové standardizovali přesnost dělením chyby přesnosti průměrnou diferencí mezi zdravými a osteoporotickými jedinci.⁽⁷⁾ Toto měřítko je ovlivněna zkreslením z kohorty vinou dvojnárodností stupně osteoporózy, která znemožňuje srovnávat standardizovanou chybu přesnosti napříč různými studii. Srovnání subjektů s a bez osteoporózy napříč sekcemi je problematické pro hodnocení dlouhodobé citlivosti z jiného důvodu: osteoporotičtí jedinci mohou mít nízký vrcholný skeletální status od počátku a proto za těchto podmínek by mohla standardizace založená na průměrné diferencí mezi zdravými a osteoporotickými jedinci nadhodnotit skutečnou dlouhodobou citlivost.

Rozšíření konceptu

Zatímco se navrhované koncepte vyhýbají mnoha problémům, o kterých jsme hovořili výše, je třeba uvést několik upozornění. Především není možné charakterizovat dlouhodobou citlivost jediným univerzálně aplikovatelným číslem účinnosti. Kritéria změny LSC (nebo TAM), intervaly kontrolního sledování MTI (nebo TAI) a standardizovaná chyba přesnosti mohou poskytnout hledanou odpověď pouze společně. Při hlubším pohledu by se mohl kritizovat navrhovaný přístup, protože nebere do úvahy, zda změna v parametru kosti, jakkoli významná, se vztahuje k relevantní změně rizika fraktury. Na tento problém poukázal Ross et al.⁽²⁰⁾ To nepředstavuje omezení, avšak vztahy mezi změnami v parametrech kosti následnou změnou rizika fraktury nebyly dosud dobře stanoveny. Zvýšený úbytek kosti může být rizikovým faktorem sám o sobě nebo kvůli extrapolovanému dlouhodobému poklesu BMD. Navíc by taková koncepte omezila relevanci měření úbytku kosti na pouhé hledisko rizika, zatímco pro klinické rozhodování a hodnocení účinnosti nebo dodržování léčby může hrát mnohem důležitější roli.

Pro použití navrhované koncepte existuje mnoho předpokladů. Předpokládá se, že související parametry kostí jsou normálně rozděleny. Pro kalkulaci dlouhodobé chyby přesnosti a míry odezvy se předpokládá, že změny jsou lineární s časem. Pro odezvu na léčení tomu tak obvykle není. Avšak navrhované koncepte by se daly snadno upravit. Nelineární změny se mohou rozdělit na částečné lineární segmenty. Ve srovnání s pozdějšími reakcemi by velká první odezva na léčení mohla vést ke kratšímu MTI (nebo TAI). Dlouhodobá chyba přesnosti by se také mohla kalkulovat z nelineárních modelů, pokud by to dávalo biologický a statistický smysl. Je teprve třeba zjistit, zda by to nabízelo výhody. Také je třeba si uvědomit, že bez ohledu na zvolený typ modelu zahrnuje standardní chyba odhadu (SEE; viz Příloha 1) dvě komponenty variability, tj. nepřesnost techniky a skutečnou odchylku od vhodného. Proto prospektivně definovaná standardizovaná chyba přesnosti nepředstavuje pouze skutečné limity techniky. V tomto ohledu může být termín "chyba přesnosti" považován za zavádějící a přednost dostává alternativní termín "dlouhodobá citlivost". Tato dvojnárodnost však pro většinu klinických aplikací nepředstavuje problém. Jestliže se například zajímáme o odhad doby sledování požadované pro zjištění úspěšnosti léčby, schopnost to zjistit bude záviset na nepřesnosti techniky i skutečné variabilitě v čase.⁽²¹⁾ Tak můžeme SEE považovat za představitele dobré aproximace celkové diagnostické, biologické i terapeutické variability.

Statistické testy jako ty, navrhované v této přednášce, mohou být vtěleny do operačního softwaru přístroje. Například v seriálovém měření může automatická indikace, zda je změna od posledního vyšetření významná, pomoci klinickému lékaři v rozhodovacím procesu.

Aplikace konceptu

Výše zmíněné výhody a nevýhody parametrů konceptu jsou demonstrovány údaji z tabulky 1. Jak můžeme vidět, výkonnost (tj. schopnost sledovat změny) techniky se nemůže

posuzovat podle neopravené chyby přesnosti, protože míry odezvy se podstatně různí. Mohou být přímo použita kritéria změny, aby se zjistilo, které změny reflektují trendy (TAM) nebo významné změny (LSC). Časy kontrolního sledování TAI a MTI a stdPE odrážejí jak chybu přesnosti, tak i citlivost na změny. StdPE je méně závislá na skupině subjektů než MTI (nebo TAI) a tak je blíže k cíli definovat jediný parametr, který charakterizuje celkovou výkonnost techniky. MTI (TAI) bude obvykle odlišný pro každou skupinu subjektů a techniku, protože má být přímým indikátorem kontrolní sledovací doby a bude záviset na skupině subjektů.

Výsledky z tabulky 1 jsou založeny na krátkodobé chybě přesnosti a tak se musí interpretovat s opatrností, protože budou pravděpodobně podhodnocovat dlouhodobou stdPE_{lt}. V tomto hypotetickém příkladu je dlouhodobá citlivost BMD_{spine} nebo BMD_{femtot} lepší než citlivost kteréhokoliv ze dvou parametrů QUS.

ZÁVĚR

Komplexní vyhodnocení dlouhodobé citlivosti techniky musí být založeno na kalkulaci kritéria změny (jako TAM nebo LSC), časovém intervalu kontrolního sledování (jako TAI nebo MTI) a standardní chybě přesnosti (stdPE). Tato tři měřítka společně umožňují dobrou charakterizaci schopnosti techniky sledovat skeletální změny: LSC je nejmenší změna považovaná za statisticky významnou, pacient musí být zván přibližně v časovém intervalu specifikovaném podle MTI (konkrétní populace) a čím menší je stdPE, tím je technika citlivější. Pro klinické rozhodování, které vyžaduje nebo umožňuje dřívější posouzení na nižších úrovních statistické významnosti, tj. trend hodnocení, může být adekvátní zkrácení časového intervalu pro následnou kontrolu o 36 % (příští návštěva po TAI místo MTT).

Byly ukázány některé z dřívějších metod standardizace přesnosti, které představovaly případy mylné aplikace (zesílení kohortového efektu) užitečné koncepce (standardizace) na parametr, který byl někdy v minulosti desinterpretován (přesnost jako parametr, který pro účely zde diskutované není hodnotný sám o sobě, ale pouze ve spojení s dobrou citlivostí). Prezentované koncepce by měly zlepšit schopnost zkoumat, charakterizovat a porovnávat schopnost technik sledovat změny ve skeletálním statutu.

OZNÁMENÍ

Rád bych poděkoval za prospěšné diskuse, které se mnou vedli: Richard Eastell, M.D., Sheffield, U.K., Ying Lu, Ph.D., San Francisco, CA, U.S.A. a Harry Genant, M.D., San Francisco, CA, U.S.A.

REFERENCE

Viz originální text.

Žádosti o výtisk zasílejte na adresu:
Claus-C. Glier
Arbeitsgruppe Medizinische Physik
Klinik für Diagnostische Radiologie
Universitätsklinikum an der
Christian-Albrechts-Universität zu Kiel
Michaelisstrasse 9
D-24105 Kiel, Deutschland

Přijato v originále 11. května 1998; v revidované formě 2. dubna 1999; schváleno 5. dubna 1999.

PŘÍLOHA 1. SLOVNÍČEK TERMÍNŮ, ZKRATEK A DEFINICÍ

LSC: Nejmenší významná změna: $LSC = 2,8 \times PE_{lt}$

Kritérium pro nejmenší změnu ve výsledku měření, která může být považována za statisticky významnou s 95 % spolehlivostí (dvojstranný test). Statistické předpoklady viz.⁽⁹⁾ Srovnej s TAM.

MTI: Časový interval sledování; LSC/medián odezvy

Časový interval sledování, po kterém se dá u většiny pacientů očekávat, že vykážou změny překračující LSC, tj. časový interval doporučený mezi kontrolními návštěvami, je-li požadována vysoká 95 % úroveň spolehlivosti (dvojstranný test). MTI je charakteristika techniky, ale závisí na skupině subjektů, např. progresi nemoci (MTI_D), odezvě na léčení (např. MTI_{cstr} nebo MTI_{vitD}). Srovnej s TAI.

PE_{st}: krátkodobá chyba přesnosti, vyjádřená procentuálně; $PE_{st} = RMS(SEE./střed_i)$

Odvozena ze dvou nebo více měření opakovaných v krátkých časových intervalech a získaná na $i = 1..m$ jedincích; viz Příloha 2.

PE_{lt}: Dlouhodobá chyba přesnosti, vyjádřená procentuálně; $PE_{lt} = RMS(SEE./střed_i)$

Odvozena z dlouhodobých studií $i = 1..m$ jedinců s minimálně třemi opakovanými měřeními na jedince v čase. Viz Příloha 2.

RMS: Průměr druhé odmocniny; průměrující metoda vhodná pro zprůměrování variací (např. chyby přesnosti), které nejsou normálně rozděleny, ale podle F-distribuce. Viz Příloha 2.

rr: Poměr odezvy: $rr(AvsR) = \text{míra odezvy (referenční technika R)}/\text{míra odezvy (zkoumaná technika A)}$, kde míry odezvy reflektují procentní změny za rok nebo procentní změnu za rok věku pro danou techniku.

SD: Standardní deviace opakovaných měření; měřítko krátkodobé přesnosti. Srovnej s SEE.

SEE: Standardní chyba odhadu; měřítko rozptylu kolem regresní čáry a tím i dlouhodobé přesnosti. Srovnej s SD.

sPE: Standardizovaná chyba přesnosti: $sPE = PE \times rr(AvsR)$

Chyba přesnosti upravená pro poměr odezvy rr referenční techniky versus zkoumaná technika A. Vyjádřena procentuálně. Jako výsledek procedury standardizace je nyní stupnice standardizované chyby přesnosti ekvivalentní stupnici chyby přesnosti referenční techniky. V důsledku toho se standardizovaná chyba přesnosti obou technik může nyní přímo porovnávat.

stdPE: Standardní chyba přesnosti: $stdPE = PE \times rr(PA-DXA_{spine}$ vs. technika A)

Standardizovaná chyba přesnosti, pro kterou bylo zvoleno PA-DXA páteře jako referenční technika. Vyjádřena procentuálně.

TAI: Interval trendu hodnocení: TAM/medián odezvy

Časový interval sledování, po kterém se dá u většiny pacientů očekávat, že vykážou změny překračující TAM, tj. časový interval doporučený, jestliže postačují poněkud volnější testy na změny. TAI je charakteristika techniky, ale závisí na skupině subjektů, např. progresi nemoci (TAI_D), odezvě na léčení (např. TAI_{cstr} nebo TAI_{vitD}). Srovnej s MTI.

TAM: Rozmezí trendu hodnocení: $TAM = 1,8 \times PE_{lt}$

Kritérium pro nejmenší změnu ve výsledku měření, která se dá považovat za statisticky významnou s 80 % spolehlivostí (dvojstranný test) nebo 90 % spolehlivostí (jednostranný test). Statistické předpoklady viz.⁽⁹⁾ Srovnej s LSC.

PŘÍLOHA 2. KALKULACE STANDARDIZOVANÉ CHYBY PŘESNOSTI

Krátkodobá chyba přesnosti se vypočítá takto: Pro jedince je absolutní chyba přesnosti dána standardní deviací (SD) opakovaných měření. Vyjádřeno procentuálně, krátkodobá chyba přesnosti $PE_{st,i}$ pro i -tého jedince je dána vzorcem:

$$PE_{st,i} [\%] = \frac{SD_i}{\bar{x}_i} = 100 \times \sqrt{\frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n_i - 1}} / \bar{x}_i$$

kde x_{ij} je parametr kosti z j -tého měření i -tého jedince a \bar{x}_i průměru n_i opakovaného měření na tomto subjektu.

Průměrná krátkodobá chyba přesnosti skupiny m jedinců není dána aritmetickým průměrem, ale průměrem druhé odmocniny (RMS) chyby přesnosti jedinců⁽²²⁾:

$$PE_{st} = \text{RMS} (PE_{st,i}) = \sqrt{\frac{\sum_{i=1}^m (n_i - 1) PE_{st,i}^2}{\sum_{i=1}^m (n_i - 1)}}$$

Dlouhodobá chyba přesnosti se může vypočítat z analýzy lineární regrese měření jednotlivce v čase. Standardní chyba odhadu (SEE_i), která reflektuje deviace opakovaných měření z upravené křivky, se může vzít jako měřítko absolutní dlouhodobé chyby přesnosti jedince. Dlouhodobá chyba přesnosti jedince, $PE_{lt,i}$ vyjádřená procentuálně, je dána vzorcem:

$$PE_{lt,i} [\%] = \frac{SEE_i}{\bar{x}_i} = 100 \times \sqrt{\frac{\sum_{j=1}^{n_i} (x_{ij} - \hat{x}_{ij})^2}{n_i - 2}} / \bar{x}_i$$

kde $\hat{x}_{ij} = a + b_{ij}$ je předpovídaná hodnota j -tého měření i -tého jedince v čase h , podle upravené linky s úsekem a a sklonem b .

Pro skupinu m jedinců je průměrná dlouhodobá chyba přesnosti PE_{lt} dána vzorcem:

$$PE_{lt} = \text{RMS} (PE_{lt,i}) = \sqrt{\frac{\sum_{i=1}^m (n_i - 2) PE_{lt,i}^2}{\sum_{i=1}^m (n_i - 2)}}$$

Standardní (nebo standardizovaná) chyba přesnosti je odvozena z chyby přesnosti definované výše nejdříve vynásobením krátkodobé nebo dlouhodobé chyby přesnosti poměrem odezvy a pak výpočtem průměrného RMS napříč všemi subjekty. Poměr odezvy, rr , se může odvodit takto:

(1) Dlouhodobé studie (preferovaný postup):

$$rr_{L,i} = \frac{(\% \text{ sklon za rok referenční techniky})_i}{(\% \text{ sklon za rok zkoumané techniky})_i}$$

% sklon je sklon regresní čáry (tj. předvídaná změna za rok) měření jedince v čase. Jako měřítko

odezvy pozorované u tohoto jedince je kalkulován pro zkoumanou techniku i referenční techniku— obojí získáno u tohoto jedince – pro výpočet míry odezvy. Pro tuto metodu, narozdíl od dalších dvou, je poměr odezvy specifický pro každého jedince a může se použít pro odhad dlouhodobé citlivosti odezvy na léčbu.

(2) Normativní data:

$$rr_N = \frac{\% \text{ sklon za rok normativních dat pro referenční techniku}}{\% \text{ sklon za rok normativních dat pro zkoumanou techniku}}$$

Zde jsou míry odezvy založeny na uspořádání napříč sekcemi změn souvisejících s věkem v normativních datech.

(3) Data napříč sekcemi (nejméně žádoucí přístup)

$$rr_C = \frac{\% \text{ sklon za rok věku pro referenční techniku}}{\% \text{ sklon za rok věku pro zkoumanou techniku}}$$

Tento přístup lze použít, když nejsou k dispozici ani dlouhodobé studie odezvy, ani normativní data. Jestliže byla zkoumaná technika a referenční technika získána z téhož jedince, můžeme odděleně pro obě techniky regresně vypočítat parametr techniky versus věk subjektu, abychom získali sklon za rok věku jako měřítko míry odezvy. Standardní chyba přesnosti je pak kalkulovaná z buď

$$\begin{aligned} \text{stdPE}_t &= PE_i \times rr_{L,i} \text{ pro dlouhodobé studie nebo} \\ \text{stdPE}_c &= PE_i \times rr_{N(\text{nebo } C)} \text{ pro studie napříč sekcemi} \end{aligned}$$

Standardní chyba přesnosti zprůměrovaná napříč skupinou jedinců stdPE je pak dána vzorcem:

$$\text{stdPE} = \text{RMS}(\text{stdPE}_i) = \sqrt{\frac{\sum_{i=1}^m (n_i - 2) \text{stdPE}_i^2}{\sum_{i=1}^m (n_i - 2)}}$$

V závislosti na typu dat, tj. krátkodobá nebo dlouhodobá chyba přesnosti, dlouhodobá studie nebo studie napříč sekcemi, se mohou vypočítat různé typy stdPE. Preferovaný přístup, při kterém je jak dlouhodobá chyba přesnosti, tak míra odezvy odvozena z dlouhodobých studií, je označena jako stdPE_{lt}:

$$\text{stdPE}_{lt} = \text{RMS}(\text{stdPE}_{L,lt,i}) = \sqrt{\frac{\sum_{i=1}^m (n_i - 2) \text{stdPE}_{L,lt,i}^2}{\sum_{i=1}^m (n_i - 2)}}$$