

Izvirni znanstveni članek ■

Možnosti primerjave skladnosti meritev med skupinami: obremenitveno testiranje kot presejalni test za zmožnost hoje po nadkolenski amputaciji

Options for Comparing Measurement Agreement between Groups: Exercise Testing as Screening for Ability to Walk After Transfemoral Amputation

Gaj Vidmar, Helena Burger, Tatjana Erjavec

Izvleček. Rehabilitacija oseb po amputaciji spodnjega uda vključuje oskrbo s protezo. Pred tem je potrebno ugotoviti, kolikšno obremenitev bo hoja z nadkolensko protezo predstavljala za pacienta. V ta namen smo pri 101 pacientu po transfemoralni amputaciji uporabili 6-minutni test hoje in telesno zmogljivost testirali z obremenitvenim testiranjem z ročnim kolesom. Želeli smo ugotoviti, pri kateri stopnji obremenitve je odgovor žilnega sistema, merjen s porabo kisika, najbližje odgovoru pri testu hoje. Uporabili smo različne statistične analize skladnosti meritev. Članek podaja vpogled v množico razpoložljivih pristopov k analizi skladnosti za številске spremenljivke s poudarkom na primerjavi skladnosti med skupinami.

Abstract. Rehabilitation of persons after lower limb amputation includes provision of a prosthesis. Before that, it should be established how much strain would walking with a prosthesis cause to the patient. For this purpose, we performed 6-minute walk test and exercise testing using hand-wheel ergometer in 101 patients after transfemoral amputation. We wanted to establish at which level of exercise power is the vascular response, as measured by oxygen uptake, closest to the response during walk test. We used various statistical methods of agreement analysis. The article reviews the numerous available approaches to the analysis of agreement for numeric variables with an emphasis on comparing agreement between groups.

■ **Infor Med Slov:** 2010; 15(2): 10-20

Instituciji avtorjev: Univerzitetni rehabilitacijski inštitut republike Slovenije – Soča (GV, HB, TE); Medicinska fakulteta, Univerza v Ljubljani (GV, HB).

Kontaktna oseba: Gaj Vidmar, URI – Soča, Linhartova 51, 1000 Ljubljana. e-pošta: gaj.vidmar@ir-rs.si.

Prejeto: 01.12.2010. Sprejeto: 14.12.2010.

Uvod

Osnovni namen rehabilitacije oseb po amputaciji spodnjega uda je omogočiti bolnikom čim boljše funkcioniranje in kakovost življenja, kar vključuje oskrbo s protezo. Hoja z nadkolensko protezo je gibalno zahtevna in bolnik pri njej porabi veliko energije, zato je pred oskrbo s protezo potrebno najprej ugotoviti, kolikšno obremenitev bo hoja s protezo predstavljala za bolnika.

Stopnja bolnikove aktivnosti pred amputacijo je pomembna pri predpisu dokončne proteze, ki jo bolnik dobi po večmesečnem obdobju uporabe začasne proteze. Bolniki po nadkolenski amputaciji žilnega vzroka so večinoma starejši, imajo številna sočasna obolenja in so manj aktivni.^{1,2} Večinoma so zmožni le gibanja po stanovanju ali v ožjem bivalnem okolju; le redki lahko hodijo z različno hitrostjo, kar je pomembno za lažje gibanje v širšem bivalnem okolju.

Kot objektivni in zanesljivi kriteriji pri odločanju za namestitvev proteze so se v dosedanjih raziskavah pokazali rezultati obremenitvenega testiranja in 6-minutnega testa hoje ter ocene z Lestvico funkcijske neodvisnosti (FIM) ob sprejemu v rehabilitacijsko ustanovo.³ Najprimernejši pokazatelj je sicer sposobnost hoje s pripomočki (bergle, hodulja), a hoje ob sprejemu na rehabilitacijo mnogi starejši bolniki ne zmorejo. To odloži odločitev o oskrbi s protezo in s tem v programu rehabilitacije, ki je časovno omejen, skrajša obdobje za vadbo hoje. Zanesljiva hoja ob zaključku rehabilitacije pa je ključen cilj rehabilitacije in napovedni dejavnik za uspešno uporabo proteze v daljšem obdobju po prevzemu.

V splošnem je energetska zahtevnost hoje odvisna od višine in vzroka amputacije, hitrosti hoje in starosti.⁴ Starejši bolniki z nadkolensko protezo v domačem okolju hodijo hitreje in bolj redno uporabljajo protezo s kolenom na zaklep kot protezo z gibljivim kolenom.⁵ Začasna proteza ima zato praviloma koleno na zaklep in pričujoča študija se nanaša le na take primere.

Za testiranje hoje najpogosteje uporabljamo 6-minutni test na ravni podlagi.^{6,7} Ob poznavanju odgovora srca na napor, ki ga predstavlja hoja z nadkolensko protezo, lahko določimo minimalno potrebno telesno zmogljivost za možnost namestitve proteze. Telesno zmogljivost testiramo z obremenitvenim testiranjem z ročnim kolesom.

Tako je bil glavni namen analiz v tem članku ugotoviti, pri kateri stopnji obremenitve je odgovor žilnega sistema, merjen s porabo kisika, najbližje odgovoru pri testu hoje. Za odgovor na to vprašanje je potrebna statistična analiza skladnosti meritev, ki pa je razmeroma mlado področje, premalo poznano med raziskovalci (ne le v zdravstvu) in torišče različnih pogledov, zato članek hkrati podaja vpogled v množico razpoložljivih pristopov in metod za statistično analizo skladnosti številskih spremenljivk.

Metode

Udeleženci

V študijo je bil vključen 101 pacient, sprejet na primarno rehabilitacijo po transfemoralni amputaciji zaradi žilnega vzroka na Univerzitetni rehabilitacijski inštitut – Soča v Ljubljani. Vzorčenje je potekalo zaporedno v letu 2009, pri čemer so bili vključeni le pacienti, ki so bili ob zaključku programa sposobni 6-minutne hoje z nadkolensko protezo s kolenom na zaklep (po želji z uporabo pripomočkov).

Postopek merjenja

Obremenitveno testiranjem so pacienti opravili ob sprejemu na bolnišnično rehabilitacijo z ročnim kolesom po diskontinuiranem protokolu submaksimalne obremenitve,³ pri čemer smo neposredne meritve z analizo plinov v zraku preko obrazne maske uspeli izvesti le pri pacientih z dobro telesno zmogljivostjo, pri ostalih pa smo VO₂ izračunali posredno po obrazcu ACSM.^{8,9} Test hoje so pacienti opravili v zadnjih treh dneh pred odpustom na razdalji 60 m v skladu s

priporočili ATS.⁶ Med hojo smo spremljali njihove metabolične parametre in EKG, pred in po zaključku hoje pa tudi krvni tlak. Vsi pacienti so bili seznanjeni s protokolom študije in so pisno pristali na sodelovanje. Študijo je odobrila Komisija za medicinsko etiko URI – Soča.

Statistične analize

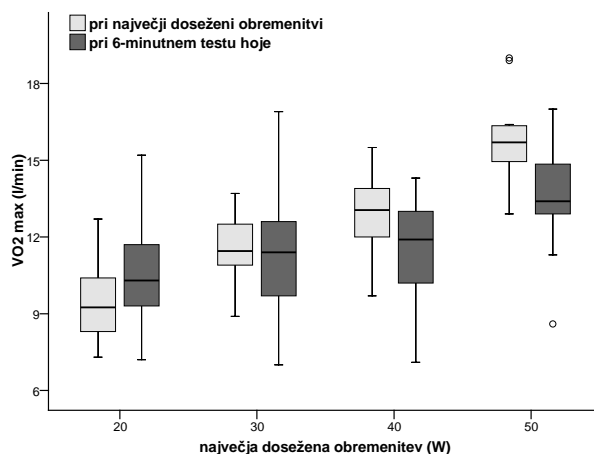
Najprej smo izračunali opisne statistike in izdelali grafični prikaz za porazdelitev največje izmerjene porabe kisika (VO₂ max) pri obremenitvenem testiranju in pri testu hoje glede na največjo doseženo obremenitev pri obremenitvenem testiranju. Potem smo izvedli regresijske analize skladnosti meritev – najprej po metodi najmanjših kvadratov skozi izhodišče, nato pa z metodami za analizo meritev z napakami. Sledil je izračun koeficienta skladnostne korelacije. Na koncu smo skladnost meritev obravnavali z analizo razlik, pri čemer smo najprej uporabili metodo meja skladnosti, nato pa smo s parametričnimi in neparametričnimi oziroma robustnimi metodami analizirali opažene razlike. Vse analize smo izvedli s statističnim programskim paketom SPSS[®] (verzija 15.0.1.1 za okolje Windows, SPSS Inc., Chicago, IL, ZDA, 2007), le analizo povprečij smo izvedli s preizkusno različico paketa Minitab[®] (verzija 15.1.30.0, Minitab Inc., State College, PA, ZDA, 2007). Regresijske analize za meritve z napakami in koeficient skladnostne korelacije smo izračunali z javno dostopnimi makri¹⁰ in preverili s preizkusno različico paketa MedCalc[®] (verzija 11.4.4.0, MedCalc Software, Mariakerke, Belgija, 2010).

Rezultati

Pregled podatkov

Vsaka analiza podatkov naj bi se začela z ustreznimi eksploratornimi prikazi podatkov in smiselnimi pregledi opisnih statistik.^{11,12} V našem primeru sta to prikaz (slika 1) in številski opis (tabela 1) pogojne porazdelitve VO₂ max pri obremenitvenem testiranju in pri testu hoje glede

na doseženo maksimalno obremenitev pri obremenitvenem testiranju. Slednja je zaradi omejitev merilne naprave ocenjena le na 10 W natančno, zato jo obravnavamo kot ordinalno spremenljivko, ki paciente razdeli v štiri skupine (20 W, 30 W, 40 W, 50 W).



Slika 1 Porazdelitev največje izmerjene porabe kisika pri obremenitvenem testiranju in pri testu hoje glede na največjo doseženo obremenitev pri obremenitvenem testiranju, prikazana s škatlastimi diagrami. Srednja črta označuje mediano, škatla interkvartilni razmik (IQR), navpične črte razpon brez osamelcev, krožci pa označujejo osamelce (dlje kot 1,5 IQR od škatle).

Tabela 1 Opisne statistike za največjo izmerjeno porabo kisika glede na največjo doseženo obremenitev.

P	20 W		30 W		40 W		50 W	
n	14		46		30		11	
	OT	Hoja	OT	Hoja	OT	Hoja	OT	Hoja
Min	7,3	7,2	8,9	7,0	9,7	7,1	12,9	8,6
Max	12,7	15,2	13,7	16,9	15,5	14,3	19,0	17,0
M	9,5	10,5	11,6	11,3	12,9	11,5	15,8	13,5
Me	9,3	10,3	11,5	11,4	13,1	11,9	15,7	13,4
SD	1,4	2,2	1,1	2,3	1,4	1,8	1,9	2,2
CV	15%	21%	10%	20%	11%	16%	12%	16%
ReRa	1,7	2,1	1,5	2,4	1,6	2,0	1,5	2,0
r	0,278		0,326		0,387		0,176	
ρ	0,300		0,424		0,425		0,218	

P = največja dosežena obremenitev; n = št. pacientov; OT = obramenilno testiranje; Me = mediana; CV = koeficient variabilnosti; ReRa = Max / Min; r = Pearsonov korelacijski koeficient, ρ = Spearmanov koeficient korelacije rangov.

Že grafični in tabelarni prikaz porazdelitev kaže na temeljno ugotovitev, ki se je potrdila skozi večino nadaljnjih analiz: pri obremenitvenem testiranju so imeli najbolj podobno porabo kisika kot pri testu hoje tisti pacienti, ki so dosegli največjo obremenitev 30 W; pri tistih, ki so dosegli le 20 W, je bila poraba kisika večja pri testu hoje, pri tistih, ki so dosegli 40 W ali 50 W, pa pri obremenitvenem testiranju. Koeficient variabilnosti in relativni razpon sta pomembna za interpretacijo regresijskih modelov za analizo meritev z napakami, predstavljeno v nadaljevanju. Enako velja za korelacijski koeficient, ki seveda ni naveden kot mera skladnosti, česar ni odveč podkrepiti s šolskim primerom, da je korelacija med dvema nizoma meritev, ki se razlikujeta za konstanto ali za konstanten faktor, popolna, skladnosti med njima pa ni.

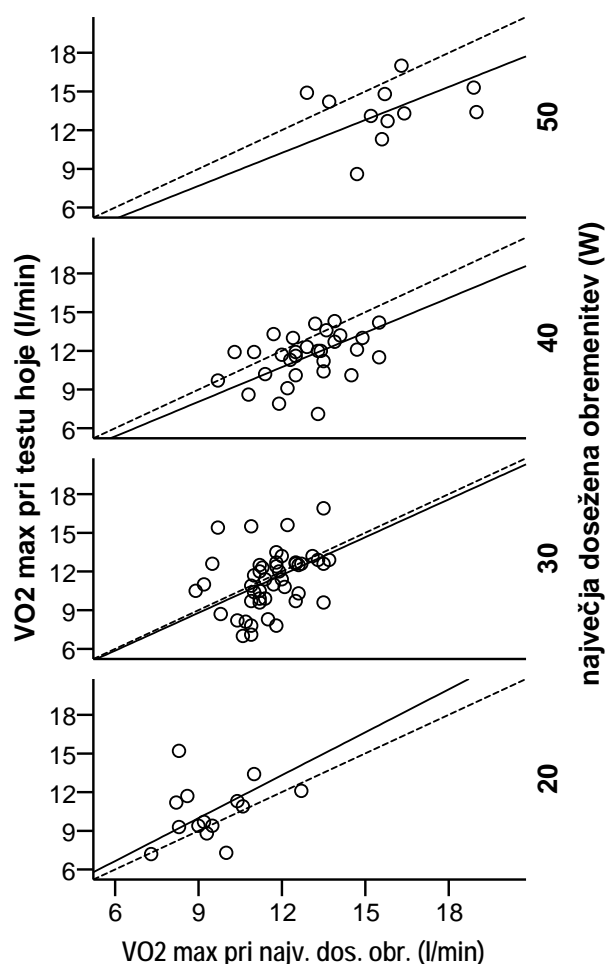
Regresijske analize skladnosti

- *Preprosta linearna regresija skozi izhodišče*

Linearna regresija skozi izhodišče je ena od najpreprostejših metod analize skladnosti meritev. Model smo ocenili znotraj vsake skupine pacientov, določene z največjo doseženo obremenitvijo. Vsi modeli so bili statistično značilno boljši od ničelnega (tabela 2). Kot je razvidno iz tabele 2 in slike 2, je bil ocenjeni naklon najbližje vrednosti 1 (ki pomeni skladnost) za paciente, ki so dosegli obremenitev 30 W.

Nato smo naklone med seboj primerjali s Chowovim testom,¹³ ki ga lahko z vsakim statističnim programskim paketom izvedemo tako, da gnezdeno primerjamo osnovni linearni model (v našem primeru brez konstante) z modelom, ki mu dodamo dvojiško indikatorsko spremenljivko za primerjalno skupino in interakcijo (v našem primeru le naklona) s to indikatorsko spremenljivko. Opravili smo le dve primerjavi s skupino 30 W, da smo se izognili problemu množstva primerjav (angl. multiple comparisons), pri čemer je bila združitev skupin 40 W in 50 W vsebinsko smiselna, hkrati pa je povečala moč testa. Razlika med naklonom skupine 30 W in

skupine 20 W je bila mejno statistično značilna ($p = 0,078$), naklon skupine 40-50 W pa je statistično značilno odstopal od naklona skupine 30 W ($p = 0,006$).



Slika 2 Panelni razsevni diagram največje izmerjene porabe kisika pri testu hoje v odvisnosti od največje izmerjene porabe kisika pri obremenitvenem testiranju glede na največjo doseženo obremenitev z vrisanimi regresijskimi premicami skozi izhodišče. Črtkana črta predstavlja enakost (premico z naklonom 1).

Tabela 2 Rezultati linearne regresije skozi izhodišče.

Najv. dos. obremenitev	<i>b</i>	SE(<i>b</i>)	<i>p</i>
20 W	1,096	0,065	< 0,001
30 W	0,974	0,028	< 0,001
40 W	0,889	0,025	< 0,001
50 W	0,845	0,047	< 0,001

b = ocena koeficienta; SE = standardna napaka ocene.

- *Regressijski modeli za analizo meritev z napakami*

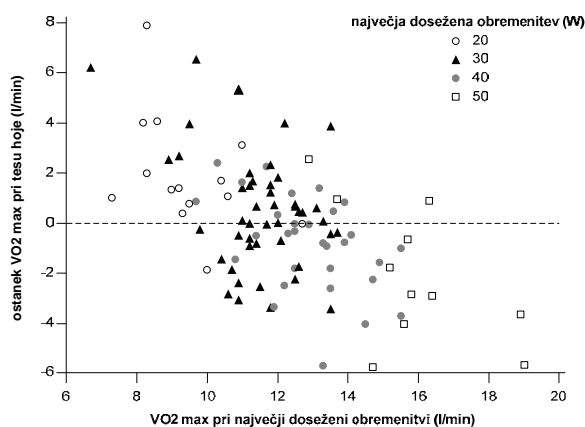
Preprosta linearna regresija predpostavlja popolno točnost spremenljivk oziroma odsotnost merskih napak, kar je pri analizi skladnosti meritev še manj realistično kot pri običajnih statističnih podatkih. Zato se v laboratorijskih vedah, meroslovju in drugod, kjer se redno srečujejo s problemom umerjanja (kalibracije), pogosto uporablja regresijske modele za meritve za napakami.¹⁴ Literature in različnih poimenovanj istih postopkov je preveč, zato naj na tem mestu le usmerimo bralca na primerjalne preglede.^{15,16,17}

Najprej smo uporabili *regresijo Passinga in Babloka*, ki temelji na Theilovem in Senovem pristopu k robustni regresiji z izračunom mediane naklonov daljic, ki povezujejo vse možne dvojice točk, določenih z danimi dvorazsežnimi številskimi podatki.¹⁸ Rezultati (tabela 2) kažejo predvsem na veliko negotovost ocenjenih modelov, saj so intervali zaupanja izredno široki (oziroma se jih v skupini 20 W sploh ni dalo oceniti). Za razliko od večine ostalih predstavljenih analiz se je ocenjena skladnost pokazala kot najnižja pri skupini 30 W, saj le pri njej interval zaupanja za oceno konstante izključuje vrednost 0 in interval zaupanja za oceno regresijskega koeficienta izključuje vrednost 1. Poleg tega je le v skupini 30 W vprašljiva linearnost povezave med obravnavanima spremenljivkama. Po drugi strani pa je porazdelitev regresijskih ostankov (slika 3) najugodnejša prav v skupini 30 W.

Tabela 3 Ocene parametrov z regresijo po postopku Passinga in Babloka.

Skupina	a	IZ(a)	b	IZ(b)	p_{CUSUM}
20 W	-8,65	/	2,00	/	> 0,10
30 W	-20,98	[-40,8; -10,4]	2,76	[1,8; 4,5]	0,05
40 W	-6,14	[-19,5; 2,2]	1,37	[0,8; 2,4]	> 0,10
50 W	-4,32	[-97,1; 12,2]	1,08	[0,1; 7,0]	> 0,10

a = regresijska konstanta; b = regresijski koeficient;
IZ = 95% interval zaupanja; / = ni ocenljivo;
 p_{CUSUM} = statistična značilnost testa linearnosti.



Slika 3 Porazdelitev ostankov pri regresiji po postopku Passinga in Babloka.

Nato smo uporabili dve metodi, ki sodita v okvir *regresije modela II* (navadna linearna regresija sodi v model I):

1. *regresija po metodi glavne osi* (angl. major axis, MA; znana tudi kot regresija po metodi glavnih komponent, angl. principal components regression; metoda najmanjših pravokotnih kvadratov, angl. least perpendicular squares; in metoda pravokotne razdalje, angl. perpendicular distance method) je osnovna in najpogostejša oblika Demingovega splošnega regresijskega modela za meritve z napakami, zato se oznaka Demingova regresija praviloma nanaša nanjo;
2. *regresija po metodi zmanjšane glavne osi* (angl. reduced major axis, RMA; znana tudi kot regresija po metodi standardiziranih glavnih komponent, angl. standardised principal components regression; metoda najmanjših zmnožkov, angl. ordinary least products, OLP; in regresija geometrijske sredine, angl. geometric mean regression, saj je ocenjeni naklon enak geometrijski sredini naklonov pri navadni linearni regresiji, ki ju dobimo z zamenjavo spremenljivk x in y) je ekvivalentna Demingovi regresiji za standardizirane spremenljivke.

Ocene parametrov (tabela 3) kažejo podobno sliko kot postopek Passinga in Babloka v smislu, da naj bi bila skladnost najnižja v skupini 30 W. Prav tako so očitni predvsem (pre)široki intervali zaupanja, ki so sicer nekoliko bolj sprejemljivi pri metodi zmanjšane glavne osi, ki ji novejša študija s področja klinične kemije, farmakologije in biokemije dajejo prednost.^{16,17}

Tabela 4 Ocene parametrov z regresijo po metodi glavne osi (MA) in zmanjšane glavne osi (RMA).

MA	a	IZ(a)	b	IZ(b)
20 W	-20,92	[-151,6; 109,8]	3,32	[-10,6; 17,3]
30 W	-44,64	[-90,7; 1,4]	4,84	[0,8; 8,9]
40 W	-14,62	[-41,4; 12,2]	2,03	[0,0; 4,1]
50 W	-41,76	[-653,8; 570,3]	3,49	[-36,0; 43,0]
RMA	a	IZ(a)	b	IZ(b)
20 W	-4,57	[-17,0; 7,8]	1,59	[0,2; 2,9]
30 W	-11,71	[-19,1; -4,3]	1,99	[1,3; 2,6]
40 W	-4,77	[-11,4; 1,8]	1,26	[0,8; 1,7]
50 W	-5,32	[-17,3; 6,7]	1,19	[0,4; 1,9]

Za pojasnila glej besedilo in tabelo 3.

Koeficient skladnostne korelacije

Linov koeficient skladnostne korelacije^{19,20} (concordance correlation coefficient, CCC) je produkt mere natančnosti (angl. precision, ocene populacijske vrednosti Pearsonovega korelacijskega koeficienta) in mere točnosti (angl. accuracy, ki predstavlja faktor popravka zaradi pristranosti, angl. bias correction factor). Tako koeficient kot obe njegovi komponenti lahko zavzamejo vrednosti med 0 in 1, pri čemer 1 pomeni popolno skladnost. Ocene po skupinah največje dosežene obremenitve so zbrane v

tabeli 5. Vrednosti so nizke oziroma zmerne,²¹ očitno pa je, da se največja izmerjena poraba kisika pri hoji vsaj v grobem sklada s porabo na obremenitvenem testiranju le pri pacientih, ki so dosegli obremenitev 30 W ali 40 W.

Tabela 5 Ocene koeficienta skladnostne korelacije.

Skupina	CCC	[95% IZ]	natančnost	točnost
20 W	0,21	[-0,22; 0,57]	0,28	0,77
30 W	0,26	[0,03; 0,46]	0,33	0,79
40 W	0,28	[0,02; 0,50]	0,39	0,72
50 W	0,10	[-0,27; 0,45]	0,18	0,58

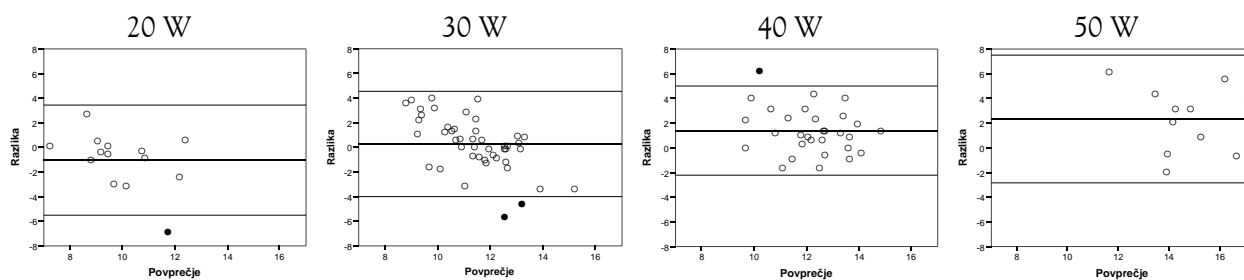
CCC = Linov koeficient skladnostne korelacije;
IZ = interval zaupanja (spodnja; zgornja meja).

Analiza razlik med meritvama

- *Metoda meja skladnosti*

Metoda meja skladnosti (angl. limits of agreement, LOA) ali pristop Blanda in Altmana²² temelji na dejstvu, da je razlika dveh spremenljivk nekorelirana z njuno vsoto (in s tem tudi z njunim povprečjem, ki polovica vsote). Razliko meritev, ki ju primerjamo, zato narišemo v odvisnosti od njunega povprečja za vsako dvojico meritev, meje skladnosti pa ocenimo kot tolerančni interval za razlike.

Tudi z metodo meja skladnosti smo meritev porabe kisika pri obremenitvenem testiranju primerjali z meritvijo pri testu hoje znotraj vsake s skupine. Rezultati (slika 4) ne kažejo jasnega neskladja pri nobeni skupini, torej ne omogočajo zanesljivega sklepa glede najprimernejše mejne obremenitve.



Slika 4 Primerjava VO₂ max med obremenitvenim testiranjem z VO₂ max pri testu hoje po skupinah z Blandovo in Altmanovo metodo meja skladnosti. Razpon obeh osi je zaradi primerljivosti enak pri vseh skupinah. Točke izven meja skladnosti so označene s polnimi krožci.

- *Analiza variance za opažene razlike*

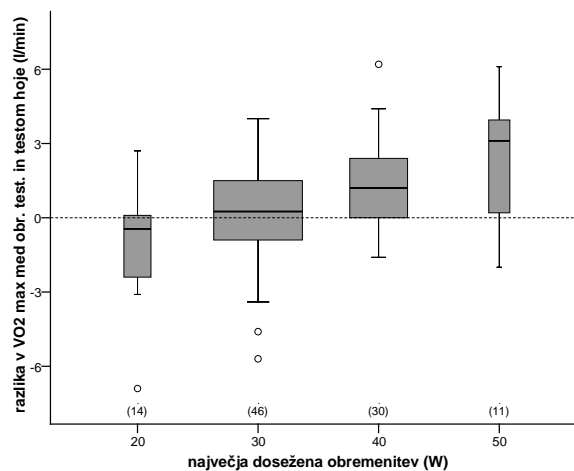
K primerjavi skladnosti dveh meritev med skupinami najpreprosteje pristopimo tako, da za vsakega merjenca izračunamo razliko med meritvama in nato te razlike primerjamo med skupinami z enosmerno analizo variance za neponovljene meritve. Opisne statistike za izračunane razlike so zbrane v tabeli 6, porazdelitev razlik pa prikazuje slika 5. Povprečna razlika je bila najbližje želeni vrednosti nič v skupini 30 W.

Predpostavka analize variance o homogenosti varianc ni bila kršena (Levenov test: $p = 0,494$). Razlike med skupinami so bile statistično značilne ($F = 6,657$; $df = 3 / 97$; $p < 0,001$), zato smo opravili naknadne (post-hoc) primerjave, za kar smo uporabili Tukeyev test HSD (po postopku Tukeya in Kramerja za neenako velike skupine). Skupina 30 W se sicer ni statistično značilno razlikovala niti od skupine 20 W ($p = 0,205$) niti od skupine 40 W ($p = 0,138$), le od skupine 50 W ($p = 0,027$), a ker se je skupina 20 W statistično značilno razlikovala od skupin 40 W ($p = 0,005$) in 50 W ($p = 0,001$) in se slednji nista razlikovali med seboj ($p = 0,585$), hkrati pa je uporabljeni test konzervativen,²³ rezultati jasno kažejo, da je bila skladnost s testom hoje največja pri skupini 30 W.

Tabela 6 Opisne statistike za razlike med meritvama največje porabe kisika po skupinah največje dosežene obremenitve.

Skupina	M(d)	SD(d)	Min(d)	Max(d)
20 W	-1,04	2,27	-6,9	2,7
30 W	0,26	2,19	-5,7	4,0
40 W	1,36	1,84	-1,6	6,2
50 W	2,33	2,64	-2,0	6,1

d = razlika v največji izmerjeni porabi kisika med obremenitvenim testiranjem in testom hoje.



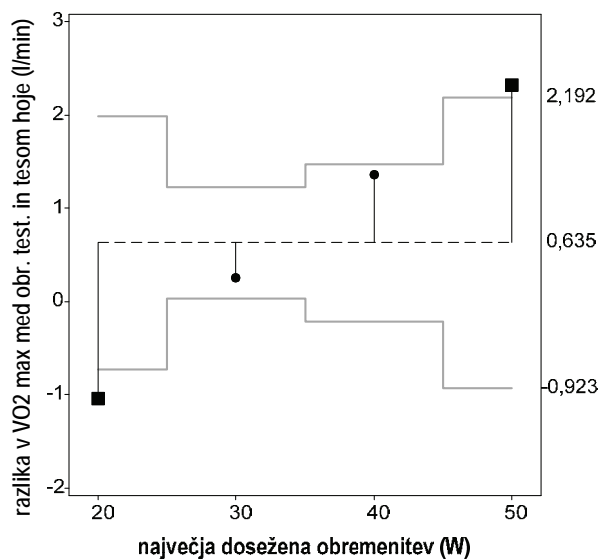
Slika 5 Porazdelitev razlike v največji izmerjeni porabi kisika med obremenitvenim testiranjem in testom hoje glede na največjo doseženo obremenitev. Tanka vodoravna črta označuje ničelno razliko. Širina škatlastih diagramov je sorazmerna z velikostjo skupine, ki je navedena v oklepaju nad oznako skupine.

- *Analiza povprečij za opažene razlike*

Manj znana alternativa (predvsem enosmerni) analizi variance je analiza povprečij (angl. analysis of means, ANOM).²⁴ V Sloveniji je žal praktično povsem neznan, čeprav je v svetu vsaj na področju industrijske statistike ter statističnega nadzora procesov in kakovosti (tudi v zdravstvu) vse bolj priznana. Njena prednost je v razumljivosti za uporabnike, saj je prikaz rezultatov vedno grafičen, in samodejni primerjavi med skupinami, s čimer odpade zapleteno področje naknadnih oziroma načrtovanih primerjav v analizi variance, ki je poleg tega polno nasprotujočih si mnenj in pristopov. V tem članku seveda ne moremo predstavljati matematičnega ozadja postopkov ANOM, ki leži v Waldovem testu zaporednega razmerja verjetnosti,²⁵ zato le opozarjamo, da so ustrezne oblike analize povprečij z dokazano robustnostjo in visoko relativno močjo že razvite za primerjavo povprečij, razmerij, deležev in varianc tako med enako velikimi skupinami kot med skupinami neenake velikosti.²⁴

Analizo povprečij za izračunane razlike med meritvama največje porabe kisika podaja slika 4. Statistično značilno odstopata skupini 20 W (v

negativni smeri) in 50 W (v pozitivni smeri). Sklep je podoben kot na podlagi koeficienta skladnostne korelacije, torej da največja izmerjena poraba kisika pri obremenitvenem testiranju ne odstopa od porabe na testu hoje pri pacientih, ki so dosegli obremenitev 30 W ali 40 W.



Slika 6 Analiza povprečij (ANOM) za razliko v največji izmerjeni porabi kisika med obremenitvenim testiranjem in testom hoje glede na največjo doseženo obremenitev. Kvadrata označujeta povprečja skupin, ki statistično značilno odstopata, kroga pa skupin, ki ne odstopata statistično značilno. Siva stopničasta črta označuje meje intervala, znotraj katerega povprečja ne odstopajo statistično značilno od povprečja izbrane skupine. Črtkana vodoravna črta označuje povprečje preko vseh skupin.

- *Neparametrična analiza opaženih razlik*

Neparametrično analizo opaženih razlik lahko za obravnavane podatke uporabimo bodisi zato, da bi se izognili predpostavki o normalni porazdelitvi ostankov pri analizi variance, bodisi zaradi že omenjene možnosti merskih napak in iz nje izhajajoče želje po robustni analizi, ki bi upoštevala le vrstni red (rang) razlik, ne pa njihove (lažno) natančne velikosti.

Če bi želeli izvesti analizo, povsem analogno opisani analizi variance, bi za opažene razlike izvedli test Kruskala in Wallisa z naknadnimi

primerjavami.²⁶ Seveda bi osnovni test potrdil, da so razlike med skupinami statistično značilne ($p = 0,002$), naknadne primerjave pa so za neparametrične alternative analizi variance izrazito problematično področje. Ustrezno programje za okolje IBM SPSS[®] sicer obstaja,¹⁰ a ustrezen izbor med razpoložljivimi možnostmi (Fisherjev test najmanjše značilne razlike – LSD, postopek Dunna in Sidaka, priredba Tukeyevega postopka, priredba testa Newman in Keulsa, primerjave vseh dvojic s testom Manna in Whitneyja ter popravki za množstvo primerjav po Bonferroniju, Holmu ali postopku FDR) in interpretacija sta še težji nalogi kot pri parametrični analizi.

Ker opisne statistike in grafični prikaz opaženih razlik jasno kažejo velikostni vrstni red skupin (ki ustreza naraščajoči največji doseženi obremenitvi), je upravičeno in mnogo preprosteje uporabiti neparametrični test Jonckheereja in Terpstre²⁶⁻²⁸ za urejene alternative. Z njim smo ničelno hipotezo o enakosti srednjih rangov opaženih razlik jasno zavrnili v prid predpostavljenega zaporedja po skupinah (eksaktni $p < 0,001$). Ker so opažene razlike v skupini 30 W najbližje ničelnim, se lahko torej odločimo za testiranje pri obremenitvi 30 W kot presejalni postopek za zmožnost uporabe nadkolenke proteze pri preučevani populaciji bonikov.

Razprava

Statistična analiza skladnosti se je začela pred stotimi leti s prvo mero za dvojiške spremenljivke,²⁹ a razmahnila se je šele pred dvajsetimi leti po izidu prve knjige, ki ji je bila v celoti posvečena.³⁰ Prva celovita sodobna učbenika sta stara celo manj kot desetletje^{31,32} in še ta sta posvečena predvsem opisnim spremenljivkam. Na področju statistične analize skladnosti številskih spremenljivk, kjer obstaja prava poplava metod in člankov, so zato uporabni tudi članki, kakršen je pričujoči, ki imajo poleg raziskovalnega tudi pregledni in pedagoški namen.

Kljub prizadevanju za celovitost in predstavitev čim več alternativ sta v članku povsem izpuščena dva možna pristopa k analizi skladnosti meritev, saj predstavljata dve kompleksni področji statistike, ki ju zaradi prostorskih omejitev nebi bilo mogoče ustrezno predstaviti. To sta bioekvivalenca in ocenjevanje zanesljivosti merjenja (zlasti z intraklasnimi koeficienti korelacije, angl. intraclass correlation, ICC).

Pri predstavljenih regresijskih analizah skladnosti meritev predstavlja največjo težavo (pre)majhen vzorec. Pri majhnem razponu, kakršnega pomeni razmerje največje in najmanjše vrednosti okoli 2 (kakršno je bilo za obe meritvi porabe kisika v vseh štirih skupinah), je namreč priporočljiva velikost vzorca več sto enot,³³ kar se v našem primeru nanaša na število pacientov znotraj vsake skupine! Omeniti je potrebno tudi nadaljnje regresijske možnosti,^{17,34-37} ki vključujejo uteženo metodo glavne osi, bivariatno različico metode najmanjših kvadriranih median, dodatne koeficiente skladnosti, grafične postopke in postopke za primerjavo z zlatim standardom (za katerega bi izbrali izmerjeno porabo kisika pri testu hoje).

Pri primerjavi opaženih razlik med skupinami je prvi možen pomislek, da bi bilo morda bolje uporabiti absolutne ali kvadrirane razlike, torej upoštevati le velikost odstopanja (v primeru kvadriranih razlik z dodatno težo za večja odstopanja) brez smeri, a zaradi namena primerjave, kjer podcenjena in precenjena pacientova zmogljivost seveda nimata enakih posledic, to nebi bilo smiselno. Dalo pa bi se prirediti zahtevnejšo metodo analize razlik, ki uporablja splošen regresijski pristop,³⁸ a to bi presegalo raven in okvir pričujočega članka.

Tudi metoda meja skladnosti seveda še zdaleč ni dokončna, univerzalna in nesporna rešitev. Razširila sta ga že sama avtorja³⁹ in nedavno je doživel izboljšave za primere, ko predpostavke osnovnega postopka niso izpolnjene,⁴⁰ hkrati pa je deležen tudi kritik v prid regresijskega pristopa za meritve z napakami.⁴¹ V vsakem primeru lahko Blanda in Altmana – tako zaradi metode meja skladnosti oziroma članka, v katerem sta jo prvič

predstavila²² in sodi med največkrat citirane znanstvene članke nasploh, kot zaradi njunih statističnih kolumn v časopisu British Medical Journal in njunih priljubljenih učbenikov – imenujemo ambasadorja medicinske statistike, zato priporočamo branje kateregakoli oziroma čim večjega števila njunih del.

Najbolj temeljno in najsplošnejšo kritiko (ali vsaj ključno dopolnilo) vsem predstavljenim metodam predstavlja bayesovski pristop.^{42,43} Ob tem Carstensen⁴³ izrecno zavrača praktično vse koeficiente skladnosti (vključno z Linovim in intraklasnimi), saj so odvisni od variabilnosti v konkretnem vzorcu, ta pa pri kliničnih študijah nikoli ne more ustrezno predstavljati variabilnosti v vseh populacijah, pri katerih se obravnavane merske metode uporabljajo.

Ne glede na vse navedeno pomeni opravljena analiza pomemben napredek v primerjavi z uvodno študijo istega problema,⁴⁴ ki je bila opravljena brez sodelovanja statistika, kar je botrovalo metodološkimi pomanjkljivostim (namesto analize skladnosti je bila opravljena le primerjava povprečij in to stratificirano po spolu, čeprav je bil vzorec še manjši). Seveda pa ostaja dovolj razlogov za nadaljnje delo tako z vsebinskega (internističnega in rehabilitacijskega) kot metodološkega (eksperimentalnega in statističnega) vidika.

Zaključek

Vprašanje statistične analize skladnosti postopkov oziroma rezultatov merjenja je eno izmed mnogih področij, ki kažejo, kako se prevladujoča znanstvena praksa razvija vse prej kot premočrtno in kumulativno. Pričujoči članek, v katerem je predstavljen primer uporabe analize skladnosti meritev na področju rehabilitacije, to izpostavlja skozi pregled različnih pristopov in možnosti analize skladnosti za številске spremenljivke.

Opravljene analize nadgrajujejo predhodna prizadevanja pri reševanju problema in ugotovitve

bodo neposredno uporabne v klinični praksi. Kljub temu ostaja še veliko prostora za nadaljnje izboljšanje statističnih analiz. Hkrati upamo, da bo članek bralce spodbudil k izobraževanju o širokem in pomembnem področju statistične analize skladnosti.

Literatura

1. Van Der Linde H, Geertzen JH, Hofstad CJ, Van Limbeek J, Postema K: Prosthesis prescription in the Netherlands: an observational study. *Prosthet Orthot Int* 2003; 27(3): 170-177.
2. Dudek NL, Khan OD, Lemaire ED, Marks MB, Saville L: Ambulation monitoring of transtibial amputation subjects with patient activity monitor versus pedometer. *J Rehabil Res Dev* 2008; 45(4): 577-585.
3. Erjavec T, Prešern M, Burger H: The diagnostic importance of exercise testing in developing appropriate rehabilitation programmes for patients following transfemoral amputation. *Eur J Phys Rehabil Med* 2008; 44: 133-139.
4. Hagberg H, Haggstrom E, Branemark R: Physiological cost index (PCI) and walking performance in individuals with transfemoral prostheses compared to healthy controls. *Disabil Rehabil* 2007; 29(8): 643-649.
5. Devlin M, Sinclair LB, Colman D, Parsons J, Nizio H, Campbell JE: Patient preference and gait efficiency in a geriatric population with transfemoral amputation using a free swinging versus a locked prosthetic knee joint. *Arch Phys Med Rehabil* 2002; 83(2): 246-249.
6. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories: ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med* 2002; 166(1): 111-117.
7. Traballese M, Porcacchia P, Averna T, Brunelli S: Energy cost of walking measurements in subjects with lower limb amputations: a comparison study between floor and treadmill test. *Gait Posture* 2008; 27(1): 70-75.
8. Glass S, Dwyer GB: Arm ergometer equation. In: ACSM's Metabolic Calculations Handbook, 7th ed. Philadelphia 2007: Lippincott Williams & Wilkins, American College of Sports Medicine, 58-60.
9. Jones CJ, Rose DJ (eds.): *Physical Activity Instruction of Older Adults*. Champaign 2005: Human Kinetics.
10. García-Granero M: Free SPSS Stuff. <http://gjyp.nl/marta/>, 2010.
11. HoaglinDC, Mosteller F, Tukey JW (eds.): *Exploring Data Tables, Trends and Shapes*. New York 1985: John Wiley & Sons.
12. Few S: *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Oakland 2004: Analytics Press.
13. Chow GC: Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 1960; 28(3): 591-605.
14. Cornbleet PJ, Gochman N: Incorrect least-squares regression coefficients in method-comparison analysis. *Clin Chem* 1979; 25(3): 432-438.
15. Linnet K: Evaluation of regression procedures for methods comparison studies. *Clin Chem* 1993; 39(3): 424-432.
16. Ludbrook J: Comparing methods of measurement. *Clin Exp Pharmacol Physiol* 1997; 24(2): 193-203.
17. Ludbrook J: Linear regression analysis for comparing two measurers or methods of measurement: but which regression? *Clin Exp Pharmacol Physiol* 2010; 37: 692-699.
18. Passing H, Bablok W: A new biometrical procedure for testing the equality of measurements from two different analytical methods. *J Clin Chem Clin Biochem* 1983; 21(11): 709-720.
19. Lin LI-K: A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; 45(1): 255-268.
20. Lin LI-K: A note on the Concordance Correlation Coefficient. *Biometrics* 2000; 56(1): 324-325.
21. Tedeschi LO: Assessment of the adequacy of mathematical models. *Agric Syst* 2006; 89(2-3): 225-247.
22. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1(8476): 307-310.
23. Hayter AJ: A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *Ann Stat* 1984; 12(1):61-75
24. Nelson PR, Wludyka PS, KAF Copeland: *Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. Philadelphia 2005: SIAM Press.
25. Wald A: *Sequential Analysis*. New York 1947: John Wiley & Sons.
26. Conover WJ: *Practical Nonparametric Statistics*, 3rd ed. New York 1999: John Wiley & Sons.

27. Jonckheere AR: A distribution-free k -sample test against ordered alternatives. *Biometrika* 1954; 41(1-2): 133-145.
28. Jonckheere AR: A test of significance for the relation between m rankings and k ranked categories. *Br J Stat Psychol* 1954; 7(II): 93-100.
29. Yule GU: *An Introduction to the Theory of Statistics*. London: Griffin, 1911.
30. Dunn G (ed.): *Design and Analysis of Reliability Studies: The Statistical Evaluation of Measurement Errors*. London 1989: Hodder Arnold.
31. Shoukri MM: *Measures of Interobserver Agreement*. Boca Raton 2003: CRC Press
32. von Eye A, Mun EY: *Analyzing Rater Agreement: Manifest Variable Methods*. Mahwah 2005: Lawrence Erlbaum.
33. Linnet K: Necessary sample size for method comparison studies based on regression analysis. *Clin Chem* 1999; 45(6 Pt 1): 882-894.
34. Choudhary PK: A tolerance interval approach for assessment of agreement in method comparison studies with repeated measurements. *J Stat Plann Infer* 2008; 138: 1102-1115.
35. Haber M, Gao J, Barnhart HX: Evaluation of agreement between measurement methods from data with matched repeated measurements via the Coefficient of Individual Agreement. *J Data Science* 2010; 8(3): 457-469.
36. Luiz RR, Costa AJ, Kale PL, Werneck GL: Assessment of agreement of a quantitative variable: a new graphical approach. *J Clin Epidemiol* 2003; 56(10): 963-967.
37. St Laurent RT: Evaluating agreement with a gold standard in method comparison studies. *Biometrics* 1998; 54(2): 537-545.
38. Carstensen B: Comparing and predicting between several methods of measurement. *Biostatistics* 2004; 5(3): 399-413.
39. Bland JM, Altman DG: Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 2007; 17(4): 571-582.
40. Woodman RJ: Bland-Altman beyond the basics: creating confidence with badly behaved data. *Clin Exp Pharmacol Physiol* 2010; 37(2): 141-142.
41. Ludbrook J: Confidence in Altman-Bland plots: a critical review of the method of differences. *Clin Exp Pharmacol Physiol* 2010; 37(2): 143-149.
42. Broemeling LD: *Bayesian Methods for Measures of Agreement*. Boca Raton 2009: Chapman & Hall/CRC.
43. Carstensen B: *Comparing Clinical Measurement Methods: A Practical Guide*. Chichester 2010: John Wiley & Sons.
44. Erjavec T, Prešern Štrukelj M, Burger H, Vipavec B, Zalar M: Telesna zmogljivost in hoja z nadkolensko protezo pri starejših bolnikih. *Rehabilitacija* 2009; 8(2): 3-9.