

VESTA 2014

Interbedömarreliabilitet för neurologiska fynd vid achillodyni

Mårten Trotzig, ST-läkare, Blackebergs Vårdcentral

April 2017

marten.trotzig@gmail.com

Klinisk handledare: Anders Nilsson, spec. i allmänmedicin Blackebergs vårdcentral

Vetenskaplig handledare: Bo Christer Bertilson, MD/PhD, Huddinge akademiska vårdcentral

Sammanfattning

Bakgrund: Klinisk bedömning av nervstörningsfynd ligger till grund för diagnos och behandling vid bland annat achillodyni. Interbedömarreliabiliteten i olika kliniska tester har tidigare studerats och visat sig variera påtagligt beroende på vilket test som använts.

Syfte: Syftet med denna studie är att undersöka interbedömarreliabilitet för kliniska neurologiska fynd vid achillodyni.

Material och metod: Vi rekryterade 33 löpare med (n=17) eller utan (n=16) smärta från hälsena eller hälben. Samtliga undersöktes av två av varandra oberoende läkare enligt ett gemensamt protokoll som innehöll 33 olika specifika undersökningar. Kappvärden på undersökningarna beräknades för att bedöma hur väl de två undersökarnas bedömningar stämde överens.

Resultat: Interbedömarreliabiliteten beräknad i form av kappvärden varierade mellan olika undersökningar. Vid bedömning av beröring och smärta erhöles högre värden ($\kappa=0,31-0,68$, SD 0,16.0,17) än vid bedömning av reflexer i patella och achilles ($\kappa=0,4-0$, respektive 0,14).

Slutsats: I likhet med tidigare studier noterades bäst interbedömarreliabilitet vid undersökningar av sensorik och sämst vid undersökning av reflexer vilket innebär att värdet av reflexbedömning är lågt.

MeSH-termer – neuralgi; observatörsvariation; primärvård; fall-kontrollstudier

Innehållsförteckning

Sammanfattning	2
Innehållsförteckning.....	3
Bakgrund	4
Syfte.....	5
Frågeställning	5
Material och metod	5
Studiedesign	5
Material	5
Metod.....	6
Statistik.....	7
Etiska överväganden	8
Resultat.....	8
Diskussion.....	9
Slutsats.....	10
Referenser	11

Bakgrund

Klinisk undersökning anses, tillsammans med anamnes, vara en central del av en läkarundersökning (1). Inom hälso- och sjukvården förekommer många situationer när man måste förlita sig på flera undersökare vid insamling av data och vid uppföljning av patienter. Då alla människor ser och tolkar fynd på olika sätt kommer det alltid att föreligga risk för brist i överensstämmelse i de data som produceras. I en vårdapparat som i allt mindre omfattning erbjuder kontinuitet i relationen mellan läkare och patient blir det viktigt att kunna förlita sig på tidigare gjorda mätningar för uppföljning. Det är då viktigt att tidigare och aktuella undersökningsresultat är jämförbara. Inom forskning är det viktigt att olika insamlare av variabler tolkar dessa lika för att kunna göra jämförande analyser mellan olika grupper.

Interbedömarreliabilitet avser nivån av överensstämmelse mellan olika undersökare (2) och har studerats i många sammanhang. I en studie av trycksår mätte man rodnad, svullnad och ulceration. Vid bedömning av trycksårens storlek och färg hade man mätverktyg för att objektivt bedöma skillnader men vid bedömning av svullnad blev bedömningen mer subjektiv (3). I en meta-analys över mikroskopiundersökningar av cellprover konstaterade man att olika bedömare bedömde cellerna på objektglasen olika (4). Vad gäller neurologiska undersökningar fann man i en studie att interbedömarreliabiliteten blev högre på direkt observerbara tecken, till exempel gånghastighet och tremor, än när man bedömde framkallade tecken som reflexer, muskelsvaghet eller vid test av sensorik (5). Bertilson et al har visat att vid undersökning av nack-/skulderbesvär gav en noggrann anamnes inklusive besvärsritning flera positiva fynd på nervpåverkan men inte en högre interbedömarreliabilitet (1).

Klinisk bedömning av nervstörningsfynd bygger på ett antal tester. Där ingår bland annat test av reflexer, grovkraft och sensorik. Dessa undersökningar utförs i allmänhet endast av en undersökare. Bedömningen ligger sedan till grund för diagnos och huruvida ett besvär är neurogent eller nociceptivt vilket kan medföra olika behandlingar. Forskning avseende relevansen av statusfynd är av stort värde för att i framtiden kunna göra mer informerade bedömningar och för att bättre kunna värdera fynden.

Achillodyn är ett samlingsbegrepp för ospecifik smärta i hälsenan där kliniska eller radiologiska tecken på inflammation inte föreligger (6). Smärtan har i likhet med andra tendalgier, tex epikondylit, traditionellt ansetts vara nociceptiv. Hypotesen att tendalgier som achillodyn snarare utgörs av nervstörning på ryggmärgsnivå har tidigare framförts och studerats (7). Studier av interbedömarreliabilitet av nervstörningsfynd vid en fall-kontrollstudie av achillodynpatienter saknas.

Syfte

Syftet med denna studie är att undersöka interbedömarreliabilitet för kliniska neurologiska fynd vid achillodyn.

Frågeställning

Vilken interbedömarreliabilitet föreligger vid klinisk undersökning av patienter med achillodyn och friska kontrollpersoner?

Material och metod

Studiedesign

Studien är en interbedömarreliabilitetsstudie i form av kappaberäkningar på en klinisk fall-kontrollstudie utförd av två bedömare som arbetat parallellt.

Material

Mellan mars och september 2014 rekryterades 33 löpare ("försökspersoner") genom annonsering till och underhandskontakter med sex löparklubbar i Stockholm (n=17) och i Wien (n=16) för klinisk neurologisk undersökning av nedre extremiteter enligt ett i förväg upprättat undersökningsprotokoll. Försökspersonerna fördelades på fall (n=17) och kontroller (n=16). Samtliga försökspersoner gav skriftligt medgivande till att ingå i studien.

Det föreliggande arbetet utgör en del i en fall-kontrollstudie för att kartlägga samband mellan upplevd smärta i hälsena och neurologiska fynd vid klinisk undersökning.

Inklusionkriterier var uni- eller bilaterala häl-/hälsenebesvär, ålder 18-55 år, kunna tala och läsa svenska eller engelska språket, samt att man till undersökningen tog med en vän utan sådana besvär som kontrollperson. Exklusionskriterier utgjorde pågående aktiv inflammation i hälområdet, tidigare operation i hälområdet och positivt Thompsons test talande för aktuell skada.

Eftersom kliniska neurologiska fynd inte bedömdes avhängit om försökspersonen var en fall- eller kontrollperson kunde samtliga försökspersoner inkluderas i interreliabilitetsstudien.

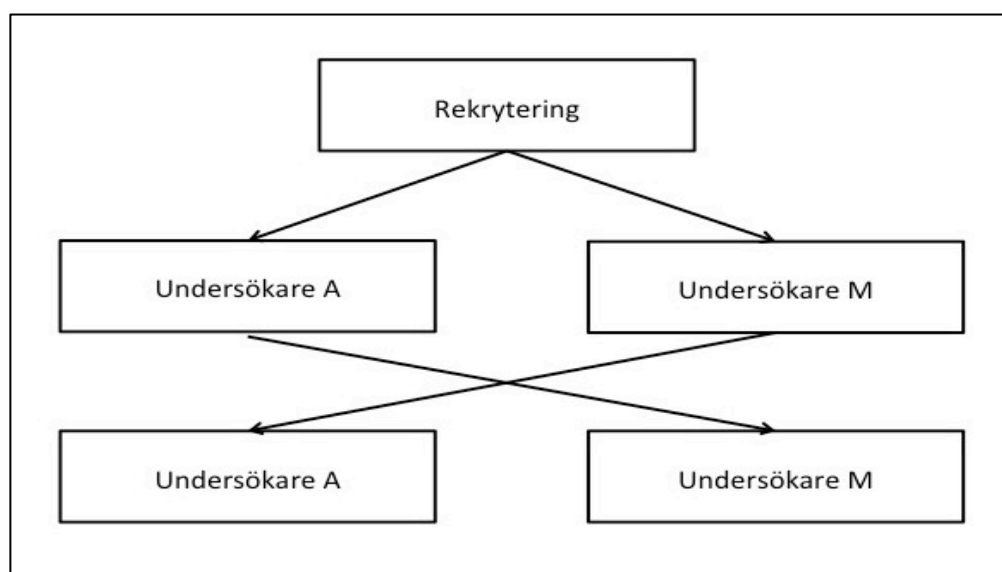
Undersökningarna i Stockholm genomfördes på Capio ArtroClinic och i Wien på Krankenanstalt Rudolfstiftung och i särskilt tillhandahållna lokaler på löparklubbens idrottsplats BFSZ Südstadt.

Metod

Frivilliga försökspersoner fick per telefon eller epost kontakta forskningssekreterare (olika i Stockholm och Wien) som kallade försökspersonerna till undersökning. Ingen randomisering gjordes avseende i vilken ordning undersökarna undersökte varje kontrollperson utan denna fördelning skedde under hand. Innan undersökningen fick försökspersonerna fylla i en besvärshitning och en hälsoenkät (EQ-5D). Undersökarna tog inte del av dessa dokument förrän efter undersökningarna. Ingen anamnes upptogs inför undersökningen.

Samtliga försökspersoner undersöktes parallellt och oberoende av varandra av två ST-läkare i allmänmedicin verksamma vid vårdcentraler i Stockholmsområdet. Fynden nedtecknades i ett i förväg upprättat protokoll och ändrades därefter inte oavsett vilken information om anamnes som tillfördes efter undersökningen. Efter att den första undersökaren var klar med undersökningen gick försökspersonen in till den andra undersökaren och undersöktes på nytt enligt samma protokoll (Figur 1).

Försökspersonen uppmanades att inte för



Figur 1 - Flödesschema

den andra undersökaren avslöja något om vilka fynd som gjorts av den första undersökaren. Efter att båda undersökningarna slutförts gavs försökspersonen visst utrymme att diskutera fynden och gavs behandlingsrekommendationer.

De kliniska undersökningarna utfördes med 33 specifika undersökningasmoment som delades upp i sex olika underskategorier. Underkategorierna var reflexstörning, kraftnedsättning, sensorisk störning, palpationsömhet spinalt, inspektion och palpationsömhet lokalt.

Fynden registrerades i ett i förväg upprättat protokoll (Figur 2). Samtliga undersökningstekniker var manuella och för undersökningarna av reflexer användes reflexhammare, för sensorik användes sporre (smärta) och kylrulle (kyla).

Achillodyn						
Namn:	Personnummer:		Vid icke patologiskt status anges 0.			
Före/efter träning:	Kodnummer:					
Reflexstörning:			Höger +/-	Vänster +/-		
L4 patellarreflex (L3-L4)						
S1 akillesreflex (S1-S2)						
Kraftnedsättning:			Höger (ja/nej)	Vänster (ja/nej)		
L4 quadriceps (sträck knä)						
L5 stortå (dra upp)						
S1 stortå (tryck ner)						
Sensorisk störning	Beröring		Stick		Kyla	
	Höger (+/-)	Vänster (+/-)	Höger (+/-)	Vänster (+/-)	Höger (+/-)	Vänster (+/-)
L4 framsida lår/insida underben						
L5 utsida lår/underben						
S1 utsida fot						
Palpationsömhet:			Spinalutskott/ paravertebralt			
			Utskott (ja/nej)	Paravert höger	Paravert vänster	
L3						
L4						
L5						
S1						
<hr/>						
Inspektion						
Akilles:						
Övrigt:						
Palpationsömhet:			Höger (ja/nej)	Vänster (ja/nej)		
Akillessena						
Calcaneus						

Figur 2 - Protokoll

Statistik

Interreliabiliteten mellan undersökare A och M uttrycks genom överensstämmelse i prevalensen av positiva fynd och genom kappakoefficient. I varje protokoll ingående i studien, noterades avseende testen reflexstörning och sensorisk störning om svaret var normalt, förhöjt eller sänkt. Inför den statistiska analysen har dessa värden dikotomiserats till positivt eller negativt utfall. Samtliga neurologiska test har

dessutom bedömts positiva om asymmetri förelåg vid undersökningen, dvs om det förelåg en skillnad i fynd mellan höger och vänster kroppshalva, vilket är vad man de facto oftast mäter kliniskt. Vid beräkning av kappakoefficient finns en tendens att den påverkas av prevalensen av positiva fynd och fungerar bäst när utfallet mellan alternativen är ungefär lika (1). Därför har kappa ej beräknats när medelvärdet av prevalensen hos båda undersökarna var över 90 % eller under 10 % eller när prevalensen hos den ena undersökaren var 0 %.

Värdering av kappakoefficienten görs efter Landis & Koch: <0 dålig överensstämmelse, 0-0,20 svag överensstämmelse, 0,21-0,40 någon överensstämmelse, 0,41-0,60 låg överensstämmelse, 0,61-0,80 god överensstämmelse och 0,81-1,00 nästan perfekt överensstämmelse (8) (min översättning).

För statistisk analys användes Stata version 13 för Mac (StataCorp LLC, Texas, USA).

Etiska överväganden

De undersökningsmetoder användes i denna studie var desamma som används vid gängse omhändertagande av patienter med häl-/hälsenebesvär och innebar inga smärtsamma moment annat än möjligen momentant övergående palpations-kompressionsömhet vid belastningstest av skadad vävnad.

Studien innebar ett intrång i försökspersonernas integritet och en risk för att deras identitet kunde offentliggöras. Dessutom behövde försökspersonerna avsätta tid för att delta i studien. För att minimera risken för att identiteterna skulle offentliggöras anonymiserades alla data och överfördes från pappersformat till en exceldatabas som sparades på lösenordsskyddade datorer. Resultat kom vidare endast att presenteras på gruppnivå och ingen personinformation kom att användas efter det att försökspersonerna hade deltagit i den fysiska undersökningen. Sedvanlig patientsekretess-/ integritetsskydd gällde för försökspersonerna, det vill säga inga andra fick besked om fynd. Etiskt tillstånd beviljades av Regionala etikprövningsnämnden i Stockholm (diarienummer 2013/2102-31/4).

Resultat

Interbedömarreliabiliteten varierar med kappavärden mellan 0,04 och 0,68. Det högsta kappavärdet uppnåddes vid beröringstest av dermatom L4 och L5 (0,68

respektive 0,52). Beröring av dermatom S1 kunde inte beräknas då medelprevalensen understeg 10 %. Bedömning av reflexer, både av akilles och patella uppvisar lägre överensstämmelse. (0,14 respektive 0,04). Sex av 14 undersökningar uppvisade ett kappavärde >0,39 (Tabell – Prevalens m.m.).

Diskussion

Denna studie är en delstudie av ett försök att försöka se om smärta i hälsena och/eller hälben kan vara kopplat till neurologiska fynd i motsvarande dermatom. Vi valde att lägga upp det med två oberoende bedömare för att få bättre styrka i studien än att bara en undersökare bedömer försökspersonerna. En kappaberäkning av interbedömarreliabiliteten är av värde för att se att de två bedömarna har noterat och tolkat fynd på samma sätt.

Prevalens av positiva fynd och interreliabilitet vid neurologisk undersökning av nedre extremitet (n=33)					
Neurologiskt test	Prevalens		Medel A+M %	Kappavärde	
	A %	M %		κ	SD
Reflexstörning					
Patella	9	27	18	0,04	0,14
Akilles	9	36	23	0,14	0,12
Kraftnedsättning					
L4	21	15	18	0,39	0,17
L5	0	21	11	–	–
S1	0	9	5	–	–
Sensorisk störning					
L4					
Beröring	21	12	17	0,68	0,16
Stick	18	12	15	0,53	0,17
Kyla	24	9	17	0,06	0,15
L5					
Beröring	21	18	20	0,52	0,17
Stick	39	21	30	0,31	0,16
Kyla	30	21	26	0,45	0,17
S1					
Beröring	9	9	9	–	–
Stick	33	9	21	0,63	0,17
Kyla	12	12	12	0,17	0,13

Tabell – Prevalens m.m.

Thaller et al visade 2014 att framprovocerade neurologiska tecken har sämre överensstämmelse än rent observerade tecken och där gjordes även en litteraturgenomgång som visade att kappavärden för "sensation" (motsvarade

sensorisk störning i vår studie) varierar mellan $\kappa=0,15$ och $\kappa=0,50$.(5,9,10).

Kappvärdet för reflexer låg i en studie på $\kappa=0,18$.(10)

Vi fann, i linje med tidigare interrelialitetsstudier, att kappvärden för neurologiska undersökningar uppvisar stor variation. En anledning till detta kan vara att de två bedömarna inte är överens om var dermatomens gränser går eller vilken vertebralkota som palperas. En annan orsak kan vara att beräkningarna utfördes inom ramen för en blindad fall-kontrollstudie där endast hälften av försökspersonerna kunde förväntas ha positiva resultat. Detta kan ha gjort undersökarna mindre benägna att finna positiva fynd då det var förväntat. De flesta liknande studier utförs utan en sådan förväntat "frisk" kontrollgrupp.

Slutsats

I likhet med tidigare studier noterades bäst interbedömarreliabilitet vid undersökningar av sensorik och sämst vid undersökning av reflexer vilket innebär att värdet av reflexbedömning är lågt. Ytterligare studier avseende interbedömarreliabilitet är önskvärda.

Referenser

1. Bertilson BC, Grunnesjö M, Strender L-E. Reliability of clinical tests in the assessment of patients with neck/shoulder problems-impact of history. *Spine*. 2003 Oct 1;28(19):2222–31.
2. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica*. 2012 Oct 15;22(3):276–82.
3. Bluestein D, Javaheri A. Pressure ulcers: prevention, evaluation, and management. *Am Fam Physician*. 2008 Nov 15;78(10):1186–94.
4. Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. *Am J Epidemiol*. 1995 Apr 1;141(7):680–9.
5. Thaller M, Hughes T. Inter-rater agreement of observable and elicitable neurological signs. *Clin Med Lond Engl*. 2014 Jun;14(3):264–7.
6. Harrasser N, von Eisenhart-Rothe R, Toepfer A. [Achillodynia]. *MMW Fortschr Med*. 2016 Aug;158(14):48–9.
7. Jensen R, Kvale A, Baerheim A. Is pain in patellofemoral pain syndrome neuropathic? *Clin J Pain*. 2008 Jun;24(5):384–94.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977 Mar;33(1):159–74.
9. Hand PJ, Haisma JA, Kwan J, Lindley RI, Lamont B, Dennis MS, et al. Interobserver Agreement for the Bedside Clinical Assessment of Suspected Stroke. *Stroke*. 2006 Mar 1;37(3):776–80.
10. Carswell C, Rañopa M, Pal S, MacFarlane R, Siddique D, Thomas D, et al. Video Rating in Neurodegenerative Disease Clinical Trials: The Experience of PRION-1. *Dement Geriatr Cogn Disord Extra*. 2012;2(1):286–97.