

Många symtomskalor håller inte måttet

Granskning av skattningsmetodiken

GUNNAR BORG, fil dr, med dr holms universitet
h c, professor emeritus i percep- gbg@psychology.su.se
tion och psykofysik, Stock-

Folk söker inte läkarvård därför att de har en viss sjukdom utan därför att de känner sig sjuka. Det var innebörden i vad Hans Dalström, överläkare i klinisk fysiologi vid Umeå lasarett, sa till mig i slutet av 1950-talet när han inbjöd mig till ett forskningssamarbete om symtomen under arbetsprov. Vi hade träffats tidigare och diskuterat hur man ska kunna bestämma något så subjektivt som en upplevelse. En upplevelse är ju inte något som kan mätas med en fysisk skala utan är osäker och privat, inte något objektivt med en väldefinierad enhet. Ändå har vi behov av att så gott som möjligt bestämma våra upplevelser: vad de handlar om och hur starka de är, både kvalitet och kvantitet. Det behövs en kvantitativ metodik, som är generellt användbar – eller gör det inte det? Kan det inte räcka med att fråga en person vad han eller hon upplever? Språket är ju det viktigaste för att berätta något.

Metodiken för att mäta upplevelser är gammal

Metodiken för att mäta upplevelser har gamla anor. Enkla skattningsskalor med 4–5 svarsalternativ började användas för flera hundra år sedan. Antalet alternativ utökades senare till 7 eller mer. Fler behövdes sällan för bestämning vid ett visst tillfälle, vilket visades redan på 1950-talet [1].

På 1800-talet formulerade fysikern och matematikern Fechner en ekvation för hur upplevelsen (R) varierar med stimulusintensiteten (S): $R = k \cdot \log S$ (jfr decibelbegreppet). Någon ny metodik utformades dock inte. Fechner kallade forskningsområdet för psykofysik. Termen är inte välkänd. En mer begriplig term är perceptometri [2]. I mitten av 1900-talet utvecklades vid Harvard en metodik för kvotskalning [3]. Den utgick från en analogi med en fysikalisk skala med en specifik enhet, ekvidistans och absolut nollpunkt.

Olika metoder prövades för estimering och produktion. Den mest använda kom att bli magnitudestimering. Personer fick fritt, utan restriktioner, uppskatta styrkan i sina upplevelser med siffror efter eget val. När nästan hela variationsområdet användes – från mycket svaga till mycket starka och inte bara styrkan vid ett tillfälle – visade det sig att en mycket stor siffervariation var nödvändig: 0–30 eller något större. Fechners logaritmfunktion visade sig inte stämma väl, ens för ljudstyrka. De flesta S–R-funktioner kunde beskrivas med en potensfunktion av det enkla slaget: $R = c \cdot S^n$, där exponenten enligt Stevens [3] varierade från ca 0,3 till 3 beroende på sinnesmodalitet.

Hans Dahlström och jag började använda kvotskalningsmetodiken för upplevd ansträngning under arbetsprov. En del patienter verkade överskatta ansträngningen, tex skogsarbetare med muskel- och ledbesvär som klagade över att de hade svårt att orka med det tunga ackordsarbetet. Jag hade då studerat upplevd hastighet vid bilkörning genom att låta folk be-

döma hastighetsförändringar vid halveringar och dubblingar av hastigheten. Det visade sig att den som kör 100 km/timme och saktar ned vid skylten 50 upplever att hastigheten minskat med mer än hälften. Redan vid 70 tycker man att det går hälften så fort. Detta kunde sedan bekräftas genom att förarna själva fick producera hastigheter som svarade mot dem som upplevdes som hälften eller dubbelt så höga. Upplevelsen varierade med kvadraten på hastigheten: $R = c \cdot S^2$. Det fanns en uppfattning att skogsarbetarna avsiktligt överdrev ansträngningen för att få förtidspension. Ofta var det inte så, menade Hans Dahlström. Mitt exempel från hastighetsupplevelsen kunde då vara ett exempel till förklaring. Upplever man att arbetet känns dubbelt så tungt kan det vara naturligt att tro att arbetsförmågan gått ned med 50 procent.

Den S–R-funktion vi bestämde var: $R = a + c \cdot S^n$, där exponenten var 1,6 och en extra parameter, a, måste läggas till på grund av förekomsten av ett brusvärde, en svag förnimmelse av trötthet eller värk redan i vila. Det här var en av de första tillämpningarna av kvotskalningsmetodiken vid bestämning av symtom [4, 5].

Problem med aktuella symtomskalor

Ett av de största och viktigaste området för symtomskattning är det som gäller smärta. Kategoriskskalor med ett ganska litet antal svarsalternativ är vanliga, 5–7 siffror förankrade med uttryck som »ingen alls«, »mycket svag«, »svag«, »måttlig«, »stark«, »mycket stark« och »maximal«. Numeriska skalor med endast ankare (förklarande uttryck) i början och slutet förekommer också. Exempel på den första typen av kategoriskskalor är de amerikanska för angina och dyspné. De ankare som valts är inte kongruenta i betydelse med siffrorna och skiljer sig mellan skalorna så att symtomjämförelser försvåras. Båda går från 0 till 4. På anginaskalan är 1 = »light, barely noticeable«, 2 = »moderate« och 3 = »severe«. På dyspnéskalan är 1 = »mild, noticeable«, »Moderate« återfinns först som 3 [6].

Ytterligare ett exempel är från den svenska översättningen av knäsymtomformuläret från International Knee Documentation Committee (IKDC), där olika symtom bestäms med olika skalor [7]. Trots att antalet skalsteg är lika (fem) följs det mellersta, »måttlig«, av »mycket« på den ena skalan men av »extremt mycket« på den andra. I samma artikel finns även ett exempel på en 11-gradig smärtskala (0–10), där 0 är »värsta tänkbara smärta« och 10 är »ingen smärta«. Detta är inte bra, eftersom en ökning bäst beskrivs med ett ökande värde. Ankare mellan 0 och 10 bör även finnas. Vidare finns behov av att ibland tvingas skatta högre än det man – med sin begränsade erfarenhet – tror är »värsta tänkbara«. Det senare är ingen bra och stabil referensnivå (se nedan).

Den vanligaste smärtskalan är VAS, visuell analogskala, som är standardmetoden i USA. Ett av många problem med VAS är att den förekommer i flera olika versioner, som ger skillnader i skattningar. Den ursprungliga skalan består av en linje av en viss längd, tex 10 cm, där början av linjen är markerad med »ingen smärta« och slutet med »högsta (eller värsta) tänkbara«. Patienten får med ett streck eller ett skjutmått »markera hur stark smärta hon eller han har. VAS fungerar i sin ursprungsversion ganska bra för direkta nivåbestämningar (trots osäkerhet vad gäller maxnivå). Några smärtaforskare menar att VAS ger resultat på en kvotskala [8]. Något klart

■ SAMMANFATTAT

En granskning av skalor för symtombestämning visar att många inte klarar de krav som måste ställas för att de ska fungera för direkta nivåbestämningar och för förändringar av och jämförelser med andra symtom.

Några aktuella exempel ges som saknar kongruens i betydelse mellan siffror och sk ankare i form av ord eller bilder. **Här redovisas** flera viktiga krav och hur de ska kunna uppfyllas för god kvalitet i sjukvården.

»Psykisk livskvalitet varierar från 'extremt dålig' till 'excellent' och lycka från 'ingen alls' till 'en hel del'. Det är tragiskt att lyckokänslan inte kan nå högre.«

stöd för detta finns dock inte, och VAS återfinns över huvud taget inte i ledande handböcker i psykofysik [3, 9].

VAS-skalar har konstruerats med tillägg av ytor, t ex en lig-gande triangel eller linjen framställd som en rektangel sön-derdelad i bitar staplade efter varandra så att den kontinuerli-ga variationen förstörs. Ytorna markeras med olika färger, siffror från 0 till 10 läggs till och linjen delas med tio streck. Ansikten läggs till, från t ex glada till ledsna. Ord placeras en-dast i rangordning, som på en vanlig kategoriskala. Det allvar-liga med alla dessa tillägg är att VAS-versionerna varierar i och med att tilläggen inte ökar i kongruens med avstånd och siffror. Detta skapar osäkerhet, individuella olikheter och teknisk felvariation. En patients smärtskattning eller resultat från en klinisk studie redovisas vanligen endast med att smärtan bedömdes med VAS. Men med vilken VAS-skala?

Tolv psykologiprofessorer fick under 2012 bedöma graden av smärta enligt Borg CR100-skalan [10]. Sex glada till ledsna ansikten (som finns på flera VAS) användes. Det ansågs vara svårt att bedöma smärtan eftersom ansiktena inte klart visa-de smärta. Om man trots brist på klara uttryck för smärta ändå försökte bedöma den (intermodalt, som genom att kra-ma en dynamometer) började den som »extremt svag« först vid det tredje till fjärde ansiktet. Svaren visade att den här ty-pen av bilder inte hör hemma i en smärtskala.

En del skalar för livskvalitet visar också brister. En vanlig sådan (QOL/Thyroid) har steg från 0 till 10. För fysiska frågor om trötthet och smärta går alla från »inga« till »allvarliga« (utan mellanankare). De mentala skalorna har däremot olika slutankare. Psykisk livskvalitet varierar från »extremt dålig« till »excellent« och lycka från »ingen alls« till »en hel del«. Det är tragiskt att lyckokänslan inte kan nå högre. Ett annat ex-empel är en skala (EORTC QLQ-C30) som patienter kan få be-svara efter behandling för prostatacancer. Endast fyra steg, »inte alls – lite – en hel del – mycket« (utan ekvidistans), ingår med frågor om huruvida man måste sitta eller ligga på dagar-na och om man har svårt att ta en kort promenad. En kort pro-menad är kanske 40 m för en patient och en halv golfgrunda på 4 km för en annan.

RPE-skalan för ansträngning

För skattning av allmän ansträngning används mest RPE-ska-lan (Ratings of perceived exertion) från 1970 [11]. Den kallas även Borgskalan, officiellt Borg RPE-skalan. Ibland används även CR10-skalan [11]. RPE-skalan ger en linjär tillväxt med belastningen vid arbetsprov av steady state. RPE utgör en to-talskattning av ansträngning, en »Gestalt« (helhetsupplevel-se) av flera integrerade symtom. De två viktigaste symtomen hos friska personer är andfäddhet och muskeltrötthet. Härtill kan komma lätta besvär vid cykling på en obekvämlig sadel, viss ledsmärta, värmekänsla samt sensoriskt brus i huvudet. Valet av Gestalt-begreppet hänger samman med såväl egna och kol-legors erfarenheter och vissa definitioner av ansträngning som med analys av mängder av svar från arbetsprov.

Intratestrelabiliteten är mycket god ($\geq 0,93$ har noterats, parallelltestkoefficienter av samma storlek och retestkorrela-tionen något lägre och mer varierande, 0,75–0,90). Höga vali-ditetskorrelationer mellan RPE och pulsfrekvens (0,70–0,85) har erhållits i ett flertal undersökningar. Predicerings av ar-betsförmåga från submaximala värden är goda från RPE och

val av träningsintensitet för motionärer, idrottare och patien-ter [11–13]. För patienter kan den vanliga andfäddheten övergå i andnöd, bentröttheten i claudicatio, och bröstsmärta kan tillkomma.

Behov i skattningsskalor

Det är flera behov som ska tillgodoses vid konstruktion av en bra skattningsskala. Det vanligaste är bestämning av symto-mets art och styrka »just nu«. Det sker ofta med hjälp av språk-et. För ett visst land och en viss kultur finns ett gemensamt språk, och enligt Wittgenstein kan det då inte finnas ett indi-viduellt språk eftersom språket är till för kommunikation.

Ett annat behov gäller bestämning av kvantitativa styrke-förändringar. Det är då viktigt att känna till relationerna mel-lan olika uttryck, t ex hur mycket starkare en stark upplevelse är än en svag. Det krävs inte bara en rangskala utan också en kvotskala. Adjektiv och adverb kan fungera som multiplikati-va konstanter och kombineras med substantiv på ett otal sätt. En intervallskala som är ekvidistant med lika stora skalsteg är bra. Men en skala bör helst även innefatta en nollpunkt, som på en meterskala. Då vet man att 4 är dubbelt så mycket som 2 och hälften av 8, och man kan bestämma en funktion för stimulus (S)–respons (R) matematiskt, t ex med potens-funktionen: $R = c \cdot S^n$. En utmärkt skala bör även innehålla en schematiserad föreställning som en intersubjektiv enhet. Man kan då göra extrapoleringar från submaximala bestäm-ningar till maximala, t ex av aerob kapacitet och styrka [11, 14].

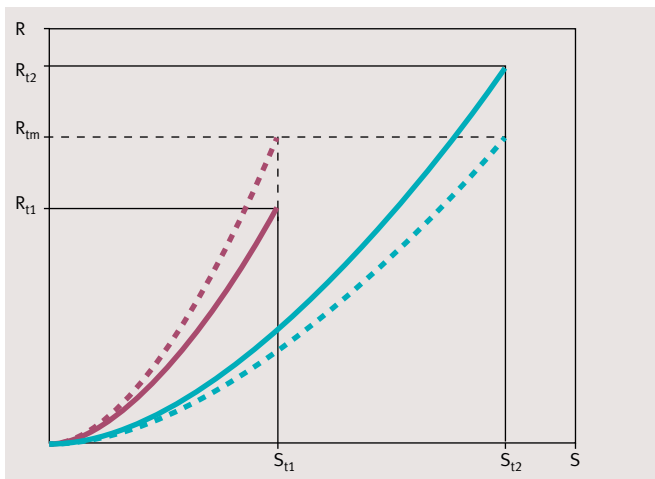
Ibland behövs två extra parametrar i potensfunktionen, a och b: $R = a + c \cdot (S - b)^n$, där a är bruset och b ett S-värde över 0. För en frisk person ökar inte ansträngningen (över eventu-ellt a-värde) av en mycket långsam, kort promenad [11]. Ytter-ligare behov gäller symptomjämförelser, t ex mellan andfädd-het och bentrötthet, mellan smärta och ansträngning, mellan patienter och friska och dessutom mellan symptom och objek-tiva fynd, t ex mellan skattad ansträngning och puls samt mel-lan muskelvärk och mjölksyra.

Stevens ställde stora krav på metodiken för kvotskalning. Den fungerade inte så bra som han önskade men fick ändå stöd av flera ledande psykofysiker, bl a den teori- och matema-tikkunnige Luce [15]. För mina första experiment med upp-levd hastighet och ansträngning fungerade den bra [11]. En unik studie som gav psykofysiologiskt stöd var den mellan smak och nervsvar [9]. En nackdel med Stevens metod var att den bara gav relationer mellan upplevelser och inte direkta ni-våer. Att 70 km/timme upplevs som hälften av 100 är intres-sant men säger inget om huruvida man upplever 70 som snabbt eller långsamt. Det beror på omständigheterna och in-dividerna.

Att utarbeta en interindividuell giltig kvotskala som också är nivåförankrad ansågs omöjligt: »Det kan ju inte finnas en giltig enhet. Det är själens obotliga ensamhet«, menade Stock-holmsprofessorn Ekman. Ett förslag på lösning av detta »olös-liga problem« var rangemodellen (från 1961 och 1962) med en schematiserad föreställning som enhet [5]. En upplevelse av extrem ansträngning av maximalkaraktär, som de flesta kan enas om, är känslan av att lyfta något som är så tungt att man nätt och jämt klarar av det. Om den är lika för olika individer med olika muskelstyrka så kan den subjektiva variationen sättas lika trots att den objektiva inte är lika. Ett barn som tycker att en sten är dubbelt så tung som en annan får kanske medhåll av sin far. Men om barnet säger att den stora stenen är

»Men en skala bör helst även innefatta en nollpunkt, som på en meterskala. Då vet man att 4 är dubbelt så mycket som 2 och hälften av 8 ...«

KLINIK & VETENSKAP ÖVERSIKT



Figur 1. Upplevd kraft för två personer vid kramande av en dynamometer. Heldragna kurvor är rådata och streckade korrigerade enligt rangemodellen, där erhållna fysiska maxvärden (S_1) satts subjektivt lika.

mycket tung, säger kanske pappan att den bara är ganska tung. Rangemodellen ger då möjlighet att justera måttkonstanten och rita in två individers S-R-funktioner i samma diagram (Figur 1). Tyvärr finns nog inte någon liknande enkel referensnivå av intersubjektiv giltighet för smärta, smak, illamående eller ångest.

Nivåkrav och kvotkrav måste kunna förenas

Stora krav måste ställas på en CR-skala (kategori-kvotskala) om den ska förena nivåkravet med kvotkravet. Skalan ska förena språkets rikedom (tex »svag« och »stark« i kombination med adverb) med talens exakthet. Den vanligaste CR-skalan är CR10-skalan, som är en generell symtomskala (Figur 2). Den används vanligen vid arbetsprov och skattning av bröstsmärta, andfåddhet och bentrötthet [11, 17].

Även om den skattade smärtan (R_{sm}) är privat och har stor individuell variation måste vissa generella fakta beaktas. En allmän rekommendation vid dosering måste utgå från en viss intersubjektiv likhet och sedan anpassas individuellt. Möjligheten att justera bestämningen kan då utnyttja intermodala jämförelser och en schematiserad enhet i form av interindividuell likvärdig ansträngning (R_{an}), dvs $R_{sm} \cdot R_{an}$.

CR10 har även använts för skattning av andra symtom, såsom oro i samband med hjärtrehabilitering, illamående efter magoperation, kissnödighet samt hunger och mättnad hos anorexipatienter. Till skillnad från RPE-skalan ger CR10 en positivt ökande funktion för ansträngning, dvs mellan den linjära för puls och den positivt ökande för mjölksyra. Reliabilitets- och validitetskorrelationer för CR10 är av samma storlek som för RPE [11, 18].

Några av de viktigaste kraven uppfylls i rangemodellen: det naturliga, subjektiva variationsområdets storlek, valet av koncisa ankare (»preciseness«), ett schematiserat huvudankare som enhet, innebörden och positioneringen av alla ankare för kongruens mellan siffror och ankare (»interpretation«), en tillförlitlig S-R-funktion, undvikande av skalbegränsning samt möjlighet för patienten att ge direkta svar och för terapeuten att ge rekommendationer enligt samma skala (det senare fungerar inte med VAS). Härtill kommer vanliga psykometriska krav. Tyvärr uppfylls inte alltid kraven, och en hel del fel och missförstånd förekommer.

Det borde vara självklart att skala och instruktion (som RPE- och CR10-skalorna med hundratals miljoner användare över hela världen) inte ändras. Men så är det inte. Särskilt gäller detta CR10-skalan. Det förekommer bl a att skalan ändras



Figur 2. Borg CR10-skalan. Copyright Gunnar Borg 1982, 1998, 2004.

för en viss patientgrupp. För KOL-patienter ändras »strong« (stark) till »severe« (vilket försvårar symtomjämförelser). Andra fel är att halva skalsteg utelämnas liksom möjligheten att skatta över 10. Instruktionen förkortas kraftigt eller utelämnas. Några klarar detta men många gör det inte. En patient tror att »vara i vila« (före cykling) ska vara 0, att »måttlig« är ca 40 procent av »max«, och »stark« 65–70 procent (eventuellt som vid normalfördelning). Det är viktigt att alla är bekanta med skalan och instruktionen före testet. En person måste också instrueras att själv kunna avbryta ett test enligt riktlinjerna i boken »Kliniska arbetsprov« [17].

En mer fingradig CR-skala är Borg CR100-skalan. Den kan ersätta CR10 och i framtiden många andra skattningsskalor [10, 18]. I samband med arbetsprov har CR100 använts även för skattning av dysymtom som illamående, brus i öronen, svettning och obehag [18]. CR100-skalan rekommenderas av allt fler, fast CR10 är mest använd [19].

Min CR-metodik innebar en första kraftig förbättring av all tidigare skattningsskalor och uppfyller därför kraven på autenticitet och originalitet. Det finns dock forskare som konstruerat andra nivåförankrade kvotskalor. Av särskilt intresse är Labeled magnitude scale (LMS), som konstruerats av Green et al 1993 och prövats vidare 1996 [20]. Den gäller dock huvudsakligen smak- och luktperception. Avsaknaden av en schematiserad enhet som referensnivå mellan »very strong« (50) och »strongest imaginable« för en viss förnimmelse (100 eller ca 95) är en stor nackdel. Den kan inte användas generellt för skattningar av symtom. LMS har gjorts mer generell genom att 100 gäller »sensation of any kind«. Ankarnas position har justerats, bl a »strong« från 33 till 34,7 och »very strong« från 50 till 52,5! Skalan är ändå inte lämplig för bestämning eller rekommendationer av styrka.

Ingen skattningsskala kan fungera som en perfekt kvotskala. Även om skalan är den bästa möjliga ska den användas av människor enligt deras egen uppfattning. Jag har ofta citerat Quines krav på sensorisk evidens och påståendet: »The requirements of intersubjectivity is what make science objective«. Intersubjektiviteten kan dock aldrig vara perfekt. Våra sinnen skiljer sig åt, liksom vår bearbetning och värdering av intryck och därmed vad vi påstår. Även om vi har likartade upp-

KLINIK & VETENSKAP ÖVERSIKT

levelser kan vi ge olika svar. En bra skala och tydliga instruktioner är förutsättningen för tillförlitliga symtombestämmingar.

Dåliga symtomskalor bör inte förekomma

De flesta läkare är bra på identifiera symtom vid undersökningar av patienter. Sättet att konstruera generella skalor är dock sämre. Dåliga symtomskalor bör inte förekomma. Flera felanvändningar har lett till felbedömning av patienter. Det behövs inte alltid olika skalor för styrkegrad. Tvärtom, där

skallighet är möjlig förbättrar denna likhet symtomjämförelser! Utformningen och användningen av symtomskalor är viktiga kvalitetsfrågor i sjukvården, i såväl forskning som kliniskt arbete.

■ *Potentiella bindningar eller jävsförhållanden: Inga uppgivna.*

REFERENSER

- Guilford JP. Psychometric methods. New York: McGraw-Hill; 1954.
- Borg G. Psychophysics and perceptometry. In: Armando M. Oliveira et al, editors. Fechner Day 2004. Coimbra: International Society for Psychophysics; 2004. p. 331.
- Stevens SS. Psychophysics: Introduction to its perceptual, neural, and social prospects. New York: Wiley; 1975.
- Borg G, Dahlström H. Psykofysisk undersökning av arbete på cykelergometer. Nord Med. 1959;62:1383-6.
- Borg G. Physical performance and perceived exertion. Studia Psychologica et Paedagogica, Series altera, Investigationes XI. Lund: Gleerup; 1962.
- ACSM. Guidelines for exercise testing and prescription. Philadelphia: Lea & Feibiger; 1991.
- Kvist J, Österberg A, Grävare Silbernagel K, et al. Översättning av IKDC-subjektive score till svenska. Idrottsmedicin. 2011;2:18-21.
- Price DD. Psychological and neural mechanisms of pain. New York: Raven Press; 1988.
- Gescheider GA. Psychophysics. The fundamentals. Mahwah, NJ: Lawrence Erlbaum; 1997.
- Borg G, Borg E. A new generation of scaling methods: Level-anchored ratio scaling. Psychologica. 2001;28:15-45.
- Borg G. Borg's perceived exertion and pain scales. Champaign, IL: Human Kinetics; 1998.
- Nobel B, Robertson R. Perceived exertion. Champaign, IL: Human Kinetics; 1996.
- Pedersen BK, Saltin B. Evidence for prescribing exercise as therapy in chronic disease. Scand J Med Sci Sports. 2006;16(Suppl 1):3-63.
- Borg E, Borg G. A demonstration of level-anchored ratio scaling for prediction of grip strength. Applied Ergonomics. 2013;44:835-40.
- Luce RD. On the possible psychophysical laws. Revisited: remarks on cross-modal matching. Psychol Rev. 1990;97:66-77.
- Borg G, Diamant HL, Ström L, et al. The relation between neural and perceptual intensity: a comparative study on the neural and psychophysical response to taste stimuli. J Physiol. 1967;192:13-20.
- Borg G. Ansträngning och andra symtom. I: Kliniska arbetsprov. Metoder för diagnos och prognos. Jorfeldt L, Pahlm O, redaktörer. Lund: Studentlitteratur; 2013. p. 237-45.
- Borg E. On perceived exertion and its measurement [dissertation]. Stockholm: Stockholms universitet; 2007.
- Mahler D, O'Donnell D. Dyspnea. Mechanisms, measurement, and management. Boca Raton, FL: Taylor & Francis; 2005.
- Green B, Dalton P, Cowart B, et al. Evaluating the »Labeled Magnitude Scale« for measuring sensations of taste and smell. Chem Senses. 1996;21:323-34.

■ SUMMARY

A medical scale should satisfy the need to determine the strength of a symptom, how it changes and how it is related to other symptoms or signs. The strength is satisfied with numbers and anchors. The variation has best been determined by »ratio scaling« without anchors. This makes it difficult to determine direct levels. Most scales do not satisfy the main requirements. A special drawback is the lack of congruence in meaning between numbers and anchors. A good scale should function as a »semi-ratio« scale and the relation between the anchors should have a similar meaning as that between numbers. Scales are constructed that satisfy these demands.