

Läs mer Engelsk sammanfattning <http://larkiv.lakartidningen.se>
En första artikel av samma författare, om analys av förändring när datamaterialet består av ordnade kategorier, publicerades i Läkartidningen nr 43/2005.

Vad är behandlingseffekt om patienten blev *bättre*, men ingen vet *hur mycket bättre*?

Statistisk metod för parade ordinaldata



ELISABETH SVENSSON, professor i statistik, spec biostatistik, institutionen för ekonomi, statistik och

informatik, Örebro universitet
elisabeth.svensson@esi.oru.se

Utvärdering av behandlingseffekt eller annan förändring när datamaterialet består av ordnade kategorier kräver extra uppmärksamhet vid val av statistisk verktygslåda. Det är inte bara bedömningar av kvalitativa variabler som oro, smärta, välbefinnande, kvalitet och funktion på någon skattningsskala som ger upphov till ordnade kategoridata eller ordinaldata, som mätnivån även kallas. Kategoriseringar av mätvärden från standardiserade instrument, som utnyttjar fysikens eller kemins lagar, och sammanfattande nivå- eller stadiindelningar av flera variabler är andra exempel på generering av ordinaldata [1-3]. När kroppslängd till exempel mäts och uttrycks i en standardiserad längdenhet, som meter, då består datamaterialet av kvantitativa kontinuerliga värden, även om de är avrundade, och statistiska metoder för analys av förändring som baseras på differenser, som Students parade t-test och Wilcoxon's teckenrangtest, är möjliga val förutsatt att övriga villkor för testen är uppfyllda. Om däremot kroppslängd kategoriseras till »kort«, »normal«, »lång« för sin ålder, så blir valet av statistisk verktygslåda ett annat [2-5].

Därför kan det bli problem med analys av ordinaldata

Karakteristiskt för ordinaldata är att datamaterialet består av kategorier eller nivåer som har en ordningsstruktur, men som saknar information om storlek och avstånd. Denna begränsning i matematiska egenskaper är mycket påtaglig vid utvärdering av förändring, eftersom beräkning av differenser inte är meningsfull när datamaterialet saknar avstånd och storlek. Detta är ett välkänt faktum som tidigare har varit svårt att helt beakta eftersom det råder brist på statistiska metoder som är avsedda för analys av beroende ordinaldata [3, 6]. De klassiska statistiska metoderna för analys av förändring i ordinaldata är samma som för dikotoma data, det vill säga teckentest och McNemars test. Det innebär att informationen från de extra kategorier som ordinaldata har går förlorad genom att datamaterialet grupperas i andelen av värdena som förändrats mot högre respektive mot lägre kategorier [4-7].

Syftet med denna artikel är att presentera en statistisk metod, som har utvecklats speciellt för analys av förändring i ordinaldata och som utnyttjar informationen från hela datamaterialet på såväl grupp- som individnivå. Förändringen kan gälla en behandlingseffekt eller annan förändring över tid. Med denna metod utvärderas den del av det observerade förändringsmönstret som kan tillskrivas gruppen separat från en eventuell förekomst av individuell heterogenitet i förändringsmönstret. En sådan uppdelning kan ge viktig information inför val av behandlings- eller rehabiliteringsstrategi, eftersom analysen kan ge svar på frågan om patientgruppen är homogen, eller om det

FAKTA 1

De statistiskt viktiga mätnivåerna hos ett datamaterial [1, 2]:
Dikotoma data (binära data, två kategorier)
Kategoridata utan ordningsstruktur (nominaldata)
Kategoridata med ordningsstruktur (ordinaldata)

Kvantitativa antalsdata (diskreta data)
Kvantitativa kontinuerliga data, indelade i icke-negativa, intervallbegränsade eller obegränsade (intervall-, kvotdata).

FAKTA 2

Begrepp
En skala (eng scale, rating scale) består av en uppsättning ordnade alternativ (kategorier) i form av siffror, verbala beskrivningar, bilder eller av en linje (visuell analog skala, VAS). Mätningen/registeringen består av en bedöm-

ning (eng assessment, judgement, rating) på skalan. När individen själv bedömer sitt tillstånd, till exempel smärta, oro, välbefinnande, livskvalitet, används ofta uttrycket självskattning (eng self-assessment).

förekommer starka individvariationer i behandlingseffekt. Möjligheten att kunna identifiera riskindivider i en patientgrupp och att kunna bedöma om åtgärdsprogram behövs på grupp- eller individnivå har både etiska och ekonomiska fördelar.

Datamaterialet

Jag väljer samma datamaterial som i tidigare artikel [7], nämligen variabeln »besvär«, som i en studie skulle kunna betyda illamående, klåda, smärta eller någon annan form av besvär. Skalan består av sex ordnade kategorier: inga besvär (A), mycket lätta besvär (B), måttliga besvär (C), påtagliga besvär (D), mycket stora besvär (E) och extrema besvär (F). Förändringen avser

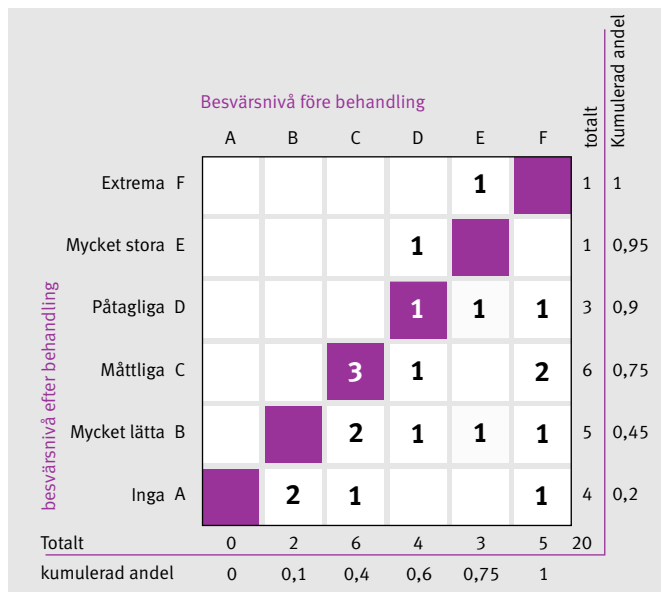
SAMMANFATTAT

Bedömningar med olika typer av skalor och bedömningsformulär ger upphov till ordinaldata. Mätvärdena i denna typ av data representerar enbart en ordningsstruktur och saknar information om storlek och avstånd mellan kategorier.

I artikeln presenteras en statistisk metod, som är speciellt utvecklad för att analysera förändring när datamaterialet

består av ordnade kategoridata, men metoden är användbar på alla datamaterial som åtminstone har en ordningsstruktur.

Tack vare metodens speciella rangteknik med parbundna ranger så är det möjligt att statistiskt beskriva och analysera den grupprelaterade förändringen separat från den individuellt avvikande från gruppens förändring.



Figur 1. Frekvensfördelningen över förändring i besvär bedömda på en skala från A (inga besvär) till F (extrema besvär) före och efter en behandling. De verbala kategoristegen är endast utskrivna vid uppföljningstillfället. Marginalernas kumulerade andelar (relativa frekvenser) är också givna.

effekt av en behandling och 20 individer har skattat sina besvär vid två tillfällen, före respektive efter behandling.

Beskrivning av förändringsmönstret

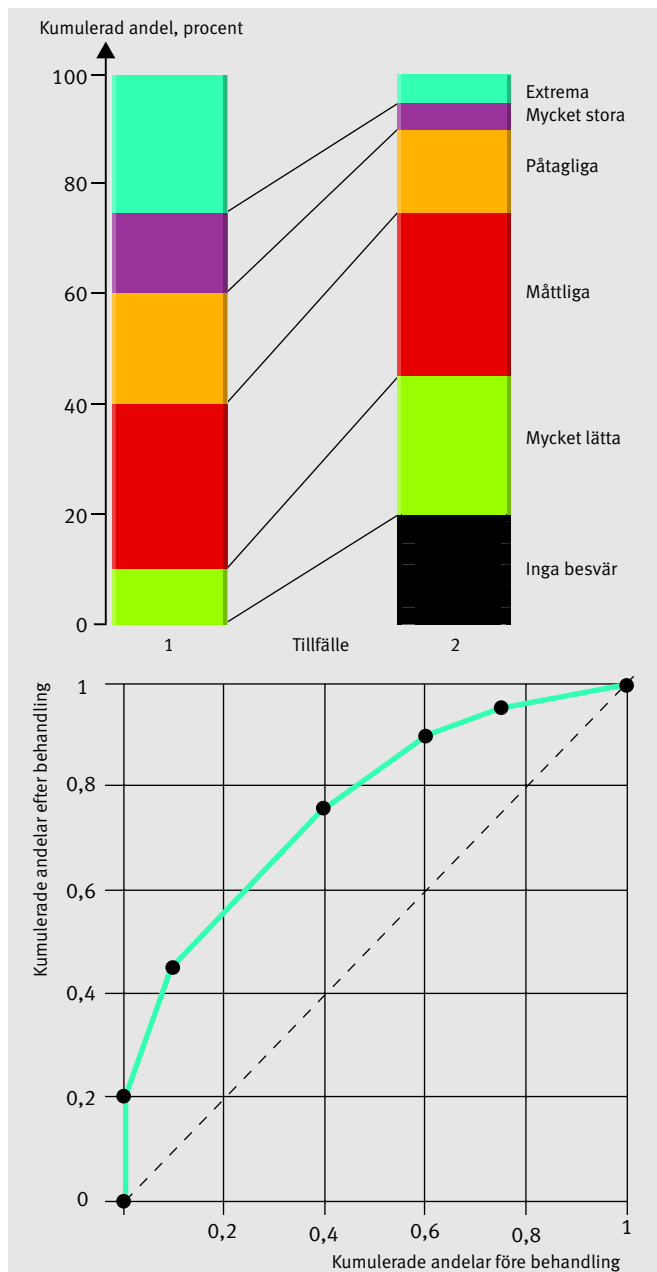
En allmän regel är att den statistiska bearbetningen av förändringsdata skall ta hänsyn till att varje individ representeras av datapar. Det individuella förändringsmönstret beskrivs av frekvensfördelningen av de 20 patienternas par av besvärsnivåer i en korstabell, även kallad kontingenstabell, över de 6×6 möjliga kombinationer av datapar från en sexgradig skala, Figur 1.

Av förändringsmönstrets huvuddiagonal framgår att tre patienter angav »måttliga besvär« (C,C) och en patient »påtagliga besvär« (D,D) vid båda tillfällena. Observationerna (D,E) och (E,F) representerar två patienter som bedömde en högre besvärsnivå efter behandlingen, och de 14 observationerna under huvuddiagonalen representerar de patienter som hade skattat en lägre besvärsnivå efter behandlingen. De fem patienter, som före behandlingen bedömde sig ha »extrema besvär« (F) återfinns i cellerna (F,D), (F,C), (F,B) och (F,A), vilket indikerar individuella variationer i behandlingseffekt.

I en tidigare artikel har jag analyserat materialet med de klassiska metoder som står till buds för par av ordnade kategoridata, nämligen teckentest för parade proportioner alternativt McNemars test [7]. Dessa test visade att denna obalans i antal observationer på båda sidor om diagonalen var statistiskt säkerställd ($P=0,006$), vilket betyder att datamaterialet styrker påståendet om att behandlingen kommer att resultera i en förändring mot mindre besvär hos en relevant population patienter.

Signifikant förändring, men hur?

En kliniskt viktig aspekt vid utvärdering av behandling och rehabilitering är att undersöka om behandlingseffekten är homogen för patientgruppen eller om de individuella förändringarna dominerar. För att utvärdera detta använder vi den statistiska metod, som är speciellt utvecklad för att analysera par av data från skattningsskalor och andra bedömningsinstrument [8-11]. I denna artikel ger jag en principiell beskrivning av metoden och måtten för grupp- och individförändringarna och tolkar re-



Figur 2. a) Stapeldiagram och **b)** Q-Q-kurva över marginalfördelningarna till datamaterialet i Figur 1.

sultaten av den statistiska bearbetningen. En statistisk presentation av måtten, en utförlig bearbetning av datamaterialet i Figur 1 och ett beräkningsprogram finns på min webbplats: <http://www.oru.se/esi/svensson>.

Diagram över systematisk förändring för gruppen

Frekvensfördelningen av dataparen i kontingenstabellen beskriver det individuella förändringsmönstret, men låt oss börja med att studera marginalfördelningarna, som visar hur datamaterialet fördelar sig på skalans kategorier vid respektive tillfälle (Figur 1). Marginalfördelningen före behandling är (0, 2, 6, 4, 3, 5), och efter behandlingen (4, 5, 6, 3, 1, 1). De kumulerade relativa frekvenserna/andelarna är också beräknade. Marginalfördelningarna är olika, vilket tyder på en systematisk förändring av gruppens besvärsnivåer över tid. Fördelningen i Figur 1 är exempel på en systematisk förskjutning av datamaterialet

FAKTA 3

Q-Q-kurva

Ett linjediagram mellan (0;0) och (1;1) som sammanbinder par av kumulerade andelar kallas allmänt för »relative operating characteristic(ROC)-kurva« eller »quantile-quantile«, Q-Q,-kurva.

Observera att båda dessa begrepp används i olika sammanhang för att beskriva relationen mellan par av kumulerade andelar, till exempel vid diagnostisk styrka och vid anpassning till kända sannolikhetsfördelningar!

mot lägre kategorier, i detta fall mot lägre besvärnivåer, eftersom de lägre kategorierna innehåller en större andel av observationerna efter behandling än före. Det är till exempel fler patienter som efter behandling bedömer sina besvär som »mycket lätta« eller »obefintliga« (5+4, det vill säga 45 procent) jämfört med före behandlingen (0+2, det vill säga 10 procent). Vidare bedömde 40 procent av patienterna sig ha »inga« till »måttliga besvär« före behandling. Motsvarande andel efter behandling var 75 procent.

Denna systematiska förändring, som tillskrivs hela gruppen, kan grafiskt illustreras med stapeldiagram med uppdelade staplar, där de relativa frekvenserna för de ordnade kategorierna staplas på varandra, med andra ord kumuleras (se Figur 2a), eller med ett linjediagram, en Q-Q-kurva, som konstrueras genom att bilda par av de två uppsättningar kumulerade frekvenser med utgångspunkt från origo, det vill säga punkten (0;0). Paren i detta fall är (0;0,2), (0,1;0,45), (0,4;0,75), (0,6;0,9), (0,75;0,95) och (1;1), se Figur 2b.

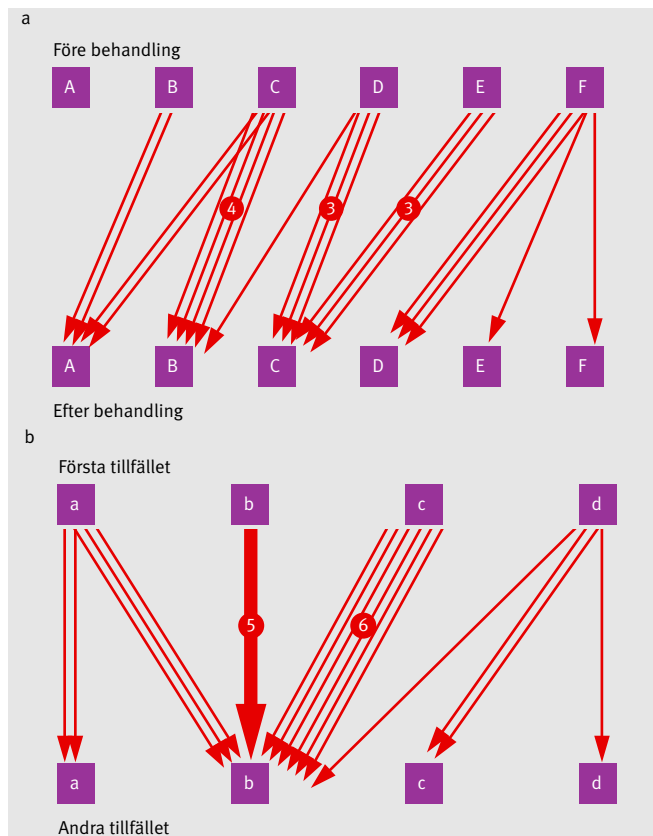
Av stapeldiagrammet, Figur 2a, framgår att kategorierna »inga«, »mycket lätta« och »måttliga« besvär upptar 40 procent av stapelns höjd vid första tillfället och dominerar (75 procent) vid andra tillfället. Gruppens mediannivå, det vill säga kumulerad andel 50 procent, var »påtagliga besvär« före behandling och »måttliga besvär« efter behandling. Kvartilerna och andra percentilkategorier kan också avläsas.

De tunna linjerna mellan de båda staplarna sammanbinder de olika besvärskategoriernas kumulerade andelar med varandra. Den nedersta linjen sammanbinder de kumulerade andelarna för »inga« besvär, det vill säga 0 procent och 10 procent, och den översta linjen representerar de kumulerade andelarna för »mycket stora besvär« (0,75;0,95). Dessa par av kumulerade andelar återfinns också i Q-Q-kurvan, som illustrerar förekomsten av systematisk förändring tvådimensionellt med kumulerad andel före behandling på den horisontella och kumulerad andel efter behandling på den vertikala axeln. Det betyder att om andelen observationer i de lägre kategorierna ökar vid andra tillfället, (0;0,2) och (0,1;0,45), kommer Q-Q-kurvan att ligga ovanför diagonalen, se Figur 2b.

En systematisk förskjutning av data mot högre kategorier ger en Q-Q-kurva under diagonalen, och lika marginalfördelningar är tecken på frånvaro av systematisk förändring i datamaterialet som kan tillskrivas hela gruppen. Då sammanfaller Q-Q-kurvan med diagonalen och stapeldiagrammet kommer att bestå av två identiska staplar. En Q-Q-kurva kan även ha ett S-format utseende (Figur 7). Det betyder att det har skett en koncentrationsförändring av observationerna snarare än en lägesförskjutning över kategorierna mellan de båda tillfällena. En blandning av systematisk läges- och koncentrationsförändring mellan de två tillfällena är vanlig.

Statistiska mått för systematisk gruppförändring

Vilka är måtten som kvantifierar det systematiska förändringsmönstret, som alltså tillskrivs gruppen? Det statistiska måttet

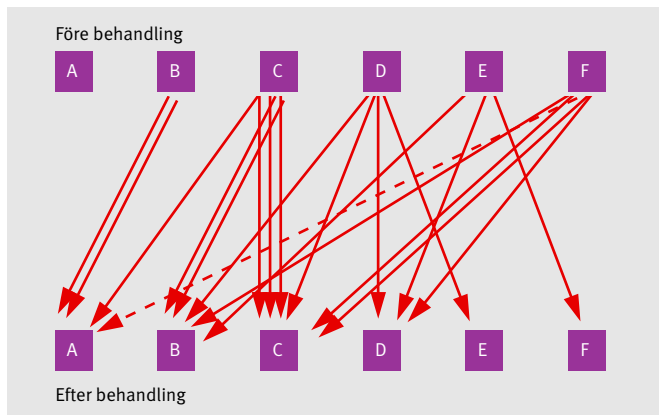


Figur 3. a) Illustration till systematisk förändring i position mot lägre kategorier på skala med de ordnade kategorierna A till D (övre figuren). **b)** Systematisk förändring i koncentration av bedömningar mot den centrala kategorin [b] på en skala med fyra ordnade kategorier (a till d) vid andra tillfället (nedre figuren). Antal observationer i båda fallen är 20.

		Upplevda besvär före behandling						
		A	B	C	D	E	F	totalt
besvär efter behandling	F						1	1
	E						1	1
	D						3	3
	C				3	3		6
	B			4	1			5
	A		2	2				4
Totalt		0	2	6	4	3	5	20

Figur 4. Det rangtransformerbara mönstret, det vill säga det förväntade förändringsmönstret utifrån marginalfördelningarna, förutsatt att individernas förändring inte medför någon förändring av deras inbördes rangordning i gruppen.

för en systematisk förskjutning i läge på skalan mellan de båda mätillfällena är *relativ position, RP*. Detta mått uttrycker skillnaden mellan i vilken utsträckning marginalfördelningen efter behandlingen är förskjuten mot högre snarare än mot lägre kategorier och det motsatta jämfört med marginalfördelningen



Figur 5a. Alternativ illustration av förändringarna i bedömda besvär på en skala med sex kategorier före och efter behandling till dataparen i kontingenstabellen, Figur 1.

		Besvärnivå före behandling					
		A	B	C	D	E	F
besvärnivå efter behandling	Extrema F					15 20	
	Mycket stora E				12 19		
	Påtagliga D				11 16	14 17	20 18
	Måttliga C			7 11	10 13		18,5 14,5
	Mycket lätta B		4,5 5,5	9 7	13 8	17 9	
	Inga A		1,5 1,5	3 3			16 4

Figur 5b. Uppsättningen parbundna ranger till frekvensfördelningen i Figur 1, där rangen för individens bedömning före behandling står ovanför rangen för individens bedömning efter behandling.

före behandlingen. En bågformad Q-Q-kurva, som i Figur 2b, illustrerar en systematisk förändring i läge. Figur 3a är en alternativ illustration.

Måttet RP fångar inte upp de fall då marginalfördelningarna vid de två tillfällena är olika koncentrerat på skalkategorierna. Q-Q-kurvan har då en mer eller mindre uttalad S-form. Det statistiska måttet för en systematisk förändring i koncentration av bedömningarna, *relativ koncentration*, RC, uttrycker skillnaden mellan i vilken utsträckning marginalfördelningen efter behandlingen är mer centrerad på skalan jämfört med fördelningen före behandling och vice versa. Figur 3b är exempel på en systematisk skillnad i koncentration av observationerna över en skalas fyra kategorier, betecknade a till d mellan måttillfällena. Vid andra tillfället återfinns 15 av de 20 observationerna i kategori b.

Möjliga värden på RP och RC är från -1 till 1. Om både RP och RC är noll finns det inga tecken på någon förändring, som kan tillskrivas hela gruppen. En systematisk förändring mot lägre kategorier efter behandling innebär negativt RP och en förändring i koncentration av data mot centrala kategorier efter behandling ger positivt RC (se Figur 3).

Utseendet på Q-Q-kurvan i Figur 2b och illustrationen i Figur 3 är tecken på en systematisk förändring mot lägre besvärnivåer efter behandling och det negativa RP-värdet, -0,52, bekräftar detta. Eftersom RP-värdet skattar skillnaden mellan

sannolikheten till ökad besvärnivå efter behandling och sannolikheten till minskad besvärnivå efter behandling jämfört med tillståndet före behandling, så kan detta resultat tolkas som att det är 52 procent större chans till minskade än till ökade besvär efter behandlingen. Osäkerheten i denna slutsats anges av det 95-procentiga konfidensintervallet för RP från -0,73 till -0,30, som ger starka indikationer för en statistiskt säkerställd, även kallad signifikant, minskning i besvärnivå för patientgruppen eftersom intervallet inte täcker noll. Motsvarande statistiska test skulle förkasta nollhypotesen om oförändrad besvärnivå på en signifikansnivå som är minst 5 procent.

Förväntat förändringsmönster på gruppnivå

Olika marginalfördelningar är tecken på en systematisk förändring för gruppen. Hur skulle fördelningen av dataparen i kontingenstabellen se ut om man enbart utgick från marginalfördelningarnas information och förutsatte att patienternas inbördes rangordning förblev oförändrad vid de två tillfällena, såsom illustreras av figur 3a? Detta marginalbestämda så kallade *rangtransformerbara mönster* konstrueras genom att bilda par av kategorier från de två marginalfördelningarna, det vill säga fördelningen av data före behandlingen (B,B,C,C,C,C,C,C,D,D,D,E,E,E,F,F,F,F,F) skall paras med motsvarande efter behandlingen (A,A,A,A,B,B,B,B,C,C,C,C,C,C,D,D,D,E,F), jämför med linjerna i Figur 3. Det lägst rankade värdet före behandling paras sålunda med det lägst rankade värdet efter behandling; därmed bibehålls ordningen. De fem paren med lägsta kategorierna blir då (B,A), (B,A), (C,A), (C,A) och (C,B), se Figur 4.

Detta rangtransformerbara mönster illustrerar den förväntade gruppförändringen om gruppen svarat homogent på behandlingen, vilket även kallas för en *populationsbaserad förändring eller behandlingseffekt*.

Marginalfördelningarna ger alltså enbart information om en systematisk förändring som kan tillskrivas hela gruppen, och om man endast nöjer sig med de statistiska måtten, som baseras på marginalerna, såsom RP och RC, så förväntas förändringsmönstret se ut som det i figur 3a och 4, som innebär att behandlingen förväntas sänka gruppens besvärnivå med högst två kategorier.

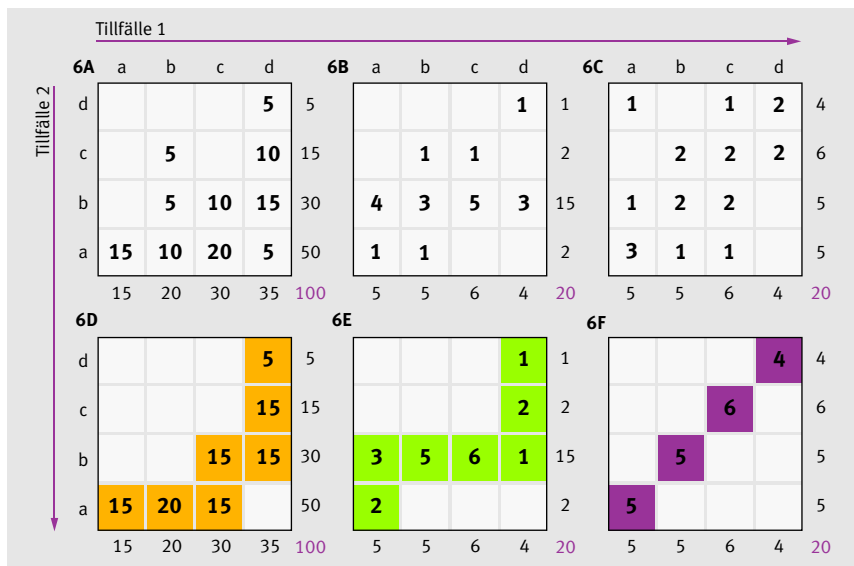
Det observerade förändringsmönstret, Figur 1, avviker från det rangtransformerbara, vilket betyder att behandlingseffekten inte är homogen för individerna. Detta faktum illustreras tydligt av figurerna 3a och 5, där Figur 3a visar dataparen i det rangtransformerbara mönstret och Figur 5a visar dataparen från Figur 1. De fyra patienterna med »påtagliga besvär« (D) före behandlingen har förändrats både till det bättre och till det sämre, vilket framgår av pilarna som spretar åt olika håll i figuren.

De individer som bryter mot det förväntade mönstret för gruppen, såsom det rangtransformerbara mönstret och Figur 3a visar, representeras av pilar som korsar varandra. Den streckade pilen i Figur 5a, som representerar den patient som bedömde sig ha »extrema besvär« (F) före behandling och »inga besvär« (A) efter behandling, korsar 12 andra pilar, vilket betyder att patienten har bytt rangordning med 12 andra patienter.

Individernas avvikelse från det rangtransformerbara mönstret, det vill säga från gruppens förändringsmönster, kan vara försumbar eller betydelsefull. Det statistiska måttet för denna heterogenitet i förändringsmönster beräknas utifrån dataparens fördelning i kontingenstabellen och baseras på att paren av bedömningar transformeras till ranger på ett speciellt sätt.

Rangmetoder

Många icke-parametriska statistiska metoder baseras på att mätvärdena transformeras till ranger. I den klassiska rangningen tilldelas varje enskild frekvensfördelning, det vill säga marginalfördelning, ranger oberoende av om fördelningarna tillhör



Figur 6. Tre olika typer av förändringsmönster av bedömningar på en skala med fyra kategorier a<b<c<d och respektive rangtransformerbara mönster.

TABELL I. De statistiska måtten med tillhörande 95-procentiga konfidensintervall (KI) för grupp- och individförändringarna för de tre exemplen i Figur 6.

	Exempel A (N=100)	Exempel B (N=20)	Exempel C (N=20)
<i>Systematisk förändring</i>			
i läge RP	-0,54	-0,19	0
(95% KI)	(-0,64;-0,44)	(-0,50;0,12)	(-0,24;0,24)
i koncentration RC	0,14	0,44	0
(95% KI)	(-0,08;0,35)	(0,09;0,74)	(-0,27;0,27)
<i>Individvariation RV</i>			
	0,12	0,083	0,17
(95% KI)	(0,05;0,21)	(0;0,29)	(0,01;0,51)
<i>Homogenitet r_a</i>			
	0,88	0,91	0,83

olika grupper, som till exempel i Wilcoxon–Mann–Whitneys rangsummetest, eller av två mätningar inom samma grupp, som vid beräkning av Spearmans rangkorrelation [3, 4]. Rangerna är alltså bundna (»tied ranks«) till varje enskild fördelning oavsett om datamaterialet kommer från oberoende grupper eller från parade data.

I min statistiska ansats för analys av parade ordinaldata är rangerna bundna till par av kategorier, så kallade *parbundna ranger*, för att ta hänsyn till att varje individ representeras av datapar, det vill säga av beroende data. Till exempel, de fyra individerna med »påtagliga, D« besvär före behandling (Figur 1) som bedömde sig ha olika grad av besvär efter behandling, representeras av olika datapar, (D,B), (D,C), (D,D), och (D,E), och tilldelas därmed olika par av ranger, (9,7), (10,13), (11,16) och (12,19), se Figur 5b. Tack vare denna rangteknik kan gruppens systematiska förändringsmönster bestämmas separat från den individuella variabiliteten.

Beräkning av individuella förändringar

Ett statistiskt mått för individuella avvikelser från det rangtransformerbara mönstret är den *relativa rangvariansen*, *RV*, som definieras utifrån variansen av skillnaden mellan de parbundna rangerna. Alternativt kan man beräkna korrelations-

koefficienten, r_a , av de två uppsättningarna parbundna ranger, som är approximativt lika med 1-RV och som är ett homogenitetsmått [12]. Om gruppens förändringsmönster sammanfaller med det rangtransformerbara mönstret är $RV=0$ och $r_a=1$, och likformigt utspridda datapar över hela kontingenstabellen ger maximalt $RV (<1)$. I vårt fall är $RV=0,33$ (95 procenta konfidensintervall 0,07 till 0,74), vilket visar att individvariationen inte är försumbar, eftersom det 95-procentiga konfidensintervallet inte täcker noll, men materialet är litet som beslutsunderlag och intervallet är mycket brett.

För ett fem gånger så stort material, $N=100$, med samma fördelning i cellerna är det 95-procentiga konfidensintervallet till $RV=0,33$ mycket smalare, 0,20 till 0,49, vilket ger en mycket starkare statistisk indikation på individuella förändringar, populärt uttryckt som statistiskt signifikant individvariation.

Tolkning av andra förändringsmönster

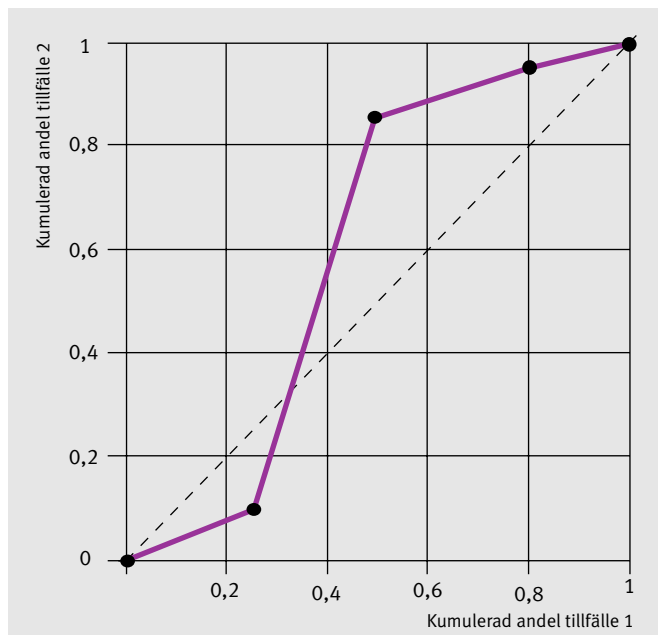
Figur 6 visar tre andra exempel på förändringsmönster när en skala med fyra ordnade kategorier, $a<b<c<d$, används. De olika marginalfördelningarna i förändringsmönster 6A är ett tecken på en systematisk förändring för gruppen mot lägre kategorier, vilket bekräftas av det rangtransformerbara mönstret, Figur 6D, och RP-värdet, som är negativt och signifikant skilt från 0, se Tabell I. De båda marginalfördelningarnas värden är något olika koncentrerade, men konfidensintervallet för RC är brett och täcker noll, så behandlingseffekten tillskrivs en gruppförändring i läge på skalan. Gruppen är inte fullständigt homogen utan det finns individvariationer som inte förklaras av gruppförändringen, vilket $RV=0,12$ visar. De fem paren i cell (b,c) ger störst bidrag till RV. Slutsatsen är att den statistiskt säkerställda gruppförändringen mot lägre kategorier är den starkaste förklaringen med en svag, men inte försumbar, individvariation i förändringsmönstret.

Förändringsmönster 6B är ett typiskt exempel på en systematisk förändring i hur bedömningarna koncentreras på skalan vid de två tillfällena; se Q–Q-kurvan i Figur 7, dataparen i Figur 3b, det rangtransformerbara mönstret i Figur 6E och det statistiskt säkerställda RC-värdet. Vid andra tillfället koncentreras bedömningarna till kategori b. Gruppen är homogen i förändring, RV är försumbart och homogenitetskoefficienten r_a är 0,91. Denna typ av förändringsmönster kan betyda att skalan inte är tillräckligt diskriminerande vid utvärdering av förändring, eller så ändras individernas bedömningar systematiskt till kategori b oberoende av nivån vid första tillfället.

Förändringsmönstret 6C förklaras helt av individuella förändringar (RV). Gruppen är alltså heterogen i förändring enligt bedömningarna på den aktuella skalan, eftersom marginalfördelningarna är lika ($RP=RC=0$). Möjliga förklaringar kan vara att skalan inte är relevant för utvärderingen eller för gruppen, egenskapen som skall mätas är olämpligt eller felaktigt operationaliserad till skalkategorierna eller till individerna som skall använda skalan. Man skulle också kunna tänka sig att det inte har skett någon förändring mellan mättillfällena, men det är föga troligt eftersom bara nio individer använder oförändrad kategori vid båda tillfällena.

Diskussion

Användningen av olika typer av skalor är mycket utbredd inom många tillämpningsområden, men medvetenheten om att da-



Figur 7. Q-Q-kurva över gruppförändringen i exempel 6B, E.

datamaterial från bedömningar på skalor saknar de matematiska egenskaper som vi är vana att hantera verkar inte vara lika utbredd. Användning av olika typer av skalor och andra bedömningsinstrument ger upphov till datamaterial som enbart har en ordningsstruktur men saknar standardiserade avstånd mellan kategorierna eller standardiserade kategoristorlekar, oavsett om kategorierna utgörs av verbala beskrivningar, siffror, bilder eller positioner på en linje, och oavsett om den bakomliggande teoretiska variabeln är kontinuerlig eller diskret, enkel eller komplex. De siffror eller andra symboler som används för att beteckna kategorierna skall inte förväxlas med matematiska tal utan är just symboler för ordningsstrukturen. Uppsättningen symboler skall kunna bytas ut utan att resultatet av den statistiska bearbetningen påverkas! Mättnivån för ordnade kategoridata brukar kallas för ordinal, men det matematiska begreppet är att ordinaldata har rang-invarianta egenskaper, vilket betyder att resultaten inte skall påverkas av hur kategorierna etiketteras. De matematiskt högre nivåerna jämfört med ordinaldata kallas ibland för intervall- och kvotdata [2, 3]. Båda dessa mättnivåer förutsätter kvantitativa egenskaper med ordning, väl definierad storlek samt lika avstånd mellan närliggande värden. Sådana egenskaper kan aldrig tillskrivas datamaterial som kommer från bedömningar på olika typer av skalor och index.

Avsaknad av matematiska egenskaper hos data från skalbedömningar blir mycket påtaglig vid analys av förändring, eftersom differens eller procentuell förändring saknar innebörd för data utan avstånd och storlek. Vid analys av förändring är det dessutom viktigt att ta hänsyn till att datamaterialet består av parade data, det vill säga två värden per individ. Dataparens fördelning i kontingenstabell alternativt spridningsdiagram ger en utmärkt beskrivning av förändringsmönstret och marginalfördelningarna ger kompletterande information om frekvensfördelningen av bedömningarna vid respektive tillfälle. De enklaste statistiska testen av förändring är teckentestet och McNemars test med komplettering av beräkning av det 95-procentiga konfidensintervallet för skillnad mellan parade proportioner [4, 5, 13]. Dessa test baseras på andelen individer med högre respektive lägre kategorinivåer vid uppföljningen jämfört med bedömningarna vid första tillfället, vilket betyder att

mycket information om det faktiska förändringsmönstret går förlorat.

I denna artikel har jag presenterat en statistisk metod, som utvecklades för att analysera parade ordinaldata. Metoden förutsätter endast ordningsstruktur hos datamaterialet och passar för alla typer av data som besitter denna enda egenskap. Det innebär att metoden även passar kvantitativa intervall och kvotdata, även om det för dessa mättnivåer finns ett stort antal alternativa statistiska metoder.

Tack vare den speciella rangmetoden med parbundna ranger är det möjligt att statistiskt mäta den grupprelaterade (systematiska) förändringen separat från de individuella avvikelserna från gruppens förändringsmönster. Gruppförändringen mäts med två mått, dels grad av förändring i läge på skalor (RP), dels i vilken utsträckning frekvensfördelningarna har olika koncentration vid de två tillfällena (RC).

Klassiska marginalmodeller tar enbart hänsyn till lägesförändringar medan det är mycket ovanligt att mäta koncentrationsförändring eftersom det oftast är matematiskt svårare. Att behandlingseffekt inte enbart är en lägesförskjutning i tillstånd är ganska uppenbart. Patienter som har en dålig kondition och sämre status vid utskrivning än andra patienter kan ha en sämre prognos eller längre tid till förbättring än de patienter som har bra utskrivningsstatus.

Att välja rätt metod är en moralisk, etisk fråga. Är det forskningsetiskt att tilldela ett datamaterial egenskaper som det inte har och utifrån dessa antaganden välja statistiska metoder för analys? Är det medicinskt etiskt att tilldela patienten symtom som han eller hon inte har och behandla därefter? Vad är då värdet av evidensbaserade beslut?

Ett beräkningsprogram finns tillgängligt på författarens webbplats: <http://www.oru.se/esi/svensson>

■ *Potentiella bindningar eller jävsförhållanden: Inga uppgivna.*

REFERENSER

- Pocock SJ. Clinical trials. A practical approach. Chichester: Wiley; 2000.
- Svensson E. Val och konsekvens, datamaterialets egenskaper bestämmer den statistiska verktygs-lådan. Läkartidningen. 2005; 102: 1331-7.
- Siegel S, Castellan NJ. Nonparametric statistics for the behavioral sciences. 2nd ed. New York: McGraw-Hill; 1988.
- Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
- Svensson E. Guidelines to statistical evaluation of data from ratings scales and questionnaires. Journal of Rehabilitation Medicine. 2001; 33:47-8.
- Dunn G, Everitt B. Clinical biostatistics. An introduction to evidence-based medicine. London: Edward Arnold; 1995.
- Svensson E. Vad är behandlingseffekt om patienten blev bättre, men ingen vet hur mycket bättre? Analys av förändring när datamaterialet består av ordnade kategorier. Läkartidningen. 2005;102:3138-45.
- Sonn U, Svensson E. Measures of individual and group changes in ordered categorical data: application to the ADL Staircase. Scand J Rehab Med. 1997;29:233-42.
- Svensson E. Ordinal invariant measures for individual and group changes in ordered categorical data. Statistics in Medicine. 1998; 17:2923-36.
- Svensson E, Starmark JE. Evaluation of individual and group changes in social outcome after aneurysmal subarachnoid haemorrhage: a long-term follow-up study. J Rehab Med. 2002;34:251-9.
- Lund I, Lundeberg T, Lönnberg L, Svensson E. Decrease of pregnant women's pelvic pain after acupuncture: a randomized controlled single-blind study. Acta Obstet Gynecol Scand. 2006;85:12-9.
- Svensson E. A coefficient of agreement adjusted for bias in paired ordered categorical data. Biometrical Journal. 1997;39:643-57.
- Swinscow TDV, Campbell M. Statistics at square one. 10th ed. London: BMJ books; 2003.