

# **Vidareutveckling av RAND-36 hälsoenkät: summaindex och kortversion**

**Charles Taft**

**Registercentrum VGR**

**Göteborg**

## Innehållsförteckning:

---

Bakgrund .....	3
Metod .....	4
Del 1: RAND-36 summaindex .....	4
Kort historik om RAND/SF-36.....	4
RAND/SF-36 begrepps- och mätmodell .....	5
RAND/SF-36 begreppsmodell .....	5
RAND/SF-36 mätmodell.....	5
RAND/SF-36 domänberäkning .....	6
RAND-HSI .....	6
Rand-36.....	6
SF-36 .....	6
Kommentarer till RAND-HSI, SF-36 och RAND-36 domänberäkningarna.....	7
RAND-36 och SF-36 summaindex beräkning – skilda vägar.....	7
SF-36 summaindex.....	8
RAND-HSI summaindex.....	10
Jämförelser mellan SF-36 och RAND-HSI summaindex .....	10
RAND/ SF-36 begrepps- och mätmodellen. En universell modell?.....	12
Skillnader mellan olika allmänna populationer? .....	13
Skillnader mellan olika patientgrupper?.....	13
Två eller tre summaindex? .....	13
Alternativa beräkningsprocedurer för summaindex.....	15
SEM-deriverade item vikter .....	15
Enkelsummering, utan vikter.....	16
Generisk patientmodell .....	17
Sammanfattning och rekommendationer .....	18
Begränsningar .....	20
Del 2: RAND-12.....	21
Inledning.....	21
Användningsområde .....	21
Tillgänglighet .....	21
Psykometrisk prövning.....	21
Vetenskaplig användning .....	21
Sammanfattning och rekommendationer .....	21
Referenser .....	22

## Bakgrund

Mätningar med patientrapporterade utfallsmått, PROM, är nödvändiga för att erhålla standardiserad och systematisk kunskap om patientens upplevelse av vårdens resultat, som grund för ständigt lärande och förbättringsarbete. Inom de nationella kvalitetsregistren har användningen av PROM ökat kraftigt på senare år, delvis på grund av en växande konsensus att PROM utgör ett viktigt hjälpmedel för att värdera vårdresultat, delvis på grund av kravet att kvalitetsregistren genomför PROM mätning för att uppnå certifieringsnivå 2 (Nilsson *et al.* 2015). Idag används flera olika generella PROM inom de nationella kvalitetsregistren, varav ett är enkäten SF-36 v1.0.

SF-36 är det internationellt mest använda generella PROM avseende hälsorelaterad livskvalitet (Garratt *et al.*, 2002) och är dessutom det mest använda PROM i randomiserad kontrollerade studier (Contopoulos *et al.*, 2009; Garratt, 2009) och kliniska prövningar (Scoggins *et al.* 2009). Enkäten togs fram av non-profit organisationen RAND Corporation i USA men idag finns det två olika upphovsrättshållare: RAND Corp, som distribuerar enkäten under namnet RAND-36 med kostnadsfri användning; och QualityMetric Inc, som distribuerar enkäten under namnet SF-36 och mot användaravgift. Enkäten är mest känd internationellt som SF-36, delvis på grund av ett tidigt internationaliseringsprojekt, International Quality of Life Assessment project (IQOLA), där SF-36 översatts och validerades samtidigt i 14 olika länder, bl. a. i Sverige (Ware & Gandek 1998). Den idag använda svenska versionen är också SF-36 (Sullivan *et al.* 2002) och är därmed avgiftsbelagd<sup>1</sup>. Denna kostnad är prohibitivt stor för ett nationellt kvalitetsregister och därför genomfördes på uppdrag av kvalitetsregisterorganisationen en svensk översättning och validering av RAND-36 i ett samarbete mellan tre registercentrum: RCSO, RC Syd och RC VGR.

Den svenska RAND-36 enkäten och tillhörande poängberäkningsalgoritm tillgängliggörs idag av RCSO/PROMcenter via webbplatsen rcs.se. Algoritmen, som är framtagen av RAND, genererar poäng på RAND-36s åtta hälsodomäner dock inte på dess två överordnade summaindex, fysisk och psykisk hälsa. Summaindexberäkning finns dock i SF-36 algoritmen samt i en tidigare RAND algoritm (Hays *et al.* 1998). Summaindexen efterfrågas alltmer eftersom de bland annat kan underlätta och förenkla resultatredovisning och statistiskanalys genom att antalet resultatvariabler minskas från åtta till två. Flera olika aggregeringsmetoder för summaindexen redovisas i den vetenskapliga litteraturen men RAND avstår dock att ta ställning till vilken av dessa är att föredra och i stället överlämnar beslutet till varje enskild forskare. Standardisering av summaindexberäkning är däremot en förutsättning för att kunna jämföra resultat mellan olika studier. Huvudsyftet med denna utredning är att kartlägga och jämföra olika metoder för summaindexberäkning som finns redovisat i litteraturen och att rekommendera ett lämpligt tillvägagångssätt för hantering av summaindexen i den svenska RAND-36 (del 1).

Ett annat syfte med utredningen är att undersöka möjligheten att kunna erbjuda en kortare version av RAND-36 (del 2). Informationsförlust och mätprecision som en kortare version kan innebära kan ibland vägas upp av minskad administrativ- och patientbörda. Inte minst i kvalitetsregister-sammanhang kan enkätens omfattning vara väsentlig vid val av PROM och i praktiskt genomförande av PROM mätning. En kort version, RAND-12, bestående av 12 frågor från RAND-36, finns redovisad i

---

<sup>1</sup> Den svenska SF-36 distribuerades kostnadsfritt och underhölls av Sektionen för vårdforskning, Sahlgrenska Akademi, Göteborgs Universitet och HRQLgruppen fram till 2008.

flera vetenskapliga artiklar och i RAND-36 HSI manualen (Hays *et al.* 1998). Den är dock inte tillgänglig via RANDs hemsida. Utredningen avser att undersöka möjligheten att tillgängliggöra en svensk version av RAND-12 för fritt användning. Målsättningen är att kunna erbjuda kvalitetsregistren lämpliga valmöjligheter i enlighet med deras speciella behov och förutsättningar.

## Metod

Kontakten togs initialt med Professor Ron Hays, UCLA, författaren till RAND-36 HSI manualen och PI för RAND-36s utvecklingsarbete, samt RAND Corporation (Ingrid Maples) för att ta reda på deras inställning till och tankar kring: 1) hantering och beräkning av RAND-36 summaindex samt 2) möjlighet att erbjuda RAND-12. Det bekräftades dels att det inte finns någon standardberäkningsprocedur för RAND-36 summaindexen och att forskare får avgöra själva och tillämpa den procedur de själva anser är mest lämplig och dels att copyright till den 12-frågorversionen tillhör QualityMetric. Därefter genomfördes litteratursökningar i tre databaser (Scopus, CINAHL och Cochrane) för att identifiera publikationer relevanta för utredningens syfte. Artikelabstrakten granskades och relevanta studier lästes i sin helhet. Sökningarna kompletterades med artiklar citerade i referenslistorna.

## Del 1: RAND-36 summaindex

---

### Kort historik om RAND/SF-36

Grunderna till RAND/SF-36 lades redan under 1970-talet inom RAND Corporations Health Insurance Experiment (HIE), en omfattande, rikstäckande randomiserad studie avsedd att jämföra olika sjukförsäkringssystem i USA med avseende på hälsostatus, sjukvårdskostnader och sjukvårdskonsumtion (Brook *et al.* 1984). Inom HIE utarbetades grundläggande metoder för utveckling och psykometrisk prövning av patientrapporterade mått och ett brett batteri av enkäter konstruerades för att mäta olika aspekter av funktion och välbefinnande i den allmänna populationen. I en uppföljande studie, RAND Medical Outcomes Study (MOS), prövades tillämpbarheten av metoderna i kliniska populationer (Tarlov *et al.* 1989). För denna studie utvecklades den 149-frågor MOS Functioning and Well Being Profile (Stewart *et al.* 1992, Hays *et al.* 1993), vilket avsåg att mäta 40 olika aspekter av fysisk och psykisk hälsa. RAND/SF-36 utvecklades senare inom RAND som en 36 frågor kortversion av MOS enkäten för att mäta åtta av de viktigaste hälsodomänerna (Hays *et al.* 1992, Ware *et al.* 1993). Enkäten publicerades 1992 under namnet SF-36 (Ware *et al.* 1992) och året därpå även under namnet RAND-36 (Hays *et al.* 1993).

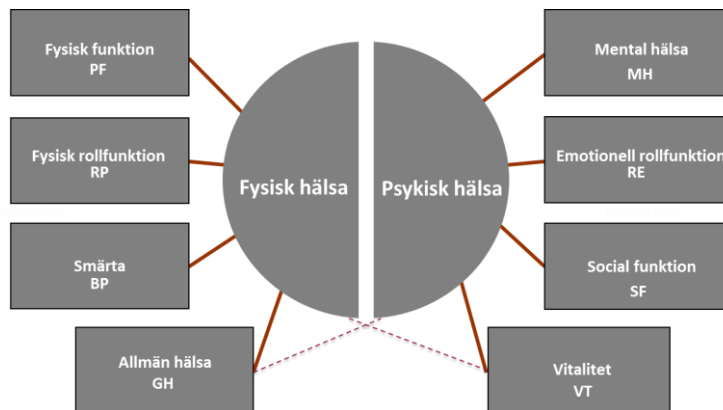
SF-36 tillsammans med poängberäkningsinstruktioner distribuerades ursprungligen av non-profit Medical Outcomes Trust men distribueras numera av QualityMetric, Inc. (idag ägs av Optum, Inc.) mot licens och användaravgift. QualityMetric har vidareutvecklat flera versioner av SF-36 under årens lopp, bl a SF-36 v2.0, SF-12 v1.0/ 2.0 och SF-8. QualityMetric hävdar copyright på alla dessa versioner och alla är avgiftsbelagda.

RAND-36 med domänberäkningsinstruktioner är däremot i public domain och, som nämnts tidigare, distribueras de med kostnadsfri användning via RANDs webbplats. RAND har inte utvecklat andra versioner av RAND-36. RAND-36 har tidigare kallats även RAND 36-Item Health Survey Inventory (HSI) och en manual med samma namn publicerades 1998 (Hays *et al.* 1998). I manualen beskrivs ett annat

domänberäkningssystem än det som är aktuell idag samt även en procedur för summaindexberäkning, vilken nedan betecknas RAND-HSI metoden.

## RAND/SF-36 begrepps- och mätmodell

**RAND/SF-36 begreppsmodell.** Enligt begreppsmodellen för RAND/SF-36 uppdelas hälsa i två huvuddimensioner, fysisk och psykisk hälsa (Figur 1) (Ware *et al.* 1993, Hays 1998). Dessa två dimensioner representeras av åtta olika basala hälsobegrepp, selekterade ur MOS studien: fysisk funktion (PF<sup>2</sup>), fysisk rollfunktion (RP), smärta (BP), allmän hälsa (GH), vitalitet (VT), social funktion (SF), emotionell rollfunktion (RE) och mental hälsa (MH). Fysisk hälsa representeras huvudsakligen och i fallande ordning av de fyra första begreppen, såsom indikeras i bilden. Psykisk hälsa består på liknande sätt främst av de fyra sistnämnda begreppen. Noterbart är att modellen har presenterats på lite olika sätt avseende domänerna GH och VT (och även SF), där de ibland representerar enbart en dimension och ibland framställs dem vara svagt relaterade till sin sekundära dimension (streckade linjer).



Figur 1. RAND/ SF-36 begreppsmodell.

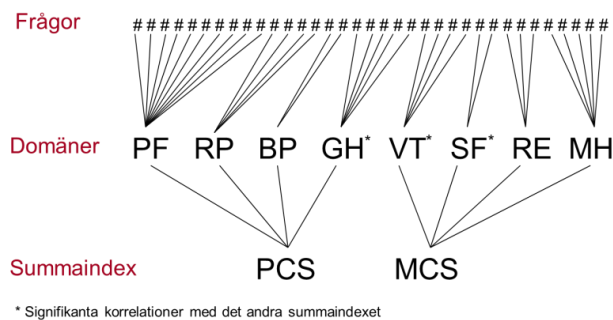
RAND-36 och SF-36 bygger till synes på samma grundläggande begreppsmodell. Det finns dock en viktig skillnad mellan dem, nämligen hur de betraktar relationen mellan fysisk och psykisk hälsa. I RAND modellen antas fysisk och psykisk hälsa vara relaterade till varandra, medan i den SF-36 modellen antas de vara distinkta och oberoende av varandra. Denna skillnad är av avgörande betydelse för hur RAND-36 och SF-36 beräknar poäng för de två summaindexen, vilket vi återkommer till senare.

**RAND/SF-36 mätmodell.** Figur 2 illustrerar taxonomin av frågor och begrepp som underbygger konstruktionen av RAND/SF-36 domäner och summakomponenter. Taxonomin har tre olika nivåer: 36 frågor; åtta domäner som aggregerar 2-10 frågor var; och två summaindex (PCS, Physical Component Summary och MCS, Mental Component Summary<sup>3</sup>) som aggregerar domänerna (Ware *et al.* 1994; Hays *et al.* 1998). Alla utom en av de 36 frågorna ingår i beräkning av domänpoäng och enligt standardiserad beräkningsmall. Mätmodellen överensstämmer väl med begreppsmodellen.

<sup>2</sup> Förkortningar motsvarar de engelska begrepp. Benämningarna för dessa begrepp skiljer sig i vissa fall mellan SF-36 och RAND-36, där t ex Mental Health (MH) i SF-36 kallas för Emotional Well Being (EWB) i RAND-36. De ingående frågorna i varje dimension är dock identiska.

<sup>3</sup> PCS och MCS är SF-36 beteckningar, motsvarigheter i RAND-HSI är Physical Health Composite (PHC) och Mental Health Composite (MHC).

Beräkningsmallen för domänaggregering skiljer sig dock mellan RAND-36, RAND-HSI och SF-36, och summaindexaggregering är också annorlunda i RAND-HSI jämfört med SF-36 (RAND-36 beräknar inte summaindex).



Figur 2. RAND/ SF-36 mätmodell.

## RAND/SF-36 domänberäkning

Eftersom summaindexen bygger direkt på och är avsedda att sammanfatta domänerna är det väsentligt att ha en förståelse för hur frågorna sammanställs till domäner. Det finns idag tre huvudsakliga metoder för aggregering av frågorna till de åtta domänerna: RAND HSI, SF-36, och RAND-36. Det är viktigt att påpeka att dessa skiljer sig i själva aggregeringsmetoden, inte i vilka frågor som bildar varje domän.

**RAND-HSI.** RAND-HSI är en tidigare metod för beräkning av RAND-36 poäng. Den togs fram av Hays på 90-talet och finns redovisat i Rand-36 HSI Manual från 1998 (Hays *et al.* 1998). Proceduren bygger på ett viktsystem som deriverades genom Raschanalys, där varje svarsalternativ för varje fråga inom samma domän tilldelas en vikt mellan 0 och 100. Vikterna representerar ungefär en kalibrering av den relativa graden/nivån av den latent egenskapen som estimeras. Vikterna för valda svarsalternativ summeras till en poäng, vilken sedan ålders/köns standardiseras med hjälp av olika normtabeller i manualen. Poängskalan varierar mellan domänerna (t ex PF skalan är 1-63 och RE är 17-55). RAND distribuerar inte poängberäkningsinstruktioner enligt denna metod, men mycket av tidigare forskningsresultat med RAND-36 enkäten baseras på metoden.

**RAND-36.** Den aktuella beräkningsmetoden för RAND-36 är egentligen en återgång till den ursprungliga metoden för domänberäkning som beskrivs i Hays, 1993. Beräknings sättet är i stort sätt detsamma som används i SF-36, dvs domänpoäng representerar medelvärdet av enkelt omkodade svar (dock ej omkalibrerade BP och GH enligt SF-36 nedan). I RAND-36 är de olika svaralternativen omkodade till värden från 0 till 100 där värden motsvarar lika intervaller på skalan (t ex svarsalternativ 1, 2, 3, 4, 5 = 0, 25, 50, 75, 100) och domänpoäng representerar medelvärdet av valda svar (0-100 poängskala).

**SF-36.** SF-36 domäner poängsätts enligt Likert metoden som innebär summering av frågor med lika vikt. Vissa antaganden måste uppfyllas för att metoden ska kunna användas och inledande analyser visade att Likert metoden kunde tillämpas för alla åtta domäner (Ware *et al.* 1993). Senare studier visade dock att frågorna i domänerna Smärta (BP) och Allmän hälsa (GH) behövde kalibreras om för att uppfylla antagandet om lika intervaller mellan svarssteg (Ware *et al.* 1993). Efter omkalibrering

ansågs domänerna kunna aggregeras som medelvärdet av valda svaren och en transformering av medelpoäng till en 0 (sämst) till 100 (bäst) poängskala användes för att underlätta tolkning.

**Kommentarer till RAND-HSI, SF-36 och RAND-36 domänberäkningarna.** Det finns uppenbara svårigheter med att använda RAND-HSI metoden, vilket delvis kan förklara varför RAND har nu övergått till en annan metod. För det första, metoden är relativt komplicerat och kräver tillgång till vikt- och normtabeller. Dessa tabeller finns publicerade, men enbart i RAND-36 HSI manualen (Hays 1993), vilket finns tillgänglig endast via författarens egen hemsida (<http://gim.med.ucla.edu/FacultyPages/Hays/surveys/R-36%20HSI%20Hays/>). Vidare, både Rasch-vikterna och normtabellerna för standardiseringen av poäng baseras på amerikanska populationsdata och det är inte säkert att de är överförbara till svenska förhållanden. Trots svårigheter att applicera själva RAND-HSI metoden har en del forskning visat att tillämpning av en Rasch-approach i sig kan producera mer valida RAND/SF-36 resultat än vad traditionell Likert metoden gör. Till exempel har flera studier visat att Fysisk Funktion (PF) domänen, beräknat enligt Rasch-metoder, kan diskriminera bättre mellan bl a ålders- och patientgrupper än Likert-baserad PF (Raczek *et al.* 1998, McHorney *et al.* 1997). Rasch-baserad poängberäkning avser dessutom att producera data på intervallnivå, vilket gör dem mer förenligt med principer för matematisk beräkning och parametrisk statistik (Hays *et al.* 2000). Det kan tilläggas dock att Rasch och Likert RAND/SF-36 poäng är högt korrelerade med varandra (Hays *et al.* 2000).

RAND-36 och SF-36 algoritmer ger exakt samma poäng på 6 av de 8 delskalor. Det finns dock en skillnad gällande Smärta (BP) och Allmän hälsa (GH), där SF-36 tillämpar ett mer komplicerat beräkningssystem. Denna skillnad är dock försumbar på gruppnivå då korrelation mellan RAND-36 och SF-36 på dessa domäner är mycket hög ( $r=0,99$ ) (Hays *et al.* 1993). Denna överensstämmelse gör att resultat från RAND-36 kan jämföras direkt med SF-36 med god säkerhet, vilket väsentligt ökar jämförelsebasen för RAND-36 resultat.

## **RAND-36 och SF-36 summaindexberäkning – skilda vägar**

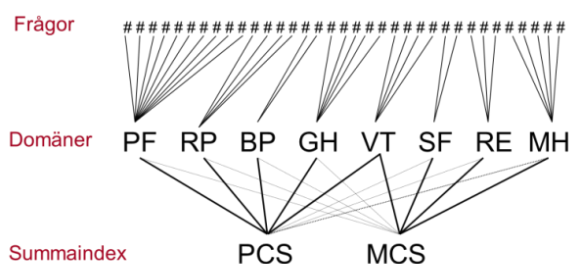
Enligt RAND/SF-36 begreppsmodellen (Fig 1) antas hälsa bestå av två huvudkomponenter, fysisk och psykisk hälsa. Indelningen har bekräftats i flera studier som konsekvent har kunnat differentiera dessa faktorer i den allmänna amerikanska befolkningen (Ware *et al.* 1994) och i andra länder (Ware, Kosinski *et al.* 1998), i olika demografiska grupper (Ware *et al.* 1995) och i olika patientgrupper (McHorney *et al.* 1993 o 1994 Osterhaus *et al.* 1994). Faktoranalytiska studier har gett stöd för att de åtta underliggande RAND/SF-36 domänerna formar två överordnade faktorer, fysisk och psykisk hälsa (McHorney *et al.* 1993, Ware *et al.* 1994, Ware *et al.* 1995). Upptäckten att dessa två faktorer kunde förklara 80-85% av variansen i de åtta domänerna ledde till att summaindexen konstruerades (Ware *et al.* 1993). Avsikten var att kunna förenkla resultatredovisning och statistikanalys genom att sammanfatta de åtta domänerna på ett övergripande sätt utan nämnvärd förlust av information eller mätprecision. Det är viktigt att understryka att RAND/SF-36 utvecklades från början för att mäta de åtta olika hälsodomänerna och att de två summaindexen representerar en senare utveckling.

Trots en likartad mätmodell skiljer sig RAND-HSI och SF-36 markant i hur summaindexen beräknas. Skillnaden grundar sig primärt i olika grundantagande avseende relationen mellan fysisk och psykisk hälsa. SF-36 antar att dessa är distinkta och oberoende av varandra medan RAND-36 utgår ifrån att

fysisk och psykisk hälsa kan differentieras men är korrelerade. SF-36 utgångspunkten har utförligt förklarats och försvarats av Ware (Ware *et al.* 1993, Ware, Kosinski *et al.* 1998, Ware *et al.* 2001,). Antagandet om att fysisk och psykisk hälsa är helt distinkta fenomen verkar dock strida mot en betydande mängd forskning som visat på ett samband mellan fysisk hälsa och psykisk ohälsa, t ex (Katon *et al.* 1991, Simon *et al.* 1991, Hays *et al.* 1994, Dirsh *et al.* 2002, Anagnostopoulos *et al.* 2009, Anagnostopoulos *et al.* 2005, Wilson *et al.* 2000, Tucker *et al.* 2013, Reed *et al.* 1998, Hann *et al.* 2008, Hays 1998, Farivar *et al.* 2007, Lacijs *et al.* 2015, Schröder *et al.* 2012, Güthlin *et al.* 2007, Hobart *et al.* 2001, Simon *et al.* 1998, Nordveldt *et al.* 2000, Schmitz *et al.* 2006, Cella *et al.* 2005, Pelle *et al.* 2013, Chang *et al.* 2007)

Skillnaden i synen på relationen fysisk-psykisk hälsa reflekteras bland annat i valet av vilken faktoranalysmetod som användes vid framtagning av RAND-HSI respektive SF-36 summaindexen. I RAND-HSI valdes en oblik rotation, vilket tillåter faktorerna att vara korrelerade medan i SF-36 användes en ortogonal rotation, vilket tvingar faktorerna att vara maximalt oberoende. Metoden för SF-36 summaindex beräkning har kritiserats i ett flertal artiklar genom åren (t ex Simon *et al.* 1998, Taft *et al.* 2001, Nordveldt *et al.* 2000, Tucker *et al.* 2013, Wilson *et al.* 2000, Farivar *et al.* 2007, Cunningham *et al.* 2003, Blanchard *et al.* 2005, Grassi *et al.* 2010). RAND-HSI metoden har däremot förskonats från direkt kritik och ibland framhävts som en alternativ beräkningsmodell för SF-36 summaindex. (Nordveldt *et al.* 2000, Farivar *et al.* 2007, Cunningham *et al.* 2003, Blanchard *et al.* 2005, Laucis *et al.* 2015)

**SF-36 summaindex.** SF-36 summaindexen togs fram genom explorativ faktoranalys (principalkomponentanalys) med en ortogonal (okorrelerad) rotation. Faktorpoäng från analysen används som vikter för varje domän vid summaindexberäkning. Metoden ger upphov till mätmodellen som illustreras i Figur 3. Figuren visar att samtliga SF-36 domäner bidrar till båda summaindex, dvs ett tydligt avsteg från RAND/SF-36 mätmodellen (Fig 1). Ett annat och betydande avsteg är att de fysiska domänerna bidrar negativt till det psykiska summaindexet (MCS) och tvärtom, de psykiska domänerna bidrar negativt till det fysiska summaindexet (PCS). I praktiken görs detta genom ett viktsystem som är avsett att hålla isär fysisk och psykisk hälsa i vilket de fysiska domänerna tilldelas positiva vikter för PCS men också negativa vikter för MCS och visa versa för de psykiska domänerna<sup>4</sup> (Simon *et al.* 1998, Taft *et al.* 2001). Notera också att domänen VT ingår i både PCS och MCS.

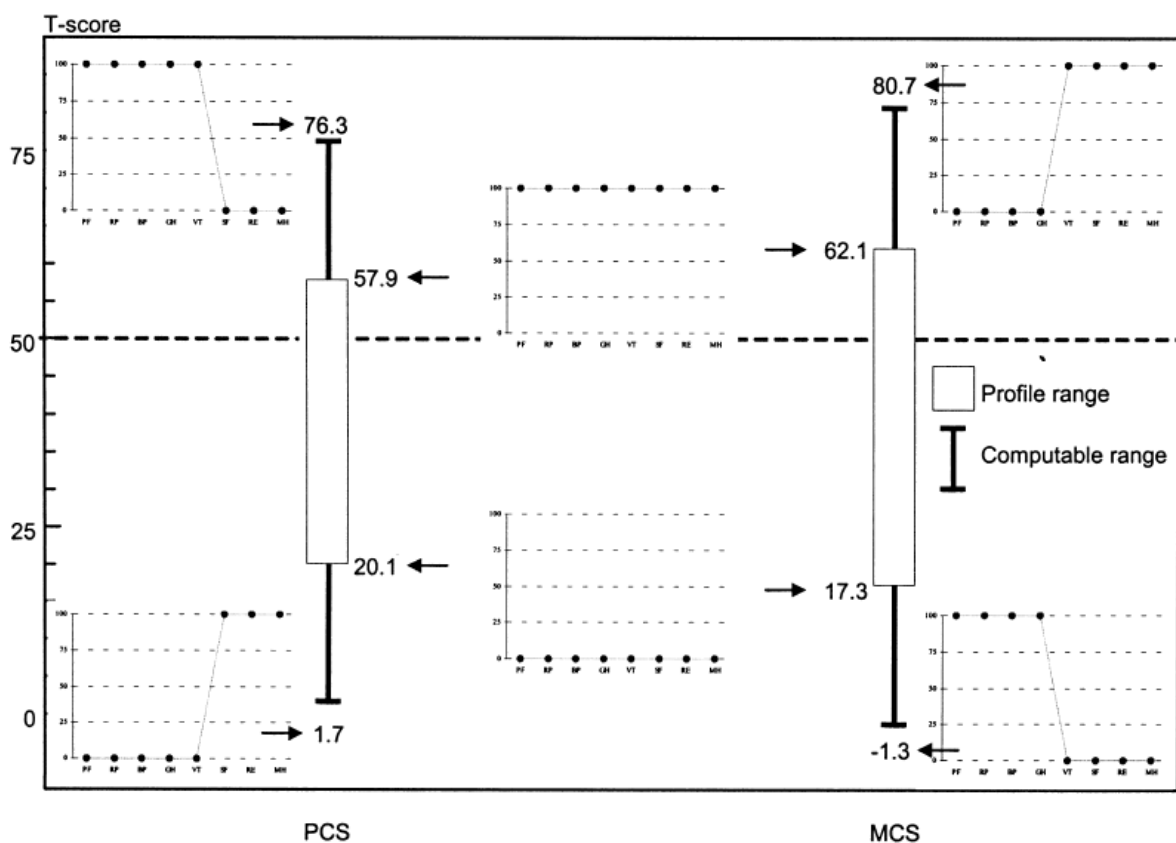


<sup>4</sup> Enligt SF-36 algoritmen först standardiseras varje domänpoäng genom en linjär z-transformering (differensen mellan populationens medelvärde och individens domänpoäng delad med populationens standardavvikelse) vilket multipliceras med vikten för domänen för varje summaindex i taget. Produkterna summeras och sedan transformeras till t-värde så att ett poäng av 50 representerar medeltalet för populationen.



Figur 3. SF-36 mätmodell i praktiken. Alla domäner ingår vid beräkning av både summaindex. Positiva vikter anges med heldragna linjer och negativa vikter med streckade linjer.

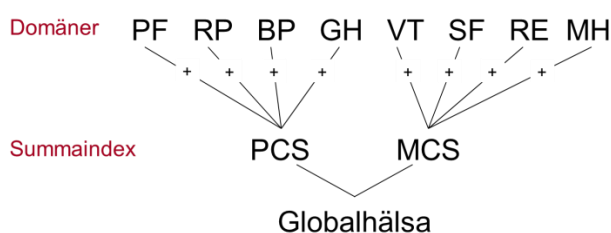
Systemet kan ge oväntade resultat beroende på blandningen av positiva och negativa vikter. Till exempel, en individ som mår dåligt fysiskt men relativt bra psykiskt (en inte ovanlig situation för kroniker) får en ännu högre MCS poäng än om hen hade i stället också mått bra fysiskt. Likaså, PCS poängen för en individ som mår bra fysiskt blir ännu högre om individen också mår dåligt psykiskt. Detta beror på att domänpoängen först normstandardiseras på så sätt att domänpoäng under populationens medelvärde blir negativa och de ovanför blir positiva; när dessa sedan multipliceras med sina vikter blir det så att negativt viktade domäner med negativa standardiserade poäng (i första exemplet de fysiska domänerna i MCS) bidrar positivt till motsatta summaindex (i exemplet MCS). Med andra ord, SF-36 beräkningssystem förutsätter inte bara att fysisk och psykisk hälsa är okorrelerade, utan också att fysisk och psykisk hälsa står i ett motsatsförhållande till varandra. Figur 5 visar effekten av de negativa vikterna på PCS och MCS poäng i relation till domänpoäng. De två graferna i mitten visar bästa möjliga domänprofil (alla domäner har maximalpoäng, dvs 100) och sämsta profilen (0 på alla domäner). Till vänster om graferna visas de genererade PCS poängerna för respektive profil (57,9 resp 20,1) och till höger de för MCS (62,1 resp 17,3). Grafen längst upp till vänster visar en profil där de fysiska domänerna är maximala medan de psykiska är i sämsta läge. Siffran till höger om grafen (76,3) är PCS poängen för denna profil. Som en direkt effekt av negativt viktade psykiska domäner är PCS poängen 18 poäng högre än PCS poängen för en optimal domänprofil, mao dåligt psykisk hälsa bidrar positivt till fysisk hälsa. Poäng i denna range är alltså en artefakt av beräkningssystemet. Effekterna finns illustrerade för extremerna för både PCS och MCS, men effekten finns på alla nivåer.



Figur 5. Illustration av effekten av SF-36 beräkningssystem på summaindex poäng. Bäst möjlig hälsa enligt domänerna visas i domänprofilen överst i mitten (100 på alla domäner), vilket motsvaras av PCS poäng på 57,9. PCS poäng blir ca 18 poäng högre genom att psykisk hälsa är i sämsta läge (profilen längst upp till vänster), dvs bästa möjlig fysisk hälsa förutsätter dålig psykisk hälsa. (figuren från Taft 2001)

Detta kontraintuitiva förhållande konstaterades empiriskt redan 1998 då patienter som behandlades framgångsrikt för depression (förbättrad MCS) visade en försämring på SF-36 PCS, trots att de fysiska domänerna inte visade någon försämring (Simon 1998). Fyndet att summaindexresultatet inte överensstämmer med domänresultat har observerats i flera studier sedan dess (Nordveldt *et al.* 2000, Schmitz *et al.* 2007, Cunningham *et al.* 2003, osv) och har gett upphov till forskning om alternativa beräkningsmetoder.

**RAND-HSI summaindex.** I likhet med SF-36 beräknas RAND-HSI summaindexen utifrån domänerna. Explorativ faktoranalys (dock principal axis) användes för att finna överordnade dimensioner. Till skillnad från SF-36 användes oblik faktorrotation i enlighet med antagandet om korrelerad fysisk och psykisk hälsa. Samma faktorer som SF-36 identifierades och domänfaktorstrukturen visades vara lika, dvs samma fyra domäner laddade högst på PCS respektive MCS. Övriga domäner laddade svagt och endast domänerna med starka positiva laddningar på fysisk respektive psykisk hälsa inkluderades i regressionsanalysen för att ta fram summaindexvikter. Därför har RAND-HSI beräkningssystemet endast positiva vikter. Poängberäkning sker på samma sätt som i SF-36, dvs varje domänpoäng standardiseras och multipliceras med sin vikt, produkterna summeras och transformeras till en t-fördelning med 50 som riktmärke för populationen. Korrelationen mellan PCS och MCS har visats vara substantiell, i enlighet med antagandet om relaterad fysisk och psykisk hälsa. (Hays manual). Till skillnad från SF-36 överensstämmer RAND-HSI mätmodellen med begreppsmodellen, dvs i modellen ingår endast de fysiska domänerna PF, RP, BP och GH i beräkningen av PCS medan enbart VT, SF, RE och MH ingår i MCS (figur 4). En annan skillnad är att i RAND-HSI modellen bildar de korrelerade PCS och MCS tillsammans en övergripande dimension, globalhälsa.



Figur 5. RAND-HSI mätmodell. Fysiska domänerna viktas positivt i PCS och psykiska domänerna viktas positivt i MCS, inga negativa vikter används.

**Jämförelser mellan SF-36 och RAND-HSI summaindex.** SF-36 summaindexen har tydliga och väl dokumenterade brister. Beräkningssystemet med en kombination av positiva och negativa domänvikter har visats producera diskrepanser mellan summaindexpoäng och de domänpoäng de avser att sammanfatta. RAND-HSI modellen utan negativa vikter bör korrigera för dessa diskrepanser. Jämförande studier mellan RAND-HSI och SF-36 tyder på att RAND-HSI fungerar bättre i detta avseende och dessutom är mer känslig för att upptäcka förändring och skillnader mellan grupper som

är kända för att ha olika hälsostatus. Till exempel, Nordvedt *et al.* (2000) fann i en studie med patienter med multipelskleros att RAND-HSI summaindex MCS bättre avspeglade resultat från de psykiska domänerna och att SF-36 MCS underskattade psykisk hälsa med mer än 7 poäng jämfört med RAND-HSI metoden. En annan studie visade att hos personer med psykiska åkommor överskattade SF-36 MCS psykisk hälsa bland personer med samtidig fysisk ohälsa (låga värden på de fysiska domänerna) och att SF-36 MCS diskriminerade sämre mellan personer med eller utan psykisk ohälsa (avspeglat i de psykiska domänerna) jämfört RAND-HSI (Schmitz 2006). Cunningham *et al.* (2003) visade att RAND-HSI summaindexen differentierade signifikant bättre än SF-36 mellan personer med och utan olika kroniska sjukdomer. Leese *et al.* (2008) fann att RAND HSI summaindexpoäng var mer reliabla än SF-36 hos patienter med schizofreni. Blanchard *et al.* (2004) fann att RAND-HSI kunde bättre urskilja och identifiera patienter med psykiska hälsoproblem i primärvården.

Det har dessutom testats om oblik rotation i sig själv kan lösa problemet med SF-36 summaindexen (Farivar *et al.* 2007). I studien replikerades metoderna som användes av Ware (1993) vid framtagning av SF-36 summaindex, dvs principalkomponentanalys av domänerna beräknade enligt SF-36 algoritmen. Endast rotationsproceduren ändrades. Analysen bekräftade en två-komponentstruktur, med en fysisk och en psykisk komponent. Som förväntat producerade rotationen en struktur där alla domäner laddade på båda komponenterna, dock var domänladdningarna på motsatt komponent mycket svaga och endast två domäner laddade negativt. Jämfört med ortogonal rotation blev antalet negativa vikter alltså betydligt färre och mindre, vilket de jämförande analyserna bekräftade ledde till bättre överensstämmelse mellan domän och summaindexresultat. Författarna konkluderade att SF-36 ortogonal-baserat resultat i fortsättningen bör presenteras tillsammans med resultat från oblik-baserade vikter. Vikterna presenteras i tabell 1.

En jämförelse av de tre viktsystemen har nyligen genomförts av Laucis *et al.* (2015). I studien har SF-36 resultat från ett stort antal ortopediska studier räknats om med RAND-HSI och SF-36 oblik-vikterna. Analysen visade att både RAND-HSI och SF-36 oblikgenerade summaindexpoäng är mer kompatibla med domänresultat än vad SF-36 ortogonal är. De redogjorde också för flera ortopediska studier där rapporterade behandlingseffekter underskattades på grund av denna SF-36 beräkningsartefakt. Intressant nog visade studien mycket snarlika resultat mellan RAND-HSI och SF-36 obliksummaindexpoäng. I en annan studie omanalyserades resultat från tre RCT-studier som undersökte effekter av antidepressiv behandling hos hjärtsjuka patienter (Pelle *et al.* 2013). I två av dessa studier visade SF-36 en förbättring i psykisk hälsa men inte en bekräftad samtidig förbättring i fysisk hälsa; däremot kunde en omanalys med oblik-baserade vikter konfirmera detta fynd. De tre modellerna har också jämförts med data från den grekiska allmänna befolkningen (Anagnostopoulos *et al.* 2009). Studien visade att en oblik modell där alla åtta domäner ingår passade något bättre än RAND-HSI modellen och att en ortogonal sådan (SF-36) visade dålig anpassning.

QualityMetric (QM) har även utvecklat ett alternativt beräkningssystem baserat på en oblik faktorlösning (Saris-Baglana *et al.* 2004). Vikterna finns angivna i Table 1 och som syns finns färre och svagare negativa domänvikter än i standard SF-36. Viktsystemet rekommenderas dock inte av företaget i nuläget och inga jämförande analyser har publicerats.

Tabell 1. PCS och MCS vikter för SF-36-ortogonal, SF-36-oblik och RAND-HSI.

	PCS				MCS			
	ortogonal	oblik			ortogonal	oblik		
	SF-36 QM	SF-36 QM	SF-36 Farivar	RAND-HSI Hays	SF-36 QM	SF-36 QM	SF-36 Farivar	RAND-HSI Hays
PF	,42	,34	,20	,27	-,23	-,11	-,02	
RP	,35	,30	,31	,35	-,12	-,02	,03	
BP	,32	,28	,23	,28	-,10	-,01	,04	
GH	,25	,24	,20	,29	-,02	,05	,10	
VT	,03	,09	,13		,24	,23	,29	,31
SF	-,01	,07	,11		,27	,26	,14	,31
RE	-,19	-,07	,03		,43	,37	,20	,16
MH	-,22	-,08	-,03		,49	,41	,35	,43

Flera saker kan noteras vid inspektion och jämförelser av vikterna i tabellen. För det första, vissa av de negativa vikterna i SF-36 ortogonalsystemet är anmärkningsvärt stora. MH och RE har negativa vikter i PCS som tillsammans är nästan lika stora som den högst viktade fysisk domänen PF, och i MCS har PF nästan lika stora negativt bidrag som VT har positivt. Detta innebär att dessa negativt viktade domäner har en betydande negativ inverkan på motsatt summaindex, vilket förvandlas till en betydande positiv effekt när de multipliceras med låga (negativa) standardiserade domänpoäng. Vidare, i de två SF-36 oblikmodellerna finns det som förväntat färre negativt viktade domäner i varje summaindex och alla domänvikter för motsatt summaindex är låga, vilket innebär att inverkan av dessa domäner på motsatt index blir försumbar. I RAND-HSI däremot ingår endast fysiska domäner i PCS och endast psykiska domäner i MCS.

Den inbördes storleksordningen av vikterna för de starkaste domänerna i varje summaindex varierar också mellan de fyra systemen, där QMs två system liknar mest varandra och likaså är Farivars och Hays system de som liknar varandra mest. Till exempel, PF som antas vara den starkaste domänen inom fysisk hälsa har högsta vikt i QMs två system medan den är den lägst viktade i Farivars och Hays system. Denna skillnad är anmärkningsvärd eftersom Farivars och QMs oblika vikter togs fram med samma metoder (principalkomponentanalys, oblik rotation). En möjlig förklaring kan vara att skillnaden beror på olikheter i studiepopulationerna i vilka viktsystemen togs fram (Hann *et al.* 2008, Anagnostopoulos *et al.* 2009).

## RAND/ SF-36 begrepps- och mätmodellen. En universell modell?

Studierna som hittills redovisats i denna rapport har konsekvent utgått ifrån den hierarkiska modell som visas på sidan 3, där hälsa anses kunna delas i två överordnade komponenter, fysisk och psykisk hälsa, vilka i sin tur representeras av åtta olika hälsodomäner som definieras av olika sammansättningar av 35 frågor. I enlighet med den hierarkiska strukturen har summaindexen genomgående alltså deriverats från domänerna, inte från de ingående frågorna. Modellen ansågs vara generell och robust även i andra länder än i ursprungslandet USA (Ware, Kosinski *et al.* 1998, Jenkinson *et al.* 1999) och i olika patientgrupper (Hays *et al.* 1995, McHorney *et al.* 1993, Ware *et al.*

1994). På senare år har modellens generaliserbarhet prövats vidare i flera olika generella populationer, i olika patientgrupper och med hjälp av andra statistiska metoder. Resultaten från dessa försök har dock inte varit helt entydiga.

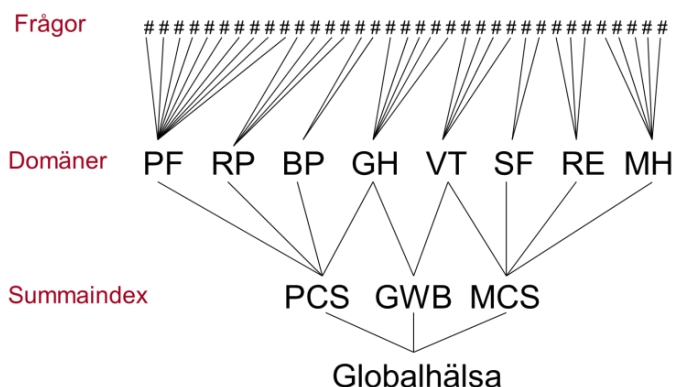
**Skillnader mellan olika allmänna populationer?** Jämförande analyser av faktorstrukturen i tio av länderna inom IQOLA-projektet (US och olika europeiska länder, inklusive Sverige) konfirmerade modellens generaliserbarhet avseende relationen frågor-domäner och domäner-summaindex. Sedan dess har det tillkommit fler än 150 översättningar av RAND/ SF-36, varav de flesta har validitetsgranskats med samma rigorösa psykometriska metoder som användes i IQOLA-projektet. Modellen har kunnat bekräftas i merparten av dessa länder, inklusive en rad olika icke västerländska populationer, som till exempel Indien (Sinha *et al.* 2013), Kina (Lam *et al.* 1998), Iran (Montazeri *et al.* 2005, Motamed *et al.* 2005) och Tunisien (Guermazi *et al.* 2012). Undantag fanns dock för ett fåtal länder, såsom Japan (Fukuhara *et al.* 1998), Turkiet (Demiral *et al.* 2006) och Taiwan (Tseng *et al.* 2003). I Japan till exempel kunde inte komponenttillhörigheten för tre av de åtta domänerna replikeras. (Fukuhara *et al.* 1998). Smärta (BP) laddade lika starkt på både PCS och MCS trots att den tillhör PCS enligt modellen; Allmänhälsa (GH) laddade på MCS i stället för PCS; och Emotionell rollfunktion (RE) tillhörde PCS i stället för MCS. Författarna konkluderade att standard PCS och MCS beräkningsalgoritmen inte kunde tillämpas i Japan och att egna landsvikter behövde utvecklas. En senare studie visade att en trekomponent modell med en tredje "role-social" komponent, innefattande RE, RP och SF, lämpade sig för Japan (Suzukamo *et al.* 2011).

**Skillnader mellan olika patientgrupper?** Även om studier på senare år har kunnat bekräfta modellen i flera patientpopulationer, såsom patienter med artrit (Leung *et al.* 2010), cancer och HIV (Chang *et al.* 2007), obesitas (Corica *et al.* 2006) och vuxna canceröverlevare (Reulen *et al.* 2006), har avvikelser från modellen också noterats. Till exempel, Hobart *et al.* (2001, 2002) fann att modellen inte kunde replikeras helt för patienter med multipel skleros respektive stroke. Domänerna Vitalitet (VT) och Social funktion (SF) i synnerhet, men även Allmän hälsa (GH), uppförde sig annorlunda i dessa patientgrupper än förväntat utifrån modellen. De fann dessutom att summaindexen kunde förklara mindre än 60% av variansen de åtta domänerna, vilket tyder på betydande informationsförlust i summaindexen. Almborg och Berg (2009) visade liknande tillhörighetsproblem med domänen Smärta (BP), också hos stroke patienter, där BP tillhörde MCS i stället för PCS. Liknande problem med RP, GH och RE har förekommit i en studie avseende patienter med funktionella somatiska syndrom (Schröder *et al.* 2012) och för patienter med knä- och höftosteoartrit, där även Social funktion (SF) laddade lika starkt på både PCS och MCS (Rannou *et al.* 2007). Bland patienter med schizofreni fann Leese *et al.* (2008) att det fanns tecken på en tredje "roll-begränsnings"-komponent bestående av RP och RE, liknande den som framkom i Japan (Suzukamo *et al.* 2011). Vidare visade Hann *et al.* (2008) att flera SF-36 domäner laddade på motsatt komponent beroende på sjukdomstillstånd (artrit, ryggsmärta, hjärtsjukdom och psykisk ohälsa). I likhet med Hobart *et al.* fann de att VTs tillhörighet till MCS och PCS varierade mellan patientgrupper och att bland patienter med hjärtsjukdomar laddade Social funktion (SF) högre på PCS än MCS och att för patienter med psykisk ohälsa laddade GH på MCS i stället för PCS. De konkluderade att SF-36 modellen inte kunde appliceras generellt i alla patientgrupper och att modellen passade särskilt dåligt för patienter med psykiska åkommor.

**Två eller tre summaindex?** Andra statistiska metoder än explorativ faktoranalys (EFA) har använts för att undersöka begreppsmodellens validitet. EFA som användes vid utveckling och validitetsprövning av RAND/ SF-36 tillämpas vanligtvis för att undersöka underliggande mönster i data (datadriven

analys), då det saknas en a priori teori eller modell om hur frågor korrelerar och bildar överordnade dimensioner. Flera forskare har hävdats att RAND/ SF-36 begreppsmodellen i stället lämpligast testas genom konfirmatorisk faktoranalys (CFA) och structural equation modellering (SEM) (de Vet *et al.* 2005, Hays *et al.* 2005, Wilson *et al.* 2000, Tucker *et al.* 2013). SEM innefattar modellanpassning och testning baserad på analys av kovariansstrukturer inom ramen för en konfirmatorisk faktoranalytisk modell och syftar således till att testa data mot en teoretisk modell (teoridriven analys).

SEM-studier har dock inte heller gett ett helt entydligt svar på frågan om RAND/ SF-36 modellen är robust. En av de första studier som analyserade modellen med SEM genomfördes med ursprungsdata från tio IQOLA länder, däribland Sverige (Keller *et al.* 1998). Generellt konfirmerades modellen vad gäller relationen frågor till domäner, dock identifierade analysen en tredje komponent, generellt välbefinnande, utöver fysisk och psykisk hälsa (figur 7). Komponenten bildades av Allmän hälsa (GH) och Vitalitet (VT). Denna tre-komponentmodell visade bättre anpassning i nästan alla av de tio länderna, med undantag för bl. a. Sverige, där den ursprungliga två-komponentstrukturen passade bäst. Det kunde också konstateras att dessa tre komponenter bestäms av en enda överordnade faktor: hälsa. Kellers modell har senare bekräftats i en grekisk studie (Anagnostopoulos *et al.* 2005); däremot fann de där att en modell med korrelerade fysisk hälsa, psykisk hälsa och välbefinnande var överlägsen Kellers okorrelerade modell. En tre-komponentmodell (dock med annat innehåll) har också konfirmerats i Japan (Suzukamo *et al.* 2011) och Taiwan (Huang *et al.* 2004) samt hos patienter med schizofreni (Leese *et al.* 2008) och för äldre personer med komorbiditeter (Meng *et al.* 2013).



Figur 7. Keller *et als* (1998) SF-36 modell med tre överordnade komponenter utifrån SEM analys. (GWB=General Well-Being)

Andra SEM-baserade studier har däremot generellt kunnat bekräfta en korrelerad två-komponentstruktur (Wilson *et al.* 2000, Tucker *et al.* 2013, Reed *et al.* 1998, Hann *et al.* 2008, Reed *et al.* 2000, Anagnostopoulos *et al.* 2009), dock har endast två av studierna jämfört två- och tre-komponentmodellerna (Hann *et al.* 2008, Anagnostopoulos *et al.* 2009). Hann *et al.* (2008) visade att den tredje komponenten (enligt Keller *et als.* modell) var högt korrelerad ( $r > .80$ ) med de övriga två och således att det inte fanns belägg för att den utgjorde en distinkt komponent. Anagnostopoulos *et al.* (2009) fann att en tre-komponentmodell inte var signifikant bättre än en två-komponent sådan.

Rasch residual faktoranalys har också tillämpats för att undersöka RAND/ SF-36 dimensionella struktur (Chang *et al.* 2007) och har kunnat konfirmera en korrelerad två-komponentmodell utifrån ingående frågorna. Generellt sätt laddade frågorna som förväntat, dvs frågorna tillhörande fysiska

respektive psykiska domäner laddade på PCS respektive MCS. Undantag fanns för vissa frågor i domänerna GH och VT: endast en GH fråga laddade på PCS medan de övriga fyra laddade på motsatt komponent, medan de fyra ingående frågorna i VT delades mellan komponenterna. Denna uppdelning kan möjligen förklara de betydande korrelationerna dessa domäner har med både komponenter, vilka antas i RAND/SF-36 modellen.

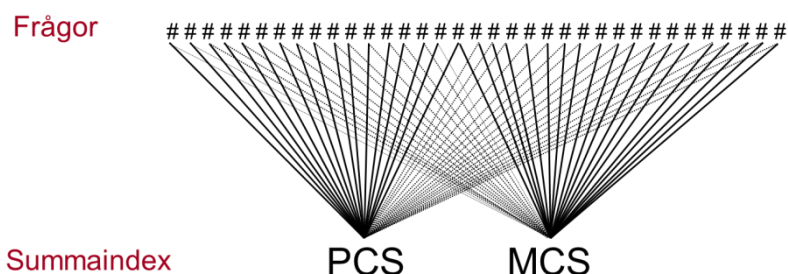
## Alternativa beräkningsprocedurer för summaindex

Utöver RAND-HSI (Hays 1993) och de andra oblikberäkningsmetoderna (Farivar *et al.* 2007, Anagnostopoulos *et al.* 2009) har relativt få studier gällande alternativa RAND/SF-36 summaindexberäkningsprocedurer kunnat identifieras. De studier som hittats har huvudsakligen haft som avstamp i sin kritik antingen att: 1) SF-36 summaindexen aggregerar och sammanfattar informationen i de underliggande domänerna på ett felaktigt sätt, främst beroende på dess antagande om okorrelerad fysisk och psykisk hälsa; eller 2) SF-36 modellen inte är allmängiltig i alla patientgrupper.

**SEM-deriverade item vikter.** I en australiensk studie undersöktes om diskrepanser mellan domän- och komponentpoäng kunde minskas med utgångspunkt i en korrelerad modell, och om SEM-analys kunde tillföra ytterligare någon förbättring (Wilson *et al.* 2000). EFA ortogonal- och oblikvikter samt SEM koefficienter deriverades ur data från Australian National Health Survey (n=23 800) och applicerades sedan på data från South Australian Health Omnibus Survey (n=3 014). En initial SEM-analys visade på dålig anpassning i modellen där komponenterna genereras utifrån de åtta domänerna och därför användes de 35 frågorna som manifesta variabler i modellen, där de fysiska och psykiska komponenterna antogs vara korrelerade. Jämförelser mellan de olika viktsystemen indikerade diskrepans mellan komponent- och domänpoäng i både standard SF-36 ortogonalsystemet och i systemet med oblik-generade vikter (där de åtta domänerna bildar komponenterna). Till exempel, trots signifikant lägre poäng på flera av de psykiska domänerna bland de äldre deltagarna genererade det ortogonala systemet ett signifikant högre MCS värde jämfört med de yngre. Det oblika systemet däremot producerade de lägsta MCS värden hos personer som inte medicinerats för varken fysiska eller psykiska ohälsa och högsta MCS hos de som medicinerats på båda fysiska eller psykiska ohälsa. SEM systemet genererade dock PCS och MCS poäng helt kongruent med domänresultaten.

I ovan nämnda studie beskrivs beräkningssystemet mycket översiktligt och utan att rapportera SEM-koefficienter eller beskriva summaindexberäkningsproceduren, men samma forskargrupp har redovisat systemet i mer detalj i en senare publikation, dock i relation till SF-36 v2.0 (Tucker *et al.* 2013). I systemet får varje fråga en vikt för både PCS och MCS, till skillnad från RAND-HSI och SF-36 där vikterna appliceras på domännivå (figur 8). Vikterna för varje fråga multipliceras med patientskattningen och sedan i likhet med RAND/SF-36 beräkning summeras produkterna som sedan transformeras till t-scores. I enlighet med RAND/ SF-36 modellen är frågorna tillhörande domänerna PF, RP, BP och GH viktade primärt på PCS medan frågorna tillhörande MH, RE, SF och VT har högsta vikter på MCS. Alla vikter på både PCS och MCS är dock positiva. Anmärkningsvärt är vikterna för frågorna tillhörande de starkaste domänerna i PCS respektive MCS enligt RAND/SF-36 modellen, dvs

PF och MH, bland de lägsta och i stället är RP frågorna högst viktade i PCS, och i MCS är det främst de som tillhör SF, men även VT och RE, som får högst vikter. Trots detta, men i likhet med den tidigare studien, konstaterade författarna att SEM-vikterna producerade summaindexpoäng som stämde bättre (ytvaliditet, face validity) med domänpoängen än vad de ortogonal- och oblik-baserade vikterna gjorde, och bättre avspeglade förväntade skillnader mellan olika ålderskategorier och personer som behandlades eller inte för fysiska och/ eller psykiska sjukdomar.



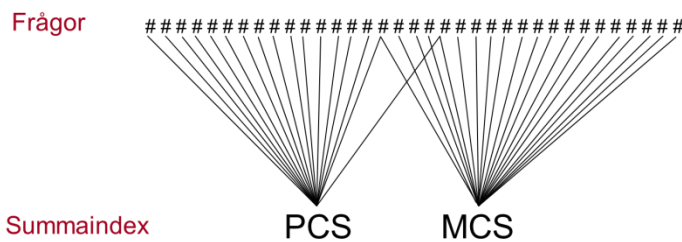
Figur 8. SEM-deriverade modell. Samtliga frågor bidrar positivt till båda summaindexen.

**Enkelsummering, utan vikter.** En annan summaindexberäkningsprocedur har nyligen publicerats (Grassi *et al.* 2010), vilket i likhet med ovan nämnda SEM-baserade system utgår ifrån frågorna i stället för domänerna vid framtagning av summaindexpoäng. Proceduren utvecklades med hjälp av data från en stor ( $n > 10\,000$ ) internationell befolkningsbaserad studie (European Community Respiratory Health Survey). Summaindexen beräknas som en enkel summering av *oviktade* svar på frågorna tillhörande PCS respektive MCS, som sedan transformeras till t-scores på samma sätt som i SEM-systemet ovan. I stället för vikter används så kallad "unit-weighting", där varje variabel kodas som 0/ 1, beroende på magnituden av faktorpoäng -antagligen med utgångspunkt i kritiken (stickprovsberoende) mot användning av faktorpoäng som vikter (McDonald *et al.* 1979; Grice *et al.* 1998). Faktorpoäng för frågorna (och därmed tillhörighet till det ena eller andra summaindex) bestämdes utifrån en EFA (principalkomponentanalys) med oblik rotation. Det kan också nämnas att EFA kördes på transformerade data där frågornas ordinala/nominala svarsskalor kalibrerades om till intervallnivå genom så kallad optimalskalning, för att säkerställa likhet med intervallskala, vilket är en förutsättning för att Likert metoden kan tillämpas.

Till skillnad mot SEM-systemet ingår frågorna i GH-domänen huvudsakligen i MCS i stället för PCS som den gör i RAND/SF-36 modellen, med undantag för två av dess fem frågor, vilka ingår i både MCS och PCS. Övriga frågor ingår i beräkningen av endast ett summaindex (figur 9), i likhet med RAND-HSI metoden där domänerna ingår i endast ett summaindex.

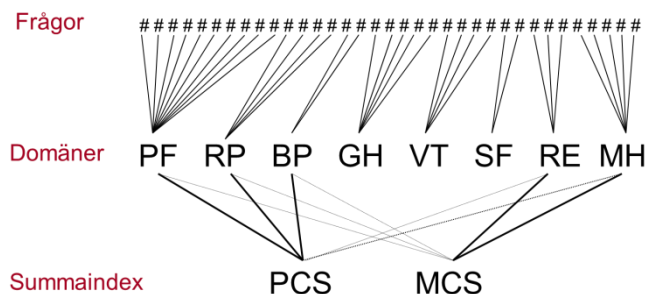
Jämförelser med summaindexpoäng räknade enligt standard SF-36 metoden med ortogonala och oblika vikter visade att samstämmigheten med domänprofilerna var bäst för enkelsummeringsmetoden och sämst för ortogonala vikter. Likaså kunde enkelsummeringsmetoden diskriminera signifikant bättre mellan patienter med olika tillstånd än de övriga två. Korrelationerna mellan poäng genererade med denna metod och poäng från oblika vikter var 0,97 för PCS och 0,96 för MCS, vilket tyder på att systemen producerar mycket snarlika poäng.





Figur 9. Enkelsummeringsmodell. Frågorna bidrar positivt till endast en summaindex med undantag för två GH frågor. I motsats till RAND/SF-36 modellen bidrar de övriga GH frågorna till MCS i stället för till PCS.

**Generisk patientmodell.** Generaliserbarheten av SF-36 modellen jämfört med flera andra modeller (bland annat SF-36 ortogonal, oblik samt Keller *et als* tre-komponentmodell) har undersökts i olika sjukdomsgrupper i en engelsk studie (Hann *et al.* 2008) med hjälp av både EFA och CFA. Upptäckten att domänsummaindexstrukturen varierade substantiellt mellan olika sjukdomstillstånd (artrit, ryggsmärta, hjärtsjukdom och psykisk ohälsa), främst avseende VT, GH och SF, ledde till att en gemensam modell för alla patientgrupperna prövades. I modellen exkluderas helt enkelt dessa tre domäner och enbart PF, RP och BP ingår i främst PCS-beräkning och RE och MH i främst MCS-beräkning, och domänvikterna är ungefär lika stora (figur 10). Domänvikter deriverades med hjälp av EFA med oblik rotation och var alla positiva på båda summaindex (Tabell 2). Modellen fungerade utmärkt i samtliga patientgrupper och analyserna visade att de tre exkluderade domänerna inte bidrog substantiellt till prediktionsförmågan eller reliabiliteten hos summaindexen.



Figur 10. Hann *et als* (2008) generell modell för olika patientgrupper. Domäner (GH, VT och SF) som varierar mellan patientgrupper i tillhörighet till fysisk och psykisk hälsa exkluderas i summaindexberäkningen.

Tabell 2. Domänvikter i Hann *et als* generiska modell.

Domän	PCS vikter	MCS vikter
PF	,33	,04
RP	,33	,11
BP	,31	,08
GH	-	-
VT	-	-
SF	-	-
RE	,13	,39
MH	,04	,47

## Sammanfattning och rekommendationer

RAND/ SF-36 utvecklades som ett multidimensionellt hälsostatusinstrument för generell användning i befolkningen. Från början mättes åtta olika hälsodomäner och resultat redovisades i form av en så kallad hälsoprofil. Summaindexen representerar en senare utveckling av instrumenten, där informationen i domänerna sammanfattas i två överordnade dimensioner: fysisk och psykisk hälsa. Modellen är således hierarkisk och har varit föremål för forskning och debatt i litteraturen i snart 20 år. Två huvudforskningsfrågor har behandlats, nämligen huruvida summaindexen sammanfattar domänerna på ett korrekt sätt och hur eventuella diskrepanser kan minskas; och mer fundamentalt, om modellen är generellt valid och applicerbar.

Ett tidigt och fortfarande aktuellt fokus i debatten är huruvida fysisk och psykisk hälsa bör betraktas som oberoende eller beroende fenomen. SF-36 utgår från det förstnämnda medan RAND-36 (RAND-HSI metoden) modellen vilar på det senare, vilket avspeglas i skillnaden mellan SF-36 och RAND-HSI i summaindexberäkningsalgoritmerna. SF-36 modellen med antagandet om två oberoende summaindex har kritiserats genom åren på både teoretiska och empiriska grunder, men trots omfattande kritik har upphovsrättshållaren QualityMetric Inc valt att behålla sin SF-36 summaindexberäkningsalgoritm. Å andra sidan har en lång rad olika explorativ- och konfirmatorisk-faktoranalytiska studier på senare år gett stöd för att fysisk och psykisk hälsa är korrelerade begrepp, och därmed har RAND-HSI modellen framförts som ett alternativ för att sammanföra domänerna till summaindex. Direkta jämförelser mellan SF-36 och RAND-HSI summaindexresultat har också visat att överensstämmelserna mellan summaindexen och domänerna de avser att sammanfatta är substantiellt bättre och mer korrekta med RAND-HSI än med SF-36. RAND-HSI är också mer känslig för att upptäcka förändring i hälsostatus och skillnader mellan grupper med olika hälsostatus. Således, i valet mellan SF-36 och RAND-HSI summaindex beräkningsprocedurerna framstår RAND-HSI som klart överlägsen.

Tillämpning av RAND-HSI proceduren är däremot inte problemfri. Möjliga hinder är bland annat att: 1) RAND-HSI summaindexvikter utvecklades baserade på domänpoäng som framräknades på ett annorlunda sätt än dagens RAND-36 domänberäkning, vilket kan inverka på vikternas giltighet för den nuvarande RAND-36; och 2) vikterna är från tidigt 1990-tal och baseras på amerikanska befolkningsdata, och är därmed inte nödvändigtvis tillämpliga idag och i Sverige. Å andra sidan är RAND-HSI ett etablerat och beprövat summaindexberäkningssystem för RAND-36 och därmed skulle dess användning underlätta och främja jämförelser med internationella studier. Erfarenheter från SF-36 IQOLA-projektet tyder dessutom på en adekvat överensstämmelse mellan olika länders viktsystem. (Ware, Gandek, Kosinski *et al.* 1998)

Ett alternativ till RAND-HSI summaindexvikter vore att använda publicerade vikter som framtagits med SF-36 data men utifrån antagandet om korrelerad fysisk och psykisk hälsa, exempelvis vikter från Farivar *et al.* (2007) eller Saris-Baglama *et al.* (2004). Dessa vikter är framräknade på domändata som mer liknar dagens RAND-36 domänpoängberäkning (undantag domänerna BP och GH) än RAND-HSI och på så sätt är det mer sannolikt att de lämpar sig för användning med den nya svenska RAND-36. Vidare är vikterna framtagna senare och därmed bör de vara någorlunda aktuella för dagens förhållanden. Det är dock oklart om vikterna har applicerats i andra studier. Nackdelen är dessutom att de är framtagna med amerikanska befolkningsdata och det är därför oklart om de är tillämpliga för Sverige. För att säkerställa vikternas giltighet för den svenska RAND-36 bör metoden replikeras

(explorativ faktoranalys med oblik rotation av domänvärde) med data från den svenska RAND-36 normdatabasen. Summaindexpoäng bör sedan framräknas med båda viktsystemen för att bedöma dess jämförbarhet. Med hög sannolikhet kommer de svenska domänvikterna dock att avvika från dessa publicerade vikter, både i absolut nivå och i relativ inbördes magnitud (jämför Farivars och Saris-Baglamas vikter i tabell 1). Explorativ faktoranalys är en datadriven metod och faktorpoäng som används som vikter är särskilt avhängig och känslig för variation i stickprovskaraktistik (McDonald *et al.* 1979; Grice *et al.* 1998; Hann *et al.* 2008). Ett möjligt alternativ till viktning för att kunna justera för eventuella små skillnader i faktorpoängvikter är att använda så kallade "unit-weighting", där varje domän antingen ingår eller inte i vardera summaindex, dvs utan viktning. Metoden har använts av Grassi *et al.* (2010) för aggregering av SF-36 frågor till summaindex, men den bör kunna tillämpas även för aggregering av domänerna. Metoden med användning av trösklar som tillämpades av Grassi skulle förmodligen innebära ett rent viktsystem liknande RAND-HSI, där enbart fysiska domäner ingår i PCS och enbart psykiska i MCS. Därmed kan de problematiska negativa vikterna som även förekommer i oblik EFA undvikas.

Även om ett oblik-baserat viktsystem skulle visa sig lämpligt för summaindexberäkning i den svenska normalbefolkningen är det ändå oklart om den RAND/ SF-36 begreppsmodellen är allmängiltig. Eftersom flera studier har visat på variation i modellen är det viktigt att värdera den också med hjälp av andra statistiska metoder och i olika patientgrupper och andra subgrupper i befolkningen. Modellen bör prövas och jämföras med andra konkurrerande modeller genom någon form av konfirmatorisk analys, t.ex. SEM analys. Det är viktigt att utröna om en modell med tre summaindex, som framkommit i litteraturen, är lämpligare än två summaindex. I den mån den svenska RAND-36 normdatabasen tillåter det, bör olika modeller testas för invarians. En särskilt angelägen uppgift blir alltså att pröva vidare om samma modell gäller i olika patientgrupper. Ifall det upptäcks substansiell variation bör man ta fram särskilda summaindexberäkningsprocedurer för olika patientgrupper, eller försöka utveckla en gemensam modell som tar hänsyn till olikheter mellan patientgrupper, liknande den som föreslås av Hann *et al.* (2008). I Hann *et al.*'s modell uteslöts domänerna som visade störst variation mellan patientgrupperna, nämligen GH, VT och SF. Dessa domäner bidrar enligt RAND/SF-36 modellen till både fysisk och psykisk hälsa och det är därför inte oväntat att graden av samhörighet till det ena eller det andra summaindexet skulle tippa åt olika håll beroende på hälsotillstånd. Det kan dock också bero på att ingående frågor i dessa domäner har lite olika inriktningar (t ex positivt jämfört med negativt ställda frågor) och inte är lika sammanhållna som antytts, vilket har indikerats i Rasch-analys för GH och VT (Chang *et al.* 2007) och även i EFA för GH (Grassi *et al.* 2010). Det är därför angeläget att modellprövningen sker på hela modellen såsom i Wilson *et al.* (2000) och Tucker (2013) och inte enbart på domännivå såsom i Hann *et al.*

**Rekommendationer.** Flera olika summaindexberäkningsalgoritmer bör tillgängliggöras för att möta olika specifika behov och forskningsändamål:

1. En generell standardalgoritm för att främja internationella jämförelser. RAND-HSI metoden är det lämpligaste valet eftersom den är den enda metoden som använts regelbundet med RAND-36. SF-36 algoritmen utgör inte ett alternativ.
2. En generell standardalgoritm för nationell bruk. Vikter kan tas fram med explorativ faktoranalys av data från den svenska RAND-36 normdatabasen, men bör prövas vidare i en

representativ delmängd från populationsdata med någon form av konfirmatorisk analys. Oblik rotation av faktorerna bör appliceras. Användning av "unit-weighting" bör övervägas.

3. Olika tillståndsspecifika algoritmer, alternativt en tillstånd-generisk algoritm. Lämpligheten av den generella modellen i olika patientgrupper bör prövas och alternativa algoritmer bör utvecklas vid behov. Data från de olika nationella kvalitetsregistren som i nuläget använder SF-36 bör utgöra en adekvat analysbas.

Det är sannolikt att de olika algoritmerna kommer att generera olika summaindexpoäng även om korrelationerna sinsemellan troligen blir höga. Det är därför viktigt att implikationer och effekter av de olika systemen tydliggörs så att forskarna kan göra ett medvetet och välavvägt val mellan systemen.

Slutligen, i enlighet med flera rekommendationer (t ex Hann *et al.* 2008, Anagnostopoulos *et al.* 2009, Taft *et al.* 2001, Simon *et al.* 1998, Farivar *et al.* 2007, Ware *et al.* 2001) bör summaindexen rapporteras tillsammans med domänprofilen. Domänerna ger viktig och nyanserad information som summaindexen inte kan helt fånga. Domänerna är också benchmarken för att bedöma hur väl summaindexen sammanfattar dess innehåll.

**Begränsningar.** Litteraturen kring RAND/SF-36 är omfattande (fler än 15 500 sökträffar), vilket innebär att trots att litteratursökningarna ämnade vara systematiskt och brett är det möjligt att relevanta studier kan ha missats.

## Del 2: RAND-12

---

**Inledning:** RAND-12 Health Survey Inventory (HSI) är en förkortad 12-frågorversion av 36-frågor RAND-36 HSI. Utvecklingen bygger på huvudinstrumentens summaindex PCS och MCS, samt även globalhälsa (GHC), där avsikten var att kunna reproducera/ mäta dessa med färre antal frågor och med någorlunda väl bibehållen precision (Hays *et al.* 1998, Ware, Kosinski, Keller 1995). Korrelationer mellan RAND-36 och RAND-12 summaindex och GHC är mycket höga ( $r_{PCS}=0,94$ ;  $r_{MCS}=0,92$ ;  $r_{GHC}=0,95$ ). I likhet med huvudenkäterna RAND-36 och SF-36 är frågeinnehåll och svarsformat i RAND-12 och QualityMetrics SF-12 identiska. Poängberäkningsalgoritmerna är dock olika mellan RAND-12 HSI och SF-12 och därför är genererade poäng inte jämförbara. En annan variant av enkäten distribueras under namnet VR-12 (tidigare Veterans RAND-12; Jones, 2001) men den tillämpar ett svarsformat liknande SF-36 version 2.0.

**Användningsområde:** RAND-12 HSI/ SF-12 utvecklades för användning primärt i större befolkningsstudier där RAND-36/ SF-36 anses vara för lång. Den avråds från användning i mindre studier eftersom den inte anses ha tillräckligt precision (Hays *et al.* 1998, Ware, Kosinski, Keller 1995).

**Tillgänglighet:** RAND Corporation distribuerar inte RAND-12 HSI på sin websida utan i stället hänvisar RAND till QualityMetric Inc som upphovsrättshållare av SF-12 ([http://www.rand.org/health/surveys\\_tools/mos/mos\\_core\\_12item.html](http://www.rand.org/health/surveys_tools/mos/mos_core_12item.html)). Detta förhållande bekräftades vid direkt kontakt med RAND, där det underströks att QualityMetric anses äga alla rättigheter till 12-frågorversionen.

Det finns ingen egen manual till RAND-12 HSI utan information om enkätens utveckling finns i en bilaga till RAND-36 HSI manualen tillsammans med själva enkäten samt poängberäkningsprocedur och tillhörande tabeller. Manualen finns tillgänglig för gratis nedladdning på författarens personliga webbsida (<http://gim.med.ucla.edu/FacultyPages/Hays/surveys/R-36%20HSI%20Hays/>). Tidigare distribuerades manualen via Psychological Corporation (nuvarande Pearson Education Inc).

**Psykometrisk prövning:** Det finns ingen egentlig psykometrisk prövning av RAND-12 HSI publicerad i den vetenskapliga litteraturen; dokumentationen på enkätens psykometriska egenskaper återfinns enbart i manualen. Det finns däremot flera vetenskapliga artiklar som redovisat psykometriska prövningar av SF-12, men skillnader i poängberäkning mellan RAND-12 HSI och SF-12 gör att resultaten inte gäller för RAND-12 HSI.

**Vetenskaplig användning:** Litteratursökningar i tre databaser (Scopus, CINAHL och Cochrane) gav endast 21 artiklar där RAND-12 HSI ingår. Med tanke på att RAND-12 HSI har funnits sedan 1998 kan användningen betraktas som mycket blygsam, i synnerhet i jämförelse med de drygt 2 700 SF-12 publikationerna under motsvarande perioden. Merparten av RAND-12 HSI publikationerna gäller dessutom metodologiska studier, oftast i form av jämförande studier med andra PROM.

**Sammanfattning och rekommendationer:** Det är viktigt att särskilja RAND-12 från RAND-36 HSI. Enligt RAND Corporation finns det ingen RAND-auktoriserad 12-frågorversion (RAND-12) av RAND-36. Å andra sidan har RAND-12 HSI funnits tillgänglig i snart 20 år. Upphovsrätten till RAND-12 HSI manualen, och antagligen även enkäten och poängberäkningsalgoritmen som publiceras däri, ägs

enligt den 1998 upplaga av förlaget Psychological Corporation. Eftersom manualen inte längre ges ut av förlaget (numera Pearson Education Inc), utan i stället distribueras fritt av författaren själv, kan man antagligen anta att upphovsrätten har upphört eller att förlaget inte längre bryr sig om den. Enligt författaren själv är RAND-12 HSI enkäten och dess poängberäkningsinstruktioner i public domain. Upphovsrättsliga frågor kring RAND-12 HSI bör dock undersökas av en jurist, förslagsvis i samarbete med utredningen som pågår i Norge avseende RAND-36/RAND-12 (Nilsson 2014).

Om RAND-12 HSI bedöms vara aktuell och utan rättsliga hinder för användning i Sverige blir följdfrågan hur poängberäkning för summaindexen och GHC skall gå till. Till skillnad från huvudenkäten RAND-36 har den vetenskapliga användningen av RAND-12 HSI varit mycket begränsad och därför är forskningsjämförbarhetsprincipen av relativt lite vikt i val av procedur. Det finns även flera skäl att inte tillämpa den nuvarande RAND-12 HSI algoritmen, bland annat: den utvecklades baserad på tämligen "gamla" data från 1993; data är från en amerikansk befolkning och proceduren har inte prövats i andra länder; poäng är normaliserade till den amerikanska befolkningen. Det rekommenderas därför att nya vikter/system för RAND-12 HSI beräkningsalgoritmen tas fram efter man har löst problemet med beräkningsproceduren för summaindexen i huvudenkäten RAND-36. Detta bör göras företrädesvis i samråd med Norge.

## Referenser

- Anagnostopoulos F, Niakas D, Pappa E Construct validation of the Greek SF-36 Health Survey. *Qual Life Res.* 2005 Oct;14(8):1959-65.
- Anagnostopoulos F, Niakas D, Tountas Y. Comparison between exploratory factor-analytic and SEM-based approaches to constructing SF-36 summary scores. *Qual Life Res.* 2009 Feb;18(1):53-63.
- Baron R, Elashaal A, Germon T, Hobart JC. Measuring outcomes in cervical spine surgery: think twice before using the SF-36. *Spine* 2006;31: 2575-84.
- Beals J , Welty TK, Mitchell CM, Rhoades DA , Yeh J-L, Henderson JA. (2006). Different factor loadings for SF-36: The Strong Heart Study and the National Survey of Functional Status. *Journal of Clinical Epidemiology*, 59, 208–215
- Blanchard CM, Cote I, Feeny D. Comparing short form and RAND physical and mental health summary scores: results from total hip arthroplasty and high-risk primary-care patients. *Int J Technol Assess Health Care* 2004; 20:230-5.
- Brook RH, Ware JE, Roger WH, *et al.* The effect of coinsurance on the health of adults: Results from the RAND Health Insurance Experiment. Santa Monica, CA: The RAND Corporation, 1984.
- Cella D, Chang CH, Wright BD, Von Roenn JH, Skeel RT. Defining higher order dimensions of self-reported health: further evidence for a two-dimensional structure. *Eval Health Prof* 2005; 28:122-41.
- Chang C-H, Wright BD, Cella D, Hays RD. The SF-36 physical and mental health factors were confirmed in cancer and HIV/AIDS patients. *Journal of Clinical Epidemiology*, Volume 60, Issue 1, January 2007, 68–72.
- Contopoulos-Ioannidis DG, Karvouni A, Kouri I, Ioannidis JPA. SF-36 outcomes in randomized trials: a systematic review. *British Medical Journal* 2009; 338:a3006.

Corica F, Corsonello A, Apolone G, Lucchetti M, Melchionda N, Marchesini G; QUOVADIS Study Group. Construct validity of the Short Form-36 Health Survey and its relationship with BMI in obese outpatients. *Obesity (Silver Spring)*. 2006 Aug;14(8):1429-37.

Cunningham WE, Nakazono TT, Tsai KL, Hays RD. Do differences in methods for constructing SF-36 physical and mental health summary measures change their associations with chronic medical conditions and utilization? *Qual Life Res*. 2003 Dec;12(8):1029-35.

Dersh J, Polatin P, Gatchel R. (2002). Chronic pain and psychopathology: Research findings and theoretical considerations. *Psychosomatic Medicine*, 64(5), 773–786.

de Vet HC, Adèr HJ, Terwee CB, Pouwer F. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. *Qual Life Res*. 2005 Jun;14(5): 1203-18.

Fuh JL, Wang SJ, Lu SR, Juang KD, Lee SJ. Psychometric evaluation of a Chinese (Taiwanese) version of the SF-36 health survey amongst middle-aged women from a rural community. *Qual Life Res*, 9 (2000), pp. 675–683.

Fukuhara S, Ware JE Jr, Kosinski M, Wada S, Gandek B. Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey. *J Clin Epidemiol*. 1998 Nov;51(11):1045-53.

Garratt AM. Patient reported outcomes in trials: are widely available, but need to be standardised and used appropriately. *British Medical Journal* 2009; 338:119-120.

Garratt AM, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal* 2002;324:1417-9.

Grassi M, Nucera A. Dimensionality and summary measures of the SF-36 v1.6: comparison of scale- and item-based approach across ECRHS II adults population. *Value Health*. 2010 Jun-Jul;13(4):469-78.

Grice JW, Harris RJ. A comparison of regression and loading weights for the computation of factor scores. *Multivariate Behav Res*, 33 (1998), pp. 221–247.

Guermazi M, Allouch C, Yahia M, Huissa TBA, Ghorbel S, Damak J, Mrad M, Elleuch MH. Translation in Arabic, adaptation and validation of the SF-36 Health Survey for use in Tunisia. *Annals of Physical and Rehabilitation Medicine* Volume 55, Issue 6, 2012, 388–403

Güthlin C, Walach H. Structural Equation Modeling to Test the Construct Validity of the Second-Order Factor Structure *European Journal of Psychological Assessment*, Vol. 23, No. 1. (2007), 15-23.

Hann M, Reeves D. The SF-36 scales are not accurately summarised by independent physical and mental component scores. *Qual Life Res* 2008; 17:413-23.

Hays RD, Prince-Embury S, Chen H. (1998). *RAND-36 Health Status Inventory*. San Antonio, TX: The Psychological Corporation.

Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. 2000 Sep;38(9 Suppl):II28-42.

Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. 1993 Oct;2(3):217-27.

Hays RD, Revicki D, Coyne KS. (2005). Application of structural equation modeling to health outcomes research. *Evaluation & the Health Professions*, 28, 295–309.

Hobart JC, Williams LS, Moran K, Thompson AJ. Quality of life measurement after stroke. Uses and abuses of the SF-36. *Stroke* 2002;33: 1348-5

Hobart J, Freeman J, Lamping D, Fitzpatrick R, Thompson A. The SF-36 in multiple sclerosis: why basic assumptions must be tested. *J Neurol Neurosurg Psychiatry* 2001; 71: 363-70.

- Huang IC, Frangakis C, Wu A. Are there different components for the Taiwan SF-36 compared to western versions—physical, mental and social? *Qual Life Res*, 13 (2004), p. 1495
- Jenkinson C, Hobart J, Chandola T, Fitzpatrick R, Peto V, Swash M. Use of the short form health survey (SF-36) in patients with amyotrophic lateral sclerosis: tests of data quality, score reliability, response rate and scaling assumptions. *J Neurol* 2002; 249(2): 178-83.
- Jones D, Kazis L, Lee A, Rogers W, Skinner K, Cassar L, et al. Health status assessments using the Veterans SF-12 and SF-36: Methods for evaluating outcomes in the Veterans Health Administration. *The Journal of Ambulatory Care Management* 2001; 24: 68–86.
- Katon W, Lin E, VonKorff M, Russo J, Lipscomb P, Bush T. Somatization: A spectrum of severity. *Am J Psychiat* 1991; 148:34.
- Lam CL, Gandek B, Ren XS, Chan MS. Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 health survey. *J Clin Epidemiol*. 1998;51:1139–47.
- Leese M, Schene A, Koeter M, Meijer K, Bindman J, Mazzi M, Puschner B, Burti L, Becker T, Moreno M, Celani D, White IR, Thonicroft G. SF-36 scales, and simple sums of scales, were reliable quality-of-life summaries for patients with schizophrenia. *J Clin Epidemiol*. 2008 Jun; 61(6):588-96.
- Leung YY, Ho KW, Zhu TY, Tam LS, Kun EW, Li EK. Testing scaling assumptions, reliability and validity of medical outcomes study short-form 36 health survey in psoriatic arthritis. *Rheumatology (Oxford)*. 2010 Aug;49(8):1495-501.
- Li L, Wang MH, Shen Y. Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalization. *J Epidemiol Community Health*, 57 (2003), pp. 259–263
- McDonald RP, Mulaik S. Determinacy of common factors: a nontechnical review. *Psychol Bull* 1979;86:297–306.
- McHorney CA, Haley SM, Ware JE Jr. Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. *J Clin Epidemiol*. 1997 Apr;50(4):451-61.
- McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993 Mar;31(3):247-63.
- McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care*. 1994 Jan;32(1):40-66.
- Meng H, King-Kallimanis BL, Gum A, Wamsley B. Measurement bias of the SF-36 Health Survey in older adults with chronic conditions. *Qual Life Res*. 2013 Nov;22(9):2359-69.
- Motamed N, Ayatollahi AR, Zare N, Sadeghi-Hassanabadi A. Validity and reliability of the Persian translation of the SF-36 version 2 questionnaire. *East Mediterr Health J*, 11 (3) (2005), pp. 349–35.
- Montazeri A, Goshtasebi A, Vahdaninia M, Gandek B. The short form health survey (SF-36): Translation and validation study of the Iranian version. *Qual Life Res*. 2005;14:875–82.
- Nilsson E, Orwelius L, Kristenson M. Patient-reported outcomes in the Swedish National Quality Registers. *J Intern Med*. 2015.
- Nilsson E. Patientrapporterade utfallsmått (PROM) i kvalitetsregister för långvariga sjukdoms-tillstånd. Generella mått och indikatorer – utvecklingsförslag. Socialstyrelsen Dnr 5.3-37022/2013, 2014.



Nordveldt MW, Riise T, Myhr KM. Performance of the SF-36, SF-12 and RAND-36 summary scales in a multiple sclerosis population. *Med Care* 2000; 38; 1022-28.

Osterhaus JT, Townsend RJ, Gandek B, Ware JE Jr. Measuring the functional status and well-being of patients with migraine headache. *Headache*. 1994 Jun;34(6):337-43.

Pelle AJ, Kupper N, Mols F, de Jonge P. What is the use? Application of the short form (SF) questionnaires for the evaluation of treatment effects. *Qual Life Res*. 2013 Aug;22(6):1225-30.

Raczek AE, Ware JE, Bjorner JB, Gandek B, Haley SM, Aaronson NK, Apolone G, Bech P, Brazier JE, Bullinger M, Sullivan M. Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: results from the IQOLA Project. *International Quality of Life Assessment. J Clin Epidemiol*. 1998 Nov;51(11):1203-14.

Rannou F, Boutron I, Jardinaud-Lopez M, Meric G, Revel M, Fermanian J, Poiraudou S. Should aggregate scores of the Medical Outcomes Study 36-item Short Form Health Survey be used to assess quality of life in knee and hip osteoarthritis? A national survey in primary care. *Osteoarthritis Cartilage*. 2007 Sep;15(9):1013-8.

Reed PJ, Moore D. SF-36 as A Predictor of Health States. *Value in Health, Volume 3, Issue 3, May–June 2000, Pages 202-207*.

Reulen RC, Zeegers MP, Jenkinson C, Lancashire ER, Winter DL, Jenney ME, Hawkins MM. The use of the SF-36 questionnaire in adult survivors of childhood cancer: evaluation of data quality, score reliability, and scaling assumptions. *Health Qual Life Outcomes*. 2006 Oct 5;4:77.

Saris-Baglana RN, Dewey CJ, Chisholm GB, *et al*. *QualityMetric Health Outcomes™ Scoring Software 4.0*. Lincoln, RI: QualityMetric Inc. 2004.

Schmitz N, Kruse J. The SF-36 summary scores and their relation to mental disorders: physical functioning may affect performance of the summary scores. *J Clin Epidemiol* 2007; 60:163-70.

Schröder A, Oernboel E, Licht RW, Sharpe M, Fink P. Outcome measurement in functional somatic syndromes: SF-36 summary scores and some scales were not valid. *J Clin Epidemiol*. 2012 Jan;65(1):30-41.

Scoggins JF, Patrick DL. The Use of Patient-Reported Outcomes Instruments in Registered Clinical Trials: Evidence from ClinicalTrials.gov. *Contemp Clin Trials*. 2009 Jul; 30(4): 289–292.

Scott KM, Sarfati D, Tobias MI, Haslett SJ. A challenge to the cross-cultural validity of the SF-36 health survey: factor structure in Maori, Pacific and New Zealand European ethnic groups. *Soc Sci Med*. 2000 Dec;51(11):1655-64.

Simon GE, Revicki DA, Grothaus L, Von Korff M. SF-36 Summary Scores – Are physical and mental health truly distinct. *Med Care* 1998; 36; 567-572).

Simon G, Von Korff M. Somatization and psychiatric disorder in the NIMH Epidemiologic Catchment Area Study. *Am J Psychiatry* 1991; 148:1494.

Sinha R, van den Heuvel WJ, Arokiasamy P. Validity and Reliability of MOS Short Form Health Survey (SF-36) for Use in India. *Indian J Community Med*. 2013 Jan;38(1):22-6.

Stewart AL, Ware JE. (1992). *Measuring functioning and well-being: The Medical Outcomes Study approach*. Durham, NC: Duke University Press.

Suzukamo Y, Fukuhara S, Green J, Kosinski M, Gandek B, Ware JE. Validation testing of a three-component model of Short Form-36 scores. *J Clin Epidemiol*. 2011 Mar; 64(3):301-8.

Taft C, Karlsson J, Sullivan M. Do SF-36 component summary scores accurately summarize subscale scores? *Qual Life Res* 10: 405-413, 2001.

- Taft C, Karlsson J, Sullivan M. Reply to Drs Ware and Kosinski. *Qual Life Res* 10: 415–420, 2001.
- Tarlov AR, Ware JE Jr, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study. An application of methods for monitoring the results of medical care. *JAMA*. 1989 Aug 18;262(7):925-30.
- Thumboo J, Fong KY, Machin D, Chan SP, Leon KH, Feng PH, *et al*. A community-based study of scaling assumptions and construct validity of the English (UK) and Chinese (HK) SF-36 in Singapore. *Qual Life Res*, 10 (2001), pp. 175–188.
- Thombs BD, Hudson M. (2008). Problems using aggregate scores of the medical outcomes study 36-item short form health survey in knee and hip osteoarthritis: Possible solutions. *Osteoarthritis Cartilage*, 16(11), 1431–1432.
- Tseng H, Lu J, Gandek B. Cultural issues in using the SF-36 health survey in Asia: Results from Taiwan. *Health and Quality of Life Outcomes*. 2003;1:72.
- Tucker G, Adams R, Wilson D (2010) New Australian population scoring coefficients for the old version of the SF-36 & SF-12 health status questionnaires. *Qual Life Res* 19(7): 1069–76.
- Ware JE, Sherbourne CD. The MOS 36-item Short-Form Survey (SF-36: I. Conceptual framework and item selection. *Med Care*, 1992; 30(6): 473-83.
- Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care*. 1995 Apr; 33(4 Suppl):AS264-79.
- Ware JE, Snow KK, Kosinski M, Gandek B. (1993). *SF-36 Health Survey: Manual and interpretation guide*. Boston: The Health Institute, New England Medical Center.
- Ware JE Jr, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J Clin Epidemiol*. 1998 Nov;51(11):903-12.
- Ware JE Jr, Gandek B, Kosinski M, Aaronson NK, Apolone G, Brazier J, Bullinger M, Kaasa S, Leplège A, Prieto L, Sullivan M, Thunedborg K. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. *International Quality of Life Assessment*. *J Clin Epidemiol*. 1998 Nov;51(11):1167-70.
- Ware JE Jr, Gandek B, Kosinski M, Aaronson NK, Apolone G, Brazier J, Bullinger M, Kaasa S, Leplège A, Prieto L, Sullivan M, Thunedborg K. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. *International Quality of Life Assessment*. *J Clin Epidemiol*. 1998 Nov;51(11):1167-70.
- Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care*. 1995 Apr;33(4 Suppl):AS264-79
- Ware JE, Kosinski M, Keller SD: *SF-36 Physical and Mental Health Summary Scales: A User's Manual* Boston, MA: The Health Institute, New England Medical Center; 1994.
- Ware JE, Kosinski M, Keller SD: *SF-12: How to score the SF-12 Physical and Mental Health Summary Scales*. Boston, MA: The Health Institute, New England Medical Center; 1995.
- Ware J, Kosinski M, Gandek B, Aaronson N, Apolone G, Bech P, Brazier J, Bullinger M, Kaasa S, Leplege A, Prieto L, Sullivan M. (1998). The factor structure of the SF-36 health survey in 10 countries: Results from the IQOLA project. *Journal of Clinical Epidemiology*, 51(11), 1,159–1,165.
- Ware J, Kosinski M. (2001). Interpreting SF-36 summary health measures: A response. *Quality of Life Research*, 10(5), 405–413.

Wilson D, Parsons J, Tucker (2000) The SF-36 summary scales: problems and solutions. *Soz-Praventivmed* 45: 239–246.

Yu J, Coons SJ, Draugalis DR, Ren XS, Hays RD. Equivalence of Chinese and US-English versions of the SF-36 health survey. *Qual Life Res*, 12 (2003), pp. 449–457.

Yu YF, Yu AP, Ahn J. (2007). Investigating differential item functioning by chronic diseases in the SF-36 health survey: A latent trait analysis using MIMIC models. *Medical Care*, 45, 851–859.