

Metodöversväganden vid analys av aggregerade data från Addiction Severity Index

Armeliu8, B-Å och Armeliu8, K.

Institutet för Klinisk Psykologi i Umeå AB, IKP

2011

Innehåll

Metodöverbåganden vid analys av aggregerade data från Addiction Severity Index	1
Bakgrund	2
Sammanfattningsmått	3
Klientskattning	3
Intervjuarskattning	3
Composite score, CS	4
Olikheter i instruktioner för skattningar mellan version 1998 och 2006	5
Empirisk undersökning av skillnader mellan version 1998 och 2006.	7
Begreppslig oklarhet kring skalsteg	9
Slutsatser och förslag till ändringar av intervjuarskattningen	10
Parametrisk och/eller icke-parametrisk statistik	11
Dubbel diagnos - Definition av missbrukare med sannolik svår psykisk störning genom ROC-analys	11
Beräkning av förändring. Statistiska problem.....	15
Konsekvenser för utfallsmåtten.....	16
Beräkning av förändring med Composite score	17

Bakgrund

Addiction Severity Index (ASI) är en semistrukturerad intervjuemetod för bedömning av alkohol och narkotikarelaterade problem. ASI översattes till svenska genom Centrum för utvärdering av socialt arbete, CUS, (Andréasson et. al., 1996) vilket resulterade i en manual och bedömningsformulär som här kallas version 1998. År 2006 gjordes en ny version av formulären till ASI, version 2006, som nu används i Sverige och en ny manual kom 2009 (Nyström, Zingmark & Jäderland, 2009). ASI finns både som en grundintervju, ASI-Grund och som en uppföljningsintervju, ASI-Uppföljning, vilken är avgränsad till att beröra de senaste sex månaderna och sista månaden i patientens liv och omfattar färre frågor. ASI-intervjun består av bakgrundsfrågor och frågor om problem inom sju olika områden: psykisk och fysisk hälsa, arbete/försörjning, alkohol och narkotika, kriminalitet och familj/umgänge. Frågorna om problem inom olika områden ställs både för tidigare problem och för problem de sista 30 dagarna. Många enheter som utreder och behandlar klienter med missbruksproblem, matar in sina data från ASI i en gemensam databas, ASI-net, där beskrivningar av klienter på individnivå tas ut i form av klientrapporter från databasen. Grupperade data har inte på samma automatiska sätt varit tillgängliga, men ett försök att åstadkomma detta har gjorts i en modul som kallas ASI-Netanalys. När data från ASI ska aggregeras uppstår olika problem och de överväganden

som legat till grund för vårt sätt att arbeta med aggregerade data i ASI-Netanalys redovisas i denna rapport.

Sammanfattningsmått

För varje område finns olika sammanfattande mått: klientskattning, intervjuarskattning och Composite Score, CS.

Klientskattning

Klienten gör två skattningar för varje område: skattning av oro/besvär och skattning av hur viktigt det är att få hjälp på följande skala:

- 0 = Inget problem eller behov av hjälp
- 1 = Litet problem eller behov av hjälp
- 2 = Måttligt problem eller behov av hjälp
- 3 = Påtagligt problem eller behov av hjälp
- 4 = Mycket stort problem eller behov av hjälp

Intervjuarskattning

Intervjuarskattning görs för varje område, där intervjuaren gör en bedömning efter manual-givna kriterier på en skala mellan 0 och 9 enligt följande:

- 0-1 = Inget problem/Hjälp krävs inte
- 2-3 = Litet problem/Hjälp krävs troligen inte
- 4-5 = Måttligt problem/Viss hjälp krävs
- 6-7 = Påtagligt problem/Hjälp krävs
- 8-9 = Mycket stora problem/Hjälp krävs absolut

För varje område finns ett antal kritiska frågor som ska utgöra underlag för intervjuarens skattning. De kritiska frågorna i grundintervjun är:

OMRÅDE/FRÅGA	BESKRIVNING
Fysisk hälsa	
1	Sjukhusvistelse under livstiden
3	Långvariga allvarliga problem
Arbete/Försörjning	
1, 2, 3	Utbildning
5	Längsta regelbundna arbete
8	Vanligaste anställningsmönster senaste 3 åren
Alkohol/Narkotika	
1-13	Missbruksbakgrund
15	Delirium tremens (DT) och överdoser av narkotika
16	Erhållen behandling
19, 20	Frivillig missbruksfrihet
Kriminalitet/Asocialitet	
3-6	Åtal
7	Fällande domar
14, 15	Förestående åtal
17	Pågående olaglig aktivitet
Familj/Umgänge	
2, 3	Stabilitet/Tillfredställelse – Äktenskaplig
5, 6	Stabilitet/Tillfredställelse – Vardaglig
8	Tillfredställelse under fritid
10-18	Problem med släktingar, partner, vänner mm under livstiden
19A+B	Allvarliga konflikter
Psykisk hälsa	
1	Sjukhusvistelse under livstiden
3-10	Aktuella symtom och symtom under livstiden

I manualerna (både 1998 och 2006) beskrivs utförligt hur intervjuarskattningen ska gå till. I det första steget koncentrerar sig intervjuaren på objektiva data inom problemområdet, främst de kritiska frågorna för att fastställa ett intervall på 2-3 skalsteg på skalan 0-9. I andra steget beskrivs hur intervjuaren ska justera sin skattning utifrån klientens skattningar av problem och hjälpbehov. Justeringen sker mot det högre eller lägre värdet inom det valda intervallet och mot klientens högsta värde på skattning av antingen oro/besvär eller av hjälpbehov.

Composite score, CS

I de senaste amerikanska versionerna av ASI är intervjuarskattningen ett tillägg som rekommenderas för kliniskt bruk, men inte för utvärdering. Utvärdering bör enligt deras mening ske med hjälp av s.k. Composite Scores, CS. CS är ett matematiskt beräknat index för varje område baserat på frågor i ASI. Man menar att CS är mer känsligt för förändring och kanske också mer objektivt i den meningen att det är en summering av flera centrala variabler inom varje område. Man menar också att en förutsättning för att intervjuarskattningarna ska vara reliabla och valida som sammanfattande mått är att intervjuarna är utbildade, vilket inte alltid är fallet i USA. Man pekar också på att uppföljningsintervjuer i USA ofta sker via telefon,

vilket försvårar användning av intervjuarskattningar vid uppföljning. Det har framförts en hel del kritik mot CS tex att klientens subjektiva skattningar av oro och hjälpbehov alltid ingår i CS, ibland som de enda variablerna för tex fysisk hälsa, vilket betyder att man kan ställa sig tveksam till objektiviteten i CS. Annan kritik som framförts av CS, är bl.a. att skalorna, trots att de skall variera mellan 0 och 1, inte är jämförbara för olika områden. Ett tredje problem är att CS är kliniskt svårtolkat.

Här koncentrerar vi oss på några viktiga sammanfattande mått som kan ligga till grund för att ge en bild av klientgruppens problem och hur problemen förändras över tid. Vi har i våra analyser av aggregerade data använt intervjuarskattningar som sammanfattningsmått för klientens problem. Det finns flera skäl för detta. Vi hade tillgång till en stor databas med ca 14000 grundintervjuer och 4000 uppföljningsintervjuer (Armelius et al 2009). Alla intervjuer var gjorda ute på fältet av tränade intervjuare i olika missbruksenheter. Data var alltså representativa för hur ASI används ute på fältet och våra resultat, analysmetoder och överväganden bör därför kunna generaliseras till klinisk verksamhet. Ett avgörande skäl för val av intervjuarskattning som huvudsakligt sammanfattande mått är att de är meningsfulla kliniska globala bedömningar, medan CS är mycket svåra att förstå och jämföra. Intervjuarskattningarna har i en svensk studie visat sig ha god reliabilitet (Nyström et al, 2010) och rimlig variation, dvs differentierar klienterna på ett konsistent sätt. Eftersom intervjuaren enligt manualen ska ta hänsyn till klientens skattningar av oro och behov av hjälp korrelerar intervjuarskattningarna högt med klientens skattningar (0.70-0.80). Intervjuarskattningarna korrelerar också högt med Composite Score, CS. Det finns också en intern konsistens mellan svaren på enskilda frågor och intervjuarskattningen inom varje område. Detta tyder på att intervjuarskattningen är ett användbart mått för att beskriva klienternas problem. Intervjuarskattningarna har dock trots sina fördelar även en del problem som vi skall beröra härnäst.

Olikheter i instruktioner för skattningar mellan version 1998 och 2006

En hel del förbättringar gjordes från version 1998 till version 2006, bl.a. lade man till frågor om hjälp eller behandling inom varje sektion. Även andra förändringar gjordes vilket man måste tänka på om man skall slå samman data från de två versionerna. Det mesta går att lösa med omkodningar, men det är två centrala variabler som fått en annan definition och därmed kanske ändrat innebörd. Både klientskattningen av hjälpbehov och intervjuarskattningen förändrades från att 1998 handla om hur viktigt det är att nu få hjälp med problemen inom varje område till att 2006 handla om hur viktigt det är att få hjälp med problemen *utöver pågående hjälp*. Redan i den ursprungliga manualen (1998) fanns dock tanken att intervjuarskattningen skulle definieras som ytterligare hjälpbehov. Detta beskrevs som:

”I ASI definieras svårighetsgrad som *hur stort behov av hjälp/behandling som föreligger där hjälp/behandling inte föreligger för närvarande – eller behov av ytterligare en hjälp eller behandlingsform* förutom den som klienten har för närvarande. Denna hjälp eller behandling noteras under Kommentarer inom respektive problemområde. Bedömningarna skall bygga på uppgifter om grad, längd och intensitet hos problemen inom ett problemområde.”

Att skattningen endast skall omfatta *ytterligare* behov av hjälp skrevs dock inte in i formuläret från 1998, vare sig för klientens eller intervjuarens skattning av hjälpbehov. Där definieras intervjuarskattningen genomgående som intervjuarens bedömning av hur stort behov av hjälp

klienten har inom området. Ovanstående definition av svårighetsgrad är också omdiskuterad och inte allmänt accepterad i USA. Att definitionen att skatta ytterligare hjälpbehov inte heller verkar ha fungerat i Sverige framgår av att det nästan aldrig anges i kommentarer vilken ytterligare hjälp som avses när intervjuaren skattat att det finns behov av ytterligare hjälp. Kravet att bedöma och notera vilken ytterligare hjälp som behövs för en klient innebär dessutom att intervjuaren har tillgång till uppgifter om vilka olika behandlingar som finns för problem inom de olika områdena samt hur bra dessa behandlingar fungerar, dvs uppgifter som man inte får i ASI-intervjun. Sammanfattningsvis kan man dra slutsatsen att både klient och intervjuare med 1998 års version av ASI skattar aktuellt behov av hjälp och inte behov av ytterligare hjälp, vilket också är enligt anvisningar i manualen och enligt 1998 års formulär, men inte i enlighet med definitionen av svårighetsgrad.

I version 2006 har man i formuläret både för klienten och för intervjuaren betonat att skattningen av hjälpbehov ska gälla behov av ytterligare hjälp. För varje område finns i version 2006 en fråga om pågående behandling, medan fältet för Kommentarer där det ytterligare hjälpbehovet skall noteras är borttaget. I manualen från 2009 för hur intervjuarskattningen skall göras med 2006 formulär framgår som i manualen 1998 att problemens grad, längd och intensitet skall ligga till grund för skattningen. Men i bedömningsformuläret definieras intervjuarskattningen som en uppskattning av ytterligare hjälpbehov för varje område och inte av svårighetsgrad. Frågan gäller alltså hur man går från problemens grad, längd och varaktighet till klientens behov av ytterligare hjälp. Definitionen av skalstegen för intervjuarskattningen i ASI ger viss vägledning då de utgår från att problemens grad, varaktighet och intensitet är direkt relaterade till klientens hjälpbehov; stora problem och stora hjälpbehov har samma skattning, liksom små problem och ett litet eller inget hjälpbehov. Eftersom detta inte alltid är fallet krävs en justering av intervjuarens första uppskattning av svårighetsgraden utifrån klientens skattning av oro/besvär och hjälpbehov. Detta görs genom tvåstegsmodellen för intervjuarskattning som beskrevs ovan, där problemnivån justeras något upp eller ner beroende på klientens högsta skattning av oro/besvär eller av hjälpbehov. Dessa två steg var tillräckliga för intervjuarskattningen enligt 1998 års version och man behövde inte blanda in kravet att bedöma ytterligare hjälpbehov. I version 2006 av formulären, när hjälpbehov explicit definieras som *ytterligare* behov av hjälp, uppstår dock problem med integrering av problem och hjälpbehov som nu kräver ett led till, nämligen en bedömning av om ytterligare hjälp krävs. I manualen från 2009 beskrivs hur man väger ihop problem/behov av hjälp om klienten *inte* har någon befintlig hjälp, men hur man gör med klienter som redan har hjälp finns inga exempel på eller anvisningar för. Hur *ytterligare* hjälp skall påverka intervjuarens skattning definieras alltså inte trots att det är en avgörande skillnad mot den tidigare manualen och de tidigare formulären.

Låt oss illustrera problemen med ett exempel. Anta att man i första steget kommer fram till att en klient har problem i intervallet 3-5. I steg 2 skall intervjuaren justera sin skattning i intervallet 3-4-5 utifrån klientens högsta skattning av oro/besvär *eller* av hjälpbehov. Klienten är mycket orolig och besvärad och skattar 4 på oro/besvär, men han har mycket pågående hjälp och skattar 0 för ytterligare hjälpbehov. Enligt manualen borde justeringen leda till en skattning av 5, då oro/besvär skattats högt. Går man däremot strikt på begreppet ytterligare hjälpbehov enligt formuläret skulle skattningen vara 0. Det blir med andra ord omöjligt att integre-

ra svårighetsgrad och hjälpbehov i detta fall om hjälpbehov definieras som ytterligare hjälpbehov och manualen ger heller inga exempel på hur man ska göra en sådan intergrering.

Problemet med intervjuarskattningens definition uppstår förmodligen därför att upphovsmännen ville skatta tre olika saker med en enda skattning 1) skatta av klientens problemnivå, 2) skatta klientens hjälpbehov och 3) skatta klientens behov av ytterligare hjälp. ASI betyder Addiction Severity Index, alltså ett index på svårighetsgraden av missbruket. I många avseende är det vad som görs i första steget: att fastställa nivån på klientens problem med hjälp av all ”objektiv” information som samlats in under varje sektion och med särskild hänsyn till de kritiska frågorna. Denna uppskattning är ännu inte påverkad av klientens skattningar av oro och hjälpbehov utan bygger endast på klientens uppgifter om sina problem. I den meningen är det en renodlad uppskattning av problemens svårighetsgrad. Men man ville också få med behov av hjälp och av ytterligare behov av hjälp som båda skulle påverka intervjuarskattningen, dvs. man ville uttrycka tre skattningar i en enda, något som försvårar för både klienter och intervjuare. En möjlig konsekvens av oklarheterna kring definitionen av intervjuarskattningen är att man får en osäkerhet när man jämför intervjuarskattningar gjorda av olika intervjuare, dvs. när man slår ihop och jämför data på gruppnivå. Detta blir särskilt aktuellt vid uppföljning, där en stor andel av klienterna kan förväntas befinna sig i behandling och behov av ytterligare hjälp borde vara lägre. Det blir då oklart om såväl klientskattningen av hjälpbehov som intervjuarskattningen endast uttrycker behov av ytterligare behandling, som det står i formuläret, eller om man följt manualen och väger ihop problemnivå och hjälpbehov enligt 1998 års modell för intervjuarskattningar.

Empirisk undersökning av skillnader mellan version 1998 och 2006.

Klient- och intervjuarskattningar av hjälpbehov i version 1998 och 2006

En fråga är om och hur definitionen av hjälpbehov i version 2006 påverkat intervjuarens skattningar. En möjlighet är att man tar instruktionen om att endast beakta ytterligare hjälpbehov ad notam och skattar 0 om inga ytterligare hjälpbehov finns även för en klient med stora problem och mycket pågående hjälp. En annan möjlighet är att man fortsatt att väga ihop all information enligt den ursprungliga modellen från 1998. Vi har analyserat detta på lite olika sätt i databasen med 13903 klienter, där 10694 har använt 1998 års version och 3209 har använt 2006 års version av grundintervjun. Vid uppföljningsintervjun hade 2653 använt 1998 års version och 296 hade använt 2006 års version. Sammanlagt finns alltså ett relativt stort antal intervjuer gjorda enligt olika förutsättningar för att vi skall få en uppfattning om hur skattningarna ser ut.

En utgångspunkt för de följande resonemangen är att klienternas problemnivåer verkar vara ungefär desamma under åren 2003-2008 då data samlades in (se Armelius & Armelius, 2011). Om man dessutom antar att klienterna har pågående hjälp i ungefär samma omfattning (den är över år snarast något ökande) vid grundintervjun över så borde klientens skattning av hjälp vara lägre i version 2006 jämfört med version 1998 om klienten skattat enligt formulärens definitioner. I tabellen nedan redovisas resultaten av klientskattning för behov av hjälp i version 1998, där klienten skattat behov av hjälp nu, och 2006, där klienten skattat behov av ytterligare hjälp, för de olika områdena.

Tabell 1. Klientskattningar behov av hjälp (version 1998) och behov av ytterligare hjälp (version 2006). Medelvärden.

	Klient N=10694	Klient N=3209
	1998	2006
Fysisk hälsa	1,3	1,3
Arbete försörjning	1,5	1,4
Alkohol	1,9	1,8
Narkotika	1,4	1,2
Kriminalitet	0,5	0,6
Familj umgänge	1,0	1,2
Psykisk hälsa	2,0	1,9

Det är ingen skillnad mellan klientens skattning av hjälpbehov i de två versionerna av grundintervjun, vilket talar för att klienterna skattat på samma sätt oberoende av vilket formulär som använts. I tabell 2 nedan visas hur de genomsnittliga intervjuarskattningarna ser ut för de två versionerna av formulär.

Tabell 2. Intervjuarskattningar av behov av hjälp (version 1998) och behov av ytterligare hjälp (version 2006). Medelvärden.

	Intervjuare N=10694	Intervjuare N=3209
	1998	2006
Fysisk hälsa	2,0	1,9
Arbete försörjning	2,7	2,5
Alkohol	3,8	3,9
Narkotika	2,9	2,8
Kriminalitet	0,5	0,6
Familj umgänge	2,5	2,6
Psykisk hälsa	3,5	3,4

Det är mycket marginella skillnader mellan intervjuarskattningar i version 1998 och i version 2006 även för intervjuarna. Den nya definitionen av intervjuarskattningen i version 2006 har inte påverkat intervjuarskattningen, vilket talar för att skattningarna inte uttrycker ytterligare hjälpbehov utan problemlnivå integrerat med hjälpbehov över huvud taget enligt 1998 års version.

Skillnad mellan klienter som har och inte har hjälp i version 2006 vid grundintervju och uppföljning

Ett annat sätt att undersöka hur bedömning av ytterligare hjälpbehov utöver pågående hjälp påverkar intervjuarskattningen för olika områden är att jämföra skattningen för de klienter som har hjälp med de som inte har någon hjälp. För klienter som har pågående hjälp borde

intervjuarskattningen av behov av ytterligare hjälp vara lägre jämfört med skattningen för de klienter som inte har hjälp. I version 2006 frågas efter pågående hjälp för alla områden. Intervjuarskattningar för klienter som har hjälp och för klienter som inte har pågående hjälp för olika områden visas i tabellen nedan.

Tabell 3. Intervjuarskattningar för klienter som har hjälp och för klienter som inte har pågående hjälp vid grundintervjun och vid uppföljningsintervjun för olika områden. Procentalen anger hur stor andel som har pågående hjälp vid intervjutillfället. Endast version 2006.

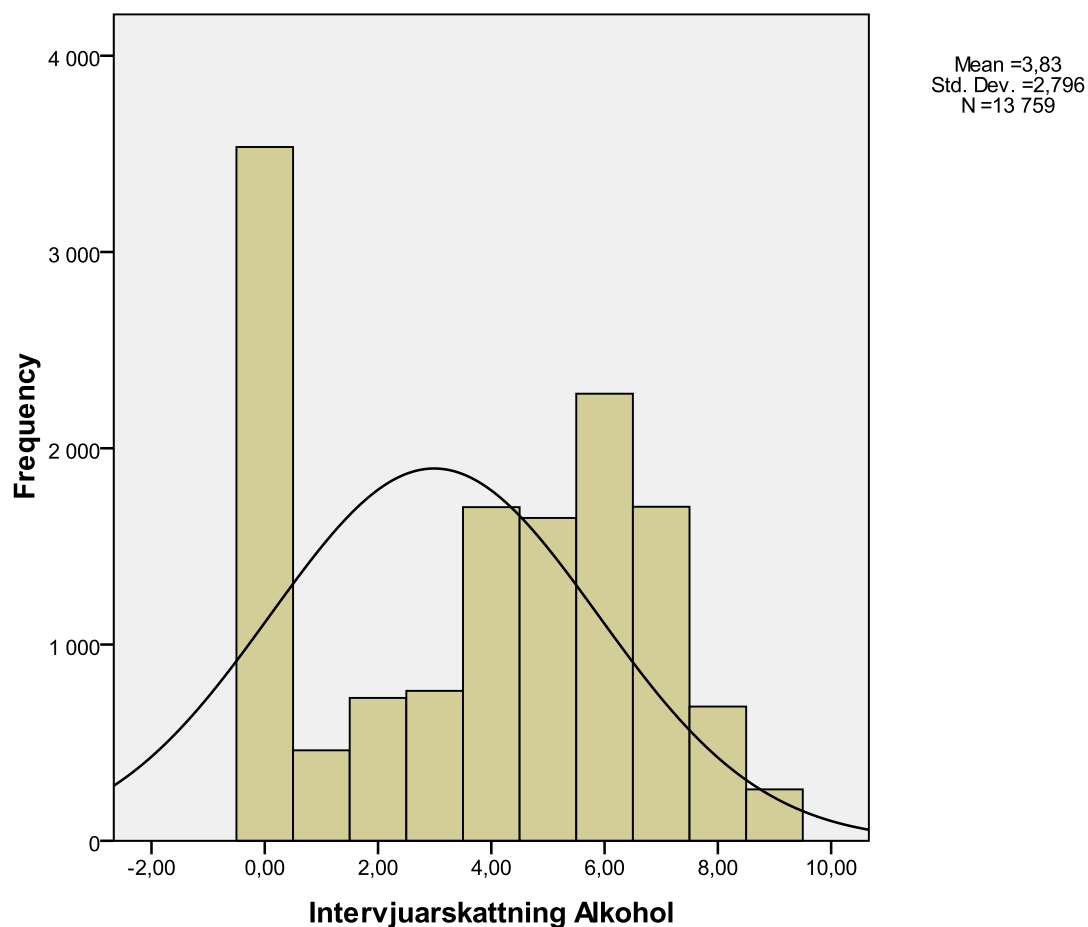
	Grund N=3209			Uppföljning N=296		
	Hjälp pågår		Hjälp pågår ej	Hjälp pågår		Hjälp pågår ej
	%	M	M	%	M	M
Fysisk hälsa	26	2,9	1,6	36	1,2	0,7
Arbete försörjning	18	3,5	2,2	24	1,6	0,9
Alkohol*	25	4,1	3,8	47	2,3	2
Narkotika*	25	2,7	2,5	47	1,1	1
Kriminalitet	8	3,2	1	8	1,1	0,3
Familj umgänge	11	4,1	2,5	10	3,1	1,3
Psykisk hälsa	31	4,4	3	20	3,3	1,7

*i version 2006 frågas efter pågående hjälp för alkoholproblem eller för narkotikaproblem vilket gör att det inte går att skilja dessa åt och andel med pågående hjälp överskattas för båda dessa områden

Både vid grundintervjun och vid uppföljningen är intervjuarskattningarna högre för samtliga områden om klienterna har pågående hjälp jämfört med om de inte har sådan hjälp. Det verkar alltså som om pågående hjälp för klienterna betyder att man har mer problem eller det är klienter med mer problem som får och har mer hjälp. Detta tyder på att intervjuarskattningen är en skattning av problemnivå och medföljande hjälpbehov utan hänsyn till om hjälp pågår eller inte. De empiriska resultaten på aggregerad nivå talar således för att såväl klienter som intervjuare inte har följt anvisningarna i version 2006 om att skatta ytterligare hjälpbehov utan snarare skattat hjälpbehov som helhet enligt 1998 års modell.

Begreppslig oklarhet kring skalsteg

Ett annat problem med intervjuarskattningarna är att skalans begreppsliga definition av skalsteg och antal skalsteg inte är entydigt. Begreppsligt definieras skalans 10 skalsteg i manualen som intervall om två skalsteg, vilket resulterar i en skala om fem steg. Att ha samma begreppsliga definition för två olika skalsteg blir särskilt problematiskt för skalstegen 0 och 1, som definierar avsaknad av problem och hjälpbehov, men där 1 ändå uttrycker något annat än total avsaknad av problem. Nollpunkten blir alltså inte entydigt definierad. Att intervallet 0-1 som har samma definition enligt manualen ändå används som två olika skalsteg framgår av frekvensfördelningarna för de olika områdena som visar att det är en stor skillnad i hur ofta 0 och 1 används. I figur 1 visas hur det ser ut för området alkohol, men det är likadana fördelningar för de andra områdena.



Figur 1. Frekvensfördelning av intervjuarskattning för alkohol vid grundintervjun.

I figuren ser man att 0 verkar användas för avsaknad av problem och 1 för lite problem. Trots den begreppsliga definitionen av 5 olika steg verkar alltså bedömarna använda skalan som om den hade 10 olika skalsteg. Det är också ovanligt att normalpersoner får en skattning som är större än 0 inom de flesta områden. Det betyder att om man inte har några problem skall det vara 0 och inte 1.

Slutsatser och förslag till ändringar av intervjuarskattningen

Det finns all anledning att tro att intervjuarna använder intervjuarskattningen som ett globalt sammanfattningsmått för klientens problemliv och hjälpbehov med utgångspunkten att problem och hjälpbehov hänger ihop. Enligt vår uppfattning var det en av utgångspunkterna för hela ASI som dock har komplicerats genom åren och det är olyckligt att man i den sista svenska versionen lyfter fram den kontroversiella definitionen som säger att svårighetsgrad är lika med behovet av ytterligare hjälp. Som väl är verkar det inte ha haft någon betydelse för de empiriska resultaten.

Det skulle förenkla användningen av ASI om man behåller en 10-gradig skala, men med en definition av varje enskilt skalsteg och där 0=inga problem. Man borde komplettera med korta

fallbeskrivningar för varje skalsteg som uttrycker problem. I framtida revision bör man också ta bort instruktionen att basera skattningen på ytterligare hjälpbehov. Man kanske också bör överväga att dela upp intervjuarskattningen i två tiogradiga delskalor: en för problemnivå och en för hjälpbehov. Man får då två relativt enkla kliniska bedömningar i stället för en komplicerad sammanvägning. När ytterligare hjälpbehov föreligger skall det beskrivas i en kommentar till varje intervjuarskattning.

Parametrisk och/eller icke-parametrisk statistik

Vi genomgående använt medelvärden och standardavvikelser för att beskriva klientgruppen och olika undergrupper med hjälp av klientenskattningar och intervjuarskattningar. Vi är väl medvetna om att särskilt intervjuarskattningen men också många andra variabler i ASI inte är normalfördelade och att det därför kan vara motiverat att använda median och kvartilmått snarare än medelvärden och standardavvikelser. Så gott som alla områden har en stor frekvens av intervjuarskattningen 0, helt enkelt för att en stor andel av klienterna inte har problem inom alla områden. Fördelningen för övriga skalvärden är dock relativt normal, vilket betyder att skillnader mellan klienter som har problem kan sägas vara hyfsat ekvidistanta på en diskret skala. Man kan tänka sig att analysera 0 för sig och 1-9 för sig för att lösa problemet. Ett annat sätt är att analysera hela skalan och se vilken konsekvens det får om man väljer att använda medelvärden (M) eller medianer (Md). För figur 1 ovan blir $M = 3.83$ och $Md = 4.0$. Skillnaden är inte så stor, men för t.ex. narkotika blir det en stor skillnad $M = 2.81$ och $Md = 0$, därför att mer än 50 % av klienterna inte har problem med narkotika och medianen blir då 0. Att använda medianvärden blir därför ett problem i synnerhet vid uppföljningen, där alla områden förutom problem med alkohol och psykisk hälsa har $Md = 0$. Skillnaden mellan medianvärden ger alltså väldigt lite och onyanserad information om förändringar. Medelvärdena är å andra sidan lite lägre än de skulle vara om vi plockat bort nollorna. Vi har ändå valt att använda M baserat på hela skalan för beskrivningar och alla signifikanstest har gjorts med både parametrisk och icke-parametrisk statistik. Skillnaden i resultat för dessa test är mycket liten, men genomgående blir det lite fler signifikanta resultat med parametriska test.

Vi har också hanterat problemen med nollor på ett annat sätt genom att göra en klusteranalys av resultaten. Vi tror också att klusteranalysens problemprofiler är ett sätt att hantera många av problemen på ett bättre sätt därför att man ser till klientens problem på alla områden på en gång och inte för en variabel i taget. Enskilda nollor får då en annan betydelse.

Dubbeldiagnos - Definition av missbrukare med sannolik svår psykisk störning genom ROC-analys

Vi har undersökt om man skulle kunna använda ASI-data för att screena klienter för svår psykisk störning som skulle kunna användas för att gå vidare med klientens psykiska problem. Detta har gjorts i en amerikansk studie (Cacciola, et al 2008) där man undersökt möjligheten att identifiera missbruksklienter med samtidig psykiatrisk diagnos (s.k. dubbeldiagnos, DD). I studien ingick en grupp om 672 klienter i missbruksvård. Enligt resultaten av SCID-III-R hade 55% av deltagarna ingen samtidig psykiatrisk diagnos, medan de återstående 45% hade en eller fler Axial I sjukdomar. Mer specifikt hade 28% en sjukdom, 12% två sjukdomar och 6% tre eller fler sjukdomar förutom sin missbruksdiagnos. De vanligaste diagnoserna var Major Depressive Disorder (28%), Posttraumatic Stress Disorder (12%), Dysthymia (7%), Social

Phobia (5%), Obsessive Compulsive Disorder (5%), Simple Phobias (4%), och Panic Disorder (4%). Färre än 2% av deltagarna fick andra diagnoser.

Receiver Operating Characteristics (ROC; Metz, 1978) analyser användes för att bestämma cut-off värden på bl.a. intervjuarskattningen av psykiska problem i ASI för att separera missbruksklienter med DD från missbruksklienter som inte hade DD. ROC analyser tillåter val av cut-off värden för beslut om diagnoser där man vet utfallet (Swets, et al. 2000). ROC kurvor genereras genom att man väljer ut en testvariabel (i detta fall intervjuarskattningen av psykiska problem). Man beräknar sedan sensitivitet (= andel korrekt identifierade DD) och specificitet (= andel korrekt identifierade ej-DD) för varje värde på testvariabeln för att predicera en tillståndsvariabel (i Cacciolas studie; om missbrukarna hade eller inte hade en samtidig Axel I diagnos). Det ”bästa” cut-off värdet skapar den bästa kvoten mellan sanna och falska DD och varierar för olika beslutssituationer (Swets 1992; Swets et al. 2000). Cut-off värdet för intervjuarskattningen i ASI bestämdes så att man fick högsta möjliga sensitivitet och samtidigt behöll en specificitet på åtminstone 50%. Med en cut-off på 3 eller mer blev sensitiviteten i den amerikanska studien 81 % och specificiteten 53 %, vilket betyder att man med en cut-off på 3 eller mer på intervjuarskattningen av psykiska problem i ASI kunde identifiera 81 % av de klienter som faktiskt hade en samtidig psykiatrisk diagnos. Specificiteten på 53% betyder att man korrekt kunde eliminera 53% av de som faktiskt inte hade en samtidig DD, men får kvar 47% som felaktigt identifieras som att de har en psykiatrisk diagnos.

Genom samarbete med Agneta Öjehagen har vi fått tillgång till ASI-data för en grupp missbruksklienter som också har en psykiatrisk diagnos enligt DSM-III-R (Schaar & Öjehagen, 2001). I anslutning till psykiatriutredningen genomfördes en multi-center studie med syfte att förbättra samarbetet mellan socialtjänst och psykiatri kring patienter med både svår psykisk störning och missbruk. Projektet pågick i olika delar av Sverige mellan 1995 och 1998. I studien ingick missbruksklienter med allvarliga psykiska störningar och 358 patienter accepterade medverkan. Svår psykisk störning definierades som diagnoserna schizofreni, andra psykoser, schizotypal personlighetsstörning eller motsvarande svårighetsgrad. Missbruksdiagnoserna var alkohol, illegala droger och/eller lugnande. Medelåldern var 40 år och det var 235 män och 123 kvinnor fördelade på fyra diagnosgrupper: Psykos (n=103) innehöll 69 med schizofreni och 34 med andra psykoser, Depression (n=61) innehöll 40 personer med unipolär sjukdom och 21 med bipolär sjukdom, Svår personlighetsstörning (n=83) innehöll 72 personer med borderline PD och 11 personer med schizotypal PD medan den sista gruppen ”lika svår störning” (n=111) innehöll andra Axel I diagnoser av liknande svårighetsgrad, mestadels ångest, depression och dysthymi. I alla grupper fanns många som också uppfyllde kriterier för personlighetsstörningar på Axel II.

Vi kan använda Öjehagens grupp för att få en uppskattning av sensitiviteten (andelen korrekt identifierade DD) för olika cut-off värden på intervjuarskattningen för psykiska problem för missbruksklienter med svår psykisk störning genom att se hur de 358 klienterna fördelar sig över de olika skalstegen i intervjuarskattningen. Hela gruppen i Öjehagens studie hade ett medelvärde på 6,5 på intervjuarskattningen. Andelen klienter som ligger över skalstegen på intervjuarskattningen visas i tabell 4 .

Tabell 4. Sensitivitet =andelen i % över cut-off för klienter i Öjehagens studie.

Cut-off större än	Andel över cut-off (N=358)
2	0,99
3	0,97
4	0,95
5	0,87
6	0,77
7	0,57
8	0,30

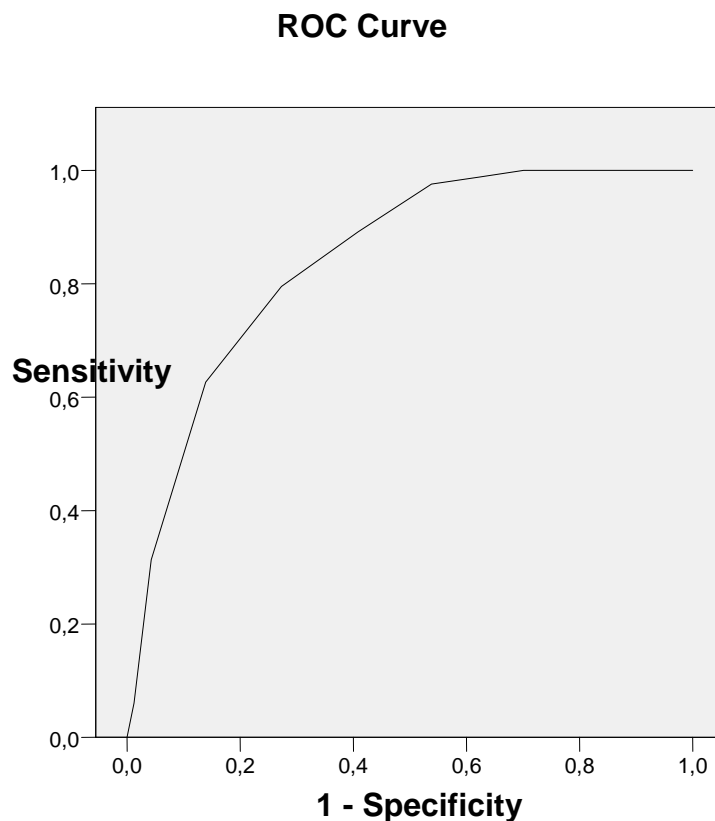
Av tabellen framgår att en cut-off på 3 eller mer ger en betydligt högre sensitivitet än i Cacciolas studie (0,97 vs 0,81), Detta förklaras av att fler klienter i Öjehagens studie hade svår psykisk störning än i Cacciolas studie. Därför hamnar nästan alla i Öjehagens studie över en cut-off på 3. I Öjehagens grupp identifieras en större andel korrekt som DD även med högre cut-off värden och med en cut-off på 5 eller mer får man en utmärkt sensitivitet på 0,87.

Eftersom det i Öjehagens studie inte finns några klienter med känd icke DD kan man inte beräkna någon specificitet. Man kan emellertid använda data från ASI-08 och Öjehagens studie för att utveckla rimliga cut-off värden för att identifiera klienter i ASI-08 som har och inte har DD. Eftersom vi inte vet hur många i ASI-08 som har och inte har en DD måste vi arbeta med rimliga jämförelser och uppskattningar. Enligt tidigare studier med ASI och psykiatriska störningar (Cacciola et al, 2008, Sallmen, 1999) varierar andelen som har DD mellan ca 50% och 80%. Uppskattningen varierar beroende på vilken grupp av missbrukare och vilken typ av psykiatriska störningar det rör sig om. I Cacciolas studie var det huvudsakligen lättare psykiatriska störningar på Axel I. I Sallmens studie var det alkoholister i LVM-vård, där andelen med samtidig psykiatrisk diagnos var 78 %. Låt oss på goda grunder anta att 50% av ASI-08s klienter skulle kunna få/har en psykiatrisk diagnos av något slag, och att betydligt färre skulle kunna få/har en allvarlig psykiatrisk diagnos motsvarande den grupp som finns i Öjehagens studie. I tabell 5 visas hur klienterna i ASI-08 fördelar sig över skalstegen på intervjuarskattningen för psykiska problem. Vi har använt beteckningen ”Sannolik DD” för att markera den andel som skulle betraktas som DD för varje cut-off värde och beteckningen ”Osannolik DD” för att markera den andel som korrekt skulle komma att identifieras som ej-DD.

Tabell 5. Andel klienter i ASI-08 som ligger över (Sannolik DD) och under (Osannolik DD) olika skalsteg på intervjuarskattningen av psykiska problem.

Cut off större än	Andel Sannolik DD	Andel Osannolik DD
2	0,70	0,30
3	0,62	0,38
4	0,54	0,46
5	0,41	0,59
6	0,27	0,73
7	0,14	0,86
8	0,04	0,96

Av tabellen ovan framgår att med ett cut-off värde på 3 blir andelen i ASI-08 som skulle bedöms höra till gruppen DD 62%, medan 38% skulle bedömas som ej DD. Vid jämförelse med Cacciolas studie, där specificiteten var 53%, framstår 38% som rimligt värde på specificiteten eftersom ASI-08 inte utgör en väldefinierad jämförelsegrupp som diagnosticerats och med säkerhet inte har en psykiatrisk diagnos. Tvärtom räknar vi ju med att ca 50% i ASI-08 skulle kunna få/har en psykiatrisk diagnos. ASI-08 innehåller således båda utfallen i tillståndsvariabeln, vilket en jämförelsegrupp i en ROC analys inte får göra. Vi kan därför anta att specificiteten underskattas och förmodligen skulle hamna närmare 50% om vi kunde sortera bort de som inte skulle få/ har en psykiatrisk diagnos från ASI-08. ROC-kurvan visas i figur 2 och Area Under Curve (AUC) är 0.8.



Figur 2. ROC-kurva.

Vi har med utgångspunkt i dessa beräkningar använt som kriterium för sannolik psykisk störning att intervjuarskattningen är 5 eller mer. Vi har också lagt till kriteriet att man någon gång haft kontakt med psykiatrin, vilket kan antas öka specificiteten.

Beräkning av förändring. Statistiska problem.

Tak- och golveffekter. Ett problem med skalor som mäter problem är s.k. tak- och golveffekter. Det betyder att klienter som fått höga skattningar från början har ett stort utrymme att förändras nedåt medan det är tvärt om för klienter som fått låga skattningar initialt. Enkelt uttryckt: har man initialt inga problem kan man inte få mindre problem och har man initialt maximala problem kan man inte få mer problem. Detta bidrar till att klienter med mycket problem tenderar att få lägre värden på intervjuarskattningarna för olika problemområden, dvs förbättras mer än klienter med låga problemnivåer, som ju inte kan bli så mycket bättre vid en uppföljning. Dessa båda felkällor påverkar beräkningar av förändringar mellan grundintervju och uppföljningsintervju. Båda drar åt samma håll, så att klienter med höga problemnivåer tenderar att ligga kvar eller få lägre problemnivå, dvs. förbättras, medan klienter med låga problemnivåer tenderar att ligga kvar eller få högre värden, dvs. försämras, vid uppföljningen. Detta är något man måste ta hänsyn till när man tolkar förändring mellan grundintervju och uppföljningsintervju såväl för aggregerade data som för enskilda klienter.

Mätfel. Skillnader i medelvärden mellan grundintervju och uppföljningsintervju som förändringsmått har ett antal mät- och statistiska problem som försvårar tolkningar av resultaten. Ett sådant problem är att varje mätning uttrycker både en faktisk förändring och ett mätfel. En klient kan därför få olika värden på grundintervjun och uppföljningsintervjun beroende på mätfel och inte beroende på en faktisk förändring. Hur mycket som kan antas vara ett sant värde i en enskild mätning uttrycks med en reliabilitetskoefficient som också kan användas för att göra beräkningar av hur stort osäkerhetsområdet är kring ett mätvärde. För ASI finns en svensk reliabilitetsstudie (Nyström et al 2010) som anger reliabiliteten för intervjuarskattningarna till ca 0,85. Det betyder att mätfelet i ASI är ca 15%. Mätfelen blir särskilt besvärliga när man gör upprepade mätningar, eftersom det rör sig om två mätningar som båda innehåller mätfel. Detta bidrar till att extrema värden blir mindre extrema, något som brukar beskrivas som regression mot medelvärdet. I vårt fall bidrar mätfelet till att klienter med höga problemnivåer kommer att få lite lägre skattningar vid uppföljningsintervjuer, medan det är tvärt om för klienter med låga problemnivåer. Begreppet reliabel förändring tar hänsyn till osäkerheten kring mätvärdena och med ASI kan en förändring på 1 eller 2 steg på intervjuarskattningen för en enskild klient betraktas som reliabel beroende på hur säker man vill vara på att förändringen är ”sann” eller pålitlig. I detta arbete har vi valt att använda 95% säkerhet vilket betyder att förändringen måste vara minst 2 skalsteg på intervjuarskattningen för att betraktas som ”sann” (se Armelius et al., 2009 för mer detaljer). En konsekvens av detta är naturligtvis att klienter som har intervjuarskattningar på 0 och 1 inte kan uppnå en reliabel förbättring. Fördelen med analyser av reliabel förändring är att man får en kontroll över reliabilitetsbrister för de klienter som uppvisar förbättringar eller försämringar.

Konsekvenser för utfallsmåtten.

När det gäller utfallsmåtten medelvärdesförändringar och reliabel förändring, så bidrar takeffekter och mätfel till att värdena vid uppföljningen blir lägre, dvs lägre problemnivåer, medan golfeffekter bidrar till att de blir högre, dvs högre problemnivåer. För dessa båda utfallsmått måste man alltså räkna med att statistiska artefakter i varierande grad bidrar till bilden av förändring. Eftersom det är av stort intresse att se hur det går för klienter som har höga problemnivåer vid grundintervjun är det takeffekter och mätfel som är verksamma och deras effekt påverkar i samma riktning och ser ut som förbättringar av klienterna. Vi menar att man bör vara försiktig med tolkningar av förbättringar för de klienter eller grupper som har de högsta initiala problemnivåerna och har därför valt att också se till förändringar i de lägsta problemnivåerna, dvs. andelen som är problemfria. Här har vi mätproblem och golfeffekter att ta hänsyn till, men små eller inga takeffekter. Det betyder att såväl golfeffekten som mätfelen bidrar till att vi kan förvänta oss en mindre andel problemfria vid uppföljningsintervjun än vid grundintervjun, något som går i motsatt riktning till effekten av takeffekter och av mätfel som ger större andel förbättrade. Vi kan med andra ord känna oss mer säkra på att eventuella förbättringar i termer av fler problemfria klienter inte är uttryck för statistiska artefakter.

Sammanfattningsvis menar vi att man kan få en uppfattning om tillförlitligheten i analysen av klienternas förändring genom att se hur de tre olika sätt att beräkna utfall sammanfaller. Om bilden av förändring blir densamma oavsett vilket av de tre sätten som används så är sannolikheten att resultaten uppkommit enbart till följd av statistiska artefakter minimal.

Beräkning av förändring med Composite score

I många sammanhang redovisas förändringar med hjälp av det sammansatta indexet Composite Score (CS). För att underlätta jämförelser med CS redovisar vi några resultat baserade på CS. Först redovisas medelvärden och standardavvikelser vid grundintervju och uppföljningsintervju.

Tabell 6. Medelvärden och standardavvikelser för CS för olika problemområden vid grundintervju och uppföljningsintervju.

	Grundintervju		Uppföljningsintervju	
	Medelvärde	SD	Medelvärde	SD
CS fysisk hälsa	0,33	0,35	0,29	0,34
CS arbete	0,56	0,24	0,54	0,25
CS alkohol	0,25	0,24	0,13	0,19
CS narkotika	0,11	0,13	0,07	0,11
CS kriminalitet	0,11	0,21	0,07	0,16
CS familj umgänge	0,27	0,20	0,20	0,15
CS psykisk hälsa	0,25	0,22	0,21	0,22

Jämförelser mellan förändringar i CS för olika grupper av klienter ger väldigt likartade resultat som jämförelser baserade på intervjuarskattningarna och redovisas inte här då det ändå är svårt att jämföra de numeriska värdena mellan CS och intervjuarskattningar. Ett sätt att ändå jämföra de två olika sätten att räkna utfall är att använda standardiserade medelvärdeskillnader, Cohen's d (Cohen, 1988), eller ES-värden. Detta återges i tabell 7 nedan.

Tabell 7. ES-värden beräknade med Composite scores (CS) och intervjuarskattningar (Isk) för de olika områdena.

	CS ES	Isk ES
Fysisk hälsa	0,12	0,11
Arbete försörjning	0,08	0,23
Alkohol	0,55	0,64
Narkotika	0,33	0,39
Kriminalitet	0,21	0,30
Familj umgänge	0,40	0,46
Psykisk hälsa	0,18	0,49

Intervjuarskattningarna visar något högre ES-värden generellt. Skillnaderna är inte så stora utom för några områden, tex Arbete och försörjning, där CS dock har mycket konstiga egenskaper samt psykisk hälsa, där CS ger betydligt lägre värden än intervjuarskattningarna.

Man kan också få en uppfattning om hur pass lika och olika resultaten blir genom att skapa ett RCI-index för CS-värdena och jämföra med RCI-indexen för intervjuarskattningarna. För detta behövs en uppskattning av reliabiliteten i CS-värdena. Vi beräknade Cronbachs alpha

som baseras på konsistensen i de ingående variablerna. Resultatet för varje område framgår av tabellen nedan tillsammans med uppgifter om reliabilitet från den svenska reliabilitetsstudien (Nyström et al 2011).

Tabell 8. Intern konsistens för composite scores beräknade som Cronbachs alfa ($N=ca$ 14000), samt reliabilitetsdata från den svenska reliabilitetsstudien, $N= 39$.

Område	CS α (ASI-08)	CS α (Nyström)	Iskatt (Nyström)
Fysisk hälsa	0,87	0,87	0,91
Arbete försörjning	0,59	0,84	0,95
Alkohol	0,82	0,74	0,74
Narkotika	0,60	0,64	0,94
Kriminalitet	0,75	0,67	0,85
Familj umgänge	0,67	0,68	0,69
Psykisk hälsa	0,82	0,92	0,86

Med tanke på att det ibland ingår mer än 10 variabler i CS är den interna konsistensen förväntansvärt låg och uppskattningen av reliabiliteten blir i allmänhet något lägre än för intervjuarskattningarna. En förklaring till de relativt höga reliabilitetskoefficienterna för svenska intervjuarskattningar antas av författarna vara att det ställs krav på utbildning av intervjuarna med kontroll av deras bedömningar. Någon förklaring till de jämförelsevis låga värdena på den interna konsistensen i CS i de svenska studierna kan vi inte finna. Eftersom olika studier ger väldigt olika uppskattningar av reliabiliteten i CS har vi i denna studie utgått från samma schablonvärde för CS som för intervjuarskattningarna, dvs 0,85. Man kan också notera att CS-variablerna har J-formade fördelningar med många nollor, men inte samma tendens till approximativ normalfördelning ovanför noll.

I tabellen nedan visas hur stor andel av olika klientgrupper som uppnår en reliabel förbättring beräknat med både CS och intervjuarskattning för några problemområden. Andelen försämrade överstiger nästan aldrig 10% och redovisas inte.

Tabell 9. Andel förbättrade klienter inom några områden beräknat med composite scores (CS) och intervjuarskattning (Isk) för olika uppdelningar av klienter i jämförelsegrupper.

	Alkohol		Narkotika		Familj/umgänge		Psykisk hälsa	
	CS	Isk	CS	Isk	CS	Isk	CS	Isk
Män	29	45	21	31	25	36	25	39
Kvinnor	33	45	23	31	31	43	27	46
Ålder – 33	17	34	36	47	29	43	28	45
Ålder 34-46	31	44	22	33	31	40	28	42
Ålder 47-	39	55	9	16	22	33	23	37
Tungt alkoholmissbruk	71	62	15	23	27	40	32	45
Tungt narkotikamissbruk	20	32	44	54	32	43	29	43

Allvarlig psykisk störning	33	50	28	42	37	50	41	64
Narkotikaproblem	11	8	49	64	30	41	29	46
Alkoholproblem	34	50	3	6	13	18	15	21
Alkohol & psykiska problem	50	64	10	20	42	61	39	61
Genomsnittligt värde	33	44	24	33	29	41	29	44

Det är uppenbart att intervjuarskattningarna generellt sett ger en större andel klienter som förbättrats reliabelt. Skillnaden för alkohol är t.ex. 11% sett över alla jämförelser, men för tungt alkoholmissbruk är det faktiskt en större andel (71%) som är reliabelt förbättrade med CS än med intervjuarskattningen (62%). Mönstret är emellertid detsamma oavsett vilket utfallsmått man använder. Klienter med mer problem och hjälpbehov förbättras mer än klienter med mindre problem och hjälpbehov och förändringen är störst inom de områden man har sitt problem. Slutsatsen av jämförelsen mellan resultat baserade på intervjuarskattningar och CS blir att resultaten med intervjuarskattningarna valideras och inte skulle förändras om man använt CS i stället för intervjuarskattningar. Intervjuarskattningarna ger lite högre uppskattningar av andelen förbättrade klienter än CS.