

Effektivitet ved psykoterapeutisk behandling gjort av viderekomne studenter

Truls Ryum

Tore C. Stiles

Nasjonalt kompetansesenter for sammensatte symptomlidelser, St. Olavs Hospital. Psykologisk institutt, NTNU
egil.a.fors@ntnu.no

Patrick A. Vogel

Det er viktig å dokumentere at behandlingen psykologstudenter gir, er god og forsvarlig. Hvilket utbytte opplever klientene i ordinær praksis?

Takk til alle psykologstudenter ved klinikken, som gjennom sin intern-kliniske praksis har samlet inn dataene som denne artikkelen bygger på. Videre takk til Sølvi Eiternes, Bjørnar Engum, Hans Nordahl, Waldemar Rognes, Kjell Røsdal og Torbjørn Tande, som også har bidratt i dette arbeidet, og som har gitt veiledning til studentene.

Innledning

Psykologutdanningen i Norge startet i sin tid med at det ble opprettet et embetsstudium i psykologi i Oslo i 1959, og senere fikk man tilsvarende studier i Bergen (1974), Tromsø (1990) og Trondheim (1995). Til nå har syv kull og totalt 168 studenter avlagt graden cand.psychol. ved NTNU i Trondheim.

Selv om profesjonsstudiet i psykologi ved NTNU er ment å skulle gi en bred og allmenn fordypning i så vel basalfag som klinisk arbeid, er det i likhet med andre profesjonsstudier i større grad enn de «frie studier» rettet inn mot spesifikk yrkesutøvelse («behandler-/psykologrolle»). Derfor har det også vært et overordnet mål at studentene skal gis en grundig opplæring i utredning, diagnostisering og behandling av ulike psykiske lidelser. Denne siden ved utdanningen har en kvalitetssikret gjennom krav til eksamener i teoretiske emner, samt godkjent praksis fra så vel intern- som eksternpraksis i klinisk rettet arbeid. I forhold til internklinisk praksis (2. avdeling) har alle studenter et semester med klinisk barnepsykologi, et semester med klinisk voksenpsykologi og et semester med klinisk nevropsykologi. Studentene får slik en bred og allmenn erfaring, men fordyper og spesialiserer seg i mindre grad enn ved for eksempel universitetene i Bergen og Oslo. Ved eksternklinisk praksis (seks måneder sammenhengende arbeid) er det imidlertid større mulighet for å velge seg praksis etter interesse og kompetanse.

Når det gjelder internklinisk praksis ved Psykologisk poliklinikk for voksne og eldre, NTNU, meldte det seg tidlig et ønske om at en skulle ha en systematisk innsamling av data vedrørende det terapeutiske arbeidet, både som ledd i den prosess- og effektforskning som foregår ved klinikken, og som hjelp til evaluering og forbedring av gjeldende rutiner. Siden 1999 har derfor alle klienter som har startet i behandling, vært nødt til å fylle ut en standardpakke med skjema både før, under og etter behandlingskontakten. Til å begynne med inngikk poliklinikken som et delsender i det nasjonale Multisenterprosjektet for studier av prosess og utfall i psykoterapi, men siden 2001 har man samlet inn data i en egen database lokalt.

Det er i hvert fall to gode grunner til at en bør undersøke nærmere den behandlingseffekt klienter opplever ved norske universitetsklinikker. Den første, mest generelle og viktigste er å kunne dokumentere at behandlingen studenter gir, er god og forsvarlig. Mye av den publiserte forskningen er her hentet fra for eksempel universitetsklinikker i USA, hvor terapeutene (studentene) som regel har lengre utdanning og er mer spesialiserte (ofte Ph.D.-kandidater) enn tilfellet er i Norge. Både rekruttering til psykologutdanningen, samt lengde og innhold i den, er også forskjellig fra land til land. En kan derfor ikke uten videre konkludere med at resultater fra ett land lar seg overføre til et annet, og det er derfor behov for forskning på egen nasjonal praksis for å kunne dokumentere og sammenligne resultatene internasjonalt.

Den andre grunnen er at det meste av den publiserte behandlingsforskningen i hovedsak er basert på randomiserte og kontrollerte studier, med de krav dette stiller til eksperimentell kontroll. Dette gir oss viktig kunnskap når det gjelder å avklare terapieffekt for en rekke psykiske lidelser, men sier lite om hvilken *effektivitet* behandlingen har (hvilket utbytte klienter opplever) i ordinær poliklinisk praksis (Lambert, Garfield & Bergin, 2004). Nå kan naturligvis «ordinær praksis» her være så mangt, og sannsynligvis er ikke dette heller noen homogen gruppe. Like fullt er det grunn til å tro at i alle fall pasientmassen og behandlingen ved en universitetsklinikk jevnt over ligger nærmere opp mot faktisk praksis fra psykisk helsevern enn ved randomiserte og kontrollerte behandlingsstudier. Derfor kan en også si at denne type forskning ligger nærmest effektivitetsforskningen, som blant annet innebærer større økologisk eller ekstern validitet.

Så vidt vi kjenner til, er det så langt bare publisert funn fra to studier som har undersøkt effekt og tilfredshet hos klienter som har gjennomgått psykoterapeutisk behandling gjort av viderekomne psykologstudenter i Norge. I den ene studien ved universitetet i Bergen konkluderte man med at selv om studentterapiene kom langt svakere ut enn et utvalg av terapier gjort av erfarne terapeuter, mente 14 av 16 klienter at behandlingen hadde hjulpet (Håland, 1986). Det var imidlertid stor forskjell i antall behandlingstimer utvalgene imellom, og de fleste klientene som fikk studentbehandling, opplevde den som for kort. I den andre studien ved universitetet i Oslo rapporterte 24 klienter (75 %) at de hadde hatt effekt av studentbehandlingen, som var inngående dynamisk psykoterapi (Gullestad, 1986). På bakgrunn av disse studiene kunne en konkludere med at studentbehandling hadde effekt, og at den var fagetisk forsvarlig. Utvalgene i begge studiene var imidlertid relativt små, og metoden var i hovedsak basert på semistrukturert intervju og spørreskjema etter avsluttet behandling. De dro dermed ikke nytte av nyere statistisk metodikk, for eksempel vurdering av statistisk signifikans, effektstørrelser eller klinisk signifikans. Resultatene kan derfor også være vanskelige å sammenligne direkte med andre empiriske studier.

Vi ønsket derfor å følge opp problemstillingen omkring hvor stor behandlingseffekt klienter opplever som følge av studentbehandling ved en universitetsklinikk med et stort utvalg klienter og bruk av kvantitativ metodikk. I en tidligere artikkel (Ryum & Stiles, 2005) har vi kun indirekte presentert funn for et lite antall klienter i et annet utvalg ved samme universitetsklinikk, og da helt kort vært inne på spørsmålet om statistisk signifikant endring. Spørsmålene våre nå var derfor:

- 1 hvorvidt behandlingen har noen effekt
- 2 hvor stor denne effekten eventuelt er
- 3 om det er ulik effekt av behandling som følge av diagnostisk tilhørighet
- 4 hvor mange klienter som opplever en klinisk signifikant endring

Metode

Psykologisk poliklinikk for voksne og eldre, NTNU tar imot henvisninger fra fastleger og distriktpspsykiatriske sentre (DPS-er) i Trondheim og omegn. Tilbudet har siden oppstarten i 1997 blitt godt kjent i byen, og mottar hvert år 100 til 150 henvisninger. De inklusjonskriteriene eller henvisningsgrunnene en har signalisert utad, har vært «vanlige problemstillinger», som ikke skulle være spesielt utsilte på forhånd. Samtidig har det vært viktig at problemene skulle være håndterlige innenfor klinikkens rammer (kompetanse, varighet og intensitet i behandling). Mer spesifikt vil dette si problemer en antar vil kunne profitere på behandling av relativt kort varighet, og som dermed også vil kunne bidra til positive lærings- og mestringsopplevelser hos studentene. Det har i hvert enkelt tilfelle vært opp til innsøkende instans og klienten å vurdere hvorvidt klinikken har vært et passende tilbud. Fra klinikkens side har en i hovedsak kunnet tilby behandling til alle henviste klienter. Alle klienter har imidlertid måtte si seg villige til at terapitimer ble tatt opp på video og observert gjennom et enveisspeil, hvor veilederen og 3–4 medstudenter fulgte behandlingen.

Klienter

For den aktuelle studien var klienter som hadde gått i behandling i tidsrommet vårsemesteret 2001 til vårsemesteret 2004, tilgjengelige for videre analyser, og det samlede antall registrerte klienter i databasen var 145. Dette er noe lavere enn det antallet klienter som totalt sett har møtt ($N = 168$), men som i det meste av psykologisk behandling faller et betydelig antall klienter fra i løpet av de aller første timene. På grunn av manglende registrering på en eller flere variabler kunne bare 117 inkluderes for videre analyser, og alle disse var fullførte behandlinger. I forhold til den totale klientmassen i denne tidsperioden utgjør dette 70 % av møtte klienter.

Før behandling ble alle klienter utredet ved hjelp av Mini Internasjonalt Nevropsykiatrisk Intervju (MINI) av den aktuelle behandleren, men endelig diagnostisering i henhold til ICD-10-systemet ble først gjort ved behandlingsslutt. Ut fra didaktiske og pedagogiske hensyn la en her vekt på å forsøke å komme fram til én primærdiagnose. Klienter som valgte å ta imot tilbud om behandling, måtte gi sitt skriftlige samtykke til at innsamlede opplysninger ville kunne bli brukt i forskningsøyemed i anonymisert form. Selv om en så langt ikke har noen konkret dokumentasjon for at så er tilfellet, er det en oppfatning at klienter har et noe lavere symptomtrykk og et høyere funksjonsnivå enn ved en ordinær poliklinikk. Alvorlig sinnslidelse og psykose, voldsutagering, alvorlig rusmisbruk og suicidalitet har ellers vært regnet som eksklusjonskriterier.

Studentterapeutene

Terapiene ble utført av viderekomne studenter ($N = 117$) ved profesjonsstudiet i psykologi ved NTNU (2. avdeling). Av disse var 34 (29,1 %) menn og 83 (70,9 %) kvinner. Behandlingen har foregått ved Klinik for voksne og eldre, NTNU som en del av den obligatoriske, internkliniske praksis. Tre til fire studenter har til sammen utgjort et behandlingsmessig team under veiledning av en erfaren psykolog, hvor hver student har utført behandlingen av sin klient. Veilederen har vært faglig ansvarlig for behandlingen, og har også gitt veiledning i gruppe til den enkelte student i omfanget én time per uke per student.

Terapiform

Det har ikke vært noe krav om at terapien skulle være manualisert i henhold til bestemte terapimodeller. Den kliniske undervisningen ved studiet er heller ikke rettet spesifikt inn mot én bestemt terapeutisk modell. En legger imidlertid vekt på at all behandling som gis, skal være basert på god klinisk forskning og praksis. Ettersom studentene som sto for behandlingen, i hovedsak var uten noen form for klinisk erfaring, synes det rimelig å vurdere terapiformen som eklektisk orientert. Aktuelle veilederes orientering var hovedsakelig kognitiv, humanistisk og dynamisk. Behandlingen har vært individualterapi én behandlingstime per uke (50 min) i inntil 40 timer. Alle klienter ble imidlertid informert om at behandlingen i utgangspunktet var begrenset til ett semester og omtrent 15 timer, og kun et fåtall av studentene har fortsatt sine terapier ut over dette etter klientens og deres eget ønske. Det har ikke vært noe krav om at medikamentell behandling måtte seponeres under den psykologiske behandlingen.

Effektmål

Effekt av terapi ble innhentet ved hjelp av to selvrappportsinstrumenter som er hyppig brukt i prosess- og effektforskning. Begge måleinstrumentene ble fylt ut gjentatte ganger i løpet av behandlingskontakten, men den aktuelle studien baserer seg på pre- og postskårer:

Symptom Checklist 90-Revised (SCL-90-R) (Derogatis, 1983) gir informasjon om klientens opplevelse av psykiske symptomer målt ved ti symptomområder, samt tre indekser for generell symptombelastning. Klienten skal krysse av 90 spørsmål som rangeres på en likertskala fra 0 (ikke i det hele tatt) til 4 (veldig mye) for hvordan han eller hun har hatt det den siste uken, inkludert i dag. Vi brukte kun indeksen for generell symptombelastning (GSI), og den norske utgaven er oversatt av Nielsen og Vassend (1994).

I tillegg gjorde vi bruk av Inventory of Interpersonal Problems (Horowitz, Rosenberg, Baer, Ureno, & Villaseñor, 1988), men i en forkortet utgave basert på arbeidet til Alden, Wiggins og Pincus (1990). Dette spørreskjemaet (IIP-64) gir informasjon om klientens opplevelse av interpersonlig fungering og problemer målt opp mot åtte delskalaer og en global skåre for generell belastning. Her skal klienten krysse av 64 spørsmål på en likertskala som rangeres fra 0 («ikke i det hele tatt») til 4 («veldig mye») for hvordan han eller hun har hatt det den siste uken, inkludert i dag. Vi brukte her den globale skåren for generell belastning av interpersonlige problemer, og den norske utgaven er oversatt ved Stiles og Høglend (1994).

Statistikk

Vi gjør her nærmere rede for de statistiske prosedyrene som er knyttet til effektstørrelser og klinisk signifikant endring.

Statistisk signifikans sier noe om sannsynligheten for å finne forskjeller mellom to utvalg, men ikke noe om hvor stor denne forskjellen er. En vanlig brukt metode for å oppgi størrelsen på en effekt er derfor å regne ut effektstørrelser. Vi regnet ut ikke-kontrollerte effektstørrelser for hvert av endringsmålene etter Cohen (1988), som vil si at vi kun brukte utvalgets egne pre- og postskårer og ikke hadde noen kontrollgruppe. For å regne ut en effektstørrelse tar en gjennomsnittlig pre-verdi, trekker fra gjennomsnittlig post-verdi og deler på gjennomsnittlig standardavvik for de to måletidspunktene.

En kan godt tenke seg en stor effekt i form av en betydelig reduksjon i opplevde symptomer fra for eksempel en alvorlig til en moderat depresjon målt ved Beck's Depression Inventory i løpet av en behandlingskontakt, men selv med moderate symptomer vil de fleste klienter oppleve betydelige plager og redusert fungering. Kriteriene for hva en regner som klinisk signifikant endring, vil i denne sammenhengen som regel være strengere enn kriteriene for statistisk signifikant endring. Mål på klinisk signifikant er imidlertid ingen erstatning til mer tradisjonell statistikk, men et komplement til dette.

Ifølge Jacobson og Truax (1991) er kjernen i klinisk signifikant endring en antagelse om at dette innebærer en retur til normal fungering, og de foreslår to kriterier for å vurdere dette: (a) klienter som får behandling, må bevege seg fra en teoretisk dysfunksjonell populasjon til en funksjonell populasjon som følge av behandlingen, og (b) endringen må være reliabel. Bevegelse til en funksjonell populasjon avgjøres gjennom å bestemme en kutteskåre, hvor en skåre under denne betraktes som mer lik en funksjonell gruppe enn en dysfunksjonell. Forfatterne foreslår tre ulike måter for å bestemme skåren på. Alle har fordeler og ulemper. Etersom populasjonene i denne studien var overlappende, baserte vi oss på kriterium c.

Reliabel endring bestemmes gjennom å kalkulere en Reliable Change Index (RCI), og hvis denne er større enn 1.96, antyder det at pre- til post-endringen ikke er et resultat av tilfeldige fluktueringer i testskårer ($p < .05$). I det følgende redegjør vi nærmere for prosedyren for kalkulering av både cut-off og RCI i denne studien.

Det ble kalkulert én kutteskåre for hvert instrument ved at pre-gjennomsnittsskåre fra vårt eget utvalg ($N = 117$) ble lagt sammen med et normalutvalg og dividerte på 2, korrigert for forskjell i varians utvalgene imellom $((SD1) (M2) + (SD2) (M1))/(SD1 + SD2)$. Det foreligger her en norsk validering av SCL-90-R (Vassend, Lian & Andersen, 1992), men ettersom denne ikke har data på test-retest-reliabilitet (som er nødvendig for å regne ut reliabel endring), valgte vi å bruke Derogatis' (1983) tidligere publiserte normer ($M = .31$; $SD = .31$; $N = 974$) på et amerikansk normalutvalg. Det kan imidlertid nevnes at den norske undersøkelsen fant godt samsvar mellom de norske og de amerikanske dataene. Det foreligger ingen norsk normering av IIP-64, så vi baserte oss også her på publiserte normer ($M = .80$; $SD = .54$; $N = 800$) fra et amerikansk normalutvalg (Horowitz, Alden, Wiggins & Pincus, 2000). Ved utregning fant vi da at kutteskåren ble 0.58 for SCL-90-R (en skåre tilsvarende eller mindre enn 0.58 ble betraktet som å falle innenfor en normalpopulasjon) og 1.03 for IIP-64 (en skåre tilsvarende eller mindre enn 1.03 ble betraktet som å falle innenfor en normalpopulasjon).

Det ble så kalkulert verdier for å regne ut Reliable Change Index (RCI) på SCL-90-R og IIP-64. Denne tilsvarende forskjellen mellom to testskårer (pre-post) dividert på standardfeilen mellom skårene, som utledes fra test-retest-reliabiliteten til målingsinstrumentet og standardavvik til preskårer til instrumentet. Standardfeil forskjell-skårene (S_{diff}) i denne studien var 0.36 på SCL-90-R ($M = 1.13$; $SD = 0.63$; $N = 117$) og 0.38 for IIP-64 ($M = 1.28$; $SD = 0.57$; $N = 117$). Disse ble utledet fra skårene ved inntak på vårt eget utvalg, samt test-retest-reliabilitetskoeffisienter for SCL-90-R og IIP-64 rapportert av henholdsvis Derogatis (1983) (0.84, $N = 94$) og Horowitz et al. (2000) (0.78, $N = 60$).

For å undersøke om forholdstallene (proporsjonene) innenfor de normale og de kliniske distribusjonene var statistisk signifikant endret fra pre til post, ble det avslutningsvis kalkulert khi-kvadrat for alle gruppene for hvert av effektmålene.

Resultater

Det ble først anvendt deskriptiv statistikk for å undersøke demografiske variabler som kjønn, alder, utdanning, sivilstatus, yrkesmessig status og diagnoser for hele utvalget. En samlet framstilling av disse dataene er gitt i tabell 1.

Tabell 1 viser at det var en overvekt av kvinner i utvalget, og at et betydelig antall klienter rapporterte å være enslige, ha utdanning ut over grunnskole og være i lønnet arbeid.

Det ble deretter valgt en prosedyre hvor vi først undersøkte hele utvalget som sådant, uten å ta hensyn til diagnostisk tilhørighet. Deretter ble klientene delt inn i grupper basert på enkeltdiagnoser (F30–39; F40–48; F60–69; komorbiditet; ingen diagnose), og de samme dataanalysene ble gjort. Her ble kategorien «andre» ekskludert, da den inneholdt kun tre klienter. Gangen i disse analysene presenteres nedenfor.

Det ble kjørt avhengige t-tester (tohalet) for å undersøke behandlingseffekten målt ved de to ulike effektmålene fra terapistart og til avslutning. Deretter ble det regnet ut ikke-kontrollerte effektstørrelser for den samme behandlingsperioden for hvert av endringsmålene. Resultatene av disse analysene er framstilt i tabell 2 og tabell 3 for hvert av måleinstrumentene.

Tabell 1. Demografiske og diagnostiske variabler for utvalget (N = 117)					
Variabel	Kategorier	Mean	(SD)	N	%
Alder		34.2	(11.4)		
Kjønn	Mann			30	(25.6)
	Kvinne			87	(74.4)
Sivilstatus	Enslig			46	(39.3)
	Gift			31	(26.5)
	Samboer			28	(23.9)
	Separeert/ skilt			12	(10.3)
Utdanning (antall år)	< 9			13	(11.1)
	9–12			50	(42.7)
	> 12			54	(46.2)
Yrkesmessig status	I lønnet arbeid			43	(36.8)
	Arbeidsløs			10	(8.5)
	Sykmeldt			25	(21.4)
	Under attføring			14	(12)
	Uføretrygdet			2	(1.7)
	Student/ skoleelev			19	(16.2)
	Pensjonert			2	(1.7)

	Husmor/husfar			2	(1.7)
ICD-10 Diagnoser	F30–39a			28	(23.9)
	F40–48b			45	(38.5)
	F60–69c			8	(6.8)
	Annen diagnose			3	(2.6)
	Komorbid			13	(11.1)
	Ingen diagnose			20	(17.1)

a = Affektive lidelser.
 b = Nevrotiske, belastningsrelaterte og somatoforme lidelser.
 c = Personlighets- og atferdsforstyrrelser hos voksne.

Gruppe	Pre		Post		Pre–post statistisk signifikant endring	ES (d)
	M	(SD)	M	(SD)		
Hele utvalget (n = 117)	1.13	.63	0.83	.63	$p < .001$.48
F30–39a (n = 28)	1.23	.62	.84	.62	$p < .001$.63
F40–48b (n = 45)	1.25	.68	.84	.67	$p < .001$.61
F60–69c (n = 8)	1.28	.53	1.15	.50	$p = .25$.25
Komorbid (n = 13)	1.18	.55	1.06	.73	$p = .39$.19
Ingen diagnose (n = 20)	.66	.44	.56	.47	$p = .23$.22

SCL-90-R = Symptom Checklist 90 Revised; Pre = før behandling; Post = behandlingsslutt; ES = Effektstørrelse (Cohen's d)
 a = Affektive lidelser
 b = Nevrotiske, belastningsrelaterte og somatoforme lidelser
 c = Personlighets- og atferdsforstyrrelser hos voksne.

Symptombelastning

For hele utvalget ble det funnet en statistisk signifikant endring mellom inntak og avslutning, målt ved SCL-90-R ($t(116) = 6.64, p < .001$). Når de diagnostiske gruppene ble analysert hver for seg, ble det funnet en statistisk signifikant endring for gruppen F30–39 ($t(27) = 4.22, p < .001$) og F40–48

($t(44) = 4.99, p < .001$), men det ikke ble funnet noen statistisk signifikant endring for de tre andre gruppene. Effektstørrelsen var på .48 for hele utvalget samlet, mens gruppene F30–39 og F40–48 lå på henholdsvis 0.63 og 0.61. De tre andre gruppene lå mellom .19 og .25.

Gruppe	Pre		Post		Pre–post statistisk signifikant endring	ES (d)
	M	(SD)	M	(SD)		
Hele utvalget (n = 117)	1.28	.57	1.07	.53	$p < .001$.38
F30–39a (n = 28)	1.32	.54	1.12	.54	$p < .05$.37
F40–48b (n = 45)	1.30	.63	1.04	.53	$p < .001$.45
F60–69c (n = 8)	1.27	.50	1.28	.51	$p = .96$	-.02
Komorbid (n = 13)	1.38	.48	1.23	.59	$p = .08$.28
Ingen diagnose (n = 20)	1.11	.61	.90	.47	$p < .05$.39

IIP-64 = Inventory of Interpersonal Problems 64; Pre = før behandling; Post = behandlingsslutt; ES = Effektstørrelse (Cohen's *d*)
 a = Affektive lidelser
 b = Nevrotiske, belastningsrelaterte og somatoforme lidelser
 c = Personlighets- og atferdsforstyrrelser hos voksne.

Interpersonlige problemer

For hele utvalget ble det funnet en statistisk signifikant endring mellom inntak og avslutning målt ved IIP-64 ($t(116) = 5.23, p < .001$). Når de diagnostiske gruppene ble analysert hver for seg, ble det funnet en statistisk signifikant endring for gruppene F30–39 ($t(27) = 2.40, p < .05$), F40–48 ($t(44) = 3.69, p < .001$) og gruppen «ingen diagnose» ($t(19) = 2.60, p < .05$). For de andre to gruppene var det ingen statistisk signifikant endring. Effektstørrelsen for hele utvalget var på .38, mens gruppene F30–39, F40–48 og «ingen diagnose» lå på henholdsvis .37, .45 og .39.

Måleinstrument	N (%) i de diagnostiske gruppene					
	Hele utvalget (n = 117)	F30–39a (n = 28)	F40–48b (n = 45)	F60–69c (n = 8)	Komorbid (n = 13)	Ingen diagnose (n = 20)
SCL-90-R						

Tabell 4. Frekvens og prosent av klienter med reliabel endring, funksjonell distribusjon pre og post, klinisk signifikant endring og forverring for de ulike diagnostiske gruppene						
RCI > 1.96	20 (17.1)	8 (28.6)	9 (20)	1 (12.5)	1 (7.7)	1 (5)
Funksjonell distribusjon pre	23 (19.7)	5 (17.9)	4 (8.9)	2 (25)	2 (15.4)	10 (50)
Funksjonell distribusjon post	50 (42.7)	11 (39.3)	21 (46.7)	1 (12.5)	4 (30.8)	12 (60)
CS	11 (9.4)	3 (10.7)	6 (13.3)	1 (12.5)	0 (0)	1 (5)
Forverring	0	0	0	0	0	0
IIP-64						
RCI > 1.96	12 (10.3)	2 (7.1)	7 (15.6)	1 (12.5)	0 (0)	1 (5)
Funksjonell distribusjon pre	41 (35)	8 (28.6)	17 (37.8)	2 (25)	4 (30.8)	9 (45)
Funksjonell distribusjon post	55 (47)	14 (50.0)	21 (46.7)	1 (12.5)	6 (46.2)	11 (55)
CS	9 (7.7)	2 (7.1)	6 (13.3)	0 (0)	0 (0)	0 (0)
Forverring	0	0	0	0	0	0
RCI (reliable change index) > 1.96 representerer antall klienter med reliabel bedring; funksjonell distribusjon pre representerer antall klienter som falt innenfor en ikke-klinisk populasjon ved behandlingsstart; funksjonell distribusjon post representerer antall klienter som falt innenfor en ikke-klinisk populasjon ved avslutning; CS (klinisk signifikans) representerer antall klienter som hadde en reliabel bedring og som falt innenfor en ikke-klinisk populasjon ved avslutning; Forverring representerer antall klienter som hadde en reliabel forverring i løpet av behandlingen a = Affektive lidelser b = Nevrotiske, belastningsrelaterte og somatoforme lidelser c = Personlighets- og atferdsforstyrrelser hos voksne; SCL-90-R = Symptom Checklist 90 Revised; IIP-64 = Inventory of Interpersonal Problems 64.						

Klinisk signifikant endring

Avslutningsvis ble det regnet ut verdier for bevegelse til normalpopulasjonen, reliabel endring og klinisk signifikant endring, og hvor mange klienter som kunne klassifiseres innenfor hver av disse kategoriene. Disse resultatene er presentert i tabell 4.

Ved avslutning befant drøye 40 % av klientene i utvalget samlet seg innenfor en ikke-klinisk populasjon, målt ved begge måleinstrumentene. For undergruppene med statistisk signifikant endring varierte dette mellom 39.3 og 55 %. En betydelig lavere andel klienter kunne klassifiseres med reliabel eller klinisk signifikant endring i alle gruppene med statistisk signifikant endring. Ingen av klienten kunne klassifiseres med reliabel forverring.

For å undersøke hvorvidt forholdstallene (proporsjonene) innenfor de normale og kliniske distribusjonene var statistisk signifikant endret fra pre til post, ble det avslutningsvis kalkulert khi-kvadrat for alle gruppene for hvert av effektmålene. På SCL-90-R fant vi da en statistisk signifikant effekt for det samlede utvalget, $\chi^2(1, N = 117) = 14.514, p < .001$ og undergruppen F40-48, $\chi^2(1, N =$

45) = 16.006, $p < .001$, men ikke for de andre gruppene. På IIP-64 fant vi ingen statistisk signifikant effekt for noen av gruppene. For flere av gruppene var det imidlertid for få observasjoner per celle (< 5 for en 2×2 tabell) til at det lot seg gjøre å kalkulere khi-kvadrat på en pålitelig måte.

Diskusjon

Vårt første spørsmål var hvorvidt behandlingen hadde effekt. Det ble her funnet en statistisk signifikant behandlingseffekt for hele utvalget, målt ved både SCL-90-R og IIP-64 fra inntak og til avslutning. Dette er i seg selv et første viktig funn, og viser at det er en opplevd reduksjon av symptomer og bedring av interpersonlig fungering i løpet av behandlingskontakten. Videre analyser ut fra diagnostisk tilhørighet avslørte imidlertid at klienter som falt inn under gruppene F60–69 («personlighets- og atferdsforstyrrelser hos voksne») eller «komorbiditet», ikke hadde noen statistisk signifikant endring pre–post på verken SCL-90-R eller IIP-64. Her var det få klienter i begge gruppene, og sannsynligheten («power») for å kunne oppdage en medium effektstørrelse ($d = 0.5$) var på beskjedne 0.15 og 0.23 i de to gruppene ($p < .05$), ifølge Cohen (1988). Utvalgene var derfor sannsynligvis for små til å kunne påvise en eventuell statistisk signifikans.

For gruppen «ingen diagnose» var det bare en statistisk signifikant endring pre–post på IIP-64. Dette henger sannsynligvis sammen med at dette var klienter som rapporterte svært lite symptomer ved terapistart, målt ved SCL-90-R. «Endringsrommet» var derfor naturlig nok også begrenset.

Vårt neste spørsmål var hvor stor endringseffekten var i løpet av behandlingen, og effektstørrelse (Cohen's d) er her et mål på dette. For gruppene med statistisk signifikant endring lå effektstørrelsene innenfor, eller tett opp mot, hva Cohen (1988) klassifiserer som medium effektstørrelser ($>.5$) i samfunnsvitenskapene målt ved SCL-90-R. Effekten var noe lavere på IIP-64, som er i tråd med tidligere rapporterte funn hvor generell symptomlette ser ut til å komme tidligere i terapi, og gå forut for senere reduksjon i for eksempel interpersonlige problemer (Horowitz et al., 1988; Kopta, Howard, Lowry & Beutler, 1994). Dette kan bero på forskjellige mekanismer eller endringsprosesser, hvor for eksempel mer uspesifikke fellesfaktorer og remoralisering av klienten (Frank & Frank, 1991) er av stor betydning for generell symptomlette. På den annen side krever endring av potensielt maladaptive interpersonlige væremåter (reduksjon av interpersonlige problemer) bruk av mer spesifikt og intensivt terapeutiske arbeid.

Det var ingen betydelige forskjeller i effektstørrelser mellom de diagnostiske gruppene med statistisk signifikant endring, hvilket betyr at for eksempel «affektive lidelser» (F30–39) og «nevrotiske, belastningsrelaterte og somatoforme lidelser» (F40–48) kom like godt ut. I forhold til SCL-90-R kan en også merke seg at effekten ligger noe høyere for F30–39 og F40–48 enn for det samlede utvalget, hvilket impliserer at det er de andre gruppene som trekker gjennomsnittseffekten noe ned for utvalget som helhet. Omregnet til prosent rapporterte 74 % av klientene målt ved SCL-90-R og 69 % av klienten målt ved IIP-64 å være statistisk signifikant bedre ved terapiavslutning enn ved terapistart, utvalget sett under ett.

Effektstørrelsene i denne studien ligger innenfor spennet ($d .4-.6$) for hva nyere metastudier klassifiserer som den generelle effektstørrelsen av psykoterapi (Lambert & Ogles, 2004). Dette estimatet baserer seg riktignok på «intent to treat», og ettersom vi kun inkluderte fullførte behandlinger i denne studien, er det grunn til å tro at studentterapiene kommer bedre ut enn Lambert og Ogles data. Effekten er imidlertid noe lavere enn hva for eksempel Wampold (2001) har rapportert, som også er basert på «intent to treat»-analyser. Sammenlignet med hva en ofte vil kunne finne i randomiserte og kontrollerte studier, er effekten betydelig lavere.

Årsaksforholdet er her sannsynligvis sammensatt. Begrensende faktorer i forhold til det å kunne oppnå høyere effekter i denne studien vil for eksempel være at utvalget av terapeuter ikke var spesielt utvalgt og trent (som i randomiserte kontrollerte studier), samt at behandlingen varte relativt kort. Det sistnevnte punktet har konsekvenser når en vet at effekten av psykoterapi er vist å være økende med antall timer (Howard, Krause & Orlinsky, 1986). Et kanskje mer avgjørende moment er også hvor mye symptomer og vansker klientene rapporterer ved behandlingsstart. På grunn av regresjonseffekter, hvor ekstreme verdier har en tendens til å regredere mot et gjennomsnitt ved repeterte målinger, vil det under ellers like betingelser generelt sett være lettere å oppnå høye effektstørrelser på et utvalg med store initiale symptomplager enn på et utvalg med mer moderate symptomplager. Jo nærmere en klient befinner seg et normalgjennomsnitt, jo mindre «endringsrom» er det, og det samlede symptomtrykket på dette utvalget kan best beskrives som «moderat».

Som et siste spørsmål ønsket vi å undersøke hvor mange klienter som ved avslutning befant seg innenfor en ikke-klinisk (normal) populasjon, samt antall klienter som kunne klassifiseres med reliabel og klinisk signifikant endring fra pre til post. Ut fra de kutteskårene vi regnet ut, fant vi at et betydelig antall klienter i de ulike gruppene befant seg innenfor et ikke-klinisk spenn ved behandlingsslutt. For de fleste gruppene var det imidlertid også et betydelig antall klienter som befant seg innenfor en normalpopulasjon allerede ved behandlingsstart, slik at den reelle bevegelsen fra dysfunksjonell til funksjonell fordeling var lav. Her var det faktisk bare gruppene «hele utvalget» og undergruppen «F40–48» som hadde en statistisk signifikant endring, målt ved SCL-90-R fra pre til post. Dette gjenspeiler seg i det faktum at det var få klienter som oppnådde reliabel eller klinisk signifikant endring. Selv om det knytter seg flere metodologiske problemstillinger til det å sammenligne disse funnene med tall fra randomiserte og kontrollerte studier, kan vi nevne at tallene ligger betydelig lavere enn hva som har vært rapportert tidligere i for eksempel NIMHs depresjonsstudie (Ogles, Lambert, & Sawyer, 1995).

De samme begrensningene som ble nevnt i forbindelse med effektstørrelser, er relevante her, i tillegg til at det vanligvis vil være vanskeligere å innfri kriteriene til klinisk signifikant endring enn statistisk signifikant endring. For et utvalg med moderate symptomer, behandling av relativt kort varighet og terapeuter med minimal klinisk erfaring ville det derfor være overraskende å finne store kliniske effekter.

Begrensinger

Det er flere potensielle begrensinger ved studien. For det første mangler den en kontrollgruppe, noe som fører til at det ikke er mulig å fastslå hvorvidt det er selve behandlingen eller utenomterapeutiske faktorer som har ført til endring. I beste fall er det *sannsynliggjort* at behandlingen fører til bedring, men noen kausalslutning i absolutt forstand kan en ikke trekke her. For det andre er det kun fullførte behandlinger som er tatt med i analysene, noe som sannsynligvis er med på å gjøre resultatene (effekten) bedre enn de ville vært om også frafall (ikke-fullførte behandlinger) hadde blitt tatt med. For det tredje har vi kun gjort bruk av to effektmål, som begge er selvrapporteringsinstrumenter. Bruk av andre effektmål ville muligens ha gitt andre resultater. For det fjerde ble ikke medikasjon «frost» i løpet av behandlingen, og endringer her kan ha påvirket resultatene. Til slutt må det også nevnes at en sannsynligvis har hatt en underdiagnostisering av lidelser, tatt i betraktning at det er lite komorbiditet og få personlighetsforstyrrelsesdiagnoser sammenlignet med hva en vanligvis finner i et ordinært poliklinisk utvalg (Alnaes & Torgersen, 1988). Den endelige diagnostiseringen skjedde også etter at behandlingen var avsluttet, og klientens opplevde effekt av behandlingen kan ha påvirket denne. En

kan imidlertid heller ikke utelukke at frekvensen av de nevnte kliniske tilstandene skyldes en reell selvseleksjon når det gjelder pasienter til klinikken.

Konklusjon

Det er grunnlag for å hevde at behandlingen som gis ved klinikken, er forbundet med moderat effektivitet, og det er ingen vesentlig forskjell i effektstørrelser mellom affektive lidelser (F30–39) og nevrotiske, belastningsrelaterte og somatoforme lidelser (F40–48). En bør imidlertid forvente lavere effekt hos klienter som ikke tilfredsstillende noen diagnose, og som rapporterer lite symptomer ved inntak. Selv om relativt mange klienter befinner seg innenfor en ikke-klinisk populasjon ved avslutning, er det få som tilfredsstillende kriteriene til reliabel eller klinisk signifikant endring. Det er ellers behov for å samle inn mer data, slik at en kan få undersøkt behandlingseffekt for gruppene «F60–69» og «komorboditet».

Studien representerer et framskritt i forhold til tidligere norske studier, for eksempel når det gjelder utvalgsstørrelse, men framtidige studier bør ideelt sett forsøke å etablere en strengere form for kontroll for å avklare spørsmålet om kausalitet.

Truls Ryum
Psykologisk institutt
Norges Teknisk-Naturvitenskapelige Universitet
7491 Trondheim, Norway
Tlf 41 60 87 35
E-post truls.ryum@svt.ntnu.no

Referanser

- Alden, L. E., Wiggins, J. S. & Pincus, A. L. (1990). Construction of Circumplex Scales for the Inventory of Interpersonal problems. *Journal of Personality Assessment*, 55, 521–536.
- Alnaes, R. & Torgersen, S. (1988). DSM-III symptom disorders (Axis I) and personality disorders (Axis II) in an outpatient population. *Acta Psychiatrica Scandinavica*, 78, 348–355.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Derogatis, L. R. (1983). *The Symptom Checklist -90 Revised: Administration, scoring and procedures manual II*. Baltimore: Clinical Psychometric Research.
- Frank, J. D. & Frank, J. B. (1991). *Persuasion and healing* (3rd ed.). Baltimore, Maryland: The John Hopkins University Press.
- Gullestad, S. (1986). Studentbehandling – forsvarlig psykoterapi? *Tidsskrift for Norsk Psykologforening*, 23, 3–12.
- Horowitz, L. M., Alden, L. E., Wiggins, J. S. & Pincus, A. L. (2000). *Inventory of Interpersonal Problems*. New York: The Psychological Corporation.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G. & Villasenor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56, 885–892.

- Howard, K. I., Krause, M. S. & Orlinsky, D. E. (1986). The attrition dilemma: Toward a new strategy for psychotherapy research. *Journal of Consulting and Clinical Psychology*, 54, 106–110.
- Håland, W. (1986). Studenter som psykoterapeuter. *Tidsskrift for Norsk Psykologforening*, 23, 365–376.
- Jacobson, N. S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.
- Kopta, S. M., Howard, K., Lowry, J. L. & Beutler, L. E. (1994). Patterns of symptomatic recovery in psychotherapy. *Journal of Consulting and Clinical Psychology*, 62, 1009–1016.
- Lambert, M. J., Garfield, S. L. & Bergin, A. E. (2004). Overview, trends and future issues. I M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of psychotherapy and behavior change* (5th ed.; ss. 805–821). New York: John Wiley & Sons.
- Lambert, M. J. & Ogles, B. M. (2004). *The efficacy and effectiveness of psychotherapy* (5th ed.). New York: John Wiley & Sons.
- Nielsen, G. & Vassend, O. (1994). SCL-90-R – Norsk oversettelse.
- Ogles, B. M., Lambert, M. J. & Sawyer, J. D. (1995). Clinical significance of the National Institute of Mental Health Treatment of Depression Collaborative Research Program Data. *Journal of Consulting and Clinical Psychology*, 63, 321–326.
- Ryum, T. & Stiles, T. C. (2005). Betydningen av den terapeutiske allianse – En studie av alliansens prediktive validitet. *Tidsskrift for Norsk Psykologforening*, 42, 998–1003.
- Stiles, T. C. & Høglend, P. A. (1994). IIP-64 – Norsk oversettelse.
- Vassend, O., Lian, L. & Andersen, H. T. (1992). Norske versjoner av NEO-Personality Inventory, Symptom Checklist 90 Revised og Giessen Subjective Complaints List. Del I. *Tidsskrift for Norsk Psykologforening*, 29, 1150–1160.
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

The aim of this study was to evaluate treatment effectiveness among out-patients with varying diagnoses receiving short-term student-conducted psychotherapy at a university clinic. One hundred and seventeen clients, all completers and diagnosed according to the ICD-10 system, were evaluated. Outcome data was collected using the Symptom Checklist 90 Revised (SCL-90-R) and the Inventory of Interpersonal Problems (IIP-64) pre- and post treatment. Treatment was not manualized according to a specific protocol, and therapists had minimal or no previous clinical experience. Results demonstrated that there were both statistical and clinical treatment effects from intake to termination, in the range of medium effect-sizes on both outcome measures. A substantial number of clients were in the non-clinical range on both measures at termination, although few patients were classified with reliable or clinical significant change.

Keywords: student-conducted therapy, effectiveness; outcome, training