

Evaluering av datakvaliteten til ”Statusrapportskjemaet”

*- et måleinstrument for pasienter i legemiddelassistert
rehabilitering*

Margit Vånar



Masteroppgave i helsefagvitenskap

Seksjon for helsefag, Det medisinske fakultet

UNIVERSITETET I OSLO

30. juni 2005

Sammendrag

Formål: Hensikten med studien var å evaluere datakvaliteten til det norske Statusrapportskjemaet, et standardisert måleinstrument for pasienter i legemiddelassistert rehabilitering (LAR). Dataene fra instrumentet blir benyttet på et sammenstilt nivå, og har vært brukt som datagrunnlag for flere evalueringsrapporter samt evaluering av regionale og nasjonale forskjeller.

Måleinstrumentet inneholder 53 variabler om pasientens aktuelle demografiske forhold og medikamentelle tilbud, kontakt med hjelpeapparatet, stoff- og alkoholbruk samt kriminell aktivitet. Skjemaet besvares av klinikere. På bakgrunn av dette ønsket en å systematisk kartlegge hvorvidt utfyllerne kjente pasientens situasjon, og estimere intertetterreliabiliteten til skårene. Utvalget besto av 20 utfyllerpar.

Metode: For å kalkulere intertetterreliabilitet og avgitt svar, ble skårene til primær- og sekundærkontakt til pasienten sammenstilt. Svarandel og samsvar mellom skårene ble kvantifisert med prosent, Cohens kappa (κ) og intraklassekoeffisienten (ICC).

Resultater: Resultatene viste at "Statusrapportskjemaet" egnet seg godt til kartlegging av demografiske data, og data om det medikamentelle behandlingstilbudet til pasientene. Generelt oppnådde disse variablene både høy svarprosent og et godt samsvar mellom skåringene. Dataene kan derfor benyttes til sammenlikninger av regional/nasjonal art, og til å følge utviklingen av LAR over tid. Data om pasientens kontakt med hjelpeapparatet, stoff- og alkoholbruk, samt kriminelle aktivitet vitnet om at mange av utfyllerne hadde begrenset kjennskap til disse forholdene. Dette kom til uttrykk gjennom hyppig bruk av svarkategorien "ukjent".

Konklusjon: Studien viste at forventningen til utfyllernes kjennskap til pasienten har vært for høye. På bakgrunn av resultatene anbefales endringer i enkeltvariabler i Statusrapportskjemaet, forbedring av dokumentasjonsrutiner, samt et veiledningsprogram for å øke kunnskap og motivasjon for utfyllerne. Dette vil forbedre datakvaliteten til variablene.

Abstract

Purpose: The objective of the study was to evaluate the data quality of the standardized Norwegian evaluation instrument “Statusrapportskjemaet” for patient’s in maintenance treatment. The instrument contains 53 variables about the patient’s demographic and medical status, their contact with various helpers, drug and alcohol use and criminal activity. The instrument was designed to be filled out by clinicians without prior training, based on their knowledge of their patients. Due to this, it was of interest to investigate how well the clinicians knew their patients, and the interrater reliability of the different scores.

Method: For calculation of interrater agreement and missing data, the scores of the primary caretaker were compared with the scores of the corresponding secondary rater. Agreement was quantified by using Cohen’s kappa (κ), Intra Class Coefficient (ICC) and percentage agreement.

Results: The results showed that the demographic data and data about the medical treatment the patients are offered are satisfactory based on methodological criteria. The variables concerning drug and alcohol use, health and psychological functioning, and criminal activity are more problematic because of the clinicians insecurity, showed in this study by a high percentage of scoring “unknown”. The measure of agreement varied from very good to poor.

Conclusion: The study showed that the instrument developers have had too high expectations of the clinician’s knowledge about the patients in maintenance treatment. An extensive training in scoring the variables and improving documentation in the patients medical journal in order to increase the data quality of the “Statusrapportskjema” is, based on this study, recommended.

Innhold

SAMMENDRAG	2
ABSTRACT.....	3
INNHOOLD	4
1 INNLEDNING	7
1.1 FORSKNINGSSPØRSMÅL	9
2 TEORI OG BAKGRUNNSSTOFF.....	12
2.1 MÅLEINSTRUMENTER	12
2.2 UTFORMING AV MÅLEINSTRUMENTER	12
2.3 VALIDITET	13
2.4 RELIABILITET.....	15
2.4.1 <i>Generelt om reliabilitet.....</i>	<i>15</i>
2.4.2 <i>Metoder for reliabilitetsmåling</i>	<i>16</i>
2.4.3 <i>Målefeil/variasjon</i>	<i>18</i>
2.5 MÅLING AV INTERRATERRELIABILITET	20
2.5.1 <i>Reliabilitetskoeffisienter.....</i>	<i>21</i>
2.5.2 <i>Prosentvis samsvar</i>	<i>24</i>
2.5.3 <i>Ulike typer variabler.....</i>	<i>25</i>
2.6 UTFYLLERDIMENSJONEN.....	27
2.7 MANGLENDE SVAR.....	29
2.8 EVALUERING AV LEGEMIDDELASSISTERT REHABILITERING (LAR).....	29
3 METODE.....	33
3.1 UTVALG OG UTVALGSSTØRRELSE	33
3.2 PROSEDYRE FOR DATAINNSAMLING	34
3.3 STATISTISKE BEREGNINGER	35
3.4 KRITERIER FOR GOD DATAKVALITET.....	35
3.5 ETISKE HENSYN.....	36
4 RESULTATER.....	38
4.1 BRUK AV INFORMASJONSKILDER	38
4.2 UTFYLLERNES EVNE TIL Å SVARE.....	39
4.3 SAMSVAR, KATEGORISKE VARIABLER	42

4.4	SAMSVAR OG BRUK AV “UKJENT”, KONTINUERLIGE VARIABLER.....	45
5	DISKUSJON	46
5.1	KJENNSKAP OG PÅLITELIGHET	46
5.2	ANBEFALINGER.....	55
5.3	STYRKER OG SVAKHETER VED STUDIEN	59
5.3.1	<i>Utvalgets størrelse</i>	59
5.3.2	<i>Utvalgets sammensetning</i>	59
5.3.3	<i>Metode</i>	60
6	OPPSUMMERING OG KONKLUSJON	61
	KILDELISTE	62
	VEDLEGG	69

TABELLINDEKS

Tabell 1 – Krav til κ - verdier	21
Tabell 2 – Eksempel på krysstabell fra Statusrapportskjemaet.....	22
Tabell 3 – Eksempel på beregning av ICC	24
Tabell 4 – Komplette svarpar, bruk av ”ukjent”	41
Tabell 5 – Samsvar for variabler der Cohens kappa (κ) lot seg beregne.....	43
Tabell 6 – Prosentvis samsvar per variabel.....	44
Tabell 7 – Komplette svarpar, bruk av ”ukjent” og intraklassekorrelasjon (ICC)	45

FIGURINDEKS

Figur 1: De to feilkomponentene i et observert måleresultat	18
Figur 2: Eksempel på nominal variabel fra Statusrapportskjemaet.....	26
Figur 3: Eksempel på ordinal variabel fra Statusrapportskjemaet	26
Figur 4: Eksempel på kontinuerlig variabel fra Statusrapportskjemaet	26
Figur 5: Bruk av informasjonskilder. Utfyllergruppe 1 (n=20) og utfyllergruppe 2 (n=20)	38

1 Innledning

Måleinstrumenter spiller en stadig viktigere rolle i medisinsk forskning. De siste tiår har den kliniske forskningen innen medisin blitt stadig mer kompleks. Disipliner som psykiatri, revmatologi og onkologi er i større grad knyttet mot livskvalitetsdimensjonen. Dette innebærer at samtlige helseprofesjoner i økende grad har fokus på livskvalitet fremfor livskvantitet. For å kunne vitenskapeliggjøre kvalitetsdimensjonen, har det vært nødvendig å utvikle metoder for å måle det en tidligere anså som ikke målbart. En viktig side ved målinger er at en ved hjelp av standardisering og forenkling skal eliminere det personlige skjønn, slik at flere personer skal oppnå de samme resultatene om de benytter seg av de samme instrumentene. Om det er stor overensstemmelse mellom de ulike personene kan målingen betraktes som pålitelig (Fjelland og Gjengedal, 1995). Det er et kjent psykometrisk prinsipp at evaluering av datakvalitet er et viktig ledd i utviklingen av nye instrumenter (Murphy & Davidshofer, 1998). Behovet for pålitelige og gyldige måleinstrumenter ble tydelig dokumentert av funnene i en kontrollert studie av 300 randomiserte schizofreni-pasienter. Studien viste at om en brukte upubliserte instrumenter, fremfor pålitelige og gyldige var det 40 % flere som svarte at behandlingen var effektiv (Marshall et al., 2000). Gode måleinstrumenter er også sentralt for å kunne bedre planleggingen av de enkelte behandlingstilbudene, samt på troverdig vis forsvare behandlingen overfor de bevilgende myndigheter og allmennheten (Lawrinson, Copeland & Indig, 2005). Måleinstrumenter kan også være med på å endre klinisk praksis ved å gjøre sammenlignende analyser, samt høyne det faglige bevissthetsnivået slik at en styrker eget utviklingsarbeid (Wade, 1992).

Måleinstrumenter blir benyttet til ulike formål innen helsefagene (Wade, 1992). I somatisk medisin kan eksempelvis blodprøver indikere en pasients infeksjonsstatus, og på den måten fortelle hvilken behandling som kreves. I psykiatrisk forskning og praksis er et vidt spekter av instrumenter implementert for å kunne måle endring eller effekt av behandling. I tillegg benyttes måleinstrumenter til diagnosesetting, utredning, vurdering av tegn og symptomer, livskvalitet, alvorlighetsgrad og flere

andre dimensjoner. Om målingene ikke er pålitelige, kan dette få dramatiske følger for pasienten. Funn basert på dårlig datakvalitet kan medføre feilslutninger med potensielt fatale følger.

En australsk studie har avdekket mer enn 300 måleinstrumenter som eksisterer for å evaluere behandling av stoffavhengige (Lawrinson et al., 2005). Av disse er Addiction Severity Index (ASI) det mest internasjonalt kjente og anerkjente. Intervjuguiden regnes som en "gull-standard" innen internasjonal forskning på stoffavhengige (Kellogg et al., 2003). ASI er et omfattende instrument som er designet både for klinisk bruk og forskning, slik at mange variabler kan virke uhensiktsmessige. Det tar derfor forholdsvis lang tid å skåre sammenlignet med andre intervjuguider. Intervjueren må også være sertifisert. Et annet og mindre omfattende instrument er det standardiserte intervjuet Brief Treatment Outcome Measure (BTOM) som helsemyndighetene i Australia har utviklet. BTOM karakteriseres som et kortfattet (intervjuet tar cirka 15 minutter), valid og reliabelt måleinstrument, som også er anerkjent av klinikerne som arbeider direkte med pasienten (Lawrinson et al., 2005). I USA har forskere utviklet The Kreek-McHugh-Schlusser-Kellogg-skalaen (KMSK). Dette instrumentet baserer seg på et fem minutters standardisert intervju som tar utgangspunkt i den perioden i livet informantene har eksponert seg selv hyppigst for stoff (Kellogg et al., 2003).

Ulike forklaringsmodeller og syn på avhengighet preger stoffbehandlingsfeltet. Det siste tiåret har en medisinsk tilnærming fått stadig større oppslutning, også i Norge. Metadon som medisin mot en kronisk opiatavhengig har lenge internasjonalt vært en veldokumentert og mye benyttet behandlingsform (Dole & Nyswander, 1965). På bakgrunn av evalueringen av to forsøksprosjekt (Skretting, 1997, Ervik, 1997) vedtok Stortinget i 1997 at det skulle opprettes et landsdekkende tilbud av det som nå kalles legemiddelassistert rehabilitering (LAR). Fra en sped begynnelse på slutten av 1980-tallet ble således substitusjonsbehandling et nasjonalt tilbud fra 1999 (Waal et al., 2001).

For å systematisk evaluere pasientgruppen som mottar LAR utviklet Seksjon for kliniske rusmiddelproblemer (SKR), Universitetet i Oslo (UiO),

”Statusrapportskjemaet for pasienter under medikamentassistert rehabilitering (MAR)”. Skjemaet kalles til daglig ”Statusrapportskjemaet” (vedlegg A).

Instrumentet ble utprøvd i en pilotstudie høsten 2001 (Kornør & Waal 2004), og ble sist revidert høsten 2003. Skjemaet er ikke testet for verken pålitelighet (reliabilitet) eller gyldighet (validitet). Siden 2002 er det blitt brukt nasjonalt som et standardisert evalueringsverktøy for pasientgruppen, og danner på et sammenstilt nivå grunnlag for sentrale vurderinger av pasientgruppen:

- Suksessmåling av norsk LAR-behandling i et internasjonalt perspektiv.
- Sammenligning mellom pasienter i ulike behandlingstilbud (LAR/ikke-LAR-sammenligninger).
- Regionale forskjeller mellom ulike LAR-tilbud.

I 2004 ble den første systematiske evalueringen av LAR-tilbudet i Norge utført på oppdrag av Sosial- og helsedirektoratet. Evalueringene er i stor grad basert på data fra Statusrapportskjemaet (Hansen, Kornør & Waal, 2004a, Hansen, Kornør & Waal, 2004b, Hansen & Waal, 2004, Moen, Hansen, & Waal, 2004, Vindedal, Steen, Larsen & Knoff, 2004). Evalueringsrapportene er offentlige publikasjoner som er forelagt myndigheter og politikere. Det er en metodesvakhet at evalueringskomiteene brukte data av ”ukjent” kvalitet. Når resultatene fra en undersøkelse blir gjenstand for offentlig oppmerksomhet, og kan danne grunnlag for politiske beslutninger, er verdien av god datakvalitet opplagt.

1.1 Forskningsspørsmål

Målet med denne studien var å undersøke datakvaliteten til Statusrapportskjemaet. Studien stiller forskningsspørsmålene:

A) Kjenner utfyllerne pasientens situasjon?

Skjemaet fylles ut av den behandleren som kjenner pasienten best. Variablene i skjemaet er valgt på grunnlag av en minimumsforventning i forhold til hva behandleren bør kjenne til av pasientens situasjon. Forholdet mellom registrerte

”statusskjemaer”, og LAR-pasienter i behandling var henholdsvis 1774/2356 (75 %) og 2168/2755 (78 %) i 2002 og 2003 (Hansen et al., 2004a). Til tross for en tilfredsstillende svarprosent for de årlige undersøkelsene var det et generelt problem at en vesentlig andel av de registrerte skjemaene ikke var komplette. Dette viste seg ved utfylte svarruter, feil utfylte svarruter og bruk av svarkategorien ”ukjent”. I 2002 og 2003 var den gjennomsnittelige andelen manglende svar per variabel henholdsvis ti og ni prosent (Hansen et al., 2004a). Slike andeler er for store til å kunne skyldes tilfeldigheter som registreringsfeil, og tyder på at mange av skjemautfyllerne ikke har den informasjonen statusrapportskjemaet etterspør.

B) Er det samsvar mellom ulike utfyllere av statusrapportskjemaet?

Manglende samsvar mellom utfyllere kan være en indikator på manglende kjennskap til pasientens situasjon, svak dokumentasjon eller manglende motivasjon/presisjon fra utfyllers side. Dette kan lede til feilaktige vurderinger av pasientgruppen, behandlingstilbudet eller dårlig ressursprioritering innenfor LAR.

Skjemaet er basert på behandlerens vurdering av pasientens situasjon, og er således ikke nødvendigvis hentet direkte fra pasienten. På grunnlag av dette kan en spørre seg om vurderingene til utfyllerne i for stor grad baserer seg på personlig skjønn, og ikke på objektive målinger.

Forskergruppen ved SKR har lenge stilt spørsmål om datakvaliteten til Statusrapportskjemaet kan vurderes som pålitelig. Eksempelvis var andelen pasienter som hadde avgitt minst en opioidpositiv urinprøve betydelig redusert fra 2001 til 2002. Dette gjaldt særlig i regionene i Øst og Vest, hvor andelen i følge H. Kornør (personlig kommunikasjon, 2005) gikk fra 29 % begge steder til henholdsvis 15 % og 18 %. Reduksjonen har ingen åpenbar forklaring, men kan sees i forhold til en mulig rivalisering som oppsto da det under Nasjonalt møte for LAR-tiltakene i august 2001 ble lagt vekt på regionale forskjeller i illegal opioidbruk blant pasientene. Generelt er det også en lav forekomst av angst og depresjon, henholdsvis 12 % i 2002 og 16 % i 2003 undersøkelsen (Hansen et al., 2004a) hos norske pasienter sammenlignet med hva en kan forvente av gruppen på bakgrunn av internasjonale studier. Internasjonalt

lider 1/3 av alvorlig depresjon og omtrent halvparten har en angstlidelse (Ward, Mattick & Hall, 1998). Under det nasjonale kontaktmøtet for LAR i september 2003 ble statusdataenes pålitelighet diskutert, og det ble konkludert med at det var ønskelig å utforske dette nærmere. Denne studien har vært et ledd i kvalitetssikringen, og er utført på oppdrag fra SKR.

2 Teori og bakgrunnsstoff

2.1 Måleinstrumenter

For å vite om et måleinstrument er anvendelig som evalueringsverktøy må datakvaliteten til instrumentet vurderes. Dette fordrer en systematisk testing av måleegenskapene til instrumentet. Måleegenskapene analyseres ved å vurdere instrumentets validitet (avsnitt 2.3) og instrumentets reliabilitet (avsnitt 2.4).

2.2 Utforming av måleinstrumenter

Før en utvikler et måleinstrument må forskeren bestemme seg for hvilke dimensjoner som skal måles, og hvordan de enkelte dimensjonene skal belyses på en tilstrekkelig og stabil måte. For å velge ut variablene, må flere aspekter vurderes (Streiner & Norman, 2003).

Det er viktig at forskergruppen som utformer et instrument undersøker hvilke variabler som er prioritert i sammenlignbare instrumenter. Wechsler (1958) diskuterte åpent hvor viktig tidligere tester var da han utviklet sin IQ-test. Variabler fra andre måleinstrument ble modifisert, og nye hentet for å imøtekomme de spesifikke kravene han stilte. I mange tilfeller var endringene små. På en lignende måte er Manifest Anxiety Scale (Taylor, 1953) basert på Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1951). Det kan være flere fordeler ved å importere variabler fra andre instrumenter. Det er tidsbesparende, og variablene er som regel blitt testet for måleegenskaper. Det er også begrenset hvor mange måter en kan spørre det samme spørsmålet (Streiner & Norman, 2003).

Bruk av pasienter og klinikere er en ofte forsømt ressurs for valg av variabler. Å bruke disse gruppene kan bidra til at variablene i større grad representerer relevante parametere for det som skal måles (Streiner & Norman, 2003). Bruk av fokusgrupper

øker kvaliteten på instrumentet, og er med på å gi instrumentet en form som oppleves som meningsfull av den aktuelle gruppen (såkalt face validity, eller overflatevaliditet). Selv om overflatevaliditeten ikke sier noe om instrumentets vitenskapelige verdi, er den med på å øke informantenes motivasjon til å besvare spørsmålene på en presis måte (Guilford, 1954). Overflatevaliditeten kan testes forholdsvis enkelt ved å be utfyllerne vurdere instrumentet ved hjelp av en skala som går fra ekstremt velegnet til irrelevant (Nevo, 1985). Lawrinson (2005) ønsket ytterligere å videreutvikle dette ved å be klinikerne som skulle benytte instrumentet om å foreta kvalitative vurderinger av BTOM i form av skriftlige tilbakemeldinger.

2.3 Validitet

Validitet oversettes som regel på norsk til gyldighet, og beskriver hvorvidt et instrument reflekterer virkeligheten presist. I litteraturen skilles det mellom tre former for validitet (Streiner & Norman, 2003).

- Content validity (innholdsvaliditet)
- Criterion validity (kriterievaliditet)
- Construct validity (begrepsvaliditet)

Innholdsvaliditet refererer til om variablene eller dimensjonene i instrumentet er representative og dekkende for det som skal måles. Om viktige aspekter ved et instrument er utelatt, vil det være vanskelig å gjøre slutninger basert på data fra instrumentet. For eksempel må en test som måler symptomer på schizofreni inkludere variabler som måler både positive og negative symptomer (Blacker & Endicott, 2003). Innholdsvaliditet vil alltid til en viss grad baseres på skjønn (Polit & Hungler, 1999).

Kriterievaliditet blir undersøkt ved å dokumentere en korrelasjon (sammenheng) mellom det aktuelle instrument og et annet instrument som måler samme fenomen, helst et instrument som vurderes som en gullstandard for å måle den aktuelle tilstand,

eller sykdom. Om testene blir utført på samme tid, og har en sterk korrelasjon, har instrumentet høy kriterievaliditet (Streiner & Norman, 2003, Bordens & Abbott, 2001). Det kan være flere grunner til å utvikle et nytt måleinstrument. Det eksisterende instrumentet kan være for tidskrevende eller kostbart å bruke, eller forskeren kan ønske å utvikle et instrument med sitt eget navn som del av en tittel... ”The Kreek-McHugh-Schlugger-Kellog-scale” ble eksempelvis utviklet fordi forskere ønsket et raskere instrument for å kartlegge grad av avhengighet (Kellog et al., 2003). I Australia ønsket forskere å implementere et enklere instrument for pasienter innen stoff og avhengighet og laget sitt eget National Minimum Data Set (NMDS) (Copeland & Conroy, 2001). På bakgrunn av NMDS utviklet de senere BTOM spesielt for gruppen som mottar medikamentell vedlikeholdsbehandling. Instrumentet ble utviklet med ønske om å redusere tidsbruk i forbindelse med kartlegging og evaluering, uten at dette gikk på bekostning av instrumentets måleegenskaper (Lawrinson et al., 2005).

Når en test er designet for å måle begreper som er varierende og/eller ikke direkte observerbare, eller som er utviklet for å forklare en gitt atferd på bakgrunn av en teori, vil det være nødvendig å vurdere begrepsvaliditeten (Bordens & Abbott, 2001). Begrepsvaliditet kan opprettes ved å sammenligne utfall fra det enkelte instrument med resultater fra andre instrumenter. En vurderer da om måleinstrumentet ”oppfører” seg som forventet i forhold til andre mål. Dette vurderes statistisk og teoretisk opp mot mål en forventer at måleinstrumentet skal korrellere med og mål testen ikke skal korrellere med. Attributter som høyde eller vekt blir definert på bakgrunn av hvordan de er målt, eksempelvis centimeter eller kilo. Når en beveger seg fra fysiske attributter til karakterer av psykologisk art som ikke lar seg observere direkte, benyttes mer abstrakte variabler. Vi kan ikke se angst eller intelligens. Angst kan uttrykkes gjennom fysiske attributter som kaldsvette, økt blodtrykk og vanskeligheter med konsentrasjonen. Disse underliggende faktorer er referert til som begrep (”constructs”) (Streiner & Norman s. 179, 2003). Et begrep karakteriseres som en ”mini-teori” for å beskrive sammenhengen mellom ulik atferd eller holdning. Begrepene stammer fra teori eller kliniske observasjoner før en evner å klassifisere de

inn i et objektivt måleinstrument. Eksempelvis vil differensieringen mellom termene ”flytende” og ”krystallisert intelligens” (Catell, 1963) være begreper som er fundert på teori. (Flytende intelligens er knyttet til evnen til å løse problemstillinger med høy grad av nyhet, den krystalliserte er i større grad knyttet til problemstillinger som kan håndteres med utgangspunkt i tilegnede ferdigheter og kunnskap, basert på erfaring). Det er spesielt to grunner for å ønske å utvikle et nytt instrument på grunnlag av begrepsvaliditet. Begrepet er nytt, og det finnes derfor ikke et egnet måleinstrument, eller forskeren kan være misfornøyd med det eksisterende måleinstrumentet, og ønsker derfor å utvikle et nytt instrument.

2.4 Reliabilitet

2.4.1 Generelt om reliabilitet

Reliabilitet oversettes på norsk til pålitelighet. Reliabiliteten til et mål omhandler dets evne til å produsere like resultater når repeterende målinger foretas under identiske forhold (Bordens & Abbott, s 122, 2001). Om det ikke inntreffer en endring i en egenskap, bør skåringen gi samme resultat uavhengig av hvem som måler, eller tidsavstand mellom målingene. Reliabiliteten til målingene avgjøres således av hvorvidt dette gir resultater på en stabil måte, ikke i hvilken grad dette ”er sant” i betydningen reflekterer virkeligheten helt presist. Et mål kan være reliabelt uten å være valid, men aldri valid uten at det er reliabelt. Dette kommer matematisk til syne ved at et instruments teoretisk maksimale validitet tilsvarer kvadratet av reliabilitetskoeffisienten (Streiner & Norman s.173, 2003). Det er derfor avgjørende at høy grad av reliabilitet etterstrebes under konstruksjon og videreutvikling av et instrument.

Krav til reliable skåringer er:

- Repeterbarhet
- Stabilitet

Med repeterbarhet menes altså at målingene skal gi de samme skårene også ved gjentatte undersøkelser. Målingen må også være stabil. Med det menes at *ulike* utfyllere skal kunne skåre det samme uavhengig av tidspunkt og i hvilket miljø undersøkelsen foretas (Streiner & Norman, 2003).

2.4.2 Metoder for reliabilitetsmåling

Det finnes ulike metoder for å måle reliabilitet (Blacker & Endicott, 2003):

- Interraterreliabilitet (IRR)
- Split-half
- Indre konsistens
- Test-retest
- Parallell-tester

Interraterreliabilitet (IRR) er et mål på samsvar mellom to eller flere individer som vurderer samme subjekt, og som bruker samme informasjon. IRR måler instrumentets stabilitet ”på tvers” av ulike informanter, eller observatører (Robson, s. 221, 1993). Om informantene kommer frem til samme skåre, tyder dette på høy eller god IRR. IRR kan også bli benyttet for skåring av en persons atferd, eksempelvis basert på video-opptak, hvor flere forsøkspersoner vurderer samme person uavhengig av hverandre i samme fysiske omgivelser.

Split-half reliabilitet beregnes ved å dele instrumentet opp i to like deler, enten ved å dele det i to på midten, eller ved å dele det opp i eksempelvis oddetallsledd og partallsledd. Skåre for de to ulike delene av instrumentet beregnes hver for seg og sammenlignes (Streiner & Norman, 2003). Et høyt samsvar mellom de to ulike delene indikerer høy reliabilitet. Dette er vesentlig fordi en i de fleste situasjoner ønsker at testen skal være homogen. Med det menes at variablene i et instrument skal berøre forskjellige sider av samme egenskap, ikke forskjellige sider av ulike karaktertrekk. Ulempen ved split-half er at en ikke kan være sikker på at de to delene er like.

Indre konsistens er et mål på samsvar mellom de individuelle komponentene i et måleinstrument. Hver enkeltvariabel blir her sett på som et selvstendig mål i en sammenhengende konstruksjon (Blacker & Endicott, s. 7 2003). Dette beregnes med Cronbachs Alpha-koeffisienten, og er på mange måter en videreutvikling av split-half reliabilitet for beregning av indre konsistens. Slike beregninger er blitt muliggjort av økt tilgjengelighet av datakraft (les: moderne computere og avanserte statistiske programpakker). Metoden går ut på å beregne alle mulige varianter av split-half-reliabilitet, og uttrykkes som gjennomsnittet av disse (Murphy & Davidshofer, s 120, 1998). Cronbachs Alpha-koeffisient gir et bilde av i hvilken grad leddene i en gitt skala korrelerer (samsvarer), og om de belyser dimensjonene på en god måte (Blacker & Endicott, 2003).

Test-retest er et mål på samsvar hvor målet krever at samme utfyller tester instrumentet på to forskjellige tidspunkt. Test-retest er ofte benyttet innen måleinstrumenter for selvrapporing. Fordelen med denne tilnærmingen er at den gir et reelt bilde av stabilitet over tid. I motsetning til eksempelvis split-half og Cronbachs Alpha, kritiseres denne for at den betinger at personen som fyller ut instrumentet ikke endrer seg nevneverdig. Tidsintervallet må være lite nok til at det er minimal, eller ingen sjanse for at endring hos individet finner sted, men stort nok til at respondenten ikke husker hva som ble fylt ut i den originale situasjon (Blacker & Endicott, 2003).

Parallele tester tar utgangspunkt i to tester som blir presentert som samme test og administrert samtidig. Responsen til de to testene er derved separert og skåret individuelt. Fordi begge testene foretas samtidig, har tidsdimensjonen ingen betydning, jamfør utfordringen til Test-retest. Nødvendigheten av å bruke alternative tester i de to halve delene introduserer imidlertid problemet med at de to testene faktisk er ekvivalente (Bordens & Abbott, 2001).

2.4.3 Målefeil/variasjon

Alle kvantitative måleresultater består av to komponenter; en som reflekterer det som blir målt og en som representerer feilkilder (Rothstein & Echternach, 1993). Dette kan illustreres som vist i figur 1.

$$\boxed{\text{MÅLERESULTAT}} = \text{SANT MÅL} + \text{MÅLEFEIL}$$

Figur 1: De to feilkomponentene i et observert måleresultat

Hensikten med reliabilitetsteori er å estimere målefeilene og foreslå metoder for å minimere disse (Murphy & Davidshofer, 1993). Målefeilene kan ha form av en systematisk variasjon eller en usystematisk variasjon. Med systematisk variasjon forstås variasjon som kommer av at den enkelte utfyller eller pasient har ulike verdinormer, karakteristika eller andre forskjeller som påvirker resultatene som oppnås. Mennesker er i kontinuerlig endring. Verdinormer og karakteristika endrer seg således over tid. Resultatene knyttet til den enkelte pasient vil derfor naturlig variere tilsvarende. Hvis resultatene for pasienten varierer på en konsistent måte, representerer dette en systematisk variasjon og kan ikke betraktes som objektive feil. Hvis det derimot oppstår ”uventede” resultater, det vil si resultater som avviker fra det en kunne forvente basert på øvrige skåringer, representerer dette en usystematisk variasjon (Streiner & Norman, 2003).

Det er i hovedsak tre faktorer som påvirker målingene og som dermed kan begrense reliabiliteten ved målingen (Rothstein & Echternach, 1993).

- Individet som måles
- Måleinstrumentet
- Personen som utfører målingen

Om individet som måles lar seg påvirke av faktorer som er vanskelig å kontrollere, får en svingninger i målingene og derved også ureliable skåringer. Den emosjonelle situasjonen når undersøkelsen blir foretatt vil også kunne variere. Er utfylleren i godt humør når skjemaet blir fylt ut, eller er han/hun i emosjonell ubalanse? Det er også viktig å reflektere over i hvilken grad den som utfører testen selv påvirker testpersonen enten ved sin væremåte eller holdninger. Dette kan være problematisk i eksempelvis standardiserte intervjuer.

Variasjon i forhold til selve måleinstrumentet henspiller seg på prosedyrene rundt selve testen, og standardiseringen av denne. Tolker ulike personer måleinstrumentet likt? Er variablene beskrevet godt nok? Har de som skårer fått tilstrekkelig informasjon om hva de enkelte variablene er ment å måle? En reliabilitets og validitetsstudie av datakvaliteten til et måleinstrument som ble designet for å vurdere kunnskap, persepsjon og praksis til helsearbeidere innen alkoholistsorgen i Frankrike viste at instrumentet verken var valid eller reliabelt eller sensitivt. Forskerne konkluderte derfor med at funn på bakgrunn av dette instrumentet kunne være misvisende (Jausent, Labarere, Boyer & Francois. 2004). Forskerne fremhevet at en for å kunne evaluere alkoholistsorgen i Frankrike, må bruke pålitelige måleinstrumenter som måler det instrumentene skal måle.

En annen feilkilde knytter seg til den som utfører testen. At ulike utfyllere har ulike verdsett, representerer i dette arbeidet en systematisk variasjon (forutsatt at verdsettet ikke endrer seg i løpet av utfyllingen). Blir prosedyrene rundt testen fulgt? Utfyllerens grad av kliniske kompetanse er også viktig. Er utfylleren novise, eller ekspert? En studie av det internasjonalt mye brukte verktøyet Global Assessment Scale (GAF), som graderer alvorligheten av symptomer og psykososialt funksjonsnivå, konkluderte med at interraterreliabiliteten til vurderingene ikke var gode nok i en klinisk kontekst (Vatnadal, Friis & Opjordsmoen, 2005). De fant at trenede klinikere var mer samstemte enn noviser på feltet. Flere studier beskriver nødvendigheten av å implementere en standardisert veiledning for alle instrumenter som blir benyttet til forskning og evaluering for å øke samsvaret mellom utfyllere og

presisjonsnivået til utfyllingen (Müller et al., 1997), (Müller & Dragicevic, 2003), (Cussick, Vasquez, Knowles & Wallen, 2005). På grunn av optimale testforhold og veiledning i forkant av studien har publiserte artikler som handler om instrumentreliabilitet en tendens til å oppnå høyere grad av reliabilitet, enn det som kan forventes i ordinær klinisk praksis (Blacker & Endicott, s. 10 2003).

Når en leser publiserte artikler om reliabilitet er det viktig å huske at den rapporterte reliabilitet av måleinstrumentet avhenger av flere faktorer som ligger utenfor selve målingen (Blacker & Endicott, s. 10 2003). Som med all statistikk er datamaterialstørrelsen av betydning. Utvalgets størrelse er av betydning for å kunne generalisere funnene til å gjelde grupper som ikke har vært med i selve studien.

Selve situasjonen utfyllingen foregår i kan også påvirke resultatene. De fysiske omgivelsene; eksempelvis om selve rommet utfyllerne sitter i er mørkt og utrivelig, eller om det er forstyrrende elementer som for eksempel telefonringing, kan påvirke resultatene (Blacker & Endicott, 2003). Reliabiliteten til måleinstrumenter kan også bli redusert på grunn av arbeidsmiljøet (McLellan, Carise & Kleber, 2003). Som en del av en større studie i USA fant de at turnover innad i personalgruppen var så høy at de stilte spørsmål både ved evnen til å dokumentere relevante pasientopplysninger, og om disse kunne vurderes som reliable (McLellan et al., 2003).

2.5 Måling av interraterreliabilitet

Interraterreliabilitet (IRR) er et mål på samsvar mellom to eller flere individer som evaluerer samme subjekt, og som bruker samme informasjon. Det er fire forskjellige mål som blir benyttet for å regne samsvar mellom to eller flere individer (Bordens & Abbott 2001).

- Cohens kappa koeffisient κ
- Intraklasse korrelasjon
- Pearsons korrelasjonskoeffisient r
- Prosentvis samsvar

2.5.1 Reliabilitetskoeffisienter

Korrelasjonskoeffisienter er et statistisk mål på den gjensidige sammenhengen mellom to eller flere variabler (Murphy & Davidshofer, s. 85, 1993). Det debatteres innen metodefaget hvilken korrelasjonskoeffisient som er mest hensiktsmessig å benytte i reliabilitetsstudier (reliabilitetskoeffisienter). Herunder presenteres de mest brukte:

Kappakoeffisienten (κ)

Cohens Kappa (κ) er den foretrukne statistiske analysen i interraterreliabilitetsstudier (Streiner & Norman, 2003). Den blir benyttet som et mål på samsvar mellom to eller flere utfylleres skårer på en *kategorisk* variabel, korrigert for tilfeldig samsvar (Bordens & Abbott, 2001). I medisinsk forskning er det mange situasjoner som kun har to mulige utfall. Eksempelvis; positiv versus negativ og normal versus abnormal. Analysen innebærer at den går inn i de to samsvarscellene (ja/ja, nei/nei) og beregner hvor mange av disse samsvarsresponsene som kun ville blitt styrt av sjansen for samsvar alene. Skalaen til kappa varierer fra minus 1 til 1. Negative verdier indikerer at det er mindre samsvar mellom utfyllerne enn en skulle forvente ved helt vilkårlige svar. En κ -verdi på 0 indikerer at samsvaret er vilkårlig. En κ på 1, indikerer perfekt samsvar (Fleiss & Cohen, 1973). Det er ingen absolutte definisjoner på hva en god κ skal være, og vurderingen må ta hensyn til hvilke krav til reliabilitet som stilles i den enkelte kliniske kontekst (Blacker & Endicott, 2003, Streiner & Norman, 2003).

Landis & Koch (1977) og Altman (s. 404,1991) har skalaer med litt ulike krav til κ -verdi. Tabell 1 viser Altman sin foreslåtte skala.

Tabell 1 – Krav til κ - verdier

K-verdi	Styrke på samsvar
< 0.20	Dårlig
0.21-0.40	Svak
0.41-0.60	Moderat
0.61-0.80	God
0.81-1.00	Svært god

Estimering av κ forutsetter at dataene kan parres og settes opp i en krysstabell/frekvenstabell (Streiner & Norman, 2003). For å foreta en kappaberegning må alle svaralternativene i en variabel være benyttet av begge utfyllerne. κ korrigerer ikke for om utfyllerne har unnlatt å svare ved å bruke svarkategorien ”ukjent”, slik at når to utfyllere er enige om at de ikke vet om pasienten er smittet (tabell 2) vil dette være med på å øke κ .

Tabell 2 – Eksempel på krysstabell fra Statusrapportskjemaet

Blodsmittestatus, HCV (n=10)		Sekundærkontakt			
		Ikke smittebærer	Smittebærer	Ukjent	Total
Primærkontakt	Ikke smittebærer	1	0	0	1
	Smittebærer	0	4	4	8
	Ukjent	0	0	1	1
	Total	1	4	5	10
$\kappa = 0,66$ $p = 0,0350$					

Tabell 2 viser at to utfyllere har vurdert 10 pasienter. De er enige om at én pasient ikke er smittebærer, og at fire er det. De er også samstemte om at de ikke fant informasjon på én pasient. I fem tilfeller tilkjennega sekundærkontakten at hun ikke kjente smittestatus, hvorav fire av disse var tilfeller der primærkontakten vurderte pasienten som smittebærer. Totalt var det enighet i seks av ti tilfeller.

For ordinale (rangerte) kategoriske variabler benyttes ofte en *vektet* kapp (Blacker & Endicott, 2003), det vil si at hvert nivå av uenighet tildeles en vekt. Fordelen med å benytte vektet kapp er at den ikke utelukkende fokuserer på absolutt enighet/uenighet mellom de ulike svarene, men også tar hensyn til avstanden mellom disse (Altman, 1991).

Pearsons korrelasjonskoeffisient r

Pearsons korrelasjonskoeffisient (Pearsons r) blir ofte benyttet i forskningssammenheng som et mål på reliabilitet mellom *kontinuerlige* variabler. Koeffisienten er basert på regresjonsanalyse, og måler hvorvidt samsvar mellom to

variabler kan uttrykkes med en rett regresjonslinje. Om de forskjellige observatørene er konsistente i sin bedømming, vil r være sterk og positiv. Pearsons r har en verdi mellom minus 1 og 1. En korrelasjon på 0 innebærer at det ikke er en lineær relasjon mellom de to raternes skårer. Pearsons r kan benyttes når utvalget er randomisert (tilfeldig) og når minst en av variablene er normalfordelt i populasjonen (Altman, s. 279, 1991). For å ha en valid tillit til r må begge variablene være normalfordelte.

Intraklassekoeffisienten

Intraklassekorrelasjon ble introdusert i 1925 (Fisher, 1925). Intraklassekoeffisienten, på engelsk Intra Class Correlation Coefficient (ICC), finnes i flere varianter, og benyttes for måling av *kontinuerlige* variabler. Koeffisienten brukes der variablene som skal sammenlignes befinner seg i samme *klasse* (jamfør intraklasse), det vil si der variablene har den samme *metriske* målestokk og samme *varians*. Hvis variablene ikke er i samme klasse, vil Pearsons r være den foretrukne korrelasjonskoeffisient – det er en *interklassekoeffisient* (McGraw & Wong, 1996). For eksempel: Hvis en ønsker å måle korrelasjonen mellom IQ-nivå og skolekarakterer, vil Pearsons r være den foretrukne korrelasjonskoeffisient (Murphy & Davidshofer, 1998). De metriske verdiene (IQ-poeng, karakterer) er ikke de samme, og det er heller ikke grunn til å tro at variansen mellom de to datasettene er sammenfallende. Hvis en ønsker å måle korrelasjonen mellom resultatene til to dommere i en dansekonkurranse, vil ICC være den foretrukne korrelasjonskoeffisient, da både de metriske variablene (poeng) og variansen er de samme.

En skiller mellom enveis- og toveis ICC-modeller, avhengig av om det er en eller to kilder til systematisk varians (Shrout & Fleiss, 1979). Videre skilles det mellom hvorvidt en ønsker å måle konsistens eller absolutt samsvar mellom utfyllerne. Om en ønsker å måle absolutt samsvar, er dette et strengere krav til samsvar mellom utfyllernes skåringer enn om en ønsker å måle konsistens (McGraw & Wong, 1996). Følgende eksempel (tabell 3) illustrerer dette: Tre utfyllere gir to pasienter følgende skåringer.

Tabell 3 – Eksempel på beregning av ICC

	Utfyller 1	Utfyller 2	Utfyller 3
<i>Pasient 1</i>	2	4	6
<i>Pasient 2</i>	4	6	8

Ved å beregne ICC ut fra et ønske om å måle konsistens, får vi en ICC-verdi på 1,0. Utfyller 1, 2 og 3 er perfekt konsistente i sine vurderinger. Ved å beregne ICC ut fra et ønske om å måle absolutt samsvar, får vi en ICC-verdi på 0,67. Utfyller 1, 2 og 3 har som vi ser ikke absolutt samsvar i sine vurderinger. ICC opererer forøvrig med samme skala som κ , og kan bli tolket som denne (Fleiss & Cohen, 1973).

2.5.2 Prosentvis samsvar

Prosentvis samsvar er et simpelt estimat for intertaterreliabilitet. Målet beregner hvor mange ganger informantene oppnådde samsvar, for så å dele dette tallet på det totale antallet observasjoner. Eksempelvis hvis utfyllerne i en studie var enige i åtte av ti observasjoner, vil prosentsamsvar bli: $\frac{8}{10} = 80\%$. Et prosentvis samsvar på rundt 70 % regnes normalt som tilfredsstillende (Bordens & Abbott, 2001). Prosentsamsvar har flere metodiske svakheter. For det første kan slike beregninger undervurdere samsvaret fordi fullstendig enighet mellom utfyllerne, eller nøyaktig like svar kreves. (Mitchell, 1979, sitert i Bordens & Abbott s. 193, 2001). Problemet kan reduseres ved å benytte et mindre strengt krav til samsvar mellom utfyllere ved for eksempel å konstatere samsvar der utfyllere har plassert svar innenfor en gitt spredning som forskeren selv har definert som tilstrekkelig. Problemet med underestimering av slik enighet vil kun være til stede i tilfeller der svarene har flere enn to mulige utfall. Dernest vil prosentsamsvar kun gi et grovt estimat av samsvar, da det i enhver undersøkelse foreligger en mulighet for at utfyllere oppnår et visst samsvar helt vilkårlig. For det tredje vil målinger som forekommer med veldig høy eller lav

frekvens kunne ha en høy grad av enighet av utfyllere. Eksempelvis vil ”vanlige” diagnoser være overrepresentert i forhold til mer sjeldne diagnoser (Rogers, 2003). Utfylleren kan også legge egne antagelser og verdier til grunn for sine svar, og disse kan vilkårlig samsvare med en annen utfyllers besvarelse hvis utfyllerne har like verdier for eksempel i en kollegagruppe. I disse tilfellene vil prosentvis samsvar overvurdere interrater samsvaret. Prosent samsvar mellom utfyllere beskriver ikke reliabilitet, fordi den ignorerer variasjon mellom subjekter, og forsømmer å regne med sjansen for samsvar (Bordens & Abbott s. 193, 2001). Til tross for metodesvakhetene er det i en del tilfeller likevel hensiktsmessig å bruke prosentvis samsvar som et estimat på interraterreliabilitet. Spesielt gjelder dette i de tilfeller der det ikke er mulig å beregne reliabilitetskoeffisienter.

2.5.3 Ulike typer variabler

En kan dele variabler i to hovedgrupper, kategoriske og kontinuerlige variabler.

Kategoriske variabler

En *kategorisk* klassifikasjon er en kvalitativ vurdering av tilstedeværelse eller uteblivelse av en gitt egenskap (Blacker & Endicott s. 7, 2003). De mest benyttede kategoriske klassifikasjonene innen psykiatrien er diagnoseverktøyene. Andre kategoriske klassifikasjonssystem kan vurdere hensiktsmessigheten av for eksempel medisinerer. Kategoriske variabler kan være todimensjonale, eksempelvis kjønn eller inneholde flere kategorier. Kategoriske variabler deles i to grupper, nominale variabler og ordinale variabler. Nominale variabler (figur 2) rangerer ikke svarkategoriene, mens ordinale plasserer svarkategorier i en rangert rekkefølge (figur 3).

A1. Beskjeftigelse <input type="checkbox"/>		
a. Yrkesstatus*		
0. Uten beskjeftigelse	2. Deltidsjobb	4. Deltidsjobb og
1. Heltidsjobb	3. Under utdanning	9. Ukjent

Figur 2: Eksempel på nominal variabel fra Statusrapportskjemaet

B7. Helhetlig vurdering av rusmestring siste 4 uker *		<input type="checkbox"/>
0. Ingen tegn til rusing verken klinisk eller ved prøver		
1. Enkeltstående tegn til rusmiddelbruk, bedømt som tilfeldig/mindre betydningsfullt		
2. Utilstrekkelig ruskontroll, utskriving ikke aktuelt*		
3. Alvorlige rusproblemer, fare for utskriving*		
4. Utskrevet pga vedvarende rusmisbruk		

Figur 3: Eksempel på ordinal variabel fra Statusrapportskjemaet

Kontinuerlige variabler

En *kontinuerlig* variabel sørger for en kvantitativ vurdering av en spesifikk attributt knyttet til et bestemt subjekt (Blacker & Endicott s. 8, 2003). Dette kan være langs et kontinuum av intensitet, frekvens eller alvorlighetsgrad. Eksempel kan være måling av funksjonsevne, livskvalitet og pasienttilfredsstillelse.

A7. Døgndose i mg	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
999. Ukjent	

Figur 4: Eksempel på kontinuerlig variabel fra Statusrapportskjemaet

Kategoriske og kontinuerlige variabler er også av betydning når en skal analysere materialet. Dette presenteres mer inngående under avsnitt 3.6, Statistisk analyse.

2.6 Utfyllerdimensjonen

I litteraturen skilles det mellom den *optimale* utfylling av et skjema, noe som er målet sett fra forskerens synsvinkel, og et *tilfredsstillende* utfylt skjema som ofte vil være et mål sett fra utfyllerens perspektiv (Krosnick, 1991). Mulige grunner til at utfylleren svarer tilfredsstillende, men ikke optimalt, kan være at personen opplever skjemaet som uforholdsmessig langt, og at det er en kognitiv utfordring å fylle det ut.

Klinikeren kan ikke hjelpe pasienten, ei heller kan forskere presentere data som er pålitelige med mindre ærlige svar er avgitt (Streiner & Norman, s. 80, 2003). En annen forklaringsmodell er at formålet med å fylle ut måleinstrumentet kan virke meningsløs og derfor irrelevant for den enkelte, og at utfyllerne derfor adopterer strategier som tillater at de fullfører skjemaet uten å legge for mye arbeid i det.

Krosnick (1991) beskriver i alt seks mulige forklaringer på hvorfor utfylleren velger å svare tilfredsstillende, og ikke optimalt:

- Personen velger det første mulige tilgjengelige svaralternativet som virker fornuftig.
- Personen er føyeelig, og gir de svarene hun tror forskeren ønsker å få.
- Utfyllerne ønsker ikke åpent å gå god for kontroversielle endringer, og svarer derfor mer konservativt enn hennes reelle standpunkter faktisk er.
- Utfylleren selekterer et svar fra det første svaralternativet, og benytter for enkelhets skyld det samme alternativet også på de resterende variablene.
- Utfylleren lar være å svare, enten ved å stille ruten tom eller ved å benytte ”ukjent” som svarkategori.
- Utfylleren kan mentalt kaste ”mynt og kron”. Dette innebærer at utfylleren velger å krysse av helt vilkårlig (Krosnick, 1991).

Å svare på et spørsmål innebærer en kognitiv utfordring. Avhengig av teori, er det fire (Tourangeau, 1984), eller fem (Schwarz & Oyserman 2001) steg som kreves for at en person skal minimalisere støydimensjonen i forhold til annen kognitiv aktivitet.

I denne studien presenteres de fem kognitive dimensjonene som Schwarz og Oyserman er kommet frem til:

- **Å tolke spørsmålet.**
Forstår og tolker utfylleren det samme med spørsmålet som forfatteren av skjemaet? Om en variabel etterspør informasjon om hvor ofte utfylleren er sint, og en av svarkategoriene er ”mindre enn en gang i året” kan respondenten oppleve at det kun er store raseriutbrudd som ønskes dokumentert, og ikke mindre frustrasjoner. Respondenten kan vurdere det som lite sannsynlig at en er sint mindre enn en gang i året.
- **Å tilbakekalle den relevante atferd eller situasjon.**
Forfattere av skjemaer har en tendens til å overvurdere menneskers evne til å tilbakekalle tidligere hendelser. Resultatene av en stor studie viste at 42 % ikke husket at de hadde vært på sykehus ett år etter hendelsen (Cannell, Fisher and Bakker, 1965).
- **Vurdering og respons**
Variablene i et måleinstrument følger sjelden det samme mønsteret som er lagret i hukommelsen til den enkelte utfyller. Dette betyr at den som skårer må bruke ulike strategier for å gjenkalle hendelsene. Overveielsen har to mulige feilkilder. For det første har de fleste en tendens til å overvurdere sjeldne hendelser. Dernest, og som en følge av dette, har de fleste en tendens til å undervurdere hvor uvanlige hendelsene var (Sudman, Bradburn & Schwarz. 1996).
- **Kode inn riktig svaralternativ.**
Etter å ha vurdert frekvensen til en gitt atferd eller handling må informanten oversette dette til den skalaen eller de svarkategoriene som er benyttet i instrumentet. Svaret representerer således en oversettelse fra respondentens egne ord til forskerens. I all oversettelse går noe av den originale presisjonen tapt.
- **Redigering av svar.**
Hva informanten faktisk tenker, og hva informanten er villig til å fortelle

forskeren, er ikke nødvendigvis samme ting. Forskeren og utfylleren har ulike perspektiver og ulike agendaer, noe som kan påvirke utfyllerens svar betraktelig (Schwarz & Oyserman 2001).

2.7 Manglende svar

Manglende svar resulterer i ukomplette datasett og er et problem som forekommer i mange typer datasett (Olinsky, Chen & Harlow, 2003). Ukomplette data reiser spørsmål om både nytten av dataene og den potensielle støydimensjonen for de som benytter dataene fordi ”ikke-responderer” ofte er systematisk forskjellig fra responderer (Montalto & Sung, 1996). Det har vært lite forskning som har konsentrert seg om manglende svar i psykologisk/medisinsk litteratur. Innen markedsføring, statistikk, økonomi og biometri er det imidlertid forsket mye på manglende data. Det er viktig at mer kunnskap og kjennskap til denne dimensjonen kommer frem, spesielt når data er savnet i et systematisk mønster (Roth, 1994). Manglende data kan redusere longitudinelle utvalgsstørrelser, samt være potensielle støybarometre. (Newman, 2003). Manglende data forårsaker to primære problemer. For det første betyr det at forskeren mister styrke i datamaterialet sitt. Denne reduksjonen innebærer at en mister mulighet til å benytte seg av statistiske analyser for å påvise en eventuell signifikant effekt av et datasett. For det andre berører dette presisjonen til de estimerte parametrene (Roth, 1994).

2.8 Evaluering av legemiddelassistert rehabilitering (LAR)

LAR ble i utgangspunktet betegnet som ”metadonassistert rehabilitering” (MAR). Metadon ble erstattet med ”legemiddel” i 2000, fordi en ønsket å ha mulighet til å også kunne benytte andre medikamenter i substitusjonsbehandling. I hovedsak er det buprenorfin (Subutex®) som blir benyttet i tillegg til metadon (Vindedal et al., 2004).

Det er åtte år siden behandling med metadon og buprenorfin ble et tilbud til heroinavhengige i Norge. I denne perioden har det vært en sterk økning i antall pasienter, fra 200 til cirka 3000 årsskiftet 2004/2005. I tillegg har mer enn 1000 pasienter vært innskrevet, og siden avbrutt eller sluttet behandlingen (Sosial- og helsedirektoratet, 2005). Tilbudet er landsdekkende og finansieres av staten, fylkeskommune og kommune (Hansen, Kornør & Waal, 2004b). Behandlingstilbudet er organisert gjennom regionale sentere tilknyttet helseforetakene. LAR ØST fungerer også som et nasjonalt kompetansesenter. Internasjonalt er behandling med metadon den mest evaluerte behandlingsform for opioidavhengighet (Farrel et al., 1994). Den dominerende konklusjon fra disse studiene er at metadon er en effektiv behandling for heroinavhengighet (Gibson, Flynn & McCarthy, 1999). Metadon og buprenorfin er kraftige opioider. Feilaktig brukt på pasienter som ikke er opioidavhengige, eller som kun har utviklet en svak avhengighet, kan resultere i at en kan påføre pasienten en livslang opioidavhengighet som personen i utgangspunktet ikke hadde. Uforsvarlig administrering og oppbevaring kan medføre risiko for at pasienten selger medisinen, eller at uvedkommende i vanvare inntar medisinen (Sosial- og helsedirektoratet, 2005). Metadon og buprenorfin binder seg til morfinreseptorene i hjernen, slik at det blokkerer for effekt av andre morfinstoffer (Waal & Frogopsahl, 2001). Norge følger en "høydosemodell". Denne er valgt på bakgrunn av en teori om at hvis metadondosen er høy, vil den euforiske effekten av heroin reduseres, samtidig som rustrang og abstinensreaksjoner forhindres. Teorien støttes av randomiserte, kontrollerte forsøk (Farrell et al., 1994, Joseph, Stancliff & Langrod, 1998, Verster & Buning, 2000).

Den norske LAR-modellen skiller seg også fra andre land ved at det er lagt større vekt på rehabiliteringsperspektivet og grad av kontroll og oppfølging (Kornør & Waal 2004). Målgruppen for LAR er "rusmiddelmissbrukere som har hatt et langvarig, opiatdominert misbruk uten at andre behandlings-, rehabiliterings- og omsorgstiltak har forhindret at tilstanden domineres av opiatavhengighet" (Sosial- og helsedepartementet, Rundskriv I- 35/2000). "Legemidler skal være ett av flere

virkemidler i et helhetlig rehabiliteringsopplegg, og målet for rehabiliteringen er å assistere pasienten med å:

- Komme seg ut av de hardt belastede misbrukermiljøene,
- Nyttiggjøre seg andre behandlings-, rehabiliterings- og omsorgstiltak,
- Redusere skadene av misbruket og faren for overdosedødsfall,
- Bedre den fysiske og sosiale funksjonsevnen, og
- Oppnå rusfrihet, bedret livskvalitet og – så langt som mulig – yrkesmessig og sosial rehabilitering” (Sosial- og helsedepartementet, s. 2, Rundskriv I-35/2000).

Statusrapportskjemaet (vedlegg A) ble utviklet i 2001 av SKR på oppdrag av og i samarbeid med de andre nasjonale tiltakene fordi en ønsket å evaluere pasientgruppen som mottar LAR. Siden 2002 er det blitt brukt nasjonalt som et standardisert evalueringsverktøy for pasientgruppen. Intensjonen var at instrumentet også skal kunne danne grunnlag for longitudinelle studier av pasienter i behandling (Hansen et al., 2004a). Statusrapportskjemaet blir fylt ut årlig av den behandleren som kjenner pasienten best og tar cirka 15 minutter å fylle ut. Om pasienten blir utskrevet, eller velger å slutte med LAR, fylles skjemaet ut når pasienten avslutter behandlingen (Hansen et al., 2004a).

Statusrapportskjemaet inneholder 53 separate variabler. Variablene er valgt på bakgrunn av en minimumsforventning til den kjennskap de behandlingsansvarlige har til sine pasienter og skal således være godt dokumentert i journal. (Hansen et al., 2004a). Variablene er ment å passe med den klientkartleggingen som i dag er praksis i stoffbehandlingsfeltet. En variabel er hentet fra Addiction Severity Index (ASI), de øvrige variablene er vurdert som relevante mål på pasientens spesielle situasjon i LAR (Vedlegg A, s. 2). Statusrapportskjemaet er delt inn i tre hovedavsnitt: A. *AKTUELL SITUASJON*, B. *SISTE FIRE UKER FØR UTFYLLINGSDATO* og C. *SISTE ÅR*. Under aktuell situasjon spørres det etter informasjon om demografiske forhold, blodsmittestatus og medikamentell behandling, samt utleveringsordning og

kontrolltiltak. I seksjon B, blir behandler bedt om å fylle ut eventuell kontakt med behandlingsapparatet. De blir også bedt om å vurdere pasientens psykiske tilstand og om pasienten har noen somatiske sykdommer som påvirker livskvalitet. Under dette avsnittet skal det også rapporteres for stoff- og alkohol bruk. I seksjon C, rapporteres det for eventuell kontakt med justisvesen, overdoser og selvmordsforsøk, samt pasientens avholdenhet fra stoff.

Statusrapportskjemaet ble utviklet som et alternativ til det internasjonalt mer kjente måleinstrumentet ASI. Dette er det mest benyttede måleinstrumentet innen evaluering av stoffavhengige (McLellan et al., 1992). Det er oversatt til en rekke språk, herunder norsk (Hidle, Lauritzen & Skretting, 1997). ASI er et 45-60 minutters standardisert intervju som blir administrert av en sertifisert kliniker eller intervjuer. Målet er at intervjuet skal få frem respondentens selvrapporterte problemer på syv områder (Mäkelä, 2003). Instrumentet er testet internasjonalt for validitet og reliabilitet med varierende grad av reliabilitet (Zanis, McLellan, Cnaan & Randall, 1994, Mäkelä, 2003). En tilsvarende studie er ikke foretatt i Norge. Det er flere ulemper med ASI. Instrumentet krever sertifisering av intervjuer og er designet for både klinisk bruk og forskning. Enkelte variabler kan derfor virke uhensiktsmessige (Greenfield, 2004).

Innen psykiatri og stoffbehandling er det generelt behov for raske og pålitelige måleinstrumenter. I Australia ble The Brief Treatment Outcome Measure (BTOM) utviklet, et standardisert intervju som tar cirka 15 minutter. Instrumentet ble utviklet etter en omfattende gjennomgang av litteratur om eksisterende instrumenter. De etablerte også en fast rådgivningsgruppe for innhold og implementering, foretok prøveintervjuer for å undersøke reliabilitet og validitet, samt testet gjennomførbarheten gjennom en 30 måneders pilotundersøkelse. BTOM vurderes som et kortfattet, valid og reliabelt måleinstrument som kan benyttes til evaluering av behandlingsresultater både for klinikken og av forskere (Lawrinson et al., 2005). Instrumentet har foreløpig ikke vært utprøvd i Norge.

3 Metode

Datamaterialet ble innhentet høsten 2004. Studien fant sted i LAR ØST sine lokaler i Oslo, og på Gjøvik sosialsenter.

3.1 Utvalg og utvalgsstørrelse

For å komme frem til utvalget, ble det tatt utgangspunkt i samtlige pasienter i LAR fra kommunene Oslo og Gjøvik. Disse var registrert i databasen Rusdata. Dette utgjorde pasientpopulasjonen det ble trukket fra. De to kommunene ble valgt fordi tidligere forskning har vist at det er forskjeller mellom pasienter på disse stedene (Kornør & Waal, 2004), og behandlingstilbudene representerer ulike tilnæringsmåter til LAR. Oslo følger en sentralisert struktur, det vil si at oppfølgingen av pasientgruppen skjer fra et LAR-senter. Behandlerne ved senteret har oppfølgings- og utredningsansvar for pasientene. På Gjøvik er tilbudet desentralisert, pasientene følges opp av forskrivende lege og kontakt ved sosialsenter.

Navnedata på pasientpopulasjonen ble deretter eksportert til SPSS som foretok et tilfeldig utvalg av 10 pasienter fra Gjøvik kommune, og 10 pasienter fra Oslo. I tillegg trakk jeg en tilsvarende reserveliste. Pasientlistene navnga primærkontakt (utfyller 1) og sekundærkontakt (utfyller 2) for pasientene i Oslo. I Gjøvik kommune ble sosialkonsulent definert som utfyller 1 og forskrivende lege som utfyller 2. Hver enkelt pasient ble vurdert av begge utfyllerne. En utfyller ønsket etter utvelgelsen å trekke seg på grunn av pasientens psykiske tilstand. Denne pasienten ble erstattet med en pasient fra reservelisten, som ga to nye utfyllere. Det endelige utvalget besto således av 20 utfyllerpar.

Utvalget i Oslo besto av sykepleiere og sosionomer, de fleste med klinisk videreutdanning i psykiatri eller rusbehandling. En utfyller hadde høyere universitetsutdanning. I Gjøvik kommune hadde sosialkonsulentene sosialfaglig utdanning på høyskolenivå og de forskrivende legene erfaring fra stoffbehandlingsfeltet.

For å fastsette størrelsen på utvalget, ble det tatt utgangspunkt i en artikkel av Cicchetti (1976) som hevdet at minimumsutvalgsstørrelse utvalg (N) kunne beregnes ved følgende formel: $N \geq 2K^2$ der K er antall svarkategorier. Formelen benyttes i tilfeller der $3 < K < 7$ (antall svarkategorier er mellom tre og syv). Vi ser således at utvalgsstørrelsen (N) skal være minimum 18 for variabler med tre kategorier (Cicchetti, 1976). Utvalgsstørrelsen på 40 skjemaer (fra 20 svarpar) ble på bakgrunn av artikkelen vurdert som tilstrekkelig stort til å kunne vurdere skjemaets datakvalitet.

Når innsamlingen av data var gjennomført, forelå det således to utfylte statusrapportskjemaer for hver pasient, (40 skjemaer) som utgjorde datamaterialet i studien.

3.2 Prosedyre for datainnsamling

Før studiens start ble det avholdt ett informasjonsmøte for forsøkspersonene på LAR ØST, samt ett for deltakerne i Gjøvik kommune. På møtene ble formålet med undersøkelsen og innholdet i studien presentert. Utfyllerparene ble deretter instruert i å fylle ut Statusrapportskjemaet for pasienten samme dag. Det ble forøvrig ikke foretatt noen trening av utfylling av måleinstrumentet i forkant av undersøkelsen. Dette fordi Statusrapportskjemaet er ment å være selvforklarende, og en beskrivelse av variablene følger hvert skjema. Det ble satt som krav at de ikke skulle samarbeide med den andre utfylleren, men bruke tilgjengelig informasjon om pasienten slik det gjøres i klinikken. De kunne rådføre seg med andre som kjente pasienten, og ble bedt om å angi bruk av informasjonskilder på eget skjema (vedlegg E).

Hver utfyller ble tildelt et løpenummer, og ført opp på en liste som ble returnert til undertegnede sammen med skjemaene. Utfyllerparene satt i samme fysiske omgivelser på arbeidsplassen mens de fylte ut Statusrapportskjemaet. For å unngå at jeg påvirket forsøkspersonene var jeg ikke fysisk tilstede under utfyllingen.

3.3 Statistiske beregninger

Målingene fra utfyller 1, ble sammenholdt med målingene foretatt av utfyller 2. Personopplysningene og dataene fra skjemaene ble kodet, og deretter lagt inn i ”Statistical Package for Social Sciences” (SPSS) versjon 12 for Windows (SPSS Inc, 2001). Hver variabel er analysert enkeltvis. De 48 kategoriske variablene ble analysert i SPSS.

Fem størrelser ble beregnet:

- Utfyllernes evne til å svare på spørsmålene, individuelt og som gruppe.
- Andel komplette svarpar: Svarpar der ingen av utfyllerne har unnlatt å svare
- Cohens Kappa (κ). Inkluderer kun komplette svarpar, og variabler hvor alle svaralternativene er brukt av begge utfyllere.
- Samsvarpar: Prosentvis andel enige par blant komplette svarpar.
- P-verdi ble beregnet i den grad materialet tillot det.

De fem kontinuerlige variablene ble analysert i SPSS ved hjelp av intraklassekoeffisienten (ICC), spesifisert med ”two way random” og ”absolute agreement”.

Fire størrelser ble beregnet:

- Utfyllernes evne til å svare på spørsmålene, individuelt og som gruppe.
- Andel komplette svarpar: Svarpar der ingen av utfyllerne har unnlatt å svare
- Intraklassekoeffisienten (ICC). Inkluderer kun komplette svarpar.
- Konfidensintervall og p-verdi ble beregnet for samtlige målinger.

3.4 Kriterier for god datakvalitet

Statusrapportskjemaet spør etter behandlers kjennskap til pasienten, og kvaliteten på dataene må gjenspeile en faktisk kunnskap om dette, ikke bare at samsvaret mellom

utfyllerne – når de faktisk besvarte variablene – var tilfredsstillende. Det var derfor også behov for å definere et akseptabelt nivå for andelen komplette svarpar som gjenspeilet det faktiske utvalget i studien. Det forelå ingen litteratur som ga klare retningslinjer for hvilke nivåer som internasjonalt anses som akseptable andeler ufullstendige svarpar. Jeg valgte å ta utgangspunkt i en forventet andel bruk av ”ukjent” på 5 % per utfyller, noe jeg anså som et maksimumsnivå for om en utfyller faktisk kan sies å ha nødvendig kjennskap til en pasients status. Når utfyllernes besvarelser ble sammenstilt, ville dette medføre at opp mot 10 % av svarparene ville bli ufullstendige ved at minst én av de to utfyllerne hadde unnlatt å svare ved bruk av ”ukjent”.

Krav til en variabel med godkjent datakvalitet ble således definert til:

- Andel komplette svarpar $> 90 \%$, og
- $\kappa > 0,60$ alternativt $ICC > 0.60$ (Altman, 1991) eller
- Samsvar $\geq 70 \%$ (Rothstein & Echternacht, 2003).

Signifikansnivået ble satt til 0,05.

3.5 Etiske hensyn

Prosjektbeskrivelsen var godkjent av Personvernombudet innen forskning, Norsk Samfunnsvitenskapelig datatjeneste (NSD) (vedlegg B). Den var også godkjent av LAR ØST (vedlegg D) samt seksjon for helsefag ved UiO. Informantene og pasientene i studien fikk muntlig og skriftlig informasjon før studien (vedlegg C). Der ble det presisert at deltagelsen var basert på frivillighet, og at utfyllerne og tredjeperson, pasientene når som helst kunne trekke seg fra studien uten at dette fikk følger for pasient eller utfyller.

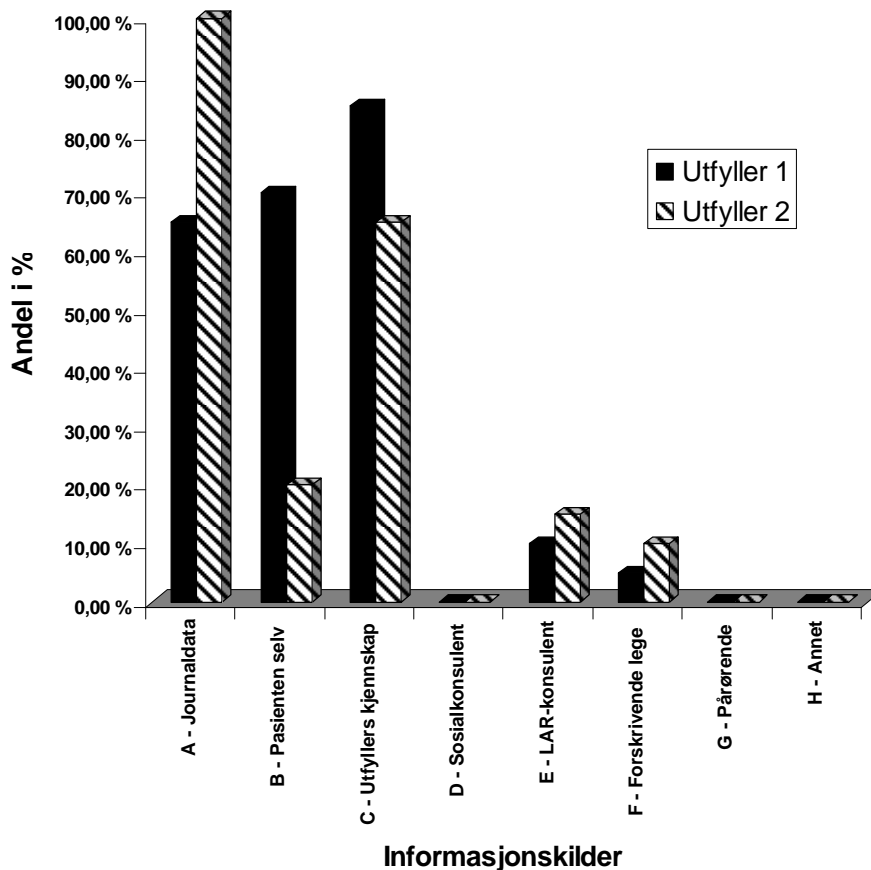
Datamaterialet ble oppbevart som anmodet fra NSD. Direkte personidentifiserende opplysninger vedrørende utfyllerne ble erstattet med et referansenummer som viste til en navneliste som ble oppbevart atskilt fra det øvrige datamaterialet. Skjemaene som

inneholdt indirekte personidentifiserbare sensitive opplysninger om helseforhold og kriminelle forhold om tredjeperson (pasientene) eksisterte kun i papirform. Ved overføring av disse opplysningene til pc, ble pasientene omkodet slik at de ikke kunne føres tilbake til enkeltpasienter.

4 Resultater

4.1 Bruk av informasjonskilder

Utfyllergruppene i studien benyttet flere informasjonskilder (figur 5). Utfyllergruppe 2 benyttet seg av journaldata i alle tilfellene, mens utfyllergruppe 1 brukte disse i 65 % av tilfellene. Pasienten selv ble også mye benyttet av utfyllergruppe 1 (14 av 20), men ble i mindre grad brukt av utfyllergruppe 2 (4 av 20). En tredje mye anvendt informasjonskilde var utfyllers egen kjennskap til pasienten som ble brukt i henholdsvis 85 % av utfyllergruppe 1 og 70 % av utfyllergruppe 2. Øvrige informasjonskilder ble i liten eller ingen grad anvendt.



Figur 5: Bruk av informasjonskilder. Utfyllergruppe 1 (n=20) og utfyllergruppe 2 (n=20)

4.2 Utfyllernes evne til å svare

For de 48 kategoriske variablene i undersøkelsen utgjorde andelen komplette svarpar 70 %. Maksimum bruk av ”ukjent” for utfyllergruppe 1 var 35 %, mens den maksimale andelen bruk av ”ukjent” for utfyllergruppe 2 var 45 %. Minimum bruk av ”ukjent” var for begge utfyllergruppene 0 %. Utfyllergruppe 1 besvarte spørsmålene med ”ukjent” i 12 % av tilfellene, mens utfyllergruppe 2 benyttet ”ukjent” i 22 % av besvarelsene. Andelen komplette svarpar varierte fra 100 % for variablene A3.

Boligforhold, A6. *LAR-medikament*, A8. *Doseringsfrekvens* og A9. *Forskrivende lege til 35 % for variabel B1h. Kontakt med kriminalomsorg i frihet siste fire uker.*

Seksjon A. *AKTUELL SITUASJON* inneholder 18 kategoriske variabler med informasjon om pasientens demografiske forhold, blodsmittestatus og medisinske behandlingstilbud hadde gjennomsnittlig 88 % komplette svarpar. Andelen varierte fra 100 % (A3. *Boligforhold*, A6. *LAR-medikament*, A8. *Doseringsfrekvens*, A9. *Forskrivende lege*), til 55 % (A5b. *Blodsmitte hepatitt C*). De syv første variablene i statusrapportskjemaet (A1a-A4) omhandler pasientens demografiske forhold. De hadde en gjennomsnittlig andel komplette svarpar på 89 %. Variabelen som etterspør informasjon om *fast fritidsaktivitet* (A1d) trekker ned gjennomsnittet, og hadde i studien en andel komplette svarpar på 60 %. For denne variabelen skåret utfyllergruppe 1 ”ukjent” ved to anledninger (10 %). Utfyllergruppe 2 skåret ”ukjent” i syv tilfeller (35 %). De to variablene som omhandler pasientens aktuelle blodsmittestatus (A5a-b) hadde i gjennomsnitt en andel komplette svarpar på 65 %. Utfyllergruppe 1 unnlot å svare i 20 % mens utfyllergruppe 2 svarte ”ukjent” i 25 % av tilfellene. Det er ni kategoriske variabler (A6-12c) som etterspør informasjon om den medisinske oppfølgingen av LAR. En oppnådde en gjennomsnittlig andel komplette svarpar for disse på 92 %.

Seksjon B. *SISTE FIRE UKER FØR UTFYLLINGSDATO* inneholder 24 kategoriske variabler. Utfyllerne hadde i gjennomsnitt 45 % ufullstendige svarpar per variabel. Minste andel var 20 % (B1d) og den største andelen var 65 % (B1h). De åtte variablene B1a-h, som spør etter informasjon om pasientens kontakt med

hjelpeapparatet, hadde i gjennomsnitt 57 % komplette svarpar. Utfyllergruppe 1 benyttet svarkategori ”ukjent” for 26 % av variablene som omhandler ”andre behandlingstiltak” (B1e – B1h). Tilsvarende benyttet utfyllergruppe 2 seg av ”ukjent” i gjennomsnitt 35 % av tilfellene for disse fire variablene (B1e-h). For de fire variablene som etterspør psykiske og somatiske plager (B2-B3), var i gjennomsnitt 60 % av svarparene komplette. For de samme spørsmålene, benyttet 15 % av utfyllergruppe 1 seg av ”ukjent”, mot 29 % av utfyllergruppe 2. For de tolv variablene som omhandler stoff- og alkoholbruk (B4 – B7), var det i gjennomsnitt 53 % komplette svarpar. Utfyllergruppe 1 benyttet seg av ”ukjent” i 15 % av tilfellene, mot 40 % i utfyllergruppe 2. Den siste variabelen i seksjon B (B7), spør etter informasjon om behandlerens kliniske vurdering av pasienten. Denne variabelen krever et svar fra utfyller; svarkategori ”ukjent” var således ikke et definert svaralternativ. Til tross for dette har utfyllergruppe 2 angitt ”ukjent” i 20 % av svarene på dette spørsmålet.

Seksjon C. *SISTE ÅR* inneholder totalt 6 kategoriske variabler. De fire første (C1a-d) spør etter informasjon om eventuell kriminell aktivitet siste år. De hadde i gjennomsnitt 75 % komplette svarpar. Utfyllergruppe 1 benyttet seg av ”ukjent” i 11 % av tilfellene, mot 18 % i utfyllergruppe 2. De to variablene som spør etter informasjon om eventuell overdose eller suicidforsøk (C2-C3) hadde en gjennomsnittlig andel komplette svarpar på 75 %. Utfyllergruppe 1 svarte ”ukjent” i 7,5 % av tilfellene, mens utfyllergruppe 2 tilsvarende svarte ”ukjent” i 23 % av tilfellene. I hele seksjon C hadde variablene en gjennomsnittlig andel komplette svarpar på 75 % med spredning fra 20 % (C1d. *Domfelt siste år*) til 30 % (C1c. *Forelegg siste år*).

Tabell 4 – Komplette svarpar, bruk av ”ukjent”

		Komplette svarpar		Bruk av ”ukjent”	
		n	%	Utfyller 1 (n=20)	Utfyller 2 (n=20)
A1a	Yrkesstatus	19	95	0	1
A1b	Arbeidstrening	19	95	0	1
A1c	Dagtilbud	18	90	1	1
A1d	Fast fritidsaktivitet	12	60	2	7
A2	Viktigste inntekt	19	95	0	1
A3	Boligforhold	20	100	0	0
A4	Sosialt nettverk	17	85	1	2
A5a	Blodsmittestatus, HIV	15	75	3	5
A5b	Blodsmittestatus, Hepatitt C	11	55	5	5
A6	MAR-medikament	20	100	0	0
A8	Doseringsfrekvens	20	100	0	0
A9	Forskrivende lege	20	100	0	0
A10a	Interagerende medicament	17	85	0	3
A10b	B-preparater	19	95	0	1
A11b	LAR-utleveringssted	19	95	0	1
A11c	Andre utleveringssteder	15	75	4	1
A12a	Urinprøveordning	19	95	1	0
A12c	Urinprøvested	17	85	2	2
B1a	MAR-konsulent	15	75	4	2
B1b	Sosial-/ruskonsulent	12	60	3	7
B1c	Ansvarsgruppe	13	65	4	3
B1d	MAR-forskrivende lege	16	80	3	1
B1e	Psykolog/psykiatrisk behandler	8	40	5	9
B1f	MAR-pasientgruppe	11	55	5	4
B1g	Ruspoliklinisk behandler	9	45	4	8
B1h	KIF	7	35	7	7
B2a	Psykiske vansker, depresjon	12	60	3	6
B2b	Psykiske vansker, angst	12	60	3	5
B2c	Psykiske vansker, psykose	11	55	4	6
B3	Somatiske plager	13	65	2	6
B4	Intravenøs injisering	13	65	3	7
B5a	Opioider, rapportert	11	55	2	8
B5b	Cannabisbruk, rapportert	12	60	0	8
B5c	Benzodiazepiner, rapportert	11	55	1	8
B5d	Sentralstimulerende, rapportert	12	60	1	7
B5e	Alkohol, rapportert	10	50	1	9
B6a	Opioider, positive urinprøver	8	40	6	9
B6b	Cannabis, positive urinprøver	10	50	5	8
B6c	Benzodiazepiner, positive urinprøver	8	40	6	9
B6d	Sentralstimulerende, positive urinprøver	8	40	6	9
B6e	Alkohol, positive urinprøver	8	40	6	9
B7	Helhetlig vurdering rusmestring	16	80	0	4
C1a	Arrestasjon/varetekt	15	75	2	4
C1b	Tiltalt/reist sak	15	75	2	3
C1c	Idømt forelegg	14	70	4	4
C1d	Idømt soning/sikring	16	80	1	3
C2	Overdose	15	75	0	5
C3	Suicidforsøk	15	75	3	4

4.3 Samsvar, kategoriske variabler

For hele undersøkelsen varierte κ fra minus 0,24 til pluss 1,00 (tabell 5).

Samsvarsprosenten varierte fra 47 % til 100 % med et gjennomsnitt på 83 % (tabell 6).

Seksjon A. *AKTUELL SITUASJON* hadde totalt 11 av 18 variabler der det var mulig å beregne κ . κ varierte fra 0,35 (A12a) til 1,00 (A1a, A9, A6, A5a). Ni variabler hadde $\kappa > 0,60$ (A1a, A1c, A5a, A5b, A6, A9, A10a, A10b og A12c). Variablene i seksjon A hadde gjennomsnittlig samsvar på 88 %. Minimum samsvar var 47 % for variabel A4. *Sosialt nettverk*. Maksimalt samsvar var 100 % for fem variabler (A1a, A1d, A5a, A6 og A9). De syv første variablene i statusrapportskjemaet (A1a-A4) omhandler pasientens demografiske forhold. Disse oppnådde et gjennomsnittlig samsvar på 86 %. De to variablene som omhandler pasientens aktuelle blodsmittestatus (A5a-b) hadde i gjennomsnitt en samsvarsprosent på 91. De ni variablene (A6-A12c) som etterspør informasjon om den medisinske oppfølgingen av LAR hadde en samsvarsprosent på 90.

Seksjon B. *SISTE FIRE UKER FØR UTFYLLINGSDATO* hadde 20 av 24 variabler der det var mulig å beregne κ . κ varierte fra minus 0,24 (B1h) til 1,00 (B6b, B6d, B6e). 10 variabler hadde en $\kappa > 0,60$ (B2b, B2c, B3, B5c, B5d, B5e, B6b, B6c, B6d, B6e). Tre variabler oppnådde negativ κ (B1f-h). Seksjonen hadde gjennomsnittlig samsvar på 77 %. Minimum samsvar var 50 % (B1e) og maksimalt samsvar var 100 % (B6b, B6d og B6e). Variablene B1a-h som spør etter oppfølging siste fire uker hadde gjennomsnittlig samsvar på 60 %. De tre variablene (B2a-c) som ønsker informasjon om psykiske vansker hadde en samsvarsprosent på 89 %. Det var 85 % samsvar på variabelen som etterspør eventuell injisering (B4). Fem variabler (B5a-e) spør etter rapportert rusmiddelbruk siste fire uker. Disse oppnådde et gjennomsnittssamsvar på 80 %. De fem variablene B6a-e med informasjon om pasientens positive urinprøver hadde gjennomsnittlig samsvar på 93 %.

Seksjon C. *SISTE ÅR* hadde 1 av 6 variabler der det var mulig å beregne κ . Denne variabelen (C1a) hadde en κ på 0,63. Seksjonen hadde et gjennomsnittlig samsvar på 95 %. Variablene C1a-d som spør etter informasjon om eventuelle lovbrudd hadde i gjennomsnitt et samsvar på 94 %. C2 og C3 spør etter eventuelt overdose/suicidforsøk siste år. Disse variablene hadde et gjennomsnittssamsvar på 93 %.

Tabell 5 – Samsvar for variabler der Cohens kappa (κ) lot seg beregne

		n	κ
A1a	Yrkesstatus	19	1,00***
A1c	Dagtilbud	17	0,64**
A5a	Blodsmittestatus, HIV	15	1,00***
A5b	Blodsmittestatus, Hepatitt C	9	0,62*
A6	MAR-medikament	20	1,00***
A9	Forskrivende lege	20	1,00***
A10a	Interagerende medikament	16	0,64**
A10b	B-preparater	18	0,83***
A11c	Andre utleveringssteder	14	0,44*
A12 a	Urinprøveordning	12	0,35
A12 c	Urinprøvested	15	0,82***
B1c	Ansvarsgruppe	10	0,42
B1d	MAR-forskrivende lege	12	0,52*
B1e	Psykolog/psykiatrisk behandler	4	0,07
B1f	MAR-pasientgruppe	6	-0,04
B1g	Ruspoliklinisk behandler	5	-0,13
B1h	KIF	4	-0,24
B2b	Psykiske vansker, angst	11	0,82**
B2c	Psykiske vansker, psykose	10	0,62*
B3	Somatiske plager	11	0,64*
B4	Intravenøs injisering	11	0,57*
B5a	Opioider, rapportert	8	0,30
B5b	Cannabisbruk, rapportert	9	0,47
B5c	Benzodiazepiner, rapportert	9	0,63*
B5d	Sentralstimulerende, rapportert	11	0,75**
B5e	Alkohol , rapportert	8	0,62*
B6a	Opioider, positive urinprøver	6	0,47
B6b	Cannabis, positive urinprøver	10	1,00**
B6c	Benzodiazepiner, positive urinprøver	7	0,71*
B6d	Sentralstimulerende, positive urinprøver	8	1,00**
B6e	Alkohol, positive urinprøver	8	1,00**
C1a	Arrestasjon/varetekt	14	0,63**

* $p < 0,05$

** $p < 0,01$

*** $p < 0,001$

Tabell 6 – Prosentvis samsvar per variabel

		n	%
A1a	Yrkesstatus	19	100
A1b	Arbeidstrening	18	95
A1c	Dagtilbud	17	94
A1d	Fast fritidsaktivitet	12	100
A2	Viktigste inntekt	17	89
A3	Boligforhold	15	75
A4	Sosialt nettverk	8	47
A5a	Blodsmittestatus, HIV	15	100
A5b	Blodsmittestatus, Hepatitt C	9	82
A6	MAR-medikament	20	100
A8	Doseringsfrekvens	18	90
A9	Forskrivende lege	20	100
A10a	Interagerende medikament	16	94
A10b	B-preparater	18	95
A11b	MAR-utleveringssted	16	84
A11c	Andre utleveringssteder	14	93
A12 a	Urinprøveordning	12	63
A12 c	Urinprøvested	15	88
B1a	MAR-konsulent	8	53
B1b	Sosial-/ruskonsulent	7	58
B1c	Ansvarsgruppe	10	77
B1d	MAR-forskrivende lege	12	75
B1e	Psykolog/psykiatrisk behandler	4	50
B1f	MAR-pasientgruppe	6	55
B1g	Ruspoliklinisk behandler	5	56
B1h	KIF	4	57
B2a	Psykiske vansker, depresjon	10	83
B2b	Psykiske vansker, angst	11	92
B2c	Psykiske vansker, psykose	10	91
B3	Somatiske plager	11	85
B4	Intravenøs injisering	11	85
B5a	Opioider, rapportert	8	73
B5b	Cannabisbruk, rapportert	9	75
B5c	Benzodiazepiner, rapportert	9	82
B5d	Sentralstimulerende, rapportert	11	92
B5e	Alkohol , rapportert	8	80
B6a	Opioider, positive urinprøver	6	75
B6b	Cannabis, positive urinprøver	10	100
B6c	Benzodiazepiner, positive urinprøver	7	88
B6d	Sentralstimulerende, positive urinprøver	8	100
B6e	Alkohol, positive urinprøver	8	100
B7	Helhetlig vurdering rusmestring	11	69
C1a	Arrestasjon/varetekt	14	93
C1b	Tiltalt/reist sak	14	93
C1c	Idømt forelegg	14	100
C1d	Idømt soning/sikring	15	94
C2	Overdose	14	93
C3	Suicidforsøk	14	93

4.4 Samsvar og bruk av "ukjent", kontinuerlige variabler

Tabell 7 viser at A7 hadde 100 % komplette svarpar, og veldig god ICC-verdi (jamfør tabell 2). Variabel A11 hadde 95 % komplette svarpar, og god ICC-verdi. A12b. *Antall urinprøver per uke*, C4b. *Antall måneder uten illegalt stoffbruk* og C4c. *Antall måneder uten heroin* hadde i gjennomsnitt 57 % komplette svarpar. Dette resulterte i en utvalgsstørrelse for disse tre variablene som var 41 % mindre enn det opprinnelige datamaterialet. A12b hadde signifikant p-verdi. For variabelen A12b brukte utfyllergruppe 1 "ukjent" i 30 % av tilfellene, mens utfyllergruppe 2 benyttet "ukjent" i 15 % som svar. For variablene C4a og C4b svarte utfyllergruppe 1 og utfyllergruppe 2 "ukjent" i henholdsvis 18 % og 28 % av tilfellene. ICC-verdien for variablene A12b og C4a var moderat. C4b hadde negativ ICC verdi.

Tabell 7 – Komplette svarpar, bruk av "ukjent" og intraklassekorrelasjon (ICC)

		N	Komplette par (%)	Bruk av ukjent		ICC	Konfidensintervall (95%)
				Utfyller 1 (n=20)	Utfyller 2 (n=20)		
A7	Døgndose	20	100	0 %	0 %	0,9988***	(0,9970-0,9995)
A11a	Antall utlev per uke	19	95	5 %	0 %	0,7837***	(0,5373-0,9078)
A12b	Antall urinprøver	12	60	30 %	15 %	0,5519*	(0,0208-0,8447)
C4a	Antall mnd uten stoff	10	50	20 %	30 %	0,4217	(-0,2117-0,8139)
C4b	Antall mnd uten heroin	12	60	15 %	25 %	-0,2122	(-0,6727-0,3806)

- * $p < 0,05$
 ** $p < 0,01$
 *** $p < 0,001$

5 Diskusjon

Tolv variabler tilfredstilte kravene til godkjent datakvalitet. Samtlige av disse var i seksjon A. *AKTUELL SITUASJON*, og beskriver pasientens demografiske data og medikamentelle behandling. Kravene til datakvalitet som ble definert i avsnitt 3.6 var: 1) andel komplette svarpar $> 90 \%$ og 2) $\kappa > 0,60$ eventuelt $ICC > 0,60$ eller 3) samsvar $\geq 70 \%$.

For hele studien var 12 % av svarene for utfyllergruppe 1 ufullstendige, og 22 % av svarene til utfyllergruppe 2 ufullstendige. Dette resulterte i at datamaterialet ble redusert med 30 prosentpoeng. Dette var mer enn for de årlige undersøkelsene som er gjennomført, der gjennomsnittet har ligget på henholdsvis 10 % (2002) og 9 % (2003) (Hansen et al., 2004a). Forskjellen mellom utfyllergruppene tyder på at mye av den nødvendige informasjonen om pasienten ikke var tilgjengelig i journal. Samtlige av utfyllergruppe 2 benyttet journal som informasjonskilde, men hadde i liten grad benyttet andre datakilder enn denne og egen kjennskap til pasienten. Studien tyder således på at forventninger til utfyllernes kjennskap til pasientene var for høye. Resultatene i seksjon B. *SISTE FIRE UKER FØR UTFYLLINGSDATO* og seksjon C. *SISTE ÅR* som omhandler rehabiliteringstilbudet, helse, stoff- og alkoholforbruket og eventuell kriminell aktivitet hadde generelt lav svarprosent.

Når utfyllerne svarte fullstendig kunne κ beregnes for 32 variabler. Disse oppnådde en gjennomsnittsverdi på 0,60. Det gjennomsnittlige samsvaret for de kategoriske variablene var 83 %.

5.1 Kjennskap og pålitelighet

De 18 kategoriske variablene i seksjon A som etterspør demografiske forhold, blodsmittestatus og det medikamentelle behandlingstilbudet hadde i gjennomsnitt 88 % komplette svarpar. Disse spørsmålene viste seg generelt å være uproblematisk for utfyllerne, med høy svarprosent, og godt samsvar mellom utfyllerne. I seksjonen var det ti kategoriske variabler som tilfredstilte kravene til god datakvalitet. Fem av

disse (A6. *LAR-medikament*, A10b. *B-preparater*, A9. *Forskrivende lege*, A8. *Doseringsfrekvens* og A11b. *Utleveringssted*) omhandler det medisinske tilbudet til pasienten. Opplysninger om det medikamentelle tilbudet til pasienten og den demografiske situasjonen synes godt dokumentert i journal. Både utfyllergruppe 1 og utfyllergruppe 2 svarte på spørsmålene, og når de svarte var de samstemte. Slike opplysninger må være godt dokumentert fordi kontrollaspektet ved LAR er vesentlig. Metadon og buprenorfin er potente opioider og er ettertraktede medikamenter på det illegale markedet. Feilaktig bruk, eller skjødesløs oppbevaring kan i verste fall føre til overdosedødsfall. Substitusjonsmedikamenter må derfor kontrolleres strengt (Sosial- og helsedirektoratet, 2005).

Fem tilfredsstillende kategoriske variabler omhandler demografisk situasjon (A1c. *Dagtilbud*, A1a. *Yrkesstatus*, A1b. *Arbeidstrening*, A2. *Viktigste inntekt*, A3. *Boligforhold*). Et av målene skissert fra Sosial- og helsedepartementet er at behandler i LAR skal bistå pasienten til å ”oppnå rusfrihet, bedret livskvalitet og – så langt som mulig – yrkesmessig og sosial rehabilitering”. Informasjon om den eventuelle yrkesmessige rehabiliteringen til pasienten viste seg å være lett tilgjengelig for utfyllerne. Dette viste seg også med høy svarprosent, og tilfredsstillende samsvar. Utfyllerne vet hvordan pasienten bor, og hva de gjør på dagtid.

For de to variablene i seksjon A som spør etter pasientens blodsmittestatus var utfyllerne usikre. Av utfyllergruppe 1 visste 20 % ikke om pasienten var smittet av HIV eller Hepatitt C, mot tilsvarende 25 % av utfyllergruppe 2. Dette resulterte i en svarandel på disse variablene på 65 %. Utfyllerne oppnådde et samsvarsgjennomsnitt på 91 % og gode κ -verdier. Selv om utfyllerne oppnådde et godt samsvarsnivå var svarandelen for liten til at disse variablene kan vurderes som pålitelige. For å kunne gi et adekvat tilbud til pasienten, bør behandlingsansvarlig vite om en eventuell behandlingstrengende somatisk grunnlidelse hos pasienten. Dette bør også komme tydelig frem i journal. Dette har selvfølgelig betydning for hvordan pasienten opplever sin rehabiliteringsprosess, også fordi å leve med en kronisk behandlingstrengende sykdom krever nøye oppfølging fra annen spesialisthelsetjeneste. Et mål for pasienter i LAR er at de skal ”komme seg ut av de

hardt belastede misbruksmiljøene” (Sosial- og helsedepartementet, Rundskriv I-35/2000). Variabelen som spør etter sosialt nettverk (A4) kan knyttes opp mot dette målet. Pasientens sosiale nettverk var ukjent for 5 % av utfyllergruppe 1, mot 10 % hos utfyllergruppe 2. Dette resulterte i et utvalg på 85 % av det opprinnelige. Når de svarte, var samsvaret 47 %.

Studien viste at det spesielt var skjemaets seksjon *B. SISTE FIRE UKER FØR UTFYLLINGSDATO*, med informasjon om rehabiliteringstilbudet, psykiske/somatiske lidelser, injiseringsproblematikk, samt stoff og alkoholbruk som dominerte blant variablene med store variasjoner mellom utfyllerne. Dette ga en høy andel ufullstendige svarpar, og resulterte i en liten utvalgsstørrelse for seksjonen.

De åtte variablene B1a-h spør etter informasjon om pasientens kontakt med hjelpeapparatet og kan således betraktes som behandlerens evne til ”å bistå de aktuelle misbrukerne med å nyttiggjøre seg andre behandlings-, rehabiliterings- og omsorgstiltak”, noe som er et av målene for LAR (Sosial- og helsedepartementet: Rundskriv I-35/2000). Av utfyllergruppe 1 svarte 22 % ”ukjent” på disse spørsmålene, mot 26 % av utfyllergruppe 2. Dette tyder på at slike opplysninger verken var journalført eller for øvrig tilstrekkelig kjent for behandlerne. Den høye andelen bruk av ”ukjent” resulterte i at datamaterialet ble redusert med 43 % i forhold til det opprinnelige datagrunnlaget. Når utfyllerne besvarte spørsmålene, var det 60 % samsvar mellom utfyllernes besvarelser. K-beregningene viste lave verdier der disse lot seg kalkulere. Ved tre tilfeller var κ negativ: B1f. *Oppfølging ved LAR-pasientgruppe*, B1g. *Oppfølging av behandler ved ruspoliklinikk* og B1h. *Oppfølging av Kriminalomsorg i frihet (KIF)*. Utfyllerne viste her stor usikkerhet i sine besvarelser både ved hyppig bruk av ”ukjent” og – når de svarte – lite sammenfallende syn på hva som var korrekte svar, da samsvaret faktisk ville vært høyere ved en vilkårlig gjetning enn utfyllernes egne svar tilkjennega. Ikke ved noen av de andre spørsmålsgrupperingene viste utfyllerne tilsvarende liten kjennskap til pasienten som nettopp for disse åtte variablene.

Det er grunn til å anta at mangelfull *journalføring* var en viktig forklaring på den store usikkerheten utfyllerne her viste. Lav frekvens av oppfølging og dertil

manglende dokumentasjon kan være en grunn. En annen mulig forklaring kan være at *svarkategoriene* ikke er tilstrekkelig presist formulert til å bli forstått. Eksempelvis kan svarkategorien ”Bruker ikke tilbudet” muligens misforstås, og vurderes som mer spesifikk enn ”Nei”. For å undersøke dette nærmere analyserte jeg svarkategoriene, ”Bruker ikke tilbudet” og ”Nei” samlet. Samsvaret økte marginalt, men de tre κ beregningene var fremdeles negative etter analysen. Dette tyder derfor på at formulering av svarkategorier ikke var spesielt viktig for disse variablene.

Tvetydighet i et spørreskjema kan vanskelig kontrolleres for ved utfylling. Det kan derfor forventes at enkelte besvarelser i noen grad kan være preget av misforståelser eller individuelle forskjeller i forståelse av spørsmål og svaralternativ. En tredje mulig forklaring kunne vært at variablene ikke var forklart nærmere i manualen. En fjerde grunn kunne vært at *variablene* i seg selv ikke var presist nok formulert. Eksempelvis kunne spørsmålet ”Behandler ved ruspoliklinikk” muligens misforstås. LAR ØST tilbyr poliklinisk behandling av opiatavhengighet. Det er det imidlertid også flere andre poliklinikker som gjør, eksempelvis Psykiatrisk ungdomsteam (PUT), og spørsmålet kan muligens forlede utfylleren til å tro at spørsmålet dreier seg om egen poliklinisk behandling ved LAR ØST.

De fire variablene om kjennskap til psykiske og somatiske plager (B2 – B3) viste seg også å være problematiske for utfyllerne, spesielt for utfyllergruppe 2. Av utfyllerne i gruppe 1 var det 15 % som svarte ”ukjent”, mens det for utfyllergruppe 2 tilsvarende var 29 % som ikke visste svarene. Dette resulterte i at bare 60 % av disse svarparene var komplette, noe som igjen gjorde datagrunnlaget lite. Å vite om pasientene har problemer av psykisk eller somatisk art er selvfølgelig betydningsfullt i en rehabiliteringsprosess. Målet, som formulert av Sosial- og helsedepartementet, er at LAR skal bistå den enkelte pasient med å ”bedre den fysiske og sosiale funksjonsevnen” (Sosial- og helsedepartementet, s. 2, Rundskriv I-35/2000). For å kunne bistå den aktuelle pasient med dette, bør opplysninger om psykisk og fysisk helse være dokumentert i journal, og enkle å finne. I motsetning til ved spørsmålene om kontakt med hjelpetiltakene, viste utfyllerne ved spørsmålene om psykiske og somatiske plager at de, når de faktisk besvarte spørsmålene, klarte å oppnå en høy samsvarsandel (88 %). For tre av disse variablene var det mulig å beregne κ ; alle

disse var gode. Den store andelen bruk av ”ukjent”, som indikerte manglende kjennskap til pasientens situasjon gjør imidlertid at en skal være forsiktig med å overtolke resultatene og representerte i seg selv et problem som viste at heller ikke disse opplysningene var tilstrekkelig kjent for utfyllerne. I manualen foran Statusrapportskjemaet står det en beskrivelse av hvordan svarkategoriene skal vurderes. Når informasjonen ikke forelå, kan det se ut som om utfyllergruppe 2 ikke spurte pasienten direkte selv (figur 5).

De tolv variablene som spør om pasientens stoff- og alkoholbruk, målt ved rapportering av pasient, urinprøvesvar og klinisk vurdering av utfyller, viste også en stor variasjon mellom utfyllergruppene. Utfyllergruppe 1 benyttet ”ukjent” i 15 % av tilfellene, mens utfyllergruppe 2 benyttet ”ukjent” i 40 % av tilfellene. Dette resulterte i at kun 53 % av svarparene var komplette, slik at utvalgsstørrelsen også for disse variablene var lav. For elleve av de tolv variablene var det mulig å beregne κ . Syv variabler oppnådde en god eller meget god κ . Gjennomsnittssamsvar var 85 % for utfyllerne for disse tolv variablene, men bruken av ”ukjent” resulterte i nesten en halvering av datagrunnlaget. Et ytterligere mål for LAR er at behandler skal bistå pasienten med å ”reduere skadene av misbruket og faren for overdosedødsfall” (Sosial- og helsedepartementet, s. 2, Rundskriv I-35/2000). For å vite om mulige skader av stoffbruken må en dokumentere pasientens inntak av stoff. Innsikt i pasientens eventuelle injisering og frekvensen av dette må derfor være godt dokumentert, og være inkludert som en del av kjennskapet til pasienten (Lawrinson, 2005). Det er en kjensgjerning at sprøytebrukere lever risikofyllt ved deling av nåler og annet brukerstyr. De er derfor i risikozonen for smitte av blodoverførte virus som eksempelvis hepatitt C, og HIV. Risikoeksponeringen inkluderer også økt fare for intramuskulære og subcutane infeksjoner. Kjennskap til, og lett tilgjengelig informasjon om, pasientens inntak av stoff er derfor vesentlig. Informasjonen er også viktig fra et medisinskfaglig perspektiv; en må vite hva pasienten eventuelt har brukt for ikke å komme i skade for å overdosere pasienten. Om kun den som kjenner pasienten best har fullstendig informasjon, gjør dette informasjonstilgangen sårbar. Pasientens rapporterte stoffbruk er også viktig å dokumentere fordi en ønsker informasjon om hvordan disse opplysningene samsvarer med urinprøveresultatene.

Det er kjent at om rapportering av illegal stoffbruk innebærer negative konsekvenser for pasienten, som for eksempel utskrivning fra behandlingstilstand, er det liten grunn til å stole på at pasienten vil være sannferdig (Darke, 1998). Bruken av informasjonskilder indikerte at utfyllergruppe 1 i større grad (70 %) benyttet pasienten selv som kilde til informasjon mot 20 % i utfyllergruppe 2. For at et måleinstrument skal kunne vurderes som reliabelt, kreves det at vurderingene skal være repeterbare og stabile. For å oppnå dette må informasjon være lett tilgjengelig. Når nøyaktig informasjon ikke foreligger, og utfyllergruppe 2 unnlot å spørre primærkilden (pasienten), var det vanskelig å oppnå stabile målinger. Det var derfor ikke overraskende at bruk av ”ukjent” varierte i så stor grad mellom utfyllergruppene.

Internasjonale studier viser at pasienter i substitusjonsbehandling reduserer kriminell aktivitet fordi de får medisin mot sin avhengighet, og at de derfor ikke trenger å begå lovbrudd for å kjøpe illegale stoffer (Sosial- og helsedirektoratet, 2005). Det er derfor relevant å måle hvorvidt pasientene har hatt slik kontakt eller ikke. De seks kategoriske variablene i seksjon C. *SISTE ÅR* spør etter kriminalitet og eventuelle overdoser/suicidforsøk siste år. Spørsmålene om kriminalitet oppnådde noe variasjon mellom utfyllerparene. Utfyllergruppe 1 benyttet seg av ”ukjent” i 11 % av tilfellene, mot 18 % i utfyllergruppe 2. Dette resulterte i 75 % komplette svarpar, og medførte også her en reduksjon i datamaterialet. Samsvaret ble 95 %, når utfyllerne svarte var de med andre ord samstemte. Det var mulig å estimere κ for variabel C1a; denne var god (0,63). For variablene som er ment å måle overdoser og eventuelle suicidforsøk var det også forskjeller mellom utfyllergruppene. Utfyllergruppe 1, benyttet ”ukjent” i 7,5 % av tilfellene mot 23 % i utfyllergruppe 2. Dette resulterte i 75 % komplette svarpar. Samsvaret for utfyllerne var i gjennomsnitt 76 %. Det er derfor grunn til å anta at primærbehandler hadde kjennskap til disse opplysningene, men at dette ikke var dokumentert, og derfor ikke tilgjengelig for sekundærbehandler. Det kan være komplisert å innhente denne typen informasjon fordi en er avhengig av at pasientene selv formidler dette i samtale. Slike sensitive opplysninger vil mange pasienter også være tilbakeholdne med å svare på. Det kan i tillegg være vanskelig både for pasienter og behandlere å skille mellom begrepene ”overdose” og ”suicidforsøk”, da

en erfaringsmessig ikke alltid kan være sikre på om en overdose var ment som et selvmordsforsøk.

De fem kontinuerlige variablene (A7. *Døgndose*, A11. *Antall utleveringer per uke*, A12b. *Antall urinprøver*, C4a. *Antall måneder uten stoff* og C4b. *Antall måneder uten heroin*) viste også variasjoner mellom utfyllergruppene. A7 og A11, som omhandler medisinsk oppfølging, oppnådde nesten 100 % komplette svarpar og tilfredsstilte kravene satt til god ICC. I skåring av antall urinprøver per uke (A12b) tilkjennega utfyllergruppe 1 at dette var ”ukjent” i 30 % av tilfellene, mens utfyllergruppe 2 svarte ”ukjent” i 15 % av tilfellene. For denne variabelen er altså utfyllergruppe 1 mer usikker enn utfyllergruppe 2. En hypotese på årsaken til dette kan være at primærbehandler i utgangspunktet ønsket at pasienten skulle ta urinprøver et visst antall ganger per uke, og dokumenterte dette antallet i journal, mens utfallet ble at pasienten ikke gjennomførte forventet antall urinprøver. Utfyllergruppe 1 ville således svart på bakgrunn av kjennskapet til denne dimensjonen (reduisert frekvens), mens utfyllergruppe 2 forholdt seg til det som var dokumentert i journal. De to variablene som spør etter *måneder siste år pasienten var uten stoff / heroin* viste også variasjoner mellom utfyllerparene. I gjennomsnitt benyttet utfyllergruppe 1 seg av ”ukjent” i 18 % av tilfellene, mot utfyllergruppe 2 sine 30 %. Utvalgsstørrelsen for å beregne ICC ble således redusert med 45 %. ICC varierte fra minus 0,21 for variablene som spør etter heroinbruk til 0,42 for variabelen som spør etter stoffbruk. Resultatene bekrefter funnene fra seksjon B hvor utfyllerne også generelt var usikre i pasientens bruk av alkohol og stoff. En kan stille spørsmål ved om variablene som spør etter antall måneder uten stoff/heroin er presist nok formulert. Eksempelvis vil pasienter som nylig er inkludert i behandling nødvendigvis bli vurdert på bakgrunn av antall måneder i behandling. Det er også tvilsomt om denne informasjonen er lett tilgjengelig for utfyllerne.

Resultatene i seksjon B. *SISTE FIRE UKER FØR UTFYLLINGSDATO* og C. *SISTE ÅR* var et godt stykke unna det en kan karakterisere som optimale fra et forskerperspektiv (Krosnick 1991). Årsakene til de mange uutfylte spørsmålene kan skyldes:

- Manglende veiledning i bruk av Statusrapportskjemaet og kompetanse hos utfyllere
- Manglende tilgang til informasjon/dokumentasjon
- Avstand i tid til hendelser
- Utfyllers manglende motivasjon

Det ble ikke foretatt noen egen opplæring i hvordan utfyllerne skulle fylle ut skjemaet før studiens start. Dette var en bevisst valgt strategi, da det heller ikke i klinikken gis slik opplæring. Statusrapportskjemaet er ment å være selvforklarende, og beskrivelsen av hvordan utfylleren skal kode variablene står foran det enkelte skjemaet. Det blir fremhevet i manualen at samtlige ruter skal fylles ut. Det er også presisert at ”det forutsettes at utfyllerne har eller skaffer seg den nødvendige informasjon, også uten spesiell kontakt med pasienten” (vedlegg A, s. 2). Utfyllerne i studien var for øvrig alle erfarne behandlere. Av dette begrensede materialet kan det se ut til at spesielt ikke utfyllergruppe 2 sikret den nødvendige informasjonen og kjennskapet som skal til for å kunne svare på spørsmålene. Det kan være en av grunnene til at bruken av ”ukjent” forekommer så hyppig i studien (22 %).

For seksjon B og C var 59 % av svarparene komplette. Det kan derfor være grunn til å anta at utfyllerne ikke leste instruksjonen tilstrekkelig godt nok til å få med seg kravene til informasjonsinnhenting. Kompetansen til utfyllerne er en annen dimensjon som kan være en forklaring på den høye ufullstendige svarprosenten (Streiner & Norman, 2003). Skjemaet er designet for personer med minimum tre års høyere utdanning innen helse og sosialfag som har et oppfølgingsansvar for pasienten. I studien hadde alle utfyllerne enten høyskolebakgrunn i form av å være sosionomer eller sykepleiere med tilleggsutdanninger, eller de var forskrivende leger. Det er således grunn til å anta at den generelle stoff- og behandlingskompetansen utfyllerne hadde var tilstrekkelig for å kunne besvare variablene i måleinstrumentet.

Manglende informasjon er en nærliggende faktor for å forklare det høye antallet ufullstendige svarpar. Studien kan indikere at de opplysningene det spørres etter i noen grad er kjent av behandler, men ikke i tilstrekkelig grad dokumentert i journal.

For å ha kontroll over hvilke informasjonskilder utfyllerne brukte, la jeg ved et eget vedlegg til statusrapportskjemaet som jeg kalte ”Bruk av informasjonskilder”. Resultatene fra disse viste at det primært var journaldata og utfyllerens egen kjennskap til pasienten som dannet grunnlaget for skåringene. Studien viste at utfyllergruppe 1 i mindre grad benyttet journal (65 %) enn utfyllergruppe 2 (100 %). Resultatene viste også at utfyllergruppe 1 i større grad baserte skåringen på egen kjennskap til pasienten enn utfyllergruppe 2. Bruken av journal i undersøkelsen var gjennomgående høy (83 %). Den lave kjennsapsgraden som er dokumentert i undersøkelsen kan således neppe tilskrives at utfyllerne ikke hadde tilgang til, eller tok i bruk journal. Det er derfor rimelig å stille spørsmålsteget ved kvaliteten på selve journalføringen av opplysningene. Det grunnleggende problemet med manglende journalføring er at informasjon ikke er gjort tilgjengelig for andre enn den som innhenter informasjonen direkte fra pasienten. En ytterligere årsak til mangelfull journalføring kan være at behandlere ikke ønsker å stille pasienten ubehagelige spørsmål. Dette kan forekomme i tilfeller der det er grunn til å anta at spørsmålene vil avstedkomme negative resultater. Behandler kan, på grunn av nærhet til pasienten, ha en bevisst eller ubevisst motivasjon for at pasienten skal fremstå i et best mulig lys. Dette kan medføre at behandler kan ha underrapportert sin pasients illegale stoffbruk. Slike tilfeller viser at forskeren og utfylleren/behandleren ofte har ulike perspektiv og ulike agendaer (Schwarz & Oyserman 2001). Statusrapportskjemaet baserer seg som kjent på utfyllers kjennskap til pasienten, ikke til pasientens direkte svar i forbindelse med intervju. Denne metoden har flere svakheter, blant annet vil ethvert ledd fra originalopplysningene fra pasientens opplevde situasjon forringe datakvaliteten. Standardiserte intervjuer som eksempelvis ASI og BTOM gir informasjon direkte fra pasienten.

Avstand i tid kan være en tredje forklaring for det lave antallet komplette svarpar. Forfattere av spørreskjemaer har generelt en tendens til å overvurdere menneskers evne til å tilbakekalle tidligere hendelser. Det er enklere å svare på nåtidsspørsmål enn på spørsmål om fortiden (Norman & Streiner, 2003). En kan nok ikke på bakgrunn av dette vurdere avstand i tid som en hovedfaktor for den lave svarandelen,

men studien peker mot en tendens at det var vanskeligere for utfyllerne å skåre seksjon *B. SISTE FIRE UKER FØR UTFYLLINGSDATO*, og seksjon *C. SISTE ÅR*.

De høye andelene ufullstendige svarpar kan også skyldes manglende motivasjon hos utfyllerne. Å skåre variablene i Statusrapportskjemaet kan virke formålsløs og irrelevant for den enkelte fordi de ikke opplever utfylling av skjemaet som en viktig del av jobben, og dermed en prioritert arbeidsoppgave. Konsekvensen av en negativ holdning kan føre til at utfyllerne adopterer strategier som tillater at de fullfører skjemaet uten å vurdere om de foretar en presis skåring av pasienten (Krosnick, 1991). Utfyllerne må heller ikke selv stå til ansvar overfor noen for hvordan de skårer, og opplever nok i liten grad en direkte personlig gevinst ved å utføre arbeidet med omhu. Samtlige Statusrapportskjemaer blir sammenstilt, slik at resultatene fra skåringen heller ikke kan gjenspeile en utvikling for den enkelte utfyllers pasienter. Utfyllerne har derfor ikke noe rasjonale for å besvare skjemaet optimalt utover at det er en obligatorisk arbeidsoppgave. Datatilsynet har foreløpig ikke tillatt å bruke skjemaet for å monitorere enkeltpasienter (Hansen et al., 2004a).

5.2 Anbefalinger

Statusrapportskjemaets datakvalitet har klare forbedringspotensialer som vil kunne forbedres ved ulike grep.

Resultatene vitner om et avvik mellom forskernes forventninger til svarkvalitet og utfyllernes faktiske muligheter for å kunne svare med tilfredsstillende presisjon. En kan i en slik situasjon velge å enten redusere forventningene til hva undersøkelsene skal gi svar på, eller øke utfyllernes evne og vilje til å gjøre en optimal utfylling. I og med at statusrapportskjemaet allerede er et viktig nasjonalt måleinstrument for LAR, noe det utvilsomt er behov for, vil jeg ikke anbefale at en reduserer ambisjonsnivået for skjemaet, men snarere setter inn ressurser og tiltak for å bedre den enkelte utfyllers evner og motivasjon for å svare mest mulig optimalt.

På *nasjonalt* nivå, det vil si for LAR ØST – nasjonalt kompetanse senter vil jeg anbefale følgende tiltak:

- Revidering av statusrapportskjemaet
- Krav til pasientmedvirkning under utfylling av statusrapportskjemaet
- Utvikling av standardisert veiledningstilbud for utfyllergruppen
- Systematisk tilbakemelding av forskningsresultater både på basis av statusrapportskjemaet, og fra øvrige nasjonale og internasjonale studier

Statusrapportskjemaet bør revideres. Denne studien tydet på at enkelte av variablene i skjemaet kunne misforstås og således bør justeres eller fjernes. Spesielt gjaldt dette spørsmål B1f – h og C4b som oppnådde negativ κ . En kan også vurdere å redusere antall variabler som spør etter informasjon om pasientens bruk av stoff og alkohol (B5 og B6), da disse i noen grad kan sies å dekke samme felt. En bør involvere utfyllerne i revideringen av skjemaet for ytterligere å sikre at skjemaet oppleves som relevant. Dette ville styrket skjemaets overflatevaliditet og kan testes forholdsvis enkelt. Før et endelig Statusrapportskjema tas i bruk bør datakvaliteten til de reviderte variablene undersøkes. Videre arbeid med Statusrapportskjemaet for å vurdere datakvaliteten må inkludere flere utfyllere enn det som var tilfelle i denne studien. Dette for å styrke generaliserbarheten, og muligheten til å kunne velge mer sofistikerte statistiske fremgangsmåter som eksempelvis κ -beregning for alle variabler.

Det er en utfordring å oppnå tilstrekkelig informasjon om pasientens situasjon når ikke primærdatakilden er tilstede. Det australske instrumentet BTOM har tatt hensyn til dette. Intervjuformen fordrer at pasienten er tilstede når instrumentet fylles ut (Lawrinson et al., 2005). En anbefaling på bakgrunn av denne studien er at utfylleren skårer variablene sammen med pasienten for å sikre datakvaliteten til skjemaet.

Gjennom obligatoriske standardiserte veiledningsprogram ville det vært mulig å heve datakvaliteten. Ved å veilede, teste og eventuelt sertifisere utfylleren ville en oppnå ting. En ville økt presisjonen på skåren fordi de skjønsmessige aspektene i større grad blir eliminert. Dernest innebærer veiledningen også en pedagogisk anledning til

å utdype hva de forskjellige variablene er ment å måle. En kan også fokusere på *hvorfor* det er viktig å fylle ut skjemaet optimalt. På bakgrunn av resultatene blir det foretatt evalueringer av gruppen, samt måling av regionale og nasjonale forskjeller. Dette kan gjøre utfyllerne mer motivert til å svare optimalt fordi de i større grad opplever at å svare optimalt betyr noe i den store sammenheng. En ulempe med å sette krav til slik kursing og eventuell sertifisering er at det fordrer økonomiske og menneskelige ressurser. En annen vesentlig ulempe er at antallet mulige utfyllere vil gå ned, da en ikke realistisk kan forvente at like mange utfyllere som en har i dag vil gjennomgå en slik veiledning.

Ved å gi systematiske tilbakemeldinger til de ulike sentra av funn fra Statusrapportskjemaet, gjerne relatert til utviklingstrekk i andre land, vil en øke forståelsen for nytten av arbeidet og heve kompetansen og motivasjonen på hele LAR-feltet. Å gjøre dette systematisk innebærer for øvrig å foreta en innholdsvaliditetsvurdering av om instrumentet måler det en ønsker å få et mål på (Streiner & Norman, 2003). Tilbakemeldinger kan relativt kostnadseffektivt gjøres gjennom bruk av hjemmesider, elektroniske nyhetsbrev og lignende. Nettbaserte diskusjonsgrupper kan også være aktuelt, og det er mange spørsmål som sikkert ville være interessante for hele behandlingsapparatet.

På regionalt nivå vil jeg foreslå følgende tiltak:

- Krav til journalføring
- Oppnevning av ”superbrukere” og ressurspersoner
- Utvikling av incentivordninger
- Tydeliggjøring av ansvar

Pålitelige data forutsetter at utfyllerne har tilstrekkelig kjennskap til pasientens situasjon, og at svarene stammer fra en datakilde som er tilgjengelig over tid for alle som skal bruke Statusrapportskjemaet. Dette forutsetter at all relevant pasientinformasjon føres systematisk i journal. Det er et ledelsesansvar å påse at forskriftene til pasientjournal blir overholdt (Lov om helsepersonell § 16). Denne studien viste at mye relevant pasientinformasjon bare eksisterte i hovedkontaktens

hode, noe som svekket datagrunnlaget vesentlig. En kan eksempelvis sikre kvaliteten i journalen ved at dette gjøres til et eget tema i medarbeidersamtaler, at det settes av dedikert arbeidstid til føring av journal, ved forenkling av rutiner og skjemaer samt ved å øke kompetanse i bruk av PC.

En kan utnevne ”superbrukere” og ressurspersoner for statusrapportskjemaet. Disse kan virke som motivatorer og samarbeide på tvers av regionene om spredning av informasjon og kunnskap. Dataene vil da bli mer synlig i den kliniske hverdagen og samtidig øke kompetansen i de ytre ledd av LAR. Disse ressurspersonene bør motiveres gjennom utsikt til faglig utvikling, men også muligheter for økonomisk gevinst eller opplevelser i form av reise i forbindelse med skoling.

Ved å utvikle et systematisk incentivprogram for utfyllerne av statusrapportskjemaet vil en trolig kunne øke motivasjonen rundt utfyllingen betraktelig. Et slikt program kan benytte seg av ulike virkemidler så som internkonkurranser, diplomer, rimelige effekter eller andre premier, trekning av reiser etcetera. Kreativitet, informasjon og synliggjøring av den enkelte vil ofte være vel så viktig som at man bruker store økonomiske ressurser på et slikt program.

Hvert regionscenter er i dag ansvarlig for at skjemaet *blir fylt ut* av behandler (Hansen et al., 2004a). Etter min vurdering bør det i skriv som følger statusrapportskjemaet og i andre sammenhenger fremgå at det enkelte regionscenter også er ansvarlig for *kvaliteten* på utfyllingen.

Utgangspunktet for disse anbefalingene har vært at en ikke bør redusere ambisjonsnivået slik at statusrapportskjemaet ikke kan fungere som et instrument for å evaluere og dokumentere LAR på nasjonalt plan. Hvis man likevel velger å gå vekk fra dette målet, men nøyer seg med et mål om å beskrive utviklingstrekk for den enkelte region, vil en eventuelt kunne gjøre ytterligere tiltak. En kan da velge å selektere de personer som utviser spesiell interesse for emnet, heller enn å bruke hele populasjonen (Streiner & Norman, 2003). Da ville en i større grad å kun få ”optimale svar” og datakvaliteten ville blitt forbedret. En åpenbar svakhet ved denne

tilnærmingen er at en ikke får et datagrunnlag som gjenspeiler hele pasientgruppen i behandling, og anbefales derfor ikke.

5.3 Styrker og svakheter ved studien

5.3.1 Utvalgets størrelse

Statusrapportskjemaet inneholder 29 variabler med to svarkategorier, og 11 variabler med tre kategorier, om en definerer ikke-avgitt svar (ukjent) vekk. På bakgrunn av dette ble utvalgsstørrelsen satt til 20 personer, noe som ble ansett å være tilstrekkelig for å besvare forskningsspørsmålene.

Utvalgsstørrelsen viste seg imidlertid å bli kraftig redusert på grunn av den høye andelen ufullstendige svarpar (30 %) og på bakgrunn av dette vil det knytte seg stor grad av usikkerhet til tolkning av resultatene, noe som får konsekvenser for om funnene kan gjøres generaliserbare. Seks av variablene hadde $n \geq 18$ og tre eller færre svarkategorier, og kan med utgangspunkt i Cicchetti (1976) sies å være allmenngyldige.

5.3.2 Utvalgets sammensetning

Utvalget ble definert på bakgrunn av tilfeldige pasienter. I LAR ØST ble hovedkontakten til pasienten definert som utfyller 1 og sekundærkontakten til pasienten definert som utfyller 2. I Gjøvik kommune ble sosialkonsulent definert som utfyller 1, og forskrivende lege definert som utfyller 2. Ved å ha representanter både fra desentralisert og sentralisert behandlingstilbud oppnådde en at utvalget var bredere sammensatt enn om det ble benyttet utfyllere fra kun ett behandlingstilbud. I utgangspunktet anså jeg det som en styrke ved studien. I etterkant kan en vurdere utvalget som for lite til å si noe om forskjeller/likheter mellom sentralisert og desentralisert modell. Det er derfor en svakhet ved studien at ikke utvalget kommer fra samme behandlingsmodell. Det er også en svakhet ved studien at ikke utfyllerne hadde samme rolle overfor pasienten.

Utfyllerne fikk beskjed om ikke å samarbeide med hverandre. En svakhet ved studien er at utfyllerne i denne studien har diskutert pasientene jevnlig på behandlingsmøter, og det derfor er mulig at utfyllerne har påvirket hverandres persepsjon av den aktuelle pasienten. Siden utfyllerne har ansvar for oppfølging og behandling av pasientene ville det vært et betydelig etisk problem å be utfyllerne la være å diskutere aktuelle pasientbehov i forskjellige fora (Hesse, 2005).

Alle utfyllerne hadde erfaring i bruk av Statusrapportskjemaet. De var kjent med at de skulle være med i en studie som omhandlet evaluering av datakvaliteten til Statusrapportskjemaet. Det kan være grunn til å undre seg om en hadde fått andre resultater om andre personer enn de som kjente pasienten skulle fylt ut. Min hypotese i etterkant av studien er at svarene i større grad ville vært preget av enda lavere svarandel, og mindre samsvar enn i denne studien.

5.3.3 Metode

Den foretrukne statistiske metode i interraterreliabilitetsstudier er κ -beregning (Streiner & Norman, 2003). I denne studien var det mulig å estimere κ for totalt 32 av de 48 kategoriske variablene. En svakhet ved κ er at metoden krever at alle svarkategoriene er benyttet. Om utvalget hadde vært større ville det vært større sannsynlighet for at flere av svarkategoriene ville vært benyttet, og det hadde vært mulig å benytte den foretrukne analyseformen hyppigere. Fordi datagrunnlaget ble så lite reduserte dette muligheten til å benytte de foretrukne statistiske beregningene. En annen svakhet ved estimeringen var at for å unngå å beregne kategorien ”ukjent” som mål på samsvar måtte en definere ut bruk av ”ukjent” fra materialet. For de fem kontinuerlige variablene var ikke metoden et tilsvarende problem, da ICC kunne estimeres for samtlige. Prosentvis samsvar ble målt for samtlige variabler. Å beregne prosent som mål på samsvar er ikke helt korrekt fordi estimatet ignorerer sjansen for tilfeldig samsvar.

6 Oppsummering og konklusjon

For å evaluere datakvaliteten til Statusrapportskjemaet ble det formulert to forskningsspørsmål. Det første var: *”Kjenner utfyllerne pasientens situasjon?”*

Utfyllerne oppnådde høy svarprosent og godt samsvar for de variablene som omhandler pasientenes demografiske og medikamentelle situasjon.

Statusrapportskjemaet er derfor godt egnet til kartlegging av disse forholdene.

Dataene kan trygt brukes til forskning for å gjøre internasjonale, nasjonale og regionale sammenlikninger og for å gjennomføre longitudinelle studier av pasienter i LAR. Når det gjaldt kjennskap til pasientenes sosiale nettverk, rehabiliteringstilbud, helse, stoff- og alkoholbruk viste utfyllerne begrenset kjennskap. Variablene knyttet til disse opplysningene tilfredsstilte ikke ordinære krav til datakvalitet. Dette viste seg ved hyppig bruk av ”ukjent”, og derved en betraktelig reduksjon av utvalgsstørrelsen. Det lave kjennskapet gjør også at generaliserbarheten til studiens øvrige funn må vurderes som meget usikker.

Det andre forskningsspørsmålet som skulle besvares var: *”Er det samsvar mellom ulike utfyllere av statusrapportskjemaet?”* Studien svarer ikke entydig på dette spørsmålet, da utvalgsstørrelsen ble kraftig redusert som en følge av uventet stor forekomst av ufullstendige svar. De svarene som ble gitt indikerer likevel at når utfyllerne først svarte, oppnådde de generelt godt samsvar for de fleste variablene.

Kildeliste

Altman, D. (1991). *Practical statistics for Medical research*. London: Chapman and Hall.

Baca-Garcia E., Blanco C., Saiz-Ruiz J., Rico F., Diaz-Sastre C., & Cicchetti D. V. (2001). Assessment of reliability in the clinical evaluation of depressive symptoms among multiple investigators in a multicenter clinical trial. *Psychiatry Research*, 102, 2, 163-173.

Blacker, D., & Endicott, J. (2003). Psychometric Properties: Concepts of Reliability and Validity. In J. Rush, H. Pincus, M. First, D. Blacker, J. Endicott, S. Keith, K. Phillips, N. Ryan, R. Smith, M. Tsuang, T. Widinger, & D. Zarin (Eds.). *Handbook of Psychiatric Measures*. Washington DC, USA: American Psychiatric Association.

Bordens K.S., & Abbott, B.B. (2001). *Research, design and methods, A process Approach* (5th ed.). NY, USA: McGraw-Hill Higher Education,

Cannell C., Fisher G.G., & Bakker, T. (1965). *Reporting on hospitalization in the Health Interview Survey. Vital and Health Statistics. Series 3, No 6*. Hyattsville, MD, USA: Public Health Service.

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: Critical experience. *British Journal of Educational Psychology*, 54, 1-22.

Cicchetti, D. (1976). Assessing Inter-Rater Reliability for Rating Scales: Resolving some Basic Issues. *British Journal of Psychiatry*, 129, 452-6.

Copeland, J., & Conroy, A. (2001). Australian National Minimum Data Set for Clients of Alcohol and other Drug Treatment Services: finding of the national pilot and developments in implementation. *Drug Alcohol Review*, 20, 295-298.

Cusick A., Vasquez M., Knowles L. & Wallen M. (2005). Effect of rater training on reliability of Melbourne Assessment of Unilateral Upper Limb Function scores. *Developmental Medicine & Child Neurology*, 47, 39-45.

Darke S. (1998). Self-report among injecting drug users: a review. *Drug and Alcohol Dependence* 51, 253-263.

Dole V.P., & Nyswander M. (1965). A medical treatment for diacetylmorphine (heroin) addiction – a clinical trial with methadone hydrochloride. *JAMA* 1965, 193, 80-4.

- Engelsk-norsk stor ordbok (2nd ed.)*. (1996). Oslo: Kunnskapsforlaget
- Ervik, R. (1997). *Evaluering av metadonprosjektet I Oslo, Del 2: Behandlingsforløp og – status*. Oslo: Statens institutt for alkohol og narkotikaforskning.
- Farrel M., Ward J., Mattick R., Hall W., Stimson G.V., & Jarlaiss D.D. (1994). Fortnightly review –methadone-maintenance treatment in opiate dependence –a review. *BMJ* 1994, 309, 997-1001.
- Fjelland R., & Gjengedal E. (1995). *Vitenskap på egne premisser, Vitenskapsteori og etikk for helsearbeidere*. Oslo: Gyldendal.
- Fleiss, J.L. (1981). Balanced incomplete block designs for interrater reliability studies. *Applied psychological measurement*, 5 (1), 105-112.
- Fleiss, J.L., & Cohen, J. (1973). Equivalence of weighted kappa and intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 33 (3), 613-619.
- Gibson D.R., Flynn N.M., McCarthy J.J. (1999). Effectiveness of methadone treatment in reducing HIV risk behavior and HIV seroconversion among injecting drug users (editorial). *AIDS* 1999, 13, 1807-1818.
- Goldberg, L.R. (1971). A historical survey of personality scales and inventories. In P. Mc Reynolds (Ed.), *Advances in psychological assessment, Vol 2* (pp. 293-336). Palo Alto, CA, USA: Science of Behavior.
- Greenfield, T.K. (2004) The ASI: Is the glass half empty or altogether brimming over? *Addiction* 99, 411-418.
- Guilford, J.P. (1954). *Psychometric methods*. NY, USA: McGraw-Hill.
- Hansen, M. B., Kornør, H., & Waal, H. (2004a). *SKR-rapport nr 4/2004 Statusrapport for pasienter i legemiddelassistert rehabilitering 2002-2003. Rapport til Sosial- og helsedirektoratet*. Oslo: Seksjon for kliniske rusmiddelproblemer.
- Hansen, M. B., Kornør, H., & Waal, H. (2004b). *SKR-rapport nr 7 2004 Bidrag til evaluering av legemiddelassistert rehabilitering i Norge Rapport til Sosial- og helsedirektoratet*. Oslo: Seksjon for kliniske rusmiddelproblemer.

- Hansen, M. B., & Waal, H. (2004). *SKR-rapport nr 5/2004. Hvor mange hvor? Antall, oversikter og regional fordeling for pasienter i legemiddelassistert rehabilitering. Rapport til Sosial- og helsedirektoratet*. Oslo: Seksjon for kliniske rusmiddelproblemer.
- Hathaway, S.R. & McKinley, J.C. (1951). *Manual for the Minnesota Multiphasic Personality Inventory (rev.)*. NY, USA: Psychological Corporation.
- Helse- og omsorgsdepartementet (2001) *LOV -1999-07-02-64 Lov om helsepersonell*.
- Helse- og omsorgsdepartementet (2001) *FOR- 2000-12-21-1385 Forskrift om pasientjournal*.
- Hesse, M. (2005): Social workers' ratings of comorbid personality disorders in substance abusers. *Addictive behaviors* 30, 1241-1246.
- Hidle, L., Lauritzen, G., & Skretting., A. (1997). *Europeisk Addiction Severity Index, EuropASI. En håndbok for opplæring og bruk av EuropASI intervjuer*. Oslo: Statens Institutt for Rusmiddelforskning.
- Jaussent S., Labarere J., Boyer J.P., & Francois P. (2004). Psychometric characteristics of questionnaires designed to assess the knowledge, perceptions and practices of health care professionals with regards to alcoholic patients. *Encephale-revue de psychiatrie clinique biologique et therapeutique*, 30 (5), 437-446.
- Joseph H., Stancliff S., & Langrod J. (1998). *Methadone maintenance treatment and other replacement therapies*. Amsterdam: Harwood Academic Publishers.
- Kellog, S., Mc Hugh, P., Bell, K., Schluger, J., Schluger, P., LaForge, S., et al. (2003): The Kreek-McHugh-Schluger-Kellog scale: a new, rapid method for quantifying substance abuse and its possible applications. *Drug and Alcohol Dependence*, 69, 137-150.
- Kilpatrick, B., Howlett, M., Sedgewick, P., & Ghodse, A.H (2000). Drug use, self report and urineanalysis. *Drug Alcohol Depend*, 58 (1-2) , 111-116.
- Kornør, H., & Waal, H. (2004). Metadondose, behandlingsvarighet og heroinbruk i medikamentassistert rehabilitering. *Tidsskrift for Den norske lægeforening* 2004, 3, 332-334.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-36.

- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Law, M., Russel, D., Pollock, N., Mc Coll, M.A., Carswell, A. & Baptiste (1997). A comparison of intensive neurodevelopmental therapy plus casting and a regular occupational therapy program for children with cerebral palsy. *Dev Med Child Neurology*, 39, no 10. p. 664-670.
- Lawrinson, P., Copeland, J., & Indig, D. (2005). Development of a brief instrument for routine outcome monitoring in opioid maintenance pharmacotherapy services: The Brief Treatment Outcome Measure (BTOM). *Drug and Alcohol Dependence*.
- Mäkelä, K. (2003). Studies of the reliability and validity of the Addiction Severity Index. *Addiction* 99, 398-410.
- Malterud, K. (1996). *Kvalitative metoder i medisinsk forskning*. Oslo: Tano Aschehoug.
- Marshall, M., Lockwood, A., Bradley, C., Adams, C., Joy, C., & Fenton, M. (2000). Unpublished rating scales: A major source of bias in randomized controlled trials of treatment for schizophrenia. *British Journal of Psychiatry*, 176, 249-52.
- McGraw, K. O., & Wong S. P. (1996). Forming Inferences About Some Intraclass Correlation Coefficients. *Psychological Methods*, 1, 1, 30-46.
- McLellan A.T., Carise, D., & Kleber, H.D. (2003). Can the national addiction treatment infrastructure support the publics demand for quality care? *Journal of Substance Abuse Treatment*, 25 (2), 117-121.
- McLellan A.T., Kushner H., & Metzger D. (1992) The 5th edition of the Addiction Severity Index. *Journal of Substance Abuse Treatment*, 9 (3), 199-213.
- Moen, S., Hansen, M. B., & Waal, H. (2004) *SKR-rapport nr 6/2004 Legemiddelassistert rehabilitering - en livslang behandling? Om retensjonen i behandlingen*. Oslo: Seksjon for kliniske rusmiddelproblemer.
- Montalto C.P., & Sung J. (1996). Multiple imputation in the 1992 survey of consumer finances. *Financial Counseling and Planning* 7 (1996), pp. 133-141.
- Müller M.J., Rossbach, W., Dannigkeit P., Müller-Siecheneder F., Szegedi A., & Wetzel H. (1998). Evaluation of standardized rater training for the Positive and Negative Syndrome Scale (PANSS). *Schizophrenia Research*, 32, 3, 151-160.

- Müller M., & Dragicevic A. (2003). Standardized rater training for the Hamilton Depression Rating Scale (HAMD-17) in psychiatric novices. *Journal of Affective Disorders*, 77, 1, 65-69.
- Murphy, K., & Davidshofer, C. (1998). *Psychological Testing Principles and Applications* (4th ed.). Toronto, Canada: Prentice-Hall International, Inc.
- Nevo, B. (1985). Face Validity revisited. *Journal of Educational Measurement*, 22, 287-93.
- Newman D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational research methods*, 6 (3), 328-362.
- Olinsky A., Chen S., Harlow L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*, 151, 1, 53-79.
- Polit, D.F. & Hungler B.P. (1999). *Assessing Data Quality, in Nursing Research. Principles and Methods* (6th ed.), J.B. Lippincott Company.
- Publication Manual of the American Psychological Association 5th Edition. (2002).*
Washington DC, USA: American Psychological Association.
- Robson, C. (1993). *Real world research. A Resource for Sosial Scientists and Practitioner –Researchers.* Padstow, Great Britain: T.J. Press Ltd.
- Rogers R. (2003). Standardizing DSM – 4 diagnosis diagnoses: The clinical applications of structures interviews. *Journal of Personality Assessment*, 81 (3), 220-225.
- Roth P.L. (1994). Missing data - A conceptual Review for Applied Psychologists. *Personell Psychology* 47, (3), p 537-560.
- Rothstein, J., & Echternach, J. (1993). *Primer on Measurement: An Introductory Guide to Measurement Issues* (Alexandria: APTA)
- Rush, J., Pincus, H., First, M., Blacker, D., Endicott, J., Keith, S., et al. (2003). *Handbook of Psychiatric Measures.* Washington DC, USA: American Psychiatric Assosiation.
- Schwarz N., & Oyserman D. (2001). Asking questions about behavior: Cognition, communications, and questionnaire construction. *American Journal of Education*, 22, 127-60.

- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing reliability. *Psychological Bulletin*, 86, 420 - 428
- Skjervheim, H. (1996). *"Deltakar og tilskodar og andre essays*. Oslo: Aschehoug.
- Skretting, A. (1997). *Evaluering av metadonprosjektet I Oslo, Del 1: Etablering, inntak av pasienter og forholdet til øvrige behandlingstiltak*. Oslo: Statens institutt for alkohol og narkotikaforskning.
- Sosial- og helsedepartementet (1998), Rundskriv I-25/1998.
- Sosial- og helsedepartementet (2000), Rundskriv I-35/2000.
- Sosial- og helsedirektoratet (2005). Oppfølging av LAR-EVALUERINGENE*. Anbefalinger fra Sosial- og helsedirektoratet til Helse- og omsorgsdepartementet.
- Streiner, D. & Norman, G. (2003). *Health Measurement Scales, a practical guide to their development and use (3rd ed)*. Great Britain: Oxford University Press.
- Sudman S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Fransisco, CA, USA: Jossey-Bass.
- Taylor, A. (1953). A personality scale of manifest anxiety. *Journal of abnormal and Social Psychology*, 48, 285-290.
- Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. Jabine, M. Staff, J. Tanur, & R. Tourangeau (Eds.) *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp 73-100). Washington DC, USA: National Academy Press.
- Turner R. R. (1996). Rehabilitation: Issues in Functional Assessment. In B. Spilker (ed.) *Quality of life and Pharmacoeconomics in Clinical Trials* (pp. 839-851). Philadelphia, USA: Lippincott-Raven Publishers.
- Vatnadal, J., Friis S., & Opjordsmoen, S. (2005). "Are GAF –score reliable in routine clinical use?". Oslo, Norway: Ullevål University Hospital, Department of Acute Psychiatry.
- Verster A., & Buning, E. (2000). *Methadone guidelines*. Amsterdam : Euro-Methwork, 2000.

-
- Vindedal B., Steen, L. E., Larsen B., Knoff R. (2004). Sosial –og helsedirektoratet. Evaluering av legemiddelassistert rehabilitering. Evalueringsrapport.
- Waal, H., & Frogopsahl G. (2001). *Veileder for helsepersonell i medikamentassistert rehabilitering*. Oslo: Aker universitetssykehus, Klinikk for psykiatri, avdeling for rusbehandling.
- Waal H., Krook, A .L., Welle-Strand, G., Espegren, O., Hole, R., Lazardis, K. B., et al. (2001). En nasjonal modell for medikamentassistert rehabilitering for opiatmisbrukere, *Tidsskrift for Den Norske Lægeforening 2001, 2301-5 ut*
- Wade, D.T. (1992). *Measurement in Neurological Rehabilitation*. Oxford, England: Oxford University Press
- Ward, J., Mattick, R.P., & Hall, W. (1998). *Methadone Maintenance Treatment and Other Opioid Replacements Therapies*. Amsterdam: Harwood Academic Publishers.
- Wechsler, D. (1958). *The measurement and appraisal of adult intelligence (4th edition)*. Baltimore, MD, USA: Williams and Wilkins.
- Willms, D.G., & Johnson, N.A. (1993). *Essentials in qualitative research: A notebook for the field*. Unpublished manuscript.
- Zanis, D.A., McLellan, A.T., Cnaan, R.A., & Randall, M. (1994). Reliability and validity of the Addiction Severity Index with a homeless sample. *Journal of Substance Abuse Treatment, 11, 541-548*.

Vedlegg

- A. Statusrapportskjema
- B. Godkjenning fra Norsk samfunnsvitenskapelig datatjeneste
- C. Informasjonsskriv
- D. Godkjenning fra LAR ØST
- E. Bruk av informasjonskilder