

**BERGENSPROSJEKTET: ”TILBAKE PÅ JOBB” –
HELSE, NYTTE OG LIVSKVALITET**

AV

ROAR GJELSVIK



**HOVEDOPPGAVE I SAMFUNNSØKONOMI
VÅREN 2003**

**INSTITUTT FOR ØKONOMI
UNIVERSITETET I BERGEN**

Forord

En lang studieperiode er nå over. Arbeidet med hovedoppgaven er desidert noe av det mest spennende jeg har gjort i løpet av den tiden, takket være god hjelp og veiledning. Jeg vil derfor rette en stor takk til mine to veiledere, Arild Aakvik og Astrid Grasdahl, for konstruktive innspill og gode kommentarer. 'Program for helseøkonomi i Bergen' (HEB) skal ha takk for finansiell støtte. Takk til Tore Pedersen for gjennomlesning. Og jeg vil i denne anledning også si takk til alle på instituttet, studenter og ansatte, for et godt miljø.

Roar Gjelsvik, Bergen 06.04.2003

1 INTRODUKSJON	3
2 HELSE, NYTTE OG LIVSKVALITET	6
2.1 ULIKE METODER FOR Å MÅLE LØNNSOMHET AV HELSETILTAK.....	7
2.1.1 <i>Cost – benefit analyse</i>	7
2.1.2 <i>Kostnads – effektivitetsanalyse</i>	9
2.1.3 <i>Kostnads - nytteanalyse</i>	10
2.2 METODER FOR MÅLING AV NYTTE OG PREFERANSER.....	14
2.2.1 <i>Kort beskrivelse av metoder for kartlegging av nytte og preferanser</i>	15
2.2.2 <i>Sammenligning av metodene</i>	17
2.3 MULTI ATTRIBUTE UTILITY (MAU) -INSTRUMENTER	18
2.3.1 <i>MAU-instrumentenes oppbygning og hva de fanger opp</i>	19
2.3.2 <i>Beskrivelse av ulike MAU-instrumenter</i>	22
2.4 UTFORDRINGER VED BRUK AV CUA SOM EVALUERINGSMETODE	28
2.4.1 <i>Metodiske utfordringer</i>	28
2.4.2 <i>Noen kjente svakheter ved MAU-instrumentene</i>	30
3 BAKGRUNN OG METODE	33
3.1 BAKGRUNN	34
3.1.1 <i>Muskel-/skjelettlidelser</i>	34
3.1.2 <i>Behandling av muskel-/skjelettlidelser</i>	35
3.2 BESKRIVELSE AV BERGENSPROSJEKTET: ”TILBAKE PÅ JOBB”	36
3.3 HVORDAN MÅLE EFFEKTEN AV ET TILTAK?	39
3.3.1 <i>Eksperimentelle og ikke eksperimentelle evalueringsmetoder</i>	39
3.3.2 <i>Frafallsseleksjon</i>	41
3.4 METODE.....	43
3.4.1 <i>Føre data inn i MAU- instrumentene</i>	43
3.4.2 <i>En vurdering av instrumentene</i>	49
3.4.3 <i>Problemer med ”mapping” metoden</i>	52
4 ANALYSE AV BERGENSPROSJEKTET: ”TILBAKE TIL JOBB”.	54
4.1 RESULTATER HELSERELATERT LIVSKVALITET	55
4.1.1 <i>Bakgrunnsdata</i>	55
4.1.2 <i>Resultater i fra MAU-instrumentene</i>	56
4.1.3 <i>Forskjellene i resultat mellom MAU-instrumentene</i>	60
4.2 FRAFALLSSELEKSJON I BERGENSPROSJEKTET.....	63
4.2.1 <i>Seleksjonsmodellen</i>	63
4.2.2 <i>Resultater Minste kvadraters metode og Full maximum likelihood</i>	67
4.2.3 <i>Spesifikasjonstest</i>	74
4.3 HETEROGENITET I BEHANDLINGSEFFEKTEN.....	75
5 OPPSUMMERING	78
REFERANSER:	82
APPENDIKS:	87

1 Introduksjon

Medisinsk utvikling og teknologiske nyvinninger har ført til at stadig flere sykdommer og plager kan behandles. I Norge som i de fleste andre vestlige land, har dette medført en stadig økende etterspørsel etter helse- og sosiale tjenester¹. Som et resultat av dette står myndighetene ovenfor et eskalerende helse- og omsorgsbudsjett. Offentlige ressurser er som kjent knappe, og mange gode formål i samfunnet konkurrerer om tilgjengelige midler. Gitt det faktum at ingen land har råd til å besørge all den helseomsorgen som kan tenkes å gi en eller annen form for fortjeneste, er det i økende grad blitt nødvendig å foreta prioriteringer også i helsesektoren. Dette har medført et behov for en mer vitenskapelig og bevisført styring av ressursene i helsesektoren, på bekostning av ad-hoc og i fra-dag-til-dag praktisering av medisinsk behandling². I en slik kontekst har økonomiske analyser blitt en viktig komponent i prioriteringsdebatten ved evaluering av nye medisiner, teknologier og tiltak.

Økonomiske analyser brukes ofte i sammenheng med randomiserte kliniske studier, som vurderes til å være den beste måten å måle effekten av nye behandlinger og tiltaksformer på (NOU 1999:20). Et problem som ofte er gjenstand for debatt i denne sammenheng, er hvilke indikatorer en skal bruke for å måle endringer i helse, når en skal vurdere kvaliteten og effektiviteten til et helsetiltak. En ser ofte positive virkninger på mennesker ut over det å få en effektiv og sikker medisinsk behandling. Verdien av økt livskvalitet og mindre ubehag for pasienten, er fordeler som det i liten grad har blitt tatt høyde for i kliniske studier. Der er det først og fremst rene medisinske utfall, som dårlig beskriver den personlige byrden et individ har ved en sykdom, som tradisjonelt har vært av interesse. Dette har ført til en økt bevissthet rundt det faktum, at for å evaluere byrden av sykdom og effekten av tiltak fra et sosialt perspektiv, holder det ikke å bare samle data på dødelighet og medisinske sykdomsspesifikke parametre. Data på økonomiske kostnader, og ikke minst tiltaks påvirkning på pasientens livskvalitet (helsestatus), er også nødvendig (Hurst og Forbes (1999)).

For å få en total oversikt over de positive og negative sidene ved et tiltak, er det til en viss grad enighet om at den beste løsningen er å måle tiltakets påvirkning på pasientens helse i termer av "helserelatert livskvalitet" (Garrat et al. (2002)). Denne helseindikatoren kalkuleres hovedsakelig gjennom helseinstrument kjent som "Multi Attributt Utility" –(MAU)

¹ Som en av flere grunner kan det her pekes på "tilbudsbasert etterspørsel" (Roemer (1961)). Underforstått innebærer dette at etterspørselen blir påvirket av tilbudet av medisinske behandlingsformer og medisiner, dvs. økt tilbud gir økt etterspørsel.

² Begrepe "behandling" og "tiltak" vil bli brukt synonymt.

Instruments, redskap som kan brukes i tilknytning til økonomiske analyser. MAU-instrumentene måler helse (gjennom spørreskjema) ved å bruke multidimensjonale enheter av "helsenytte". Metoden måler helsen til et individ som en subjektiv oppfatning av "velvære" eller "livskvalitet", som ikke bare varierer med helsen, men også med erfaringer, kultur og sosioøkonomiske karakteristika (Hurst and Forbes (1999)). Slike helseindikatorer er ment til å fange opp et bredere spekter av helserelaterte forhold enn tradisjonelle medisinske mål, og tillater en valid sammenligning av helseeffekter på tvers av forskjellige sykdommer og helsetiltak (NOU 1999:20).

Målsetningen med denne oppgaven er å analysere helseeffekten av et rehabiliteringstiltak fra et samfunnsøkonomisk perspektiv. Analysen gjennomføres ved å kartlegge om det er en eventuell økning/gevinst i gjennomsnittlig helserelatert livskvalitet³ for en gruppe pasienter som fikk behandling med et intensivt, multidisiplinært behandlingsopplegg for muskel-/skjellet lidelser, sammenlignet med en gruppe som fikk "vanlig" behandling gjennom primærhelsetjenesten. Dataene for analysen er hentet fra Bergensprosjektet: "Tilbake på jobb". Dette er et prosjekt som ble gjennomført som et randomisert kontrollert studie på 90-tallet, for å prøve ut et nytt tverrfaglig rehabiliteringstiltak for langtidssykemeldte personer med smertefulle og funksjonshemmede muskel- / skjelettlidelser. Ved å ta utgangspunkt i en samling spørreskjema fra studiet, der deltakerne svarte på spørsmål angående egen helse ved inntaket og oppfølgingen ett år senere, kartlegges helserelatert livskvalitet for de to gruppene. Her vil vi implementere disse dataene i de fire MAU-instrumentene The EuroQol Classification System (EQ5D), Quality of Well-Being (QWB) Scale, Health Utilities Index Mark III (HUI3) og Quality of life (15D). Resultatene instrumentene genererer, vil vi deretter bruke til å evaluere helsegevinsten i de to gruppene opp mot hverandre.

Siktemålet med rehabiliteringstiltaket som ble prøvd ut i Bergensprosjektet, var å få folk tilbake i jobb. På kort sikt ble ikke denne målsetningen oppnådd (publiserte resultater basert på 1-2 års oppfølging). Den medisinske- og psykologiske evalueringen (Håland Haldorsen og Ursin (1998), Strand (1998)) brukte tradisjonelle medisinske objektive mål for å måle den medisinske effekten av tiltaket. Disse konkluderte med at selv om behandlingsgruppen ikke kom raskere tilbake til jobb, rapporterte denne gruppen i noen grad bedre subjektiv helse, mindre smerte, økt trivsel og økt fysisk aktivitet. De samfunnsøkonomiske evalueringene⁴ av

³ Helseindikatoren/helsemålet vil gå under "MAU-score" eller "helserelatert livskvalitet".

⁴ Dette var ikke evaluering av økonomisk lønnsomhet, men kun en effektevaluering.

Bergensprosjektet (Bratberg et al. (1998) Grasdahl (2001a)), fokuserte på sannsynligheten for å vende tilbake til jobb og effekt på inntekt ved å delta i behandlingsgruppen. Disse fant heller ingen signifikante positive effekter av å ha deltatt i behandlingsgruppen.

Data fra Bergensprosjektet er så langt ikke blitt brukt til å evaluere rehabiliteringstiltaket med helserelatert livskvalitet som utfallsmål. I følge Guzman et al. (2001) er dette blitt lite gjort på ryggpasienter i det store og hele. Spørreskjemaene som ble brukt til å registrere helse og helserelatert livskvalitet, er ikke identiske lik MAU-instrumentene som vanligvis brukes. Spørsmålene og svaralternativene som ble brukt, er likevel egnet til å konstruere måleinstrument tilnærmet lik noen av de standardiserte MAU-instrumentene som vanligvis brukes. Vi ønsker å se om dette lar seg gjøre for fire ulike MAU-instrumenter. Gjennom analysen ønsker vi også å sammenligne de ulike MAU-instrumentene, og se om bruken av helserelatert livskvalitet generert gjennom MAU-instrumenter kan fange opp noen helsemessige forbedringer, som ingen av de andre brukte utfallsmålene har klart.

Oppgaven er organisert på følgende måte:

- Kapittel 2 gjør rede for de metoder som samfunnsøkonomisk teori foreslår for å evaluere lønnsomheten til helsetiltak. Metoder og verktøy som benyttes i tilknytning til måling av helserelatert livskvalitet presenteres her. Teorien danner grunnlaget for analysen som utføres i kapittel 4.
- Kapittel 3 tar for seg Bergensprosjektet: "Tilbake på jobb" sin metodiske fremgangsmåte, målsetning og resultater, samt teori og metoder for hvordan en kan måle effekten av et tiltak. Kapitlet viser og i detalj implementeringen av de kliniske dataene fra Bergensprosjektet på MAU-instrumentet EQ-5D. En diskusjon rundt problemene ved denne metoden vil og bli ført.
- Kapittel 4 presenterer gjennomsnittlig helserelatert livskvalitet for behandlings- og kontrollgruppen generert gjennom MAU-instrumentene. Regresjonsanalyse benyttes for å se om det foreligger en positiv behandlingseffekt av tiltaket med dette utfallsmålet. I tillegg blir det korrigert for frafallsseleksjon for å vurdere om dette påvirker resultatene. Til slutt undersøkes om behandlingen virker forskjellig for ulike undergrupper.
- Kapittel 5 oppsummerer oppgaven ved å se på noen gjennomgående resultater.

2 Helse, nytte og livskvalitet

Innenfor økonomisk teori er man opptatt av å få mest mulig nytte ("utility") eller velferd ut av de ressurser en har disposisjon. Kostnadseffektivitet er blant flere kriterier (som for eksempel alvorlighetsgraden ved sykdom, effekt av behandling) sentralt ved prioritering mellom flere helsetiltak. Prioritering innen helse med kostnadseffektivitet som mål, betyr at ressursene settes inn i de tiltak som gir mest helse for ressursene. For å identifisere hvilke helsetiltak som er kostnadseffektive, er det tatt i bruk samfunnsøkonomiske lønnsomhetsanalyser i flere land, også i Norge. Samfunnsøkonomiske lønnsomhetsanalyser er verktøy til å kartlegge gevinst og kostnadssiden ved helsetiltak, for å gi en oversikt over hvilke alternative muligheter som finnes for helsetiltak innenfor gitte økonomiske rammer. Målsetningen med slike analyser er å bedre informasjonsgrunnlaget for beslutningstakingen, slik at en sikrer at ressursene benyttes til tiltak som kan øke velferden i samfunnet.

Det vil i dette kapittelet bli gjort rede for de vanligste samfunnsøkonomiske lønnsomhetsanalysene. I dette studiet er kostnadsnytteanalyse (CUA) sentral. Denne metoden vil derfor bli beskrevet mer utførlig. Motivasjonen bak bruken av livskvalitet i CUA har vært drevet av: (1) Et ønske om å kunne vurdere og måle alle aspekter ved en behandlingseffekt. Bruken av livskvalitet gir således et enkel oppsummerende mål på pasientenes helserelaterte livskvalitet. (2) Ønske om å kunne sammenligne resultat på tvers av områder innenfor helse, for en bedret og mer rasjonell fordeling av ressurser på tvers av ulike helseprogram. Mål på livskvalitet kan brukes i kostnadsnytteanalyser som kvalitetsvekter ved kalkulering av QALY's (helserelatert livskvalitet multiplisert med tid).

Imidlertid finnes det en rekke ulike direkte avsløringsmekanismer (f.eks. Standard Gamble, Time Trade-Off), som har til hensikt å kartlegge og måle individers preferanser for ulike helsetilstander. MAU-instrumenter brukt i analysen bygger på en eller flere av disse avsløringsmekanismene, og er alle forskjellige i oppbygning. Dette medfører at de gir forskjellige resultat for samme helsetilstand (Nord (1999a), Hansen (2002)). Spørsmålet er om noen av instrumentene kan sees på som bedre enn de andre, og hvilke implikasjoner dette har for bruken av MAU-instrumenter i økonomiske analyser. Dette gjør det nødvendig å få en oversikt over hvilke metoder, modeller og MAU-instrumenter som finnes for å måle helse.

2.1 Ulike metoder for å måle lønnsomhet av helsetiltak

Ved samfunnsøkonomiske vurderinger tilknyttet helsespørsmål finnes det flere ulike analysemetoder. Selv om disse behandler kostnadssiden tilnærmet likt, er de veldig forskjellige med hensyn til hvordan de måler gevinsten som oppnås i det enkelte tiltaket. Hvilken analysetype som brukes avhenger av typen problemstilling en er ute etter å besvare, og type data en har tilgang til⁵. I det følgende skal vi se på de grunnleggende prinsippene ved de tre vanligste økonomiske analysene, cost-benefit analyse (CBA), kostnadseffektivitetanalyse (CEA) og kostnadsnytteanalyse (CUA), og illustrere anvendelsen av de med noen eksempler.

2.1.1 Cost – benefit analyse

“Cost–benefit” analyse (CBA) går ut på at den økonomiske verdien av en medisinsk gevinst og behandlingkostnadene tallfestes og veies opp mot hverandre. Monetær nytte måles i termer av individets betalingsvillighet for å oppnå resultatet (helseforbedring), eller ved samfunnets produksjonsgevinst ved at personer lever lenger eller er friskere. Kostnadene (costs) ved tiltaket skal i prinsippet gjenspeile verdien av hvor mye en *må gi opp* av andre ting for å gjennomføre tiltaket. Gevinsten (benefits) skal i prinsippet gjenspeile hvor mye en er *villig til å gi opp* av andre ting for å gjennomføre tiltaket. Dersom gevinsten overstiger kostnadene vil det være lønnsomt for samfunnet at tiltaket blir gjennomført

Nettonåverdi (NNV) kriteriet sier at et tiltak er samfunnsøkonomisk lønnsomt, når nåverdien av strømmen av fremtidige nettogevinster er positiv. Dette kan vises med likning (2.1). B_t og C_t er henholdsvis verdien av samfunnets gevinst i periode t og verdien av samfunnets kostnader i periode t . r er samfunnets diskonteringsrate, og T er tiltakets totale varighet med hensyn til behandlingseffekten (Folland et al. (2001)).

$$NNV(\text{Nettonåverdi}) = \sum_{t=1}^T \frac{(B_t - C_t)}{(1+r)^t} \quad (2.1)$$

Dersom den medisinske gevinsten overstiger de medisinske kostnadene, tilsier dette at tiltaket bør gjennomføres. Dvs. hvis $NNV > 0$ er det aktuelt å iverksette tiltaket. Ved valg mellom flere behandlingsformer hvor gevinsten er større enn kostnadene, velges den

⁵ Drummond et al. (1997) og (NOU 1999: 20) presenterer en detaljert beskrivelse av hvordan økonomiske evalueringer av helsetiltak utføres i praksis.

behandlingsformen som har størst differanse mellom gevinst og kostnader⁶. Et detaljert eksempel på CBA finnes i Håland Haldorsen et al. (2002).

En alternativ måte er å bruke internrente metoden, hvor en ser på avkastningen per krone, i motsetning til det å bare se på krone verdien. Hvis et prosjekt genererer en strøm av gevinster (B_t) og kostnader (C_t) over T perioder, er internrente avkastningen (p) definert som p som løser likning (2.2).

$$\sum_{t=1}^T \frac{(B_t - C_t)}{(1+p)^t} = 0 \quad (2.2)$$

Internrente avkastningen er diskonteringsraten som vil gjøre den nåværende verdien av prosjektet lik null. Internrente metoden sier at en skal akseptere et prosjekt hvor p overstiger mulighetskostnadene, eller har høyest p av to gjensidige utelukkende prosjekter. Ulempen med denne metoden er at hvis en har prosjekter med forskjellig skala, kan internrente metoden gi ulike motstridende svar, idet at den totale avkastningen i kroner mellom to prosjekt kan gi et annet resultat.⁷

CBA analyser byr på en rekke utfordringer. For det første finnes det ingen entydig fasit for hvilke gevinst- og kostnadselementer som skal inkluderes i analysen, og hvor disse skal inngå. Det er en iboende tvetydighet ved sammenligning av cost-benefit forhold, fordi gevinsten alltid kan anses som negative kostnader og omvendt. Derfor kan en ved klassifisering av gevinst og kostnader gjøre alle typer mulige prosjekters cost-benefit forhold vilkårlig høy (Rosen (2002)). En annen og spesiell utfordring, er hvordan en skal verdsette sentrale utfallsmål som menneskeliv, helse og livskvalitet. Det er stor usikkerhet forbundet med det å måle folks betalingsvilje for helsetjenester, siden det å tilegne en monetær verdi på menneskers liv og helse er forbundet med metodologiske og etiske problemer. Disse går blant annet ut på å bestemme hvem som skal vurdere hvor mye et liv og en helseforbedring er verd.

⁶ Dette forutsetter at kostnadseffektiviteten er eneste kriterium for prioritet og en utelukker andre kriterier som etikk, ønskelighet, osv. ved tiltaket.

⁷ Eksempel: En investerer 100NOK i et prosjekt (A) og får økt profitt et år senere på 110NOK. Avkastningen vil her være 10%. Dvs ved å finne p som løser likningen $100 + (110 / (1+p)) = 0$, får en $p = 0,10$. I prosjekt (B) investerer en 1000NOK og får 1080NOK året etter. Det gir en $p = 0,08$. En velger følgelig det prosjektet med størst p , altså prosjekt (A). Ved å bare se på den totale profitten snues dette bildet, idet at prosjekt (A) gir bare 10NOK i avkastning, mens prosjekt (B) gir 80NOK i avkastning. Følgelig bør en velge prosjekt (B) ut i fra et slik kriterium.

Bruken av dette prinsippet kan føre til uønskede sosiale skjevheter, idet at prioriteringer av tiltak som gir samfunnet høyest mulig produksjonsgevinst, vil føre til at tiltak for eldre og andre svake grupper nedprioriteres.⁸ CBA kan være en del av beslutningsgrunnlaget i prioriteringssituasjoner, men resultatene bør tolkes med forsiktighet (NOU 1997: 18).

2.1.2 Kostnads – effektivitetsanalyse

Den tradisjonelle form for analyse innen evaluering av helsetiltak kalles kostnadseffektivitetsanalyse (CEA). Analysen involverer en ikke-monetær verdsetting av en behandling. Det innebærer at benyttede effektivitetsmål (gevinstsiden) uttrykkes i sin naturlige enhet (også kalt harde endepunkt), hvis det uttrykker det relevante utfall av behandlingen. Dette kan være for eksempel x antall reduserte hjerteinfarkt, x antall leveår vunnet, osv. Effektivitetsmålene settes i sammenheng med kostnadene ved behandling, og gir resultater i form av en kostnadseffektivitetsrate som for eksempel kostnad (målt i kroner) per x unngåtte hjerteinfarkt. Ulike behandlingsformer for å unngå x antall hjerteinfarkt kan følgelig sammenlignes, hvor en kan velge den behandlingen som gir lavest kostnad per x antall unngåtte hjerteinfarkt. I CEA søker man å finne frem til det tiltaket som minimerer kostnadene (gir mest effektivt resultat) ved å oppnå en gitt helsemålsetting.

En kan også uttrykke resultatene i CEA ved hjelp av antall personer som må behandles ("Number needed to treat" (NNT)) (NOU 1997:18). NNT angir hvor mange personer som må behandles for å unngå ett sykdomstilfelle, eller oppnå et symptomfritt individ. Hvis sannsynligheten for å unngå sykdom (S) for personer gitt en *medisin 1* viser seg å være 50%, p_1 , mens sannsynlighet for å unngå (S) for personer gitt *medisin 2* viser seg å være 90%, p_2 , vil absolutt risikoreduksjon (ARR) ved å gå fra *medisin 1* til 2 være: $p_2 - p_1 = 40\%$ poeng. Ved å ta den inverse av dette tallet kommer vi frem til NNT. I dette tilfellet betyr dette at $1/0,4 = 2,5$ personer må behandles for å unngå et ekstra tilfelle av S .

$$NNT = 1 / ARR = 1 / p_2 - p_1 \quad (2.3)$$

CEA er lett å forstå ved at en får resultatene uttrykt i naturlige enheter. Metoden kan være nyttig når en sammenligner prosjekter med like enheter. Problemet med metoden er at når

⁸ Andre utfordringer vil være verdsetting av gevinster og kostnader, tiltakets levetid og bruken av diskonteringsrenter.

måleenheten for effektivitet ikke er den samme, er det vanskelig å sammenligne resultater på tvers av sykdommer og behandlinger. Dermed utgjør dette et for svakt informasjonsgrunnlag i prioriteringsdebatten. Et annet problem som ble nevnt innledningsvis, er at bruken av slike utfallsmål ved enkelte sykdommer og lidelser er dårlig relatert med det kliniske utfallet, siden det ofte forekommer flere utfall av interesse. Dermed klarer en ikke å fange opp alle sider ved en behandling. Et tredje problem er at CEA ikke skiller klart mellom utfall med ulik verdi eller betydning.

2.1.3 Kostnads - nytteanalyse

En mulig løsning for de tre overnevnte problemene i CEA, er å bruke en kostnadsnytteanalyse (CUA) som representerer en spesialvariant av CEA. CUA bidrar på samme måte som CEA til å kartlegge og synliggjøre konsekvensene av ulike tiltak. Forskjellen ligger i metoden en bruker til å måle de medisinske resultatene eller nyttegevinsten av tiltaket. I CUA oppgir en effekten av behandling ved en økning i livskvalitet (QoL)⁹, i form av en helseindikator, istedenfor ved det direkte helseutfallet som i CEA.

I CUA bruker man et flerdimensjonalt mål på nytten av behandling, hvor sykdom vanligvis tillegges en nytteverdi på en skala mellom 0 (død) og 1 (frisk). Metoden kvantifiserer helseutfall langs flere dimensjoner inn i en helseindikator ofte kalt ”helserelatert livskvalitet” (HRQoL)¹⁰. Denne helseindikatoren er et mål på helse. Verdien av en medisinsk behandling beregnes således etter hvor stor økning på nytteskalaen (økning i helserelatert livskvalitet) en kan forvente som følge av behandlingen, og hvor lenge en kan nyttiggjøre seg den økte livskvaliteten. Denne størrelsen kalles gjerne QALY (Quality Adjusted Life Years), som er helserelatert livskvalitet multiplisert med antall forventede gjenværende leveår. QALY er en vanlig måleenhet i CUA, hvor resultatet som regel oppgis i kostnad per QALY vunnet.

Økt livskvalitet og nytte

En grunn til at man ønsker å bruke livskvalitet som resultatmål, er at et behandlingstiltak ofte har flere formål og gir flere forskjellige former for medisinske gevinster. En annen grunn kan også være at behandling ikke påvirker selve sykdommen eller sykdomsutviklingen nevneverdig, som for eksempel i noen tilfeller innen kreftbehandling. Livskvalitet (nytte) er i

⁹ QoL er forkortelse for ”Quality of life”.

¹⁰ HRQoL er forkortelse for “Health related quality of life”.

denne sammenheng relatert til forhold som hvordan en person takler og gjennomfører daglige gjøremål, og hvordan personen vurderer sin egen situasjon. Verdien av økt livskvalitet og mindre ubehag for pasienten, er fordeler som det i liten grad har blitt tatt høyde for i medisinske kliniske studier (Aakvik (2001)). Dette til tross for at økt livskvalitet kanskje er den viktigste effekten av mange tiltak (Hawthorne et al. (2000)). De tradisjonelle målene for behandlingseffekt gjelder gjerne endringer i sykelighet og dødelighet, i form av myke¹¹ (for eksempel blodtrykk) eller harde endepunkter (for eksempel dødelighet eller sykelighet), som nødvendigvis ikke oppleves av pasienten. For eksempel vil en behandling av depresjoner både kunne gi medisinsk gevinst i form av bedret mental tilstand, gi økt nattesøvn og muliggjøre gjeninntreden i arbeid. En ser således at behandling ofte påvirker helse i flere dimensjoner, og at en har flere utfall av interesse.

En måler den medisinske effekten av et helsetiltak som påvirker helse i flere dimensjoner, ved hjelp av ulike direkte avsløringsmekanismer og MAU-instrument (disse forklares nærmere i kapittel 2.3 og 2.4). Metodene kartlegger subjektive oppfatninger av helsegevinsten slik de oppfattes av pasienten, og tilordner pasientens helsetilstand et nyttenivå. Gjennom å stille en rekke spørsmål angående daglige gjøremål og allmenntilstand (for eksempel bevegelighet, smerte, syn, hørsel, osv.), og gi verdier alt ettersom hvordan responsen er, måler MAU-instrumenter pasientens helse ved å systematisk kartlegge dennes preferanser for egen livskvalitet og livslengde. Ved å benytte en slik (kardinal) helseindikator får en knyttet en (nytte) verdi til et helseutfall på en skala mellom 0 og 1. Verdien uttrykker nytten av helsetilstanden til pasienten. Nytte kan i denne sammenheng sees på som jo mer foretrukket et utfall er, desto mer nytte er assosiert med det (Drummond et al. (1997)). Mer presist kan det sies at begrepet ”nytte” brukes som et mål på graden av tilfredsstillelse av konsumentens preferanser, eller en indeks over styrken på en persons preferanser (Hawthorne et al. (1999))¹². En helseindikator som ”helserelatert livskvalitet” innebærer en tallfesting av en persons preferanser for ulike helsetilstander, som er en samlingsverdi for endring i helse på bakgrunn av ulike helsetiltak, uttrykt i form av en endring både på livskvalitet og livslengde. Dette gjør det mulig å kunne måle effekten av et tiltak, hvor livskvalitet måles i form av en vurdering av hvordan individets helserelaterte livskvalitet påvirkes. Ved å sammenligne denne

¹¹ Myke endepunkter sier ikke uten videre hvilken effekt en medisinsk intervensjon har på reell sykelighet og dødelig slik som de harde gjør, derav navnet myke endepunkt (NOU 1997:18).

¹² Den konvensjonelle nytteteorien som ligger til grunn for måling av livskvalitet og individuell nytte, bygger på Von Neumann og Morgenstern (NM) (1953) sin normative modell for å beskrive preferansene til et individ. Modellen består av et sett med nytteaksiomer som foreskriver hvordan rasjonelle individer ”bør” (anbefaling, ikke en beskrivelse) tenke før de fatter en avgjørelse ved usikre utfall. (Beskrives i Drummond et al. (1997))

verdien med tilsvarende verdi for et individ som ikke har mottatt behandling, kan behandlingens gevinst oppgis i form av eventuell økning i helserelatert livskvalitet eller QALY.

Det må bemerkes at det foreligger en del metodiske problem knyttet til måling av livskvalitet og individuell nytte, som nok er grunnen til at mange helseøkonomer vegrer seg mot å bruke denne metoden (Sanders et al. (1998)). Det sees allikevel på som ingen umulig oppgave å tallfeste verdien av økt livskvalitet, selv om det må gjøres en del forutsetninger for hvordan livskvalitet kan måles og sammenlignes mellom individer. Disse forutsetninger går blant annet ut på at nytte er målbart, kan sammenlignes og legges sammen.

QALY

Som nevnt innledningsvis er det vanlig at resultatene oppgis i kostnad per kvalitetsjusterte leveår, QALY (Quality adjusted life years). QALY er sammensatt av helserelatert livskvalitet (HRQoL), og tiden individet får nyttiggjøre seg av denne helseforbedringen. Dette kan uttrykkes enten ved likning (2.4) eller (2.5). I likning (2.4) antar en at helserelatert livskvalitet er tidskonstant og ikke kan endre seg over tid. I likning (2.5) kan helserelatert livskvalitet endre seg over tid som i den grafiske fremstillingen av QALY i figur 2.1. Med den tro at helse kan endre seg over tid, er følgelig likning (2.5) den rette å bruke.

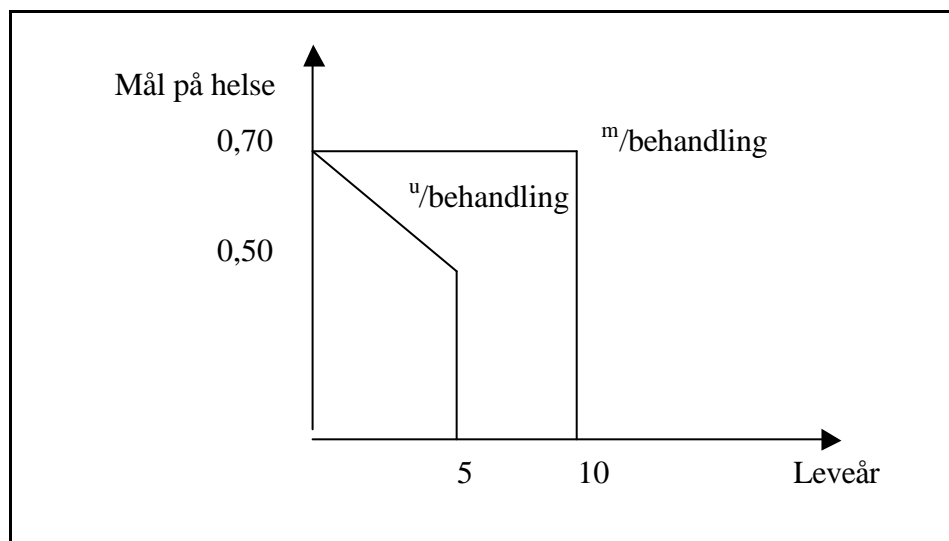
$$QALY = HRQoL * [T - t_0] \quad (2.4)$$

$$QALY = \sum_{t_0=1}^T HRQoL_t \quad (2.5)$$

QALY representerer nytte på forholdstallsnivå¹³ ved at en nytteøkning fra 0.2 til 0.3 anses å være likeverdig med en økning i fra 0.8 til 0.9, samt at QALY verdien for forskjellige individer kan legges sammen og sammenlignes med verdier fra andre individer. Det er en

¹³ Nytte kan måles og beskrives på forskjellige nivå. Målte preferanser kan være på ordinal, kardinal eller forholdstallsnivå. På ordinalnivå rangeres utfallene i fra minst foretrukket til mest foretrukket. En kan tenke seg at en spør et individ spørsmålet: Hvilke foretrekker du av disse to (gir deg størst nytte), A eller B? På kardinalnivå settes det et nummer til utfallet som forsøker å representere styrken til preferansen for det utfallet relativt til de andre. Dette gjør det mulig å måle avstanden mellom utfallene, men siden nullpunktet er tilfeldig kan ikke forholdet mellom ulike verdier sammenlignes. På forholdstallsnivå har en i tillegg til måleenheten et absolutt nullpunkt som gjør det meningsfylt å sammenligne forholdet mellom tall som tilhører forskjellige nivå. Et eksempel på dette er alder hvor det gir mening å si at 20 år er det dobbelte av 10 år.

fordel at en baserer beregningene av helserelatert livskvalitet på kliniske studier, og at dataene blir samlet inn under det kliniske studiet. Er ikke dette mulig kan det gjøres retrospektivt som i analysen i kapittel 4, ved å ”mappe” effektivitetsdataene med MAU-instrumentene. Effektivitetsdata fra klinisk forsøk danner grunnlaget for vurderingen av hvordan en behandling påvirker livskvalitet og levealderen uttrykt ved QALY. Den vanligste metoden er å kartlegge helserelatert livskvalitet på en gruppe pasienter som mottar behandling og en gruppe pasienter som ikke mottar behandling. Dette beskrives mer omfattende i kapittel 3.3. Eksempel på bruken av QALY vises i figur 2.1.



Figur 2.1 Grafisk fremstilling av QALY

Eksempel: En person med en lidelse har fått tilegnet verdien 0,70 gjennom å ha svart på et MAU-instrument, og har 10 forventede leveår igjen med behandling. Ved å multiplisere verdien med de 10 forventede leveårene til personen, finner en at denne har $0,70 * 10 = 7$ kvalitetsjusterte leveår igjen. Uten en behandling av lidelsen vil personen få redusert antall forventet leveår til 5, hvor denne i løpet av den tiden vil få redusert helsetilstanden til 0,50 før hun/han dør. Følgelig vil denne personen ha $[(((0,70 - 0,50)*5)/2) + (0,50*5)] = 3$ kvalitetsjusterte leveår igjen. Forskjellen mellom det innerste og ytterste arealet utgjør kvalitetsjusterte leveår vunnet med bakgrunn i behandling. I dette eksempelet vil dette utgjøre $7 - 3 = 4$ kvalitetsjusterte leveår vunnet.

Fordelene med CUA

CUA er den mest omfattende analysetypen av de tre overnevnte, og gir ideelt sett den beste sammenstillingen av økonomiske fordeler og ulemper ved et tiltak. Fordelen med CUA er at

en ved å konvertere effektivitetsdata til en felles måleenhet som QALY, klarer å inkorporere simultant både endringer i sykkelighet (livskvalitet) og dødelighet (livslengde). Dermed kan en kostnadsøkning for et tiltak bli sammenlignet med den tilhørende helseforbedringen for det samme tiltaket. En kan i tillegg vekte resultater med større betydning mer enn de andre, noe som ikke går ved de andre analysetypene. Ved å angi kostnadseffektiviteten til behandlingen i form av kostnad per QALY, løser en også problemet med ulike enheter mellom ulike sykdommer og behandlinger. Dette muliggjør en sammenligning av kostnadseffektivitet mellom ulike tilstander og intervensjoner. Dermed kan en rangere prosjekter der en har helseutfall langs flere dimensjoner (dette kan være sparte leveår, helseforbedringer, økt livskvalitet). En fullstendig CUA gir en samfunnsøkonomisk lønnsomhetsrangering av ulike tiltak, og kan brukes til prioriteringer mellom ulike mål. Imidlertid er det forbundet mange problemer og stor usikkerhet av etisk og metodemessige art ved å måle nytte, at CEA ofte foretrekkes i flere tilfeller (NOU 1999:20).

Det å kunne se på flere ulike behandlingsformer under ett, for på den måten kunne sammenligne og sette dem opp mot hverandre, vil kunne lette jobben for beslutningstakere. At det er kostnadsnytteanalyser som bør brukes er det til en hvis grad enighet om, noe som et stort antall studier bekrefter (se for eksempel Hawthorne et al. (2000), Krabbe et al. (1997)). Enigheten stopper dersom en retter søkelyset mot de ulike metodene for å måle helse, samt de tilhørende MAU-instrumentenes betydning og påvirkning når det gjelder utfallet av slike analyser. Det er påvist at bruken av alternative MAU-instrumenter gir ulike resultater for en og samme diagnose (Nord (1999a)). Beslutningstakere står dermed ovenfor et problem, idet at det dermed ikke er gitt hvilket tiltak som er reelt sett det mest kostnadseffektive¹⁴, når resultatene mellom MAU-instrumentene varierer. Dette medfører at resultatene i fra CUA må tolkes med en viss forsiktighet, idet at valget av MAU-instrument er blitt påvist å ha en innvirkning på om et tiltak blir funnet kostnadseffektivt eller ikke.

2.2 Metoder for måling av nytte og preferanser

For å bestemme nytten knyttet til forskjellige helsetilstander og behandlingsutfall, kan analytikere bruke en av tre strategier: (1) bruke ekspert uttalelser, (2) bruke verdier skaffet til veie i fra relevant litteratur, eller (3) utføre direkte målinger gjennom intervjuer og

¹⁴ Kostnadseffektivitet tenkes her målt som antall QALY vunnet ved en behandling.

spørreundersøkelser. På grunn av den potensielle skjevheten assosiert ved de to første (for eksempel ekspert vurdering er feil, eller publisert litteratur er uskikket), sees generelt den tredje strategien på som den beste, og er den som er relevant i kontekst av kliniske forsøk.

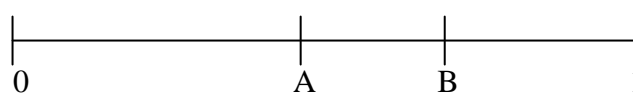
Det eksisterer imidlertid flere forskjellige teknikker (avsløringsmetoder) for å kartlegge individers preferanser av forskjellige helsetilstander eller behandlingsutfall (måle kardinal nytte), alle med sitt teoretiske rammeverk og begrepsapparat. Valget av metode er viktig siden de produserer forskjellige verdier for identiske beskrivelser av livskvalitet (Doland et al. (1996), Nord (1999a)). Det er flere grunner til denne variasjonen, hvor alle ikke er fullt ut forstått (Brooks (1995), Dolan et al. (1996), Nord (1999a)). Hovedproblemet er at ingen av disse metodene kan sees på som en ”gullstandard”. I det følgende beskrives de sentrale teknikkene som benyttes til å kartlegge individers preferanser og syn på prioritering, samt en diskusjon over hva litteraturen mener om de.

2.2.1 Kort beskrivelse av metoder for kartlegging av nytte og preferanser

De tre mest brukte teknikkene for å måle individers preferanser for forskjellige helsetilstander eller behandlingsutfall, er Rating scale (RS) og dens varianter, Time trade-off (TTO) og Standard Gamble (SG). Undersøkelsene går ut på at individ blir stilt ovenfor forskjellige scenarier og blir bedt om å rangere disse. Resultater og verdier fra målinger av preferanser fra en statistisk gruppe i befolkningen kan benyttes videre i MAU-instrumenter. Teknikkene er kort oppsummert nedenfor.

Rating scale (RS)

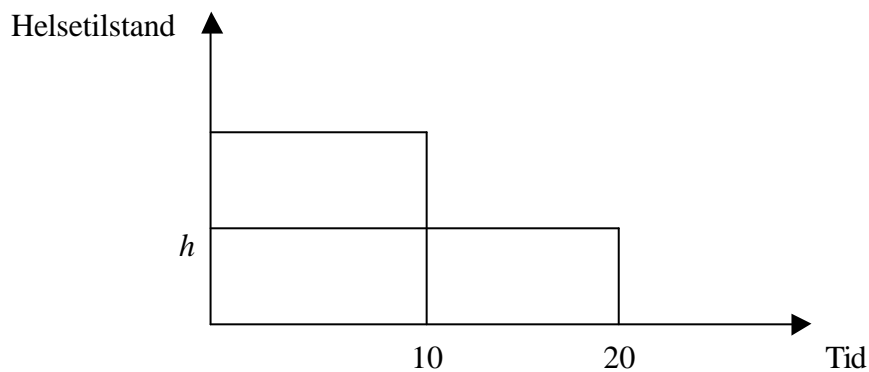
Respondenten blir her spurt om å indikere styrken på sine preferanser for en helsetilstand A, på en rett linje kalibrert mellom 0 (død) og 1 (fullgod helse). Avstanden over og under den valgte verdien representerer styrken på preferansen relativ til full helse og død. Fordelen med metoden er muligheten til å kunne sammenligne helsetilstander direkte, her representer ved en helsetilstand B.



Figur 2.2 Rating Scale

Time trade off (TTO)¹⁵

I TTO blir respondentent spurt om å indikere hvor mye av sitt gjenværende liv (eller forventet livslengde utnevnt av intervjuer) de er forberedt på å gi opp, før å bli returnert i fra en hel livstid (eller utnevnt tid) i helsetilstanden som måles, til full helse. Andelen livslengde som blir gitt opp, indikerer den negative nytten ved denne helsetilstanden. For eksempel hvis en person er villig til å gi opp 20% av sin gjenværende livslengde for å unngå en helsetilstand, vil den negative nytten knyttet til denne helsetilstanden tilsvare 0,2 og nytten vil tilsvare 0,8. Dette kan visualiseres i figur 2.3 hvor h er en helsetilstand med for eksempel nyresvikt. Hvis et individ er villig til å oppgi 10 leveår med nyresvikt av sine 20 antatte gjenværende leveår, mot å kunne leve som frisk i 10 år, vil helsetilstanden h vurderes til å ha verdien $10/20 = 0,5$



Figur 2.3 Time Trade-off

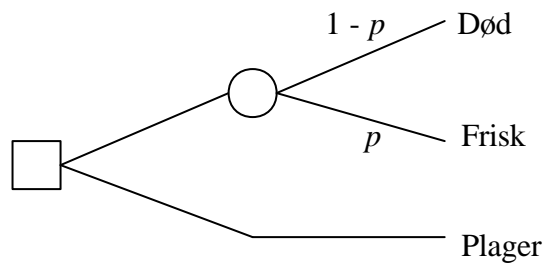
Standard Gamble (SG)¹⁶

I SG blir respondentent bedt om å velge mellom en livstid (eller et valgt antall år) i helsetilstanden som skal måles, eller et "lotteri" hvor det er en sannsynlighet p for at en vil leve med full helse, og en sannsynlighet $(1-p)$ for at en vil nesten umiddelbar dø. Sannsynligheten p varieres inntil individet er indifferent. I dette punktet representerer sannsynligheten p betydningen vekt eller indeks av nytte. Dette resultatet inntreffer når nytte verdien av full helse (U_1) og død (U_0) settes respektivt lik 1 og 0. Fra dette følger det at nytte verdien ved sykdom (U_s) kan uttrykkes som:

$$U_s = p * U_1 + (1 - p) * U_0 = p * 1 + (1 - p) * 0 = p.$$

¹⁵ Time trade off ble utviklet av Torrance, Thomas og Sakett (1972) til bruk innenfor helsesektoren.

¹⁶ Standard gamble er den klassiske metoden for å måle kardinal nytte og ble først presentert av Von Neuman og Morgenstern (1953).



Figur 2.4 Standard Gamble

2.2.2 Sammenligning av metodene

De to mest brukte metodene er i følge litteraturen SG og TTO. Dette bunner i at valgbaserte metoder som SG og TTO sees på som best i forhold til skalerings baserte som RS. Fordelen med de valgbaserte metodene, er at det å gjøre valg er en naturlig menneskelig oppgave hvor vi har mye erfaring i motsetning til det å skalere noe (Drummond et al. (1997)). I tillegg er disse metodene observerbare og verifiserbare i motsetning til RS.

Enkelte hevder at fremtidig helse er forbundet med usikkerhet, og at preferanser målt under usikkerhet (som SG) dermed er de mest "riktige" (se ref. i Dolan et al. (1996)). I tillegg hevder flere helseøkonomer at SG teoretisk sett er den beste metoden, siden SG baseres direkte på veletablerte og fundamentale deler av Von Neumann og Morgenstern sine aksiomer i forventet nytte teori. Dette har ført til at SG har lenge blitt sett på som en "gullstandard". På den andre siden er det funnet bevis for at folk bevisst systematisk bryter med aksiomene, og at det ikke finnes noe empirisk grunnlag for dem (Doland et al. (1996), Richardson (1997)). Dermed faller mye av grunnlaget til SG vekk.

TTO har i følge Krabbe et al. (1997) en posisjon innenfor fagfeltet, hvor den blir sett på som mer gjørlig/mulig enn SG, og mer "realistisk" enn RS. En del helseøkonomer hevder at en kombinasjon av levetid og livskvalitet, målt i form av ett enkelt helsemål som kan bli brukt i nyttekostnadsanalyser, blir best ivaretatt gjennom TTO. Dette fordi TTO gir mer valide og reliable resultater enn SG, ettersom individene i denne metoden (TTO) bruker en direkte sammenligning av utfallene. Det kreves dermed at respondenten eksplisitt tar i betraktning bytteforholdet mellom kvantitet og kvalitet, noe som de andre metodene ikke krever.

Siden oppgaven et individ utfører i hver av de tre metodene er ulike, bør en ikke være overrasket over at preferansescoren funnet i empirien fra de ulike metodene er forskjellige. I henhold til Drummond et al. (1997) og Krabbe et al. (1997), er vanlige funn for metodene at

SG gir en høyere preferansescore enn TTO. Dolan et al. (1996) finner derimot det motsatte i sin undersøkelse. Han konkluderer med at folk er mer villig til å ta en usikker gamble som involverer risiko for å dø, enn å ofre en hvis mengde av livslengden sin. Hvorfor er det slike forskjeller? Dette temaet er komplisert ved det faktum at verdsettingen av helsetilstander ved valgbaserte metoder, er en funksjon av både alvorlighetsgraden til helsetilstanden og innholdet i valget. Responsen fra SG spørsmål er sannsynligvis påvirket av individets forhold til risiko. Responsen til TTO spørsmål er trolig påvirket av livslengde og tidspreferanser. Det konkluderes derfor med at grunnen til forskjellen mellom SG og TTO, er at SG fanger opp risiko adferd, og at de høyere verdiene for SG som en ofte finner kan forklares ut i fra risikoaversjon og diskontering (Brooks (1996)). Forskjellen mellom RS og TTO (samt SG) er antatt å ligge i forskjellen mellom det å velge og skalere noe¹⁷.

Diskusjonen over viser at det er vanskelig å velge mellom metodene basert på teoretisk grunn alene (og tydeligvis på empirisk grunn også (se Krabbe et al. (1997)), siden ingen av verdiene i fra metodene automatisk kan antas å ”mappe” direkte til nytte (Doland og Roberts (2002)). Metodenes forskjellige utgangspunkt kan være en viktig forklaring av flere bakenforliggende årsaker til hvorfor en oppnår forskjellige resultater for de ulike metodene. Dette er en faktor som er med på å gjøre det vanskelig å si hvilke av metodene som kan sees på som de mest ”riktige”. Videre vil en tro at disse forskjellene har en innvirkning på resultatene MAU-instrumentene genererer, og er med på å bygge oppunder de forskjellige svarene som MAU-instrumentene gir for samme helsetilstand.

2.3 Multi Attribute Utility (MAU) -instrumenter

Metodene beskrevet i kapittel 2.2 er generelle metoder for å kartlegge preferansene til individer. MAU-instrumenter bygger på resultater og verdier fra målinger av preferanser fra statistiske grupper i befolkningen, generert gjennom en eller flere av disse teknikkene. MAU-instrumenter måler helse i flere dimensjoner, og er således en ”systematisering” av RS, TTO eller SG på mange dimensjoner. I det følgende skal vi beskrive MAU-instrumentene brukt i den empiriske analysen, og deretter peke på noen kjente problemer ved disse.

¹⁷ Donald og Roberts (2002) fant at alder, kjønn og sivilstand er de viktigste karakteristika for hvordan folk verdsetter en helsetilstand.

2.3.1 MAU-instrumentenes oppbygning og hva de fanger opp

MAU-instrumentenes oppbygning

MAU-instrumenter er instrumenter utviklet for å måle individers helse i flere dimensjoner. Instrumentene er formet slik at de har en persons velvære (og ikke sykdom) som utgangspunkt ved måling av helse. MAU-instrumentene består av to deler: (1) Et deskriptivt system som ofte er utformet som et spørreskjema angående forskjellige helsedimensjoner (syn, hørsel osv.), og et variabelt antall svaralternativer/nivå (god, middels, dårlig osv.). Individet indikerer enten på egenhånd, eller i et intervju, i hver dimensjon hvilke svaralternativ/nivå som best korresponderer til deres helsetilstand. Denne beskrivelsen danner individets helseprofil. (2) Instrumentet består videre av et scoringssystem hvor hvert svaralternativ er gjennom metode og modellkalibrering¹⁸ gitt en verdi på en skala mellom 0 og 1, som uttrykker en objektiv verdsetting av helsemålet (se for eksempel Hawthorne (2000)). Responsen vektes og kombineres ved hjelp av utviklete tabeller og matematiske formler (additiv, multiplikativ, multilinear), som omregner helseprofilen til en enkel tallverdi. Tallverdien beskriver individets nytte ved å være i en bestemt helsetilstand. Tabellene og de matematiske formler bygger på statistiske analyser av tidligere populasjonsdata, hvor en har kartlagt preferanser i den generelle populasjonen ved hjelp av en eller flere av metodene beskrevet i 2.2. Kartleggingen viser hvor mye de forskjellige dimensjonene ved helse skal vurderes relativt til hverandre. Dette gir objektive nivåsmål, dimensjonsmål og helsemål i MAU-instrumentet.

Gjennom sin oppbygning tar MAU-instrumenter mål av seg å måle livskvalitet i kontekst av helse. For de fleste mennesker refererer termen "livskvalitet" (QoL) til hvor godt, ønskelig og behagelig livet som helhet føles (Nord et al. (2001)). Med andre ord individets følelse av glede eller tilfredsstillelse, som typisk reflekterer en global bedømmelse av alle aspekter ved livet (O'Connor (1993)). Siden det er pasienten som gir sin subjektive vurdering om egen helse, gjennom selv å fylle ut spørreskjemaet i MAU-instrumentet, refereres livskvalitet ofte til som "subjektiv velvære" (subjective well-being) (Fitzpatrick et al. (1998)). Hva som eksakt menes med livskvalitet er et kontroversielt tema, som har medført at det i litteraturen er et fravær av en presis og samstemt definisjon på livskvalitet i kontekst av helse (Muldoon

¹⁸ RS, TTO og SG er generelle metoder for å kunne kartlegge individers preferanser for ulike helsetilstander. Disse preferansene blir kalibrert for å finne individers helsemål gjennom et MAU-instrument. En slik kalibrering skjer gjennom en vekting av de nivåsmål og/eller dimensjonsmål som de ulike metodene har generert.

(1998)). Begrepet ”livskvalitet” brukes derfor ofte vagt og uten en klar definisjon (Sanders et al. (1998))¹⁹. Mangel på en klar definisjon er med på å komplisere hvilke elementer som skal inngå ved måling av helse. Det må i denne sammenheng påpekes at hver helsedimensjon representerer en latent variabel, sådan gjør det globale konseptet livskvalitet. En latent variabel er en uobservert variabel som er postulert som en konstruksjon og er definert av et sett av dimensjoner. Dvs. at livskvalitet eller nytte ikke er noe som vi kan observere direkte. Men et MAU-instrument tar for seg, (a) observerbare karakteristikk, som (b) regnes om til subjektive svar. Dette er den beste metode foreløpig gitt at vi kan måle nytte på intervall nivå.

Helserelatert livskvalitet

Den enkle tallverdien som MAU-instrumentene regner ut i fra individens helseprofil, er en indeks verdi som representerer et estimat over styrken på preferansene for forskjellige helsetilstander. Verdien blir vanligvis tolket som et kvantitativt mål på helserelatert livskvalitet (Torrance et al. (1995)). Et slik nyttemål er designet til å fange opp ”alle” aspekter omkring en helsetilstand, både de positive og negative, og representerer verdien til den helsestatusen (Hawthorne (2000)). Termen ”helserelatert livskvalitet” (HRQoL) brukes som oftest som en distinksjon i fra termen ”livskvalitet” blant helseøkonomer. Livskvalitet måles i form av en vurdering av hvordan individets helserelaterte livskvalitet påvirkes, dvs. individets evne til å leve et normalt liv, klare seg selv, grad av smerte og angst osv. Gjennom å bruke en preferanse basert konstruksjon ved måling av konseptet helserelatert livskvalitet, får en som nevnt et estimat på styrken på preferansene for forskjellige helsetilstandene. Dvs. en helsetilstand som representerer en større helserelatert livskvalitet for et individ enn en annen helsetilstand, vil bli foretrukket av det individet. Sant å si er det vanskeligere å ellers kunne se hva livskvalitet kunne bety. Folk kan være veldig forskjellige med hensyn til hvordan de ser på livskvalitet assosiert ved forskjellige utfall. Imidlertid, innenfor en individuell helsetilstand som bedømmes til å være bedre enn en annen i termer av livskvalitet, er den som foretrekkes. Wagstaff (1991) påpeker at selv om nytteteori hyppig brukes i avledningen av livskvalitet scores, så er det bare brukt til å måle folks helse snarere enn nytten de oppnår fra den. Derfor må det påpekes at nytten vi snakker om her ikke er individets direkte nytte (helserelatert livskvalitet er ikke = nytte), men en indikator på nytte. Derfor burde en gjerne bruke begrepet ”helserelatert nytte” istedenfor ”nytte”. Vi holder oss foreløpig til begrepet ”nytte”.

¹⁹ Se diskusjon i Patrick et al. (1998)

Nytte (livskvalitet) består av mange hundre komponenter/dimensjoner. Dette kan uttrykkes i form av en vektor med n -element som i likning (2.6). Et MAU-instrument tar gjerne bare for seg 10-15 av disse komponentene, og gjerne mindre, som for eksempel EQ-5D som tar for seg 5 dimensjoner. EQ-5D med sine 5 komponenter kan uttrykkes som i likning (2.7). Teoretisk sett hadde det beste vært å ta med seg alle komponentene fordi dette hadde vært det mest nøyaktige. Men en slik strategi hadde vært alt for omfattende, noe som betyr at en må ta utgangspunkt i noen få. Spørsmålet blir således hvilken av disse komponentene som er viktigst ved måling av helse. Dette er et spørsmål hvor utviklerne av MAU-instrumentene strides, og er en av flere grunner til at en har så mange ulike instrumenter som påstår å måle konseptet "helsereelatert livskvalitet".

$$\text{Livskvalitet} = f(x_1, x_2, \dots, x_n) \quad (2.6)$$

$$\text{MAU} = f(x_1, x_2, \dots, x_5) \quad (2.7)$$

Et annet spørsmål er hvordan en skal veie disse dimensjonene opp mot hverandre avhengig av deres betydning²⁰. Fra det utilitaristiske synspunktet menes det at alle disse dimensjonene summeres. Det impliserer jo høyere sum, desto mer nytte. Så enkelt er det ikke. Dimensjonene vektet forskjellig avhengig av hva som er viktigst (Nord (1999a)). For eksempel vil en tro at smerte veier mer enn tap av syn i den totale nyttefunksjonen til et individ. MAU-instrumentene må kunne avspeile samfunnets preferanser for å valid kunne måle helse.

Helsereelatert livskvalitet er et relativt løst konsept som refererer til sub-sett av livskvalitet som kan bli påvirket av en persons helsestatus. Ettersom et helsetiltak potensielt kan påvirke alle aspekter ved livet, kan det argumenteres for at helsereelatert livskvalitet er et identisk konsept til livskvalitet. Distinksjonen brukes vanligvis for å implisere at helsereelatert livskvalitet ikke måler omgivelsene en person lever i, eller deres sosioøkonomiske status (Hawthorne et al.

²⁰ Multiattributt nytte teori, en utvidelse av konvensjonell nytteteori av Keeny og Raiffa (1976) kan anvendes for å modellere preferansescorene for helsetilstander definert ved et multiattributt-helsestatus klassifiseringssystem. Typen preferanseuavhengighet mellom attributtene/dimensjonene avgjør nødvendig type preferansefunksjon. Man skiller mellom tre ulike modeller for å kombinere MAU-instrumentenes ulike dimensjoner. (1) Additive: Helsemålet blir kun bestemt ut i fra individets individuelle svar på de ulike dimensjonene, og ikke på hvordan de ulike dimensjonene kombineres. Dvs. ingen interaksjon mellom dimensjonene kreves. (2) Multiplikative: Krever interaksjon mellom individets preferanser i enkelte spørsmål, men at disse er uavhengige av andre spørsmål. (3) Multilineære: Impliserer full interaksjon mellom individets preferanser for ulike dimensjoner (Drummond et al. (1997)).

(1999)). Nord et al. (2001) setter spørsmål ved om en ikke gjør dette ved måling av helse, og om ikke helsereelatert livskvalitet er identisk til livskvalitet. Selv om det ikke eksisterer noen presis eller samstemt definisjon på livskvalitet (eller helsereelatert livskvalitet sådan), er det til en hvis grad enighet blant helseøkonomer om at helsereelatert livskvalitet er et ”multidimensjonalt” konsept som omfatter fysiske, emosjonelle og sosiale komponenter assosiert med sykdom eller behandling (Sintonen (1994), O’Connor (1993)). Gjennom MAU-instrumentene søker en ved dette konseptet essensielt to typer informasjon i fra individet. (1) den funksjonelle statusen til individet, og (2) pasientens vurdering, eller vurderingen til folk generelt, av hvordan helsen påvirker han eller hennes livskvalitet (Muldoon (1998)).

2.3.2 Beskrivelse av ulike MAU-instrumenter

Her beskrives de fire MAU-instrumentene EQ-5D, QWB, HUI3 og 15D, som blir brukt i den empiriske analysen i kapittel 4. Instrumentene ”bygger over samme lest”, derfor blir kun EQ-5D beskrevet inngående her for å gi en grunnleggende forståelse av hvordan et MAU-instrument fungerer. Alle instrumentene er beskrevet i detalj i Drummond et al (1997), unntatt 15D som er nøye beskrevet i Sintonen (1994) og Sintonen (1995)

The Euroqol Classification System (EQ-5D)

EQ-5D²¹ er et generisk instrument for måling av helsereelatert livskvalitet. Instrumentet ble designet for en sammenligning av individers helse på tvers av de europeiske land, hvor nyttevektene i instrumentet ble oppnådd fra et representativt utvalg fra populasjonen i UK ved bruk av TTO teknikk. EQ-5D sitt deskriptive system består av 5 dimensjoner, *gange*, *personlig stell*, *vanlige gjøremål*, *smerte/ubehag* og *angst/depresjon*. Hver dimensjon har tre svaralternativ/nivå over målet på helse, som er betegnet som *ikke noe problemer*, *noen problemer*, eller *ekstreme problemer*. De fem dimensjonene, hver med tre nivå, gir teoretisk sett mulighet for $3^5 = 243$ ulike svarkombinasjoner. Instrumentet brukes ved at et individ blir bedt om å markere det nivået som er mest beskrivende for deres nåværende funksjonsnivå i hver av de fem dimensjonene. Ved å kombinere responsen fra hver av de fem dimensjonene, får en dannet en 5-sifret kode som utgjør individets helseprofil. De ulike helsemålene kan bli funnet ved å bruke tabell 2.1 og likning (2.8). Dimensjonene leses vertikalt, mens responsverdiene/nivåene leses horisontalt. Responsverdien representerer unytten ved den respektive helsetilstanden.

²¹ Se også Brooks (1996) for detaljert beskrivelse av EQ-5D.

Dimensjon/nivå	Ikke noen problemer	Noen problemer	Ekstreme problemer
Gange	0	0,069	0,314
Personlig stell	0	0,104	0,214
Vanlige gjøremål	0	0,036	0,094
Smerte/ubehag	0	0,123	0,386
Angst/depresjon	0	0,071	0,236
		Konstant (K) = 0,081	Konstant (C) = 0,269

Tabell 2.1 TTO gjennomsnitt²²

EQ-5D tilhører en undergruppe av additive helsefunksjoner og utregningen av helsemålet er vist i likning (2.8). Dimensjonsmålet for *vanlige gjøremål* for svaralternativ 2 kan skrives som $k_3 u_3(x_3=2)=0,036$. Konstantene K og C skal kun tillegges en gang, og bare hvis henholdsvis svaralternativ 2 eller 3 oppnåes. Dermed kan helserelevanter livskvalitet i EQ-5D regnes ut med følgende likninger.

$$DU(x)^{EQ-5D} = \sum_{j=1}^5 k_j u_j(x_j) + \mathbf{b}K + \mathbf{g}C \quad (2.8)$$

$$\sum_{j=1}^5 k_j = 1 \quad (2.9)$$

$$\mathbf{b} = \begin{cases} 1 & \text{hvis } k_j u_j(2) \neq 0 \\ 0 & \text{ellers} \end{cases} \quad (2.10)$$

$$\mathbf{g} = \begin{cases} 1 & \text{hvis } k_j u_j(3) \neq 0 \\ 0 & \text{ellers} \end{cases} \quad (2.11)$$

For å finne helserelevanter livskvalitet gjennom EQ-5D, kan en se på et individ med for eksempel helseprofilen 21111. Helseprofilen tilsier at individet har litt problemer med å gå omkring, men har ellers ingen plager. Ut i fra en oppslagtabell som følger med EQ-5D kan en konvertere individets helseprofil til et individuelt mål på helse, som i dette tilfellet vil være 0,850. Uten oppslagtabell brukes likning (2.8) og tabell 2.1, hvor EQ-5D TTO scorene kalkuleres ved å subtrahere de relevante koeffisientene (unytten) fra 1 som i likning (2.12).

²² Tabell og likninger bygger på Drummond et al. (1997). Notasjon: $u_j(x_j)$ er en enkel attributt nyttefunksjon for attributt j . $DU(x)$ er unytten ("disutility") forbundet med helsetilstand x , representert ved en n -element vektor. k og k_j er modellparametere hvor k er en skaleringskonstant og k_j er dimensjonsvekt for dimensjon j . $U(x)$ er nytten forbundet med helsetilstand x .

$$\begin{aligned}
U(x)^{EQ-5D} &= 1 - DU(x)^{EQ-5D} \\
&= 1 - (0,081 + 0,069 + 0 + 0 + 0 + 0) = 0,850
\end{aligned}
\tag{2.12}$$

Dette gir det samme resultatet som oppslagtabellen.

Quality of Well-being (QWB) Scale

QWB fokuserer på påvirkningen av sykdom og svakheter på funksjon og observerbar adferd, som utførelsen av sosial rolle, evnen til å bevege seg rundt i samfunnet, og fysisk funksjon. Dvs det gjøres et skille mellom funksjonell evne og funksjonell utførelse. Istedenfor å be individet om å rapportere antagelser om hva som muligens kunne blitt utført, rapporteres utført aktivitet. Altså fokuserer QWB på funksjonsutførelse (eller hva individet faktisk gjorde). MAU-instrumentet bruker preferansevektorer til å integrere de tre dimensjonene *Mobilitet (MOB)*, *Fysisk aktivitet (PAC)* og *Sosial aktivitet (SAC)* som hver har tre svaralternativ/nivå, samt en *symptom-/problem (CPX) kompleksitet liste* med 23 svaralternativ, inn til et enkelt nummer mellom 0 og 1. Nummeret klassifiserer helse relatert livskvalitet for individet. QWB tilhører gruppen av additive instrumenter som kan representeres ved likningene (2.13) og (2.14). Utregningen skjer ved likning (2.15) som er den samme som for (2.12). Leddet w_t er de preferansevektede målene for hvert nivå i dimensjonene:

$$DU(x)^{QWB} = \sum_{j=1}^4 k_j u_j(x_j) \tag{2.13}$$

$$\sum_{j=1}^4 k_j = 1 \tag{2.14}$$

$$\begin{aligned}
U(x)^{QWB} &= 1 - DU(x)^{QWB} \\
U(x)^{QWB} &= 1 - [(CPXw_t) + (MOBw_t) + (PACw_t) + (SACw_t)]
\end{aligned}
\tag{2.15}$$

QWB genererer til sammen 1024 ulike svarkombinasjoner. CPX-listen er spesielt for dette MAU-instrumentet. Denne ble inkludert med tanke på at ”subjektive klager er en viktig komponent til et generelt helse mål, fordi de relaterer dysfunksjonalitet til et spesifikt problem”. I tillegg representerer de 23 CPX- kompleksiteten ”alle mulige klager som kan hindre funksjon” (se ref. i O’Connor (1993)). Selv om disse symptomene ikke nødvendigvis

påvirker pasientens fysiske funksjon, fører de til en lavere livskvalitet (Kaplan (1989)). De standardiserte spørsmålene brukes til å klassifisere individet inn til et av nivåene i hver av dimensjonene, mens individene klassifiseres i CPX-listen ved den typen symptom eller problem som plager dem mest. MAU-instrumentet er basert på en ad-hoc undersøkelse i USA ved hjelp av RS teknikk. Instrumentet har et minimumsmål på 0. Dvs. at QWB respondenter ikke anser noe helsemål, oppnåelig gjennom instrumentet, som verre enn døden.

15-D

MAU-instrumentet 15D består av 15 dimensjoner, *bevegelighet, syn, hørsel, pust, søvn, spising, tale, vannlating/avføring, vanlige aktiviteter, mental funksjon, ubehag og smerter, depresjon, stress, livskraft og seksuell aktivitet*. Til hver av de 15 dimensjonene er det 5 svaralternativ/nivå, som til sammen gir muligheten for opptil 30 517 578 125 ulike svarkombinasjoner. Dimensjonsmålene er produsert gjennom en undersøkelse av preferanser blant et representativt utvalg i den finske befolkning, ved hjelp av RS teknikk. Til hvert svaralternativ i instrumentet hører det til en "verdinøkkel" som tallfester hvert svaralternativ i hver dimensjon, slik at en får dimensjonsmålet (u_j) direkte. Ved å kombinere dimensjonsmålene gjennom følgende additive funksjon, finner en individets helserelevante livskvalitet:

$$DU(X)^{15-D} = \sum_{j=1}^{15} k_j u_j(x_j) \quad (2.16)$$

$$\sum_{j=1}^{15} k_j = 1 \quad (2.17)$$

15D tilhører gruppen av additive MAU-instrumenter og har verdien 0,10 som minimumsmål. Dette er en tilstand som er marginalt bedre enn døden, men betraktelig høyere enn minimumsmålet til EQ-5D som er på -0,59. Som vi skal se senere i kapittel 4 kan forskjellen i minimumsmål mellom MAU-instrumentene, ha betydning for målingen av gjennomsnittlig helserelevante livskvalitet.

Health Utility Index Mark III (HUI3)

Det finnes tre modeller av HUI. Modell 3 (HUI3) består av åtte dimensjoner *syn, hørsel, tale, mobilitet, førlighet (i hender), emosjonalitet, kognisjon*²³ og *smerte*, med fem til seks ordnede svaralternativ/nivå for hver dimensjon. Dimensjonene inkludert i HUI3 er valgt på basis av hva folk synes er viktig (Feeny et al. (1995)). HUI3 tillater 972.000 forskjellige svarkombinasjoner for mål på helsetilstander. Preferansescoren i instrumentet er basert på SG- og VAS-²⁴ teknikk. Disse ble oppnådd fra et randomisert representativ utvalg fra den generelle populasjonen i Canada eldre enn 16 år, og er basert direkte på Von Neuman/Morgensten nytteteori, og utvidelsene som er gjort for å imøtekomme multiple attributter (Torrance et al. (1995)). HUI3 fokuserer på kapasitet i motsetning til utførelse, altså det omvendte av QWB. Instrumentet adopterer en "within the skin" definisjon av helsestatus, hvor det fokuseres på de viktigste dimensjonene ved fysisk og emosjonell helse (Fenyet et al. (1995)). Instrumentet bruker en multiplikativ modell, her representert ved likning (2.18) og (2.19), for å kombinere individets dimensjonsmål. Modellen innehar full strukturell uavhengighet, noe som betyr at hver mulig kombinasjon av nivå mellom dimensjonene er logisk mulig. For å kalkulere helserelatert livskvalitet kombineres responsen til et individ fra hver av de åtte dimensjonene i HUI3. Dette danner en 8-sifret kode som utgjør individets helseprofil, for eksempel 21121213. Individets helserelaterte livskvalitet regnes ut ved å kombinere tilhørende dimensjonsmål (se Feeney et al. (1995)), samt ved å bruke likning (2.20).

$$DU(x)^{HUI3} = (1/k) \left[\prod_{j=1}^8 (1 + k k_j u_j(x_j)) - 1 \right] \quad (2.18)$$

$$k = \prod_{j=1}^8 (1 + k k_j) - 1 \quad (2.19)$$

$$U(x)^{HUI3} = 1,371 * (0,98 * 1 * 1 * 0,93 * 1 * 0,95 * 1 * 0,90) - 0,371 = 0,70 \quad (2.20)$$

HUI3 tillater negative verdier hvor den laveste tillate verdien for en helsetilstand er $-0,36$. Individet vil i dette eksempelet få foreskrevet en helserelatert livskvalitet på $0,70$ på en $-0,371$ til 1 skala.

²³ Kognisjon: I betydningen psykisk prosess som har å gjøre med tenkning og bevissthet.

²⁴ Visual analogue scale (VAS) likner et termometer med endepunkt på 100 (beste tenkelige helsetilstand) på toppen og 0 (verste tenkelige helsetilstand) på bunnen. Dette er en enkel metode for å erverve en egen klassifisering (rating) av nåværende helserelaterte livskvalitet ved å generere en score. Er tilnærmet det samme som RS teknikken.

Forskjeller mellom MAU-instrumentene

Tabell 2.2 viser en kortfattet oversikt over MAU-instrumentene som er beskrevet. Som tabellen og utledningen forut viser er MAU-instrumentene forskjellige på nesten alle punkter, noe som er med på å vanskeliggjøre en direkte sammenligning mellom dem.

MAU-Instrument	Måle Metode	Modell	Minimums Mål	Antall dimensjoner	Antall respons
EQ-5D	TTO	Additiv	-0,59	5	3
HUI3	SG	Multiplikativ	-0,36	8	5-6
QWB	RS	Additiv	0	3 (4)	3-5(23)
15D	RS	Additiv	0,10	15	5

Tabell 2.2 Egenskaper ved de ulike MAU-instrument

Noe av det mest iøynefallende er forskjellen i antall dimensjoner mellom instrumentene. Denne forskjellen bygger på at det er liten enighet blant utviklerne om hvor mange og hvilke dimensjoner som skal inkluderes for å representere de fysiske, emosjonelle og sosiale komponenter ved måling av helserelatert livskvalitet. Dvs. det er liten enighet om hvordan en skal operasjonalisere komponentene som målbare dimensjoner. Torrance (se ref. i O'Connor (1993)) mener at den fysiologiske og emosjonelle funksjonen bidrar direkte til livskvalitet, og satt sammen utgjør disse to helserelatert livskvalitet. Sosial funksjon (sosial rolle og sosial kontakt) er noe som han mener ligger utenfor omfanget til helserelatert livskvalitet, og skal derfor ikke være med. På den andre siden bruker Kaplan (1989) termen helserelatert livskvalitet til å referere til påvirkningen av helsetilstanden på funksjon, men inkluderer sosial rolle, selv om han mener at helserelatert livskvalitet kan være uavhengig av livskvalitet relevant til arbeid, innlosjering, eller lignende faktorer. Mangelen på en klar definisjon av hva helserelatert livskvalitet innebærer, er som vi har sett reflektert i de mange instrumentene som har blitt foreslått for å måle den, og en mangel på gullstandard mellom dem (Sanders et al. (1998), Sintonen (1994)). En annen iøynefallende ting i tabell 2.2 er forskjellene i minimumsmål mellom instrumentene, som representerer laveste oppnåelige helsemål ved hvert instrument. Negative helsemål er ikke et problem i seg selv. Men derimot det at MAU-instrumentene opererer med forskjellige minimumsmål vil vanskeliggjøre sammenligning mellom de. Dette gjelder ikke for den øverste delen av skalaen siden 1 er toppverdien alle bruker. Uansett så påpeker Hansen (2002) at jo lavere minimumsmål et MAU-instrument har, jo høyere behandlingseffekt tenderer instrumentet å gi. Forskjellene mellom instrumentene og hva det kan føre til, er et tema som vi kommer nærmere tilbake til under analysen, og under diskusjonen av resultatene.

2.4 utfordringer ved bruk av CUA som evalueringsmetode

2.4.1 Metodiske utfordringer

Helseøkonomiske modeller som CEA og CUA kan bidra til bedre beslutninger angående prioriteringer, slik at en får kostnadseffektiv bruk av tilgjengelige ressurser. Men en bør merke seg det at å maksimere helseforbedring imidlertid ikke er det eneste formålet med helsetjenesten. Selv om ”mest mulig helse for pengene” er et kraftfullt slagord, viser det seg at det kan komme i konflikt med andre hensyn som for eksempel rettferdig fordeling av godet (NOU 1997:18). QALY er et mål på helseforbedring som vi gjerne ønsker mest mulig av. Men en maksimering av QALY er ikke nødvendigvis den beste løsning på prioriteringsspørsmålet (Wagstaff (1991)). Samfunnets syn på etikk, fordelinger og rettferdighet er faktorer som også må tas med.

Det eksisterer mange forskjellige oppfatninger om hva som er rettferdig fordeling av godet (helseforbedring). Kort kan det nevnes at hovedskillet går mellom det utilitaristiske synet på rettferdighet som Williams (1992) argumenterer for, og det egalitristiske synet som Nord (1999a) argumenterer for. I det utilitaristiske synet maksimeres helseforbedringen uten noe føringer på hvordan fordelingen skal skje. Altså er det ingen motsetning mellom effektivitet og rettferdighet. Tanken her er at den som får mest mulig helsenytte igjen for behandlingen, også får høyeste prioritet. I det egalitristiske synet er det andre hensyn enn mest mulig helseforbedring som også skal telle. Dette kan være alder, sykelighet osv. De to synene illustrerer at nytte kan defineres på forskjellige måter, hvor det første synet er mye kritisert for å ikke ta hensyn til en rettferdig fordeling (se Wagstaff (1991)). Det siste synet gir støtte til å inkorporere andre fordelingsmålsetninger i tillegg til helseforbedring. Rawls (1974) går enda lengre i denne retningen, og mener at en helseforbedring har ingenting å si, kun fordeling hvor alle har en lik rett til godet. Med det mener han at omfordeling er mye viktigere enn effektivitetstap. For utilitaristene blir dette en avveining. Dette betyr ikke at en nødvendigvis skal gi opp nyttemaksimeringsprinsippet i helseøkonomiske modeller. En relativ beregning av helseforbedring i forhold til for eksempel det antall QALY et individ totalt kan forvente, innebærer en utvidelse av nyttebegrepet som bryter med den rendyrkede utilitaristiske definisjonen (Norheim og Bringedal (2001)).

CUA har blitt kritisert for å ikke ta nok hensyn til rettferdighet og fordeling i forhold til det som er allmenn oppfatning (Nord (1999a), Wagstaff (1991)). Nord argumenterer for at folk

har sterke preferanser for rettferdighet i fordelingen av helsetjenester og for de svakest stilte, og at enkelte synes det er mer verdifullt å behandle svakstilte grupper, selv om behandlingseffekten er lavere for denne gruppen enn andre. Han mener derfor at QALY er et mål på individuell nytte som ikke er riktig grunnlag å bruke i prioriteringsdebatten. QALY er fordelingsnøytral, og folk er det ikke (Richardson (1997)). Kriteriene i prioriterings spørsmål må reflektere folks syn på hva som er riktig å gjøre ved for eksempel fordelings- og rettferdighetshensyn (Ubel (1999))²⁵. Dette er et eksplisitt verdisyn som kan inkorporeres i analysen. Som løsning foreslår Nord å bruke Person Trade-Off²⁶ metoden for å vekte fordeling og rettferdighetshensyn i større grad enn hva TTO, RS og SG gjør. Metoden kartlegger folks preferanser i forhold til hvem som fortjener behandling. Spørsmålene som stilles går ut på å sette lidelser opp mot hverandre, for å finne ut hvor mange tilfeller av den ene lidelsen som må behandles, for at de skal være likeverdige med å behandle et tilfelle av den andre lidelsen. Metoden har blitt kritisert for å ilegge rettferdighet for mye vekt, med det argumentet at det er lett å bry seg om andre når en selv ikke berøres av beslutningen (Doland et al. (2002)). Dette viser at CUA har en metodisk utfordring med hensyn til hvem sine preferanser en skal kartlegge, pasientens eller en representativ konsument. I tillegg har en spørsmålet om hvilke metoder en skal bruke for å kartlegge disse preferansene, hvilke kriterier som skal legges til grunn, og hvordan en skal veie preferansene opp mot en annen.

Det finnes mange faktorer en kan trekke inn i CUA, og som kan være et viktig moment i en prioriteringsdebatt. Først og fremst er det forhold som har med behandlingseffekt og behandlingsbehov som dominerer debatten, men i noen problemstillinger trekkes også andre moment inn (se for eksempel Aakvik (2001) for diskusjon rundt temaet). Kriterier som tilstandens alvorlighet, tiltakets nytte og tiltakets kostnadseffektivitet er alle relevante, men vektleggingen av hver av dem vil variere i forhold til hvilke mål som legges til grunn. Med dette som bakgrunn ser en at samfunnsøkonomiske analyser ikke kan erstatte verdibaserte valg, etiske grunnregler, klinisk skjønn eller sunn fornuft. Andre faktorer som likhet, rettferdig fordeling, tilgjengelighet og individuelle preferanser er også viktige. Likevel vil samfunnsøkonomiske analyser være nyttige i en totalvurdering av tiltaks funksjon i samfunnet. Gjennom å kartlegge og synliggjøre konsekvensene av ulike tiltak, kan

²⁵ Et offentlig utvalg (NOU 1998:16) har påpekt at QALY metoden ikke løser de vanskelige fordelingsproblemene som alltid vil være tilstede når knappe ressurser skal fordeles innenfor helsevesenet. Og at det derfor sees på som lite realistisk og ønskelig at ulike grupper skal prioriteres strengt etter en enkel indikator som QALY. Lønning utvalget (NOU 1997:18) viser til at livskvalitet er svært vanskelig fenomen å måle, og at QALY derfor ikke kan være eneste kriterium når en skal vurdere ønskeligheten av et helsetiltak.

²⁶ Se Nord (1999a) for detaljer omkring metoden.

samfunnsøkonomiske analyser hjelpe beslutningstakere med vanskelige og viktige valg angående ressursfordelingen innenfor helsesektoren, selv om resultatene i fra slike analyser gjerne ikke kan brukes alene i beslutningsprosessen.

2.4.2 Noen kjente svakheter ved MAU-instrumentene

Et problem ved MAU-instrumenter i praktisk bruk, er at det ennå ikke finnes noen "gullstandard". Underforstått betyr det at det ikke finnes noe MAU-instrument som peker seg ut som det beste og mest riktige. Det har vist seg at samme helsetilstand kan gi forskjellige resultater ved bruk av forskjellige MAU-instrumenter. Det er i tillegg også problemer forbundet med betydningen av verdiene MAU-instrumentene genererer. Dette er faktorer som er med på å gjøre sammenligninger av resultater fra ulike CUA usikre (Nord (1999c)). Mange av årsakene til dette har sin bakgrunn i det som ble diskutert i kapittel 2.4.2. Vi vil komme tilbake til dette temaet mer utfyllende i kapittel 4.1.3 i sammenheng med resultatene fra analysen. Her skal vi se litt nærmere på andre sider ved MAU-instrumentene som kan være et problem i sammenheng med CUA.

Merkverdige sammenstillinger i MAU-instrumentene

Et problem som har blitt påpekt ved MAU-instrumenter, er enkelte av disse har merverdige sammenstillinger av helseproblem. Et eksempel som kan trekkes frem, er i QWB hvor noen av verdiene på skalene ikke er helt i samsvar med hva som oppfattes som "naturlig". Eksempelvis er den høyeste negative vekten (unytten) assosiert med Fysisk aktivitet 0,077. Denne verdien korresponderer til "*I rullestol, beveger eller kontrollerer ikke den ved uten hjelp av noen andre, eller i seng, stol eller sofa hele dagen (helse relatert)*". På den samme siden har en at CPX-11 "*hoste, hvesing, kortpustet, med eller uten feber, fryser eller ondt over alt*" har vekten -0,257. Fordi hvert av elementene fra QWB dimensjonene adderes til 1 ($U(x)^{QWB} = 1 - (CPXwt) - (MOBwt) - (PACwt) - (SACwt)$), er CPX-11 estimert til å ha 3 ganger så mye å si for velværen enn det mest alvorligste fysiske begrensningen på aktivitet. Dette medfører at en person som er bundet til en rullestol uten andre begrensninger eller symptomer, vil få en MAU-score på 0,923 for den totale helsefunksjonen, mens en person med "*hoste, hvesing, kortpustet*" som ellers er frisk vil få en MAU-score på 0,743. Noe som virker helt absurd fordi en gjerne tror at "*hoste, hvesing, kortpustet, med eller uten feber, fryser eller ondt over alt*" går over etter en stund, mens det å sitte i rullestol ikke går over. Det er således blitt funnet at relativt små CPX-komplekser kan ha en kjempe effekt på den totale

MAU-scoren. Utviklerne av QWB forsvaret dette med at slike symptom-/problemvekter tillater helsetilstandsindeksen til å bli ”veldig sensitiv til små øvre ende variasjoner i helsetilstanden”. Istedenfor virker det som om de ”små” justeringene har en kapasitet til å overvelde hele funksjonsskalaen (O’Connor (1993)). Dermed får en det at behovet for å måtte gå med briller (-0,101), skaper en større reduksjon i velvære enn det å måtte sitte i en rullestol (-0,060 eller -0,077). Dette er ikke nødvendigvis et problem hvis andre aspekter i funksjonsskalaen også bidrar i slike tilfeller. Men faktumet gjenstår det at skalaene er ment til å være uavhengig og lineær addert i QWB.

Merkverdige sammensetninger av liknende karakter er funnet i enkelte av de andre instrumentene. En stor del av debatten her går på hvordan de enkelte dimensjonene i instrumentene skal veies opp mot hverandre. Sintonen (1995) påpeker at EQ-5D ofte er det motsatte av intuitiv, idet logisk verre tilstander i enkelte tilfeller får høyere score enn logisk bedre tilstander, og at dimensjonsvektene derfor ikke helt står i samsvar med hverandre. Dette kan ha en innvirkning på om et tiltak finnes lønnsomt eller ikke, på bakgrunn av hvilket instrument som brukes. Det at enkelte tilstander vektet ulogisk og ”feil” i instrumentene i forhold til hva som er av allmenn oppfatning i samfunnet, medfører et problem for CUA og dens bidrag i prioriteringsdebatten.

Sensitivitet

For enhver sykdom ønsker en å finne det instrumentet som måler helsen, på bakgrunn av sykdommen, på en mest mulig representativ måte. Dette tilsier at det ikke er nok å bare se på antall dimensjoner, men at en bør velge det instrumentet som best dekker sykdomsbildet. Undersøkelser som spør en person om dimensjoner som ikke angår en sykdom, kan komme i skade for å gi denne et MAU-score som er for høy. Dette skjer fordi denne personen får dimensjonsmål som tilsvarer friske personers mål på slike spørsmål, og hans reduserte dimensjonsmål kan dermed forsvinne²⁷. Ved kun å stille spørsmål som har direkte med sykdommen å gjøre, kan en muligens komme nærmere det ”sanne” helsemål. På en annen side vil en mengde med ”spesialist” instrumenter for hver sykdomstilfelle, føre til at en mister oversikten og gyldigheten av de ulike instrumentene og dermed validiteten til CUA.

²⁷ En person som er fullstendig sengeliggende vil oppnå en helserelatert livskvalitet på 0.336 i EQ-5D, 0.833 i QWB, 0.932 i 15D og 0.424 i HUI3.

Det er i litteraturen delte meninger omkring sensitiviteten til MAU-instrumenter med hensyn til deres evne til å fange opp små endringer i helsen. Hurst og Forbes (1999) mener at MAU-instrumenter er mer enn sensitiv nok i forhold til ”spesialist” instrumenter. Torgerson og Raftery (1999) påpeker at MAU-instrumenter, muligens på grunn av sitt brede omfang, ikke er tilstrekkelig egnet til å reflektere forandringer i helsetilstanden til en populasjon med en spesifikk sykdom. Og at de således ikke kan være sensitive nok i forhold til å fange opp forandringer i de dimensjonene som er viktige for pasienten. Garrat et al. (2000) mener på sin side at ”spesialist” instrumenter er MAU-instrumenter overlegen, idet at disse fokuserer på aspekter ved helsen som er viktige ved det spesifikke helseproblemet. Konsekvensen av de ulike meningene har medført at de fleste anbefaler at den diskriminerende evnen til MAU-instrumentet, dvs. evnen til å skille mellom de individene som får en helseforbedring og de som ikke får det, bør evalueres for hver sykdom. I tillegg bør gjerne MAU-instrumenter kombineres med ”spesialist” instrumenter for å evaluere deres sensitivitet opp mot disse.

Ved valg av riktig instrument til riktig lidelse eller sykdom, er det ikke bare spørsmål om riktige dimensjoner måles, men og hvordan dimensjonene er vektet i forhold til hverandre. Hovedbekymringen for pasienter med for eksempel smerter i korsryggen er trolig deres smerte. Det er og rimelig å tro at samfunnet vekter smerte mer enn andre aspekt i helsen (Suarez-Almazor et al. (2000)). Et MAU-instrument som tar innover seg dette, vil bedre måle endringer i helse når hovedmanifestasjonen til en spesiell tilstand er forbundet med smerte. EQ-5D og HUI3 har store unyttevekter (disutility) tilskrevet det verste nivået for smerte i forhold det verste nivået til de andre dimensjonene. Denne vektingen gir støtte til at smerte er oppfattet å ha en større påvirkning på livskvaliteten hos den generelle populasjonen, enn andre faktorer/dimensjoner. Suarez-Almazor et al. (2000) konkluderer i sin undersøkelse av pasienter som har smerter i korsryggen, med at EQ-5D og HUI3 er sensitive nok til å skille mellom de som fikk en helseforbedring og de som ikke fikk det, selv om dette skillet ikke var like klart som ved ”spesialist” instrumentene. Dette styrker antagelsen om at MAU-instrumentene brukt i denne analysen er sensitive nok til å måle livskvaliteten til deltakerne i Bergensprosjektet. Dette kan også sjekkes ved å se nærmere på endringer innefor de forskjellige dimensjonene. Dette vil imidlertid være en omfattende oppgave og blir derfor ikke gjort her.

3 Bakgrunn og metode

Hvilken type utfallsmål en bruker for å måle endring i helse, avhenger som vi har sett mye av sykdommen og tiltaket en ser på. Noen typer helseforbedringer kan måles direkte, som det å bli kvitt forkjølelsen, eller at beinbruddet har grodd. For andre typer formål er det grunn til knytte helseforbedringen til nytte eller individuell velferd, for eksempel ved bruk av MAU-instrumenter. En ønsker ikke da å bare måle helseforbedringen i seg selv, men en ønsker også å måle hva helseforbedringen genererer av forbedret livskvalitet. Som vi har sett så er en grunn til at en ønsker å bruke livskvalitet som resultatmål, at behandlingstiltak ofte har flere formål og gir flere forskjellige former medisinske gevinster. En annen grunn kan være at behandlingen ofte ikke påvirker selve sykdommen eller sykdomsutviklingen nevneverdig

Mye taler for at bruk av livskvalitet som resultatmål på pasienter med muskel- og skjelletlidelser er relevant, siden dette er en sammensatt lidelse hvor en behandling sannsynligvis vil gi gevinster langs flere ulike dimensjoner (se Garrat et al (2002), Feeney et al. (1995) for diskusjon rundt dette).

Helserelatert livskvalitet blir primært sett på som særlig nyttig som et supplement til objektive kliniske eller biologiske mål i kliniske undersøkelser, der en sammenligner en behandlingsgruppe med en liknende gruppe, eller en gruppe som ikke mottar noen form for behandling i det hele tatt (Carr og Higginson (2001), Cox et al. (1992)). I denne sammenheng skal vi se litt nærmere på randomiserte eksperiment, og hvorfor denne sees på som ”gullstandard” (NOU 1997:27) innefor effektevaluering av helsetiltak.

Bergensprosjektet: ”Tilbake på jobb” sin målsetning, fremgangsmåte og resultater, samt en beskrivelse av muskel-/skjelletlidelser vil kort bli fremstilt her. Metoden brukt for å implementere de kliniske dataene fra Bergensprosjektet på MAU-instrumentene blir beskrevet inngående for MAU-instrumentet EQ-5D. Problemer omkring dette arbeidet og metoden vil følgelig diskuteres før resultatene presenteres i kapittel 4.

3.1 Bakgrunn

3.1.1 Muskel-/skjelettlidelser

Muskel-/skjelettlidelser eller belastningslidelser²⁸, er en fellesbetegnelse for smerter i rygg og/eller nakke, og/eller skuldre, og/eller ubehag i muskler, sener og ledd som varer sammenhengende i mer enn noen dager²⁹. Hvordan en person rammes varierer i forskjellig grad. En skade eller slitasje i rygg eller nakke, kan variere mellom en skiveprolaps til ømme eller stive muskler. Andre lidelser som går mer på generell skade/slitasje i muskulatur som fibromyalgi, har symptomer som i varierende grad kan føre til redusert styrke og utholdenhet, samt redusert bevegelighet, prikking i fingrene, nummenhet, ømhet, hovne senefester og ledd. Felles for alle er at lidelsene i sin tur kan føre til redusert funksjonsevne, som kan bli så plagsomt at en person kanskje ikke kan utføre sitt arbeid lenger. Noe som kan resultere i en sykemelding, og eventuelt senere uføretrygd. Årsakene til belastningslidelser er ofte sammensatte, og kan være av en fysisk og psykisk art. Ensformige bevegelser utført i et høyt tempo, fastlåste arbeidstillinger, tunge løft, stress og vantrivsel kan ligge bak, og kan være bundet opp til arbeidet vårt eller fritidssystemer. Muskel-/skjelettlidelser henger ofte sammen med andre subjektive symptomer som trøtthet, nervøsitet, depresjon, angst og irritable innvoller og blærer som er med på å forverre, samt komplisere sykdomsbildet.

Risikogrupper

Tilstedeværelsen av muskulære smerter uten objektive somatiske funn er ganske høy i en ellers så sunn befolkning. Så mye som 50% av kvinner i arbeid, og tilnærmet 40% av mennene, rapporterer at de har opplevd muskulære smerter i en eller annen form de siste 30 dagene. Andelen med alvorlige smerter er mindre, hvor ca 8-12% av kvinnene, og ca 3% av mennene rapporterer at de er seriøst plaget (Håland Haldorsen og Ursin (1998)). Så mye som 36% av de som mottar sykepenger i Norge har oppgitt muskulære plager/ belastningslidelser som sin hovedplage, som ofte er generalisert. Ca 45% av sykemeldingene som varer mer enn i 8 uker kan tilbakeskrives til denne lidelsen (Grasdahl (2001a)). Muskel- og skjelettlidelser er den hyppigste forekommende diagnosegruppen blant nye uførepensjonister, hvor over 35% har denne lidelsen (NOU 2000:27). Belastningslidelser koster ifølge NAVFs sin utredning samfunnet 20 mrd. kr per år (Håland Haldorsen og Ursin (1998)). I følge Statistisk sentralbyrå

²⁸ Begrepene ”belastningslidelser” og ”muskel-/skjelettlidelser” vil bli brukt synonymt.

²⁹ Kilde: <http://www.helsenytt.no/artikler/belastningslidelser.htm>

er det i alle aldersgrupper flest kvinner som rammes av muskel-/skjelettlidelser. Kvinner får ofte kroniske smertesyndrom som fibromyalgi, mens menn rammes ofte av kronisk vonde rygger. Tabell 3.1 viser prosentandelen av befolkningen 16 år og over med smerter i minst en arm, skuldre, nakke, rygg, knær eller føtter i en 14-dagsperiode, etter alder og kjønn i prosent. En ser her at det er kvinner som er mest plaget av smerter i alle aldersgrupper, og at andelen personer med smerter øker med alderen for både kvinner og menn. Det er og et skille for begge grupper rundt 45 år, hvor prosentandelen personer med smerter øker drastisk.

	Alle	16- 24 år	25- 44 år	45- 66 år	67- 79 år	80 år -
Menn	20	13	17	25	21	31
Kvinner	29	19	24	37	30	46

Tabell 3.1 Prosentandel av befolkning med smerter i minst en arm, skuldre, nakke rygg, knær eller føtter i en 14-dagsperiode. (Kilde, se fotnote³⁰)

3.1.2 Behandling av muskel-/skjelettlidelser

Det finnes ingen standardisert måte å diagnostisere og behandle pasienter med slike lidelser i dag (Håland Haldorsen og Ursin (1998)). Behandlingsformen inkluderer vanligvis en oppfølging hos primærlege, kombinert med behandling hos fysioterapeut og eventuelt kiropraktor. Mangel på tid og kunnskaper, særlig i primærhelsetjenesten, fører ofte til mangel på tid til å diagnostisere pasientene riktig. Det finnes heller ikke gode nok behandlingstilbud for den enkelte pasient. En enkel sykehistorie med smerter i korsryggen kan føre til 20-30 forskjellige diagnoser fra allmennpraktikerne og spesialistene (Håland Haldorsen et al. (1998)). Håland Haldorsen og Ursin (1998) påpeker at resultatet av dette er at pasientene gjerne får et tilfeldig behandlingsopplegg som ikke er vitenskapelig dokumentert, og som kan være tilfredstillende for en pasient, men uheldig for en annen. De påpeker videre at det er et problem er at det ikke finnes noen felles kriterier for legene å følge, for å avgjøre om en person trenger sykemelding eller ikke. Legene fordeler seg 50-50 når det gjelder å skrive ut sykemelding for en person med muskel-/skjelettlidelser. Det medfører at personer med nytte av en sykemelding og behandling kan bli oversett slik at problemene kan utvikle seg fra galt til verre. Kort oppsummert så er muskel-/skjelettlidelser kilden til den største sykefraværsdelen her til lands. Men i tillegg er den og den lidelsen som er svakest dekket når det gjelder behandling og forebygging.

³⁰ Kilde: <http://www.ssb.no/emner/03/00/helseund/tab-2001-04-03-06.htm>

Et offentlig utvalg (NOU 2000: 27) for å vurdere spørsmål knyttet til sykefravær og ny tilgangen til uførepensjon, konkluderte med at mangel på kunnskap ved muskel- og skjelettsykdommer ikke uten videre kan erstattes med et bredspektret tilbud. Muskel- og skjelettplager er mange og sammensatte. En differensiert tverrfaglig behandlingsmodell vil være samfunnsøkonomisk lønnsomt i form av redusert sykefravær med pasienter med dårlig og middels god prognose. De påpekte at for å lykkes er det nødvendig å tilpasse tiltakene til situasjonen til den enkeltes forutsetninger, og understrekte at det er viktig å utvikle retningslinjer og rutiner for samarbeid med den kommunale helse- og sosialtjenesten, slik at personer med sammensatte behov gies et mer helhetlig og sammenhengende tilbud.

3.2 Beskrivelse av Bergensprosjektet: ”Tilbake på jobb”

I 1993 ble det etablert et prosjekt i Bergen rettet mot belastningslidelser hos yrkesaktive. Prosjektet skulle prøve ut et nytt behandlingstilbud (Åremodellen) utviklet i Sverige, som ble videreutviklet til den medisinske modell for ”Bergensprosjektet: Tilbake på jobb”. Helseinvest Bergen A/S gjennomførte behandlingsopplegget. Målsetningen til prosjektet var å få langtidssykemeldte og personer med hyppig sykefravær med belastningslidelser og kroniske muskelsmerter tilbake i arbeidslivet, etter å ha gjennomgått en strukturert medisinsk utredning, behandling, trening, samt oppfølging etter Åremodellen. Forsøket hadde som mål å søke å dokumentere en samfunnsøkonomisk gevinst av prosjektets intervensjon, sammenlignet med det tilbudet som ellers ble gitt til denne gruppen. Bak disse målene hadde en forventning om at pasientene ville kunne bli friskmeldte etter å ha deltatt i prosjektets alternative behandlingsopplegg, dersom det kunne påvises/dokumenteres en objektiv bedring av pasientenes funksjonsnivå og en reduksjon av deres subjektive symptomer.

Prosjektet ble gjennomført som en randomisert kontrollert studie. I tillegg ble det opprettet en oppfølging av hele pasientgruppen som ble invitert til å være med i studiet. Personer med langtidssykemeldinger på 8 uker eller mer, som tilfredsstilte inklusjonskriteriene om å bo i Bergen³¹ og ha fast jobb, ble invitert til å delta i prosjektet. De som deltok ble undersøkt (pretest) ved Fysioterapihøyskolen i Bergen, og deretter randomisert til enten kontrollgruppen eller til behandlingsgruppen ved Helseinvest rehabiliteringsklinikk. Alle som deltok ble invitert til en ny identisk oppfølgingsundersøkelse (posttest) ett år senere. Deltakerne gikk

³¹ Grunnet rekrutteringssvikt ble prosjektområdet senere utvidet til å omfatte omegnskommunene.

igjennom en grundig tverrfaglig utredning ved pre- og posttest. De ble undersøkt med standardiserte fysikalske undersøkelser, og dessuten bedt om å fylle ut en samling ulike spørreskjemaer (Fellespakken) som omfattet psykososiale forhold, smerteopplevelse, funksjonsevne og livskvalitetsspørsmål. Det er disse dataene som ligger til grunn for analysen her. For mer detaljerte opplysninger omkring designet og utførelsen av ”Bergensprosjektet: Tilbake på jobb”, henvises det til Kronholm (1998) og Bomann-Larsen (1998).

Behandlingsgruppen

Alle i behandlingsgruppen fikk hvert sitt individuelle tilpassede program syv timer, fem dager i uken, i en måned. Programmet besto av kognitiv intervensjon, fysikalsk opptrening, kroppsbevissthetstrening og undervisning for å motivere til å endre livsvaner for egen helse. Målsetningen var å bevisstgjøre og dernest ansvarliggjøre pasienten hva angår egen helse og livssituasjon, avdramatisere smerten, og lære pasientene å mestre hverdagen på en bedre måte både psykisk og fysisk. Etter at behandlingen var ferdig ble pasientene sendte videre til primærlegen med en kopi av sin individuelle ”Rehabiliteringsplan”, med råd om videre behandling. I tillegg ble de regelmessig fulgt opp i ett år av behandlingsteamet³² ved klinikken. Målet ved dette behandlingsopplegget var å se om denne gruppen kom seg fortere og i større grad tilbake til jobb sammenlignet med kontrollgruppen. Det totale antall pasienter som besvarte pretest skjemaet var 350 i Helseinvest gruppen. Av disse var det 327 (93%) som møtte opp og besvarte posttest skjemaet.

Kontrollgruppen

Deltakerne i kontrollgruppen ble etter inntaket henvist tilbake til primærlegen, og behandlings tilbudet der som gies til sykemeldte med muskel- og skjelettplager. Dette inkluderer vanligvis oppfølging hos primærlege kombinert med behandling hos fysioterapeut, eventuelt kiropraktor som beskrevet i kapittel 3.1.2. Det totale antall pasienter som besvarte pretestskjemaet var 173 i kontrollgruppen. Kun 106 (62%) av disse møtte opp og svarte på posttestskjemaet. Dette utgjør et stort frafall for denne gruppen i forhold til behandlingsgruppen. Med frafall menes det at deltakerne ikke har fylt ut Fellespakken (kan ha fullført behandling) ved posttest, slik at en mangler kliniske oppfølgingsdata for disse individene.

³² Behandlingsteamet besto av en lege (nevrolog), en psykolog, fysioterapeut og sykepleiere.

Resultater fra tidligere evalueringer av Bergens prosjektet

Det ble ikke funnet noen statistiske positive effekter i retning av å komme raskere tilbake til jobb, eller i større utstrekning for behandlingsgruppen i forhold til kontrollgruppen alle diagnoser sett under ett, etter 1 år (Håland Haldorsen og Ursin (1998)), eller etter 2 år (Bratberg et al. (2002)). Det ble heller ikke funnet noen signifikante forskjeller mellom diagnosegruppene, annet enn for de med diffuse muskelsmerter som syntes å ha en svak behandlingseffekt i form av økt grad av friskmelding. Friskmeldingssannsynligheten varierte noe med bakgrunnsdata som alder, kjønn, diagnose osv. De "vanskelige" tilfellene som gift, kvinne, lav utdanning og diffuse muskelsmerter ble svakt påvist å ha den beste effekten av behandlingen (Bratberg et al. (1998.))

Resultatene ble sett på som bemerkelsesverdige med tanke på pasientenes egen evaluering i Fellespakken. Så å si alle viktige behandlingsmål registrert gjennom egnevalueringen var statistisk signifikant oppfylt i forhold til kontrollgruppen (Håland Haldorsen et al. (1998)). Behandlingsgruppen hadde i oppfølgingen mindre bruk av fysikalske tjenester, bedret arbeidsforhold, bedre trivsel, bedre subjektiv helse (pasientenes egenvurdering av helsen), bedre søvn og pasientene var fornøyd med behandlingen (Håland Haldorsen og Ursin (1998)). Fysioterapeutiske pre- og posttester ga ingen signifikante forskjeller mellom gruppene med hensyn på kondisjon, oksygenopptak, ventralfluksjon (fingertupp gulv avstand) og arbeidsevne. Men behandlingsgruppen hadde signifikant bedre resultat med hensyn aktive bevegelsesutslag, avspenningsevne, ledighet og smerte (Håland Haldorsen et al. (1998)). Psykologisk og medisinsk-fysiologisk evaluering avhengig av posttestresultatene, korrigerende aldri for frafalsseleksjon da de mente at dette ikke påvirket resultatene (Haldorsen og Ursin (1998)). Strand (1998) konkluderte med at de små forskjellene kunne tyde på at kontrollgruppen også hadde fått et godt behandlingstilbud, og at behandlingsopplegget mest sannsynlig virket forskjellig i forskjellige undergrupper. Det ble videre konkludert med at det ikke var nødvendig å utføre noen fullstendig kostnadsnytteanalyse av prosjektet, siden den gjennomsnittlige gevinsten (få folk tilbake på arbeid) av prosjektet ikke var signifikant forskjellig fra null, og at resultatet derfor måtte bli negativt (Strand (1998), Bratberg et al. (1998)).

3.3 Hvordan måle effekten av et tiltak?

Ved effektevaluering av helsetiltak er det ikke tilstrekkelig å bare observere de som utsettes for et eller annet medisinsk tiltak, kommer godt ut av det på et eller annet suksessmål (Bratberg et al. (2001)). Det må også kunne slås fast at utfallet er forskjellig fra hva det ville vært uten tiltaket. En trenger svar på det kontrafaktiske spørsmålet: "Hva ville utfallet for en person som gjennomgikk tiltaket vært, dersom denne personen hadde vært foruten dette tiltaket?". Dette er et grunnleggende problem ved identifisering av kausalsammenhenger i studier av samfunnsprosesser.

Evaluering av helsetiltak kan i prinsippet ta utgangspunkt i flere typer suksesskriterier. En mulighet er å vurdere om det er målbare forskjeller i helsetilstanden mellom behandlingsgruppen og en sammenligningsgruppe en tid etter gjennomført behandling. Grovt sett går hovedskillet blant metodene for å estimere behandlingseffekter av tiltak mellom de eksperimentelle og ikke eksperimentelle evalueringemetoder.

3.3.1 Eksperimentelle og ikke eksperimentelle evalueringemetoder

Eksperimentell evalueringemetode

Den eksperimentelle metoden regnes blant enkelte å være den beste metoden/idealsituasjonen for å kartlegge om en kausalrelasjon eksisterer mellom behandling og utfall (Sibbald og Roland (1998), NOU 1997:27). Når en ønsker å studere effekten av en eller annen behandling, vil en gjerne gjennomføre dette som et randomisert eksperiment. Først bestemmes målgruppen en tror vil ha effekt av behandlingen. Deretter randomiseres denne målgruppen (for eksempel ved loddtrekning) ut i en behandlingsgruppe og en kontrollgruppe. Behandlingsgruppen mottar behandlingen som en ønsker å måle effekten av, mens kontrollgruppen mottar ingen spesifikk behandling, eller den behandlingen som behandlingsgruppen ville mottatt uten det spesielle tiltaket som skal evalueres. Det siste var tilfellet for kontrollgruppen i Bergensprosjektet. Randomiseringen fører til at det er helt tilfeldig hvem som ender i de to gruppene. Dette sørger for (i teorien) at de observerbare og ikke-observerbare karakteristika i behandlingsgruppen og kontrollgruppen i gjennomsnitt er like. Gjennomsnittet for utfallsvariabelen i kontrollgruppen kan derfor tjene som et forventningsrett estimat for gjennomsnittet for utfallsvariabelen i behandlingsgruppen, hvis denne ikke hadde mottatt behandling i tiltaket som blir evaluert. Utfallet til kontrollgruppen

reflekterer således det kontrafaktiske utfallet, og behandlingseffekten er gitt ved forskjellen i gjennomsnittsutfall mellom de to gruppene.

Likning (3.1) viser bruken av gjennomsnittsutfall (U_i , $i=B,K$) til å estimere behandlingseffekten av et tiltak mellom en behandlingsgruppe (U_B) og en kontrollgruppe (U_K).

$$U_B - U_K = \text{Behandlingseffekt} + (E_B - E_K) \quad (3.1)$$

Differansen i gjennomsnittlig utfall mellom behandling (U_B) og kontrollgruppen (U_K) gir et forventningsrett estimat (eksperimentelt estimat) av gjennomsnittlig behandlingseffekt, hvis egenskapene (E_i , $i=B,K$) til de to gruppene er like. Det medfører at uttrykket i parenteser i likning (3.1) er lik null. Hvis egenskapene (E) til de to gruppene er ulike ($E_B \neq E_K$), vil ikke uttrykket i parenteser være lik null. Det er kun når denne er null at sammenligningen av gjennomsnittsutfallene vil gi et pålitelig estimat av behandlingseffekten.

Ikke eksperimentell evalueringsmetode

For å vurdere om en behandling har effekt, må man kunne sammenligne de som har fått behandlingen, med noen som ikke har fått den. I større sosiale eksperimenter, er det gjerne ikke mulig å bruke en kontrollgruppe (eksperimentelle data ikke tilgjengelig). Da brukes det ofte en sammenligningsgruppe som er mest mulig lik behandlingsgruppen, bortsett fra at denne ikke får behandling. En ser således på utfallet i sammenligningsgruppen som et bilde på hva utfallet ville ha vært i den kontrafaktiske situasjonen for de som fikk behandling. En stor utfordringen her, er å finne en slik gruppe. Spørsmålet er om en kan få et pålitelig estimat av behandlingseffekten ved en direkte sammenligning av gjennomsnittsutfall, på samme måte som ved den eksperimentelle metoden. Generelt er svaret nei. Det er rimelig å anta at gruppens utfall avhenger av en rekke egenskaper som kan være vanskelig å observere som motivasjon, intelligens, helse osv. (E i likning (3.1)). Ulempen med en sammenligningsgruppe er at man generelt ikke får et pålitelig estimat av behandlingseffekten ved en direkte sammenligning av gjennomsnittsutfallene mellom behandlingsgruppen og sammenligningsgruppen, fordi disse ikke kan utelukke muligheten for at assosiasjonen er skapt av en tredje faktor som er linket til både intervensjonen og utfallet (Heckman og Hotz (1989)). Dvs. parenteser i likning (3.1) er ikke null. Bare når parenteser er lik null, vil sammenligning av gjennomsnittsutfall gi et pålitelig estimat for behandlingseffekten. Siden en

ikke kan være sikker på om de observerbare og ikke-observerbare karakteristika mellom gruppene i gjennomsnitt er like, brukes regresjonsanalyser for å korrigere for mulige gjenværende forskjeller mellom gruppene.

Grad av kontroll?

Det må presiseres at begrepet "randomisering" ikke er synonymt med "kontroll". Kontroll innebærer at en er i stand til å nøytralisere alle andre faktorer som kan være med på å påvirke utfallet til eksperimentet, enn den som er gjenstand for studium. Det er mange ting som kan være med på å påvirke resultatene, og en vet ikke alt en person blir eksponert for. Dette innebærer at en har mindre grad av kontroll enn ved naturvitenskapelige eksperiment. Innsamling av data på observerbare kjennetegn er med på å gi oss kontroll med omgivelsene rundt personen i forsøket. Når eksperimenter omfatter mennesker, er disse i alminnelighet klar over at de deltar i et eksperiment. Bevisstheten om dette kan i seg selv ofte være med på å påvirke utfallet, ved at pasienter som er i kontrollgruppen oppsøker for eksempel "alternativ" ekstra behandling. Det er dette en prøver å korrigere for i eksperimenter hvor en i tillegg til en behandlingsgruppe trekker ut en kontrollgruppe som får et placebo, et virkningløst stoff.

Det er vanskelig å etablere sikker kunnskap om tiltakseffekter selv i randomiserte forsøk med rimelig stor grad av kontroll over utenforliggende faktorer som også kan tenkes å påvirke utfallet av tiltaket. "Mangel på kontroll" kan en ha i både de eksperimentelle og ikke eksperimentelle metodene. Jo mer kompleks "behandlingen" som evalueres er, desto vanskeligere er det å kontrollere for utenforliggende faktorer. Forskjellen mellom de to metodene, er at ved en randomiseringen prøver å oppnå at behandling og kontrollgruppen er lik for det de blir eksponert for i omgivelsene, og at observerbare og ikke-observerbare egenskaper dermed fortsatt er lik mellom gruppene etter eksponeringen. I alle tilfeller har det vist seg at uansett hvilken type metode en bruker, må en være forsiktig med tolkningen av resultater, fordi det i denne typen evalueringer er vanskelig å etablere uomtvistelige kunnskaper om tiltakseffekt.

3.3.2 Frafallsseleksjon

Det må påpekes at uttrykket i parentes i likning (3.1) ikke nødvendigvis trenger å være lik null selv i randomiserte eksperimenter. Flere ting kan påvirke resultatene i form av en eller

annen type seleksjon³³. Det finnes flere typer metoder for å kontrollere og korrigere for seleksjon, men en hovedklassifisering følger av hva en har tilgjengelig av data. Har en bare data etter behandlingen, eller har en data fra både før og etter behandlingen? Her skal vi konsentrere oss om den siste, og se nærmere på frafallsseleksjon og hva det innebærer.

Et vanlig problem ved sosiale eksperimenter som varer over tid, er at noen deltakere faller fra underveis. Spesielt har dette en tendens til å skje oftere blant deltakerne i kontrollgruppen enn blant deltakerne i behandlingsgruppen (se ref. i Grasdahl (2001a)). Frafall av observasjoner skaper et "missing-data" problem. I tillegg til dette kan en ha "missing-data" på grunn av ufullstendige svar. I Bergensprosjektet mangler en oppfølgingsdata og informasjon på helse for nærmere 1/3 av deltakerne, grunnet manglende oppmøte til oppfølgingsundersøkelsen. Frafallet var her størst i kontrollgruppen. Et slik frafall fører til at en bare har tilgjengelig informasjon på utfallet for de som fullførte forsøket (fylte ut Fellespakken ved posttest). Hvis frafallet er ikke tilfeldig og korrelert med utfallsvariabelen, kan dette påvirke den målte behandlingseffekten, siden karakteristika ved de to gruppene en til slutt sammenligner ikke nødvendigvis er like lenger.

Hvis en tenker seg et kontrollert randomisert eksperiment som begynner ved tidspunkt T_0 , skal randomiseringen føre (i teorien) til at egenskapene ($E_i, i=B,K$) til begge gruppene er helt like (ved $T_0: E_B = E_K$). Hvis frafallet fra tidspunkt T_0 til T_1 er av overnevnt type, vil en risikere at uttrykket i parenteser i likning (3.1) ikke lenger er lik null ved tidspunkt T_1 (ved $T_1: E_B \neq E_K$). Dvs. det kan det være forskjeller i observerbare og ikke-observerbare karakteristika ved de gjenværende deltakerne i gruppene. Dette gjør det nødvendig å undersøke om den observerte forskjellen i utfall mellom behandlings- og kontrollgruppen ved oppfølging representerer en sann behandlingseffekt, eller om den er generert av frafallsseleksjon. En

³³ Til tross for eksperiment metodens intuitive appell har en det at jo mer omfattende "behandling" en studerer, jo vanskeligere er det å kontrollere for andre årsaksfaktorer. Medisinsk og yrkesmessig atferd er eksempler på dette. Heckman og Smith (1995), Sibbald og Roland (1998) og en del andre nevner noen mulige problemer knyttet til sosiale eksperimenter, hvor de understreker at to antagelser må holde for at utfallet til en eksperimentell kontrollgruppe skal kunne korrespondere med utfallet som deltakerne ville ha hatt hvis de ikke hadde deltatt i programmet. 1) Randomiseringsskjevheter: Randomiseringen må ikke endre seleksjonsprosessen inn i programmet, slik at de som deltar i eksperimentet er av en annen type som normalt ville ha fått behandlingen hvis dette ikke hadde vært et eksperiment. 2) Substitusjonsskjevheter: Det kan tenkes at noen av de som er med i kontrollgruppen oppsøker nære substitutter for behandlingen andre steder. Dermed for en at utfallet til kontrollgruppen ikke lenger korresponderer med utfallet som deltakerne i behandlingsgruppen ville hatt hvis de ikke hadde deltatt i programmet. Konsekvensen er at forskjellen i gjennomsnittsutfall mellom gruppene ikke lenger gir et korrekt estimat av behandlingseffekten.

vanlig metode for å undersøke dette er å utføre en analyse med en regresjonsmodell der en prøver å korrigere mulige forskjeller mellom gruppene som følge av frafallet.

3.4 Metode

Grunnlaget for analysen er spørreskjemaet (Fellespakken) som deltakerne i Bergensprosjektet fylte ut ved pre- og posttest. Ved å finne spørsmål som deltakerne svarte på i Fellespakken som er like eller likner de spørsmålene som er i MAU-instrumentene, kan en la svarene i Fellespakken indikere hva en tror deltakerne ville ha svart i instrumentene. Dermed er det mulig å kalkulere gjennomsnittlig helserelatert livskvalitet for deltakerne i behandlingsgruppen og kontrollgruppen gjennom de fire MAU-instrumentene, som siden kan benyttes til å regne ut gjennomsnittsindividets kvalitetsjusterte leveår, QALY.

3.4.1 Før data inn i MAU- instrumentene

Før det er mulig å implementere dataene i MAU-instrumentene, må dataene bearbeides litt for å passe inn. Det vil her med et detaljert eksempel bli demonstrert hvordan dette har blitt gjort for mobilitetsdimensjonen i The Euroqol Classification System (EQ-5D). De fire andre dimensjonene i EQ-5D vil bli forklart litt lettere. Tanken bak det hele er å gi et totalbilde av fremgangsmåten for et helt instrument. Samme fremgangsmåte er brukt i alle de fire instrumentene ved overføringen av kliniske data fra Fellespakken til MAU-instrumentene. Hvordan overføringen av de kliniske dataene er blitt gjort i MAU-instrumentene QWB og HUI3 er vist i appendiks.

Et eksempel på utførelse av ”mapping” i EQ-5D

EQ-5D består av de fem ulike dimensjonene, *mobilitet*, *egenomsorg*, *vanlige aktiviteter*, *smerte/ubehag* og *angst/depresjon*, hvor hver dimensjon har tre svaralternativ/nivå under der igjen. Følgende er blitt gjort for å få dataene i Fellespakken til å passe inn i EQ-5D.

1) Mobilitet:

Tabell 3.2 viser mobilitetsdimensjonen i MAU-instrumentet EQ-5D, og dens tre nivåer som gir en verdi på hvor godt/mye en person kan bevege seg i hverdagen. Ved å finne liknende spørsmål/dimensjoner i Fellespakken som dreier seg om mobilitet, kan en overføre data derfra og inn i instrumentet.

Mobilitet:

1. Ingen problemer med å gå.
2. Noen problemer med å gå.
3. Bundet til sengen.

Tabell 3.2 Mobilitetsdimensjonen i EQ-5D

I Fellespakken har deltakerne i spørsmål 29, tabell 5, krysset av for ”*hvor mye smerte er det gjennomsnittlig forbundet med hver dag for forskjellige beskrevne aktiviteter*”. Tabell 3.3 viser et utdrag fra den opprinnelige tabellen. Der er det tatt med aktiviteter som er valgt til å inngå som et gjennomsnitt for å bestemme nivået i mobilitetsdimensjonen, samt et eksempel på et mulig svar. For at bevegelighets- /mobilitetsdimensjonene mellom MAU-instrumentene skulle ha mest mulig samsvar, og ikke ble bestemt av ett alternativ i et instrument og så av flere i et annet, ble det valgt å inkludere de samme alternativene i EQ-5D som ble brukt i mobilitetsdimensjonen i 15D og HUI3. Hovedgrunnen til dette var at denne dimensjonen var veldig lik mellom MAU-instrumentene, i motsetning til andre dimensjoner³⁴.

	Ingen Smerte	Noe Smerte	Sterk Smerte	Nærmest Utålelig Smerte	Utålelig Smerte, Uforenelig med aktivitet
1. Gå		X			
2. Sitte			X		
3. Bøye seg		X			
7. Løpe/jogge				X	
8. Gå i trapp		X			
17. Sportsakt			X		

Tabell 3.3 Utdrag fra tabell 5 i Fellespakken

Et problem her er at det ikke er et likt antall svaralternativer/nivå mellom EQ-5D sin mobilitetsdimensjon og tabell 5 i Fellespakken. Valget her var å enten stryke noen alternativer i tabell 5 fra Fellespakken, eller slå sammen noen svaralternativer/nivå før de kliniske dataene ble overført til instrumentene. Den siste strategien ble valgt, og i samarbeid med andre medstudenter kom vi frem til sammenslåingen som vist i tabell 3.4.

³⁴ Se også appendiks for utførelse av ”mapping” i QWB og HUI3.

	Nivå 1	Nivå 2		Nivå 3	
Nivå i EQ-5D	1. Ingen problemer med å gå	2. Noen problemer med å gå		3. Bundet til sengen	
Alternativ fellespakken	1. Ingen smerte	2. Noe Smerte	3. Sterk smerte	4. Nærmest utålelig smerte	5. Utålelig smerte, uforenelig med aktivitet

Tabell 3.4 Nivåsammenligning mellom EQ-5D og Fellespakken i mobilitetsdimensjonen

Gjennomsnittet av hva en person har svart i tabell 3.3, avgjør hvilket nivå i mobilitetsdimensjonen personen vil få. I tabell 3.3 har en som følger at ingen smerte = 1, noe smerte = 2, etc. I eksempelet her er gjennomsnittet (= gj.sn.) i tabell 3.3 (= tab.3.3) lik: $2 + 3 + 2 + 4 + 2 + 3 = 16/6 = 2,67$. Dette gir alternativ 2 på mobilitetsskalaen i EQ-5D i henhold til tabell 3.5 som viser nivåintervallene, siden verdien 2,67 ligger i intervallet 1,51- 3,5. Ideen her er å runde av til nærmeste heltall.

Nivåene i EQ-5D	Intervallene i EQ-5D
<u>Mobilitet:</u>	
1. Ingen problem med å gå.	Hvis gj.sn. i tab.3.3 er mellom 1 og 1,5.
2. Noen problem med å gå.	Hvis gj.sn. i tab.3.3 er mellom 1,51 og 3,5.
3. Bundet til sengen.	Hvis gj.sn. i tab.3.3 er mellom 3,51 og 5.

Tabell 3.5 Nivåintervallene i mobilitetsdimensjonen

2) Egenomsorg:

Tabell 3.6 viser egenomsorgsdimensjonen og dens tre nivåer, som gir en verdi på hvor godt en person klarer å ta vare på seg selv i form av egent stell, av og påkledning. Tabell 3.7 viser alternativet fra spørsmål 30, tabell 5 i Fellespakken; *'min evne til å ta vare på meg selv'*. Tabell 3.8 viser nivåsammenligningen mellom disse to. I eksempel her vil en som har svart *"tilfredsstillende"* i Fellespakken, få tilskrevet nivå 2 i egenomsorgsdimensjonen.

<u>Egenomsorg:</u>
1. Ingen problemer med egenomsorg.
2. Noen problemer med å vaske og kle på seg selv.
3. Klarer ikke å vaske eller kle på seg selv.

Tabell 3.6 Egenomsorgsdimensjonen i EQ-5D

	Svært utilfredsstillende	Utilfredsstillende	Verken god eller dårlig	Tilfredsstillende	Svært tilfredsstillende
Min evne til å klare meg selv er				X	

Tabell 3.7 Utdrag fra tabell 5 i Fellespakken

	Nivå 1	Nivå 2		Nivå 3	
Nivå i EQ-5D	1. Ingen problemer...	2. Noen problemer med å...		3. Klarer ikke å...	
Alternativ fellespakken	5. Svært tilfredsstillende	4. Tilfredsstillende	3. Verken god eller dårlig	2. Utilfredsstillende	1. Svært utilfredsstillende

Tabell 3.8 Nivåsammenslåing mellom EQ-5D og Fellespakken i egenomsorgsdimensjonen

3) Vanlige aktiviteter:

Tabell 3.9 viser vanlige aktiviteterdimensjonen i EQ-5D. Dimensjonen tar for seg hvor godt en person kan utføre vanlige dagligdagse aktiviteter som arbeide, studere, husarbeid, fritidsaktiviteter osv. Fra spørsmål 29, tabell 5 i Fellespakken, er det brukt et utvalg av spørsmål som omhandler *’hvor mye smerte er det gjennomsnittlig forbundet med hver dag for de forskjellige beskrevne aktiviteter’*. Her er det tatt utgangspunkt i aktiviteter som er nevnt i EQ-5D. Disse er fremstilt i tabell 3.10, og snittet av disse bestemmer nivået for vanlige aktiviteterdimensjonen. Nivåsammenslåingen og intervallinndelingen er som for tabell 3.4 og 3.5 i mobilitetsdimensjonen, og vises derfor ikke her. I eksempelet her vil personen få tildelt nivå 2 akkurat som i mobilitetsdimensjonen.

Vanlige aktiviteter:

1. Ingen problemer med å utføre vanlige aktiviteter.
2. Noen problemer med å utføre vanlige aktiviteter.
3. Klarer ikke å utføre vanlige aktiviteter.

Tabell 3.9 Vanlige aktiviteterdimensjonen i EQ5D

	Ingen Smerte	Noe Smerte	Sterk Smerte	Nærmest Utålelig Smerte	Utålelig Smerte, Uforenelig med aktivitet
11. Kjøre bil		X			
13. Lese			X		
14. Se på TV		X			
15. Husarb.				X	
16. Hagestell		X			
18.Arbeid			X		

Tabell 3.10 Utdrag fra tabell 5 i Fellespakken

4) Smerte/ubehag:

Denne dimensjonen som er vist i tabell 3.11, tar for seg hvor mye smerte/ubehag en person er plaget med til vanlig. Fra Fellespakken er det valgt å bruke snittet fra et utvalg av forskjellige smertetyper for å bestemme nivået i denne dimensjonen. Disse er hentet fra tabellen "helseproblem siste 30 døgn". Smertetyperne som er valgt er de man som oftest forbinder med belastningslidelser, og er vist i tabell 3.12. Tabell 3.13 viser nivå sammenslåingen for disse to og nivåintervallene. I eksempelet her vil personen få tilskrevet nivå 2. Dette som følge av at $1 + 2 + 2 + 0 + 1 + 1 = 7/6 = 1,17$ og ligger i intervallet 0,51-2,50.

<u>Smerte/ubehag:</u>
1. Ingen smerte eller ubehag.
2. Moderate smerter eller ubehag.
3. Ekstreme smerter eller ubehag.

Tabell 3.11 Smerte-/ubehagsdimensjonen i EQ-5D

Nedenfor nevnes noen alminnelige helseplager	0. Ikke plaget. 1. Litt plaget. 2. En del plaget. 3. Alvorlig plaget.
4. Hodepine	1
5. Nakkesmerter	2
6. Smerter øverst i ryggen	2
7. Smerter i korsryggen	0
8. Smerter i armer	1
9. Smerter i skuldre	1

Tabell 3.12 Utdrag fra Helseproblem siste 30 døgn i Fellespakken

	Nivå 1	Nivå 2		Nivå 3
Nivå i EQ-5D	1. Ikke plaget...	2. Moderate smerter/...		3. Ekstreme...
Alternativ Fellespakken	0. Ikke plaget	1. Litt plaget	2. En del plaget	3. Alvorlig plaget
Intervall	0 – 0,50	0,51 – 2,51		2,51 - 3

Tabell 3.13 Nivå sammenslåing og nivåintervall i smerte-/ubehagsdimensjonen

5) Angst/depresjon:

Tabell 3.14 viser angst-/depresjonsdimensjonen og dens tre nivåer, som tar for seg hvor mye angst/depresjon en person har til daglig. Fra spørsmål 38, tabell 9 i Fellespakken, er det valgt å bruke snittet av de to spørsmålene om følelsene ”nervøsitet” og ”nedtrykthet, tungsindighet”. Disse er vist i tabell 3.15. Tabell 3.16 viser nivå sammenslåingen mellom EQ-5D og Fellespakken, samt nivåintervallene. Eksempelet her vil gi nivå 2, siden $2 + 2 = 4/2 = 2$ ligger i intervallet 1,51 – 3,50.

<u>Angst/depresjon:</u>
1. Ikke nervøs eller deprimert.
2. Moderat nervøs eller deprimert.
3. Ekstremt nervøs eller deprimert.

Tabell 3.14 Angst-/depresjonsdimensjonen i EQ-5D

	Ikke plaget	Litt plaget	Ganske mye plaget	Veldig mye plaget
4. Nervøsitet		X		
17. Nedtrykt, tungsindighet		X		

Tabell 3.15 Utdrag fra tabell 9 Fellespakken

	Nivå 1	Nivå 2		Nivå 3
Nivå EQ5D	1. Ikke nervøs...	2. Moderat nervøs...		Ekstremt...
Alternativ Fellespakken	1. Ikke plaget	2. Litt plaget	3. Ganske mye plaget	4. Veldig mye plaget
Intervall	1 – 1,50	1,51 – 3,50		3,51 - 4

Tabell 3.16 Nivå sammenslåing og nivåintervall i angst-/depresjonsdimensjonen

6) Helserelatert livskvalitet:

De fem svarene i eksempelet danner en 5-sifret kode som utgjør individets helseprofil. Denne koden: 2 2 2 2 2, betyr at individet har moderate problemer i alle helsedimensjonene. Koden

sjekkes opp mot en tabell hvor helsereelatert livskvalitet er ferdig utregnet for hver mulig kodekombinasjon, eller en kan regne det ut med likningene 2.7 – 2.11 (se Drummond et al. (1997)). I eksempelet her vil personen få regnet ut en helsereelatert livskvalitet på 0,516 på død/perfekt helse skalaen (0=død, 1= perfekt helse).

Denne ”mapping” metoden er brukt i alle fire MAU-instrumentene. Dermed har vi fått kalkulert deltakernes gjennomsnittlige helsereelaterte livskvalitet ved pre- og posttest i alle instrumentene, avhengig av hvordan de har svart på alle de nødvendige spørsmålene som må til for å få regnet ut en MAU-score, samt om de møtte opp til posttest. Disse verdiene benyttes således til å evaluere de to behandlingsformene opp mot hverandre, for å se om det ”alternative” behandlingsopplegget har en større effekt på den helsereelaterte livskvaliteten enn det ”vanlige” behandlingsopplegget for denne lidelsen/sykdomsgruppen. Underforstått betyr det at det er en behandlingseffekt på det ”alternative” behandlingsopplegget en er på jakt etter.

3.4.2 En vurdering av instrumentene

Det finnes sikkert noen som vil være uenige i og har kommentarer med hensyn til hvordan ”mappingen” er og kan utføres i de fire MAU-instrumentene. Det finnes rom for ulike valg som kan diskuteres. Det hele bunner ut i hvilke preferanser en selv har for hvordan en tolker ting opp mot hverandre, som for eksempel rimeligheten og passformen ved sammenslåingen av spørsmål/dimensjoner mellom MAU-instrumentene og Fellespakken. Dette kan bli veldig subjektivt, og derfor er nivåsamenslåingen gjort sammen med andre medstudenter for å prøve å unngå nettopp dette problemet.

Når en gjennomfører en ”mapping” er det mange spørsmål som melder seg hele veien. I denne prosessen er det to som peker seg ut: 1) Passform/samsvar: Hvor godt passer det en har funnet i Fellespakken med det i MAU-instrumentet, spør de om noenlunde det samme? 2) Sammenslåing: Er nivåsamenslåingen god/riktig? I denne sammenheng er det tatt med et lite delkapittel med en egen vurdering av MAU-instrumentene og deres ”totale passform” med Fellespakken for å belyse disse to spørsmålene.

Et annet punkt ut av mange som er med på å indikere hvor godt et instrument er til å dekke denne lidelsen, er hvor sensitivt³⁵ et instrument er i form av evnen til å skille mellom individer og grupper med forskjellige helsetilstander (diskriminering). Og i tillegg påvise forandringer hos individer eller grupper over tid (respons til forandringer i helsetilstanden) (Sintonen (1994)).

Vurdering av The Euroqol Classification system (EQ-5D)

I EQ-5D spørres det noenlunde om de samme spørsmålene som i Fellespakken, slik at det ikke er noen problem å "klart" finne igjen spørsmål i EQ-5D som likner de i Fellespakken. Dermed kan en si at EQ-5D har en god passform/samsvar med Fellespakken med hensyn til spørsmålsformuleringer. Et gjennomgående problem er at det er tre svaralternativer/nivå i hver dimensjon, mens for spørsmålene i Fellespakken er det alltid fire eller flere (som oftest fem). Dette medfører det problemet at begge endepunktene (best og verst tilstand) som for eksempel "*ingen smerter*" og "*utålelige smerter*", er like for EQ-5D og Fellespakken, mens det som er imellom må slås sammen på en eller annen måte som vist i kapittel 3.4.1. Alle nivå sammenslåingene i EQ-5D virker gode i den forstand at sammenslåingene virker logiske og klare. Unntaket er egenomsorgsdimensjonen, hvor spørsmålet er om alternativet "*tilfredsstillende*" (og "*verken god eller dårlig*") kan settes i forbindelse med nivå 2 i EQ-5D; "*noen problemer med å vaske og kle på seg selv*" (se tabell 3.8). Det er blitt eksperimentert en del med nivåinndelingen her, og funnet at det har minimalt å si for det totale resultatet hvordan en deler inn nivåene i denne dimensjonen. En annen positiv ting med dette instrumentet er at det ikke er blitt gjort noen antagelser på noen av dimensjonene, slik at de alle blir utnyttet i motsetning til for eksempel HUI3³⁶.

Vurdering av Quality Of Well-Being (QWB)

I QWB spørres det også noenlunde om de samme spørsmålene som i Fellespakken. Dermed er det heller ikke her er noen problem å finne igjen spørsmål i QWB som likner de i Fellespakken. Det konkluderes derfor også her med at det er en god passform/samsvar mellom MAU-instrumentet og Fellespakken med hensyn til spørsmålsformuleringer. Videre har en også her problemet med at det ikke er like svaralternativ/nivå mellom instrumentet og Fellespakken. Dette har blitt løst som i EQ-5D (se appendiks). Alle nivå sammenslåingene i

³⁵ Andre indikatorer på hvor godt et MAU-instrument er, er gjørlighet/mulighet og generell anvendbarhet, reliabilitet (pålitelighet) og validitet (gyldighet) (se Sintonen (1994), Hawthorne et al. (2000))

³⁶ Se appendiks og vurdering av HUI3 for nærmere belysning av dette problemet.

QWB virker gode i den forstand at disse også her virker logiske og klare. Unntaket er sosial aktivitetsdimensjonen, hvor spørsmålet er om alternativet ”*tilfredsstillende*” (og ”*verken god eller dårlig*”) kan settes i forbindelse med nivå 2 i QWB; ”*Begrensninger i noe/det meste av rolle aktiviteter*” (se appendiks tabell A6). Her trekkes den samme konklusjonen som i EQ-5D. Det er heller ikke her blitt gjort noen antagelser på dimensjonene i instrumentet.

Vurdering av Health Utilities Index Mark 3 (HUI3)

I HUI3 blir det spurt om dimensjonene, *syn, hørsel, språk, førlighet og kognisjon*, som er dimensjoner det ikke blir spurt om i Fellespakken (se appendiks). Dette har ført til at det er blitt gjort den antagelser at alle individene i forsøket har perfekt score på disse dimensjonen, som utgjør 5 av de 8 dimensjonene i HUI3. For de resterende 3 dimensjonen var det ikke noe problem å finne igjen spørsmål som liknet de i Fellespakken. Alle nivå sammenslåingene i disse virker logiske og klare. Uansett er det rimelig å si at HUI3 har en dårlig passform/samsvar med Fellespakken, siden det er blitt gjort så mange antagelser med hensyn på dimensjonene. Videre er det og rimelig å tro at HUI3 ikke klarer å være noen særlig god indikator på en pasients helsetilstand på grunn av dette. HUI3 passer nok bedre inn i et annet sykdomsbilde enn det for belastningslidelser, siden så få dimensjoner/sider ved denne lidelsen blir dekket av instrumentet. Dermed er det grunn til å trekke resultatene i fra HUI3 i tvil, selv om de dimensjonene som det er blitt gjort antagelser trolig har liten innvirkning på behandlingen (og omvendt).

Vurdering av Quality of life (15D)

I 15D blir det spurt om dimensjonene, *tale, hørsel, syn og spising*, som det ikke blir spurt om i Fellespakken. De samme antagelsene er blitt gjort på disse dimensjonene som i HUI3. Antagelsene utgjør 4 dimensjoner av en total på 15. Det kan sies å være innenfor rimelighetens grenser siden dette er så få dimensjoner i forhold til totalen, samt at en vil tro at disse har liten innvirkning på selve behandlingen. Dette sammen med det faktum at det ikke er noen problemer å finne igjen spørsmålene for de resterende dimensjonene i Fellespakken, kan det konkluderes med at det er godt samsvar mellom 15D og Fellespakken med hensyn til spørsmålsformuleringer. Alle nivå sammenslåingene i 15D virker gode i den forstand at sammenslåingene også her virker logiske og klare. Det kan dermed konkluderes med at 15D skal gi et valid resultat basert på dette.

3.4.3 Problemer med ”mapping” metoden

Ved siden av de direkte problemene ved overføringen av de kliniske dataene til MAU-instrumentene, er det en del andre problemer ved denne metoden som bør nevnes. Disse problemene har ikke noe med selve ”mapping” prosessen å gjøre, men er av det praktiske problemet at analysen blir utført retrospektivt. Dette er noe som kan ha en innvirkning på resultatene.

Det optimale i denne analysen hadde selvsagt vært at deltakerne i Bergensprosjektet selv hadde svart på MAU-instrumentene ved pre- og posttest, slik at en hadde fått ”reine/direkte” svar i fra dem. Små formuleringsforskjeller i spørsmålene og svaralternativene mellom MAU-instrumentene og Fellespakken, kan være av en slik art at noen deltakere kanskje ville ha svart annerledes i MAU-instrumentene i forhold til Fellespakken. Nivåsammenligningen som vist i kapittel 3.4.1 vil til en viss grad fange opp dette problemet. Samtidig er mange nivåsammenligninger gjerne en indikator på at MAU-instrumentet kanskje ikke er sensitivt nok til å måle små endringer i helse. Det å måtte slå sammen intervaller i Fellespakken for å få det hele til å passe inn i et instrument, kan medføre at vi mister en del informasjon på veien. Dette tilsier at en ikke kan være 100% sikker på at det som en deltaker har svart i Fellespakken, vil være det samme som han/hun ville ha svart i MAU instrumentet. Om dette vil utgjøre en positiv eller negativ påvirkning på resultatene her er vanskelig å si.

Et annet problemet er at enkelte deltakere ikke har svart på alle spørsmålene som inngår for en dimensjon i Fellespakken. Resultatet av det er at en får manglende (missing) verdier i for eksempel tabell 3.2, hvor gjennomsnittet der bestemmer nivået i mobilitetsdimensjonen i EQ-5D. I denne tabellen er det valgt å bruke gjennomsnittet for opptil to manglende svar, fordi dette øker antall observasjoner i MAU-instrumentet uten å medføre en betydelig skjevhet i gjennomsnittlig helserelatert livskvalitet. De deltakerne som har mer enn to manglende svar her, vil ikke få regnet ut en verdi på denne dimensjonen. Disse får dermed heller ikke regnet ut en MAU-score for dette instrumentet. Det kan være mange grunner til at en person ikke svarer på et spørsmål. En grunn kan være at personen rett og slett har oversett spørsmålet, eller ha unnlatt å svare fordi det var så mange spørsmål. En tredje grunn kan være at personen har fått et spørsmål av typen hvor mange barn han/hun har. Hvis denne personen ikke har noen barn, kan det hende at han/hun ikke svarer noen ting på dette spørsmålet. Liknende tilfeller kan og tenkes seg for andre typer spørsmål.

Det at det blir gjort forandringer i MAU-instrumentene og at analysen gjøres retrospektivt, gjør at vi får kvasiinstrument med påfølgende kvasimål. Det bør derfor legges for minnet at når det videre i studiet refereres til helserelatert livskvalitet generert i fra MAU-instrumentene, er det ikke de ekte instrumentene og livskvalitetsmålene vi snakker om her. Men snarere en tilnærming eller estimat av det sanne gjennomsnittet.

4 Analyse av Bergensprosjektet: ”Tilbake til jobb”.

I dette kapittel presenteres variablene og datamaterialet som ligger til grunn for den empiriske analysen ved hjelp av deskriptiv statistikk. Resultatene MAU-instrumentene kommer frem til av gjennomsnittlig helserelatert livskvalitet for behandling- og kontrollgruppen når de kliniske dataene fra Bergensprosjektet overføres til instrumentene, presenteres også her. Dette er estimat som kan brukes til å evaluere de to behandlingsformene opp mot hverandre. Analysen er basert på ”intensjon til å behandle” (”intention to treat”). Som resultatene vil vise, varierer gjennomsnittlig helserelatert livskvalitet for de to gruppene for ulike MAU-instrument. Dette gjør det nødvendig å ha en liten diskusjon rundt hva som muligens kan drive disse forskjellene mellom instrumentene.

Frafall av deltakere er et problem som ofte plager eksperimenter som varer over en lengre tidsperiode. En god del av deltakerne i Bergensprosjektet falt fra underveis i forsøket, og møtte ikke opp til posttest. Frafallet var særlig stort blant deltakerne i kontrollgruppen, med den fare at kontrollgruppen ikke lenger kan betraktes som representativ for hva behandlingsgruppen ville vist uten behandling. Dermed er det mulig at differansen i helserelatert livskvalitet mellom behandlings- og kontrollgruppen, er påvirket av seleksjonen forbundet med frafall til oppfølgingsundersøkelsen. Det medfører at det er nødvendig å utføre en analyse av datasettet ved hjelp av en regresjonsanalyse, hvor en forsøker å korrigere for frafallsseleksjon. Dette gjøres her med en Full Maximum Likelihood modell. Denne modellen vil således også bli utledet og forklart her.

Et annet tema som vil bli tatt opp her er spørsmålet om hvor valide resultatene i fra regresjonsmodellene er, og hvor mye en kan stole på estimatene når en forsøker å beregne behandlingseffekten. Forskjellige tester kan være en hjelp på veien for å avgjøre om modellspesifikasjonen er god nok slik at estimatoren(e) er forventningsrett(e).

Til slutt vil det bli sett nærmere på om tiltaket kan ha virket forskjellig i enkelte undergrupper. Dette kan belyses nærmere ved å tilføye interaksjonsledd for behandling og ulike bakgrunnskarakteristika i regresjonsmodellen.

4.1 Resultater helse relatert livskvalitet

4.1.1 Bakgrunnsdata

Bakgrunnsdataene til det fulle utvalget for behandlings- og kontrollgruppen ved pre- og posttest, er oppsummert i tabellene 4.1 og 4.2. Variablene som vises i tabellene, er de som senere i kapittel 4 vil bli brukt i analysen ved korreksjon av frafallsseleksjon. Tabell 4.1 viser at gjennomsnittlig alder for begge gruppene ved pretest er på rundt 43 år. For variabelen kjønn er kvinne brukt som base (Kvinne = 0, mann = 1). Tabellen viser at det er en overvekt av kvinner i begge gruppene, hvor andelen kvinner utgjør ca 63%. Gjennomsnittlig inntekt før inntak er på rundt 187000 kr. Deltakerne har i gjennomsnitt vært sykemeldt i litt over 3 måneder (12 uker) før de ble med i forsøket. Ca 86% av de bor i Bergen.

Variabler	Behandlingsgruppen (n=350)			Kontrollgruppen (n= 173)		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
Alder	43,496	10,681	20 / 66	43,809	10,669	20 / 67
Kjønn	0,369	0,483	0 / 1	0,369	0,484	0 / 1
Inntekt*	1,879	0,740	0,217/7,480	1,810	0,769	0,064/4,877
Mndsyk	3,397	1,961	1 / 20	3,409	2,989	1 / 33
Bergen	0,869	0,338	0 / 1	0,861	0,347	0 / 1

Tabell 4.1 Bakgrunnsdata for hele utvalget ved pretest (*Inntekt/100.000)

Variabler	Behandlingsgruppen (n=327)			Kontrollgruppen (n= 106)		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
Alder	43,537	10,550	20 / 65	43,028	10,381	21 / 64
Kjønn	0,355	0,479	0 / 1	0,377	0,487	0 / 1
Inntekt*	1,873	0,714	0,217/7,480	1,843	0,724	0,482/4,877
Mndsyk	3,412	2,008	1 / 20	3,510	2,258	1 / 15
Bergen	0,878	0,328	0 / 1	0,934	0,250	0 / 1

Tabell 4.2 Bakgrunnsdata for hele utvalget ved posttest (*Inntekt/100.000)

Fra pre- til posttest falt en del av deltakerne fra. Dette kan ha en innvirkning på resultatene i eksperimentet, og er tema som vil bli analysert og diskutert mer grundig senere i kapitlet. Men ut i fra de to tabellene kan det være mulig å få en pekepinn på hvem som ikke møtte opp til posttest. Tabellene viser at gjennomsnittlig alder ikke endres fra pre- til posttest. Andelen kvinner i behandlingsgruppen øker med ca.1%, mens andelen kvinner i kontrollgruppen reduseres med ca. 1% fra pre- til posttest. Videre ser en at gjennomsnittlig inntekt endres lite og at det samme gjelder for antall måneder syk før inntaket. Andelen som bor i Bergen øker for begge gruppene, men mest for kontrollgruppen hvor antallet går i fra 86% til 93%. En slik økning kan tolkes som et tegn på at folk bosatt utenfor Bergen kommune unnlot å møte opp til

posttest. Bakgrunnsvariablene er oppsummert som i tabell 4.1 og 4.2 for hvert av instrumentene i appendiks. Resultatene der støtter oppunder det jeg har funnet her.

4.1.2 Resultater i fra MAU-instrumentene

Deltakerne i Bergensprosjektet besvarte som forklart i kapittel 3.2 et skjema med fellesbetegnelsen "Fellespakken", som angikk psykososiale forhold, smerteopplevelse, funksjonsevne og livskvalitetsspørsmål. Tabellene 4.3 og 4.4 viser resultatene av "mappingen" mellom de fire MAU-instrumentene og Fellespakken. Tabell 4.3 viser den gjennomsnittlige livskvaliteten for deltakerne i behandlings- og kontrollgruppen ved pretest, basert på hva deltakerne svarte ved inntaket før behandling. Tabell 4.4 viser posttest resultatene basert på hva deltakerne svarte ett år senere, etter at de respektive behandlingene i de to gruppene var ferdig.

Variabler	Behandlingsgruppen			Kontrollgruppen		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
EQ-5D	0,627	0,149	.002/.919	0,582	0,207	-.056/.919
QWB	0,577	0,051	.473/.701	0,569	0,055	.473/.701
15D	0,812	0,059	.609/.949	0,806	0,059	.624/.930
HUI3	0,509	0,180	-.017/.931	0,519	0,182	.043/.863

- Antall observasjoner i behandlingsgruppen: EQ-5D = 297, QWB =284 , 15D = 212, HUI3 =298 .

- Antall observasjoner i kontrollgruppen: EQ-5D = 136, QWB =134 , 15D =94 , HUI3 =144.

Tabell 4.3 Gjennomsnittlig MAU-score for behandlings- og kontrollgruppen ved pretest

Variabler	Behandlingsgruppen			Kontrollgruppen		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
EQ-5D	0,681	0,161	-.108/.919	0,658	0,159	.137/.919
QWB	0,596	0,070	.473/.701	0,587	0,067	.473/.701
15D	0,845	0,077	.611/.984	0,836	0,072	.635/.984
HUI3	0,516	0,164	.167/.931	0,493	0,191	.047/1

- Antall observasjoner i behandlingsgruppen: EQ-5D =249 , QWB =239 , 15D =226 , HUI3 =255 .

- Antall observasjoner i kontrollgruppen: EQ-5D = 85, QWB =80 , 15D =78, HUI3 =92.

Tabell 4.4 Gjennomsnittlig MAU-score for behandlings- og kontrollgruppen ved posttest

Noen generelle kommentarer til resultatene

Tabell 4.3 viser at det ved pretest er forskjell i gjennomsnittlig helsereelatert livskvalitet mellom behandlings- og kontrollgruppen. MAU-instrumentene viser en gjennomgående dårligere helse for kontrollgruppen enn for behandlingsgruppen, med unntak av HUI3. Selv om disse forskjellene er små, skal det rent teoretisk ikke være noen forskjeller mellom

gruppene ved randomisering i henhold til Heckman (1979) og LaLonde (1986). Grunnen er som forklart i kapittel 3.3, at alle egenskaper skal være like for de to gruppene ved en randomisering. MAU-instrumentene viser her at det i praksis ikke nødvendigvis trenger å være slik. I den videre analysen må en ta høyde for denne forskjellen ved beregning av behandlingseffekten av den multidisiplinære behandlingsformen. Men med den antagelsen at de andre egenskapene/karakteristikkene for de to gruppene foreløpig er like. Ut i fra disse tallene kan en få mistanke om at kontrollgruppen kan ha vært mer ”sykelig” enn behandlingsgruppen. Men disse forskjellene er for små til å at en kan trekke noen klare konklusjoner om dette. En annen tanke er at det muligens kan være slik at de individene som ikke har svart på alle spørsmålene i et MAU-instrument, og som dermed er utelatt av analysen, kan ha vært med på å ”svekke” randomiseringen. Vi vil komme tilbake med en test på om observerte forskjeller i MAU-instrumentene er systematiske.

Tabellene 4.3 og 4.4 viser at det er store forskjeller i gjennomsnittlig helserelatert livskvalitet mellom MAU-instrumentene både ved pre- og posttest. Det kan være verd å merke seg at laveste verdi er $-0,108$ (en helsetilstand som vurderes verre enn døden) for behandlingsgruppen i EQ-5D ved posttest, mens laveste verdi for 15D er $0,611$. Negative verdier er som nevnt ikke mulig i 15D, men kan forekomme i de andre MAU-instrumentene. Et helsemål som er verre enn døden er ikke et problem i seg selv siden det er fullt mulig å bruke negative mål på helse for en persons helsetilpasning i en helsefunksjon³⁷. Problemet er derimot at ulike MAU-instrument opererer med forskjellige minimumsmål og at en sammenligning dem imellom dermed blir vanskeligere. Særlig gjelder dette for helsemål i den nedre del av skalaen (Hansen (2002)). Disse variasjonene i gjennomsnittlig helserelatert livskvalitet og i minimumsmål, kommer av forskjeller i oppbygningen av instrumentene og hvilken nytteverdi som er satt til en helsetilstand i de forskjellige instrumentene. Dette er et tema som vi vil komme tilbake til i kapittel 4.1.3.

Selv ved disse iboende fundamentale forskjellene gir alle MAU-instrumentene et inntrykk av ganske lave gjennomsnittlige verdier, kanskje med unntak av 15D. I følge Nord (1999a) viser HUI3 og QWB i tabell 4.4 en gjennomsnittlig helserelatert livskvalitet for deltakerne, som blir sett på som en mellomting av ”*svært alvorlig problem*” og ”*fullstendig ufør*”. Verdiene til EQ-5D og 15D ligger seg på ”*svært alvorlig problem*” og ”*moderat problem*”. Nord påpeker at

³⁷ Totur kan være et eksempel på noe som er verre enn døden.

enkelte MAU-instrumenter, inkludert de som brukes i denne analysen, ikke har tilstrekkelig kompresjon av verdier i den øverste delen av skalaen. Dette mener Nord er nødvendig for å fange opp samfunnets sosiale preferanser for de sykeste. Dermed kan de lave resultatene kanskje tolkes som mer alvorlige enn de er, og at gruppene kan i realiteten være friskere enn hva tallene tilsier i henhold til samfunnets sosiale preferanser. Dette er et problem som medfører at behandlingen av lette sykdommer ofte vil bli favorisert på bekostning av tyngre problem hvis tallene fra disse instrumentene brukes i en full CUA. Som løsning innfører han "Rules of thumb" som viser hvilke nytteverdier som bør være ordnet til de forskjellige problemnivåene³⁸.

Det er en svak sammenheng mellom størrelsen på MAU-scoren og antall observasjoner. En vil tro at dette er mer en tilfeldighet enn en realitet med tanke på diskusjonen om forskjeller mellom MAU-instrumentene som kommer senere i kapittelet, samt det som ble nevnt i forrige avsnitt.

Antall observasjoner varierer også mellom instrumentene, og bestemmes mye av hvor mange dimensjoner det er i instrumentet. Her er det en klar sammenheng som kan forklares ut fra det at jo flere dimensjoner det blir spurt om, desto større er sannsynligheten for at enkelte individer kan ha utelatt å svare på noen av spørsmålene. Dermed får en manglende verdi på MAU-scoren for disse individene.

Behandlingseffekt av programdeltakelse?

Vi er primært interessert i å finne ut om gjennomsnittlig helserelatert livskvalitet økte for deltakerne fra pretest til posttest. Begge gruppene i forsøket fikk hver sin type behandling som ga sin respektive gjennomsnittlige behandlingseffekt. Spørsmålet er dermed om det multidisiplinære behandlingsopplegget er bedre enn standard behandling.

I tabell 4.5 viser alle instrumentene en generell helseforbedring over tid for begge gruppene. Unntaket er HUI3 hvor kontrollgruppen har fått et dårligere gjennomsnittlig helserelatert livskvalitet fra pre- til posttest. Det kan være mer grunn til å tro at dette skyldes skjevheter i utfallsmålet, noe vi kommer tilbake til, enn negative virkninger av behandlingen de fikk. En annen tanke kan være at dette gir uttrykk for at det er de friske som har uteblitt i fra posttest,

³⁸ For videre diskusjon rundt dette temaet se Nord (1999a) side 47-50, og Nord (1999b).

og at det er de syke som har møtt opp. Resten viser en generell forbedring. Behandlingsgruppen i EQ-5D viser en økning i helse relatert livskvalitet på 0,054 for sin type behandling, mens kontrollgruppen viser en økning på 0,076 av sin behandling. Ut i fra disse to differansene kan det nesten tyde på at kontrollgruppen har fått en bedre behandling enn behandlingsgruppen. Det vil i så fall være en for kjapp konklusjon her. De tre andre MAU-instrumentene støtter nemlig ikke slik konklusjon. QWB og 15D viser at behandlingseffekten er ganske lik for de to behandlingsformene, hvor økningen i helse relatert livskvalitet er på mellom 0,009 og 0,006. Dette kan tyde på at forbedringene i helse er svakt bedre for behandlingsgruppen enn for kontrollgruppen, men dette kan og være et resultat av seleksjon.

	Behandlingsgruppen	Kontrollgruppen	Differanse ^b
EQ-5D ved posttest	0,681	0,658	0,023
EQ-5D ved pretest	0,627	0,582	0,045
Differanse ^a	0,054	0,076	-0,022

	Behandlingsgruppen	Kontrollgruppen	Differanse ^b
QWB ved posttest	0,596	0,587	0,009
QWB ved pretest	0,577	0,569	0,008
Differanse ^a	0,019	0,018	0,001

	Behandlingsgruppen	Kontrollgruppen	Differanse ^b
15D ved posttest	0,845	0,836	0,009
15D ved pretest	0,812	0,806	0,006
Differanse ^a	0,033	0,030	0,003

	Behandlingsgruppen	Kontrollgruppen	Differanse ^b
HUI3 ved posttest	0,516	0,493	0,023
HUI3 ved pretest	0,509	0,519	-0,010
Differanse ^a	0,007	-0,026	0,033

^a score posttest – score pretest = gj.sn. økning i HRQoL fra pre til posttest

^b score behandlingsgruppen – score kontrollgruppen = differanse i HRQoL mellom behandling og kontrollgruppen ved pre og posttest.

Tabell 4.5 Forskjellige differanser

Når det gjelder økning i helse relatert livskvalitet fra pre- til posttest, er differansen mellom behandling og kontrollgruppen på henholdsvis 0,1 og 0,3 prosentpoeng for QWB og 15D³⁹. Differansene er noe større for de andre instrumentene, hvor differansen i HUI3 er på 3,3 prosentpoeng.. De små differansene mellom behandlings- og kontrollgruppen, tilsier at det ikke er noe nevneverdig større helseeffekt av å få den multidisiplinære behandlingsformen i forhold til den ordinære. Dermed ser det ut til at resultatene foreløpig gir støtte til tidligere

³⁹ QWB: 0.009 – 0.008 = 0.001, 15D: 0.009 – 0.006 = 0.003

funn i dette eksperimentet. At det kan tyde på at de to behandlingsformene gir en noenlunde lik behandlingseffekt, slik som det blir konkludert med i Strand (1998) og Bratberg et al. (2001). Det er interessant å merke seg i denne sammenheng at selv om instrumentene gir forskjellige resultat, virker det som om de alle klarer å fange opp tilnærmet den samme differansen mellom gruppene. Dvs det er ingen store forskjeller her selv om MAU-instrumentene er ganske så ulike i oppbygning og form⁴⁰.

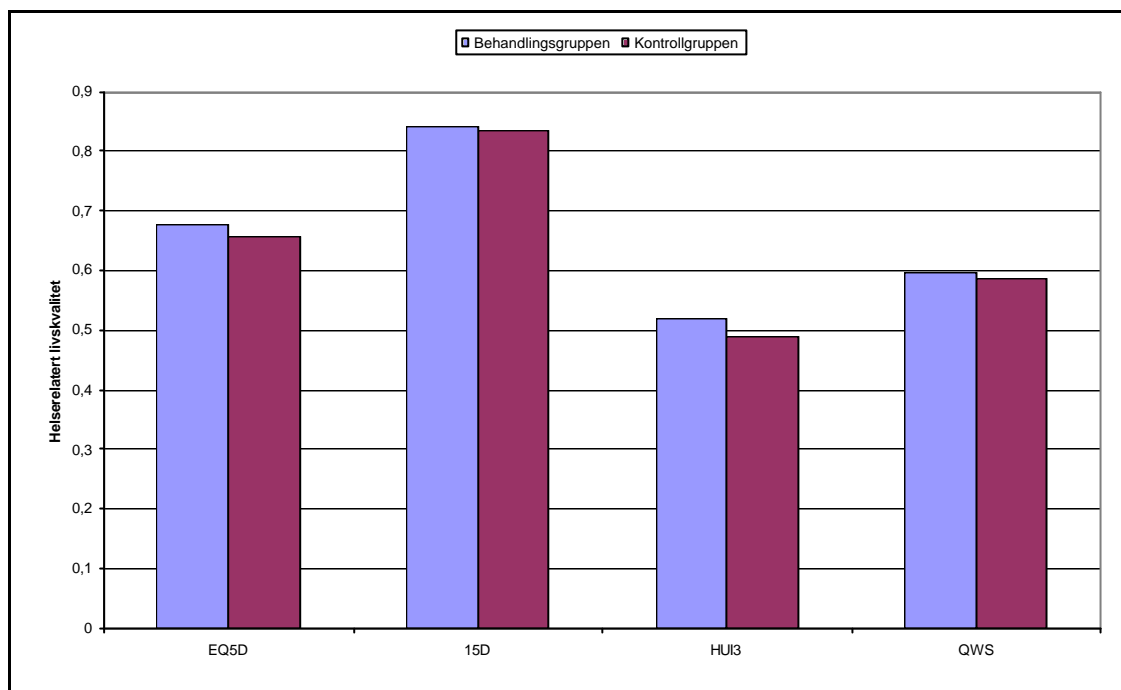
Vi har funnet at menn har en ørliten bedre MAU-score enn kvinnene i alle aldersgrupper, og at gjennomsnittlig helserelatert livskvalitet er lavere for begge kjønn ved høyere alder. Dette stemmer overens med det at det er generelt kvinner som er mest plaget og har mest smerter i følge Statistisk sentralbyrå⁴¹ og Bratberg et al. (1998), og at dette øker med alderen for begge kjønn. Videre tyder det på at kvinner muligens har et bedre utbytte av den multidisiplinære behandlingen enn menn, men differansene her er så små at en kan ikke signifikant konkludere med at kjønn eller alder har noe å si for behandlingseffekten. Vi vil senere teste denne hypotesen ved bruk av interaksjonsledd i regresjonsanalysen.

4.1.3 Forskjellene i resultat mellom MAU-instrumentene

Resultatene i tabell 4.5 og kurven i figur 4.1, viser at gjennomsnittlig helserelatert livskvalitet mellom MAU-instrumentene varierer veldig. Denne variasjonen avhenger ikke av antall observasjoner og dimensjoner mellom instrumentene, men av mer fundamentale ting som instrumentene er tuftet på.

⁴⁰ To studier med fokus på rehabilitering av ryggplager (Moffett et al. (1999) og Kaplan et al. (1992)) brukte MAU-instrumenter som utfallsmål på en gruppe pasienter som er ganske like de i Bergensprosjektet, men som etter beskrivelser er gjennomsnittlig litt bedre. Studiene fant gjennomsnittlig helserelatert livskvalitet ved bruk av HUI3 og QWB på 0,72 og 0,61 for behandlingsgruppen ved posttest. Selv om studiene er ulik Bergensprosjektet på flere punkter, kan dette brukes som en indikator på at tallene vi har funnet er representative for pasientene i det to instrumentene, og at mappingen er ”korrekt” utført idet at resultatene våre ikke er totalt ulike.

⁴¹ <http://www.ssb.no/emner/03/00/helseund/tab-2001-04-03-06.htm>



Figur 4.1 Grafisk fremstilling av forskjellene i resultat mellom MAU- instrumentene

De variable funnene i analysen er ikke unik. Nord (1999c) og Hawthorne (2000) påpeker at det er store forskjeller mellom MAU- instrumentene ved måling av samme helseprofil. Tabell 4.6 fra Nord (1999b) viser forskjellene mellom instrumentene med hensyn til problemnivå.

Instr./ problemnivå	Alvorlig	Betydelig	Moderat
QWB	0,45- 0,55	0,65- 0,70	< 0,80
HUI ⁴²	0,40	0,70	0,90- 0,94
EQ- 5D	0,20	0,60	0,70
15D	0,77	0,86	0,91- 0,93

De tre problem nivåene ble beskrevet som følgende:

- Alvorlig: Sitter i rullestol, har smerter det meste av tiden, ute av stand til å jobbe.
- Betydelig: Bruker krykker til å gå, har lette smerter av og til, ute av stand til å jobbe.
- Moderat: Har noen problemer ved å bevege seg utendørs, har et lite ubehag, men er i stand til å gjøre noe arbeid og har bare mindre problemer hjemme.

Tabell 4.6 Forskjell i helserelatert livskvalitet mellom MAU- instrumenter

Tabell 4.6 viser at variasjonen mellom instrumentene og hvilken nytteverdi som skal tilskrives for samme helseprofil er ganske stor. Differansen mellom EQ-5D og 15D for et individ med et alvorlig problemnivå, er på hele 0,57 i helserelatert livskvalitet⁴³. Nord (1999b) og Hawthorne et al. (2000) påpeker at disse forskjellene kommer av fundamentale problem som

⁴² Er opprinnelig HUI2 I artikkelen. HUI2 og HUI3 blir ofte sett på som så like at de blir behandlet som likeverdige innen fagfeltet. Se for eks Hawthorne et al. (2000). Denne tilnærmelsen brukes også her.

⁴³ $0,77 - 0,20 = 0,57$

mangel på klarhet i basiskonsepter som livskvalitet og helse relatert livskvalitet som diskutert i kapittel 2.3. I tillegg er det en uenighet mellom utviklerne av MAU-instrument når det gjelder instrumentets oppbygning, i form av aspekter rundt helse relatert livskvalitetsmålingene (RS, RG eller TTO?), vektene i dimensjonene, eller kombinasjonsreglene brukt til å regne ut nytte i instrumentet. Utviklerne legger ofte forskjellig mening i tallene sine og er som oftest ikke enige om hvilken verdi som skal tilskrives forskjellige helsetilstander. Dette kan ha gitt et utslag i resultatene våre ved at et individ som har verdien $-0,108$ i EQ-5D ved posttest, har henholdsvis verdiene $0,695$ i 15D, $0,518$ i QWB og $0,529$ i HUI3. Dvs verdier som ligger seg henholdsvis på ”*svært alvorlig*” for EQ-5D og 15D, og ”*alvorlig*” for HUI3 og en mellomting av ”*alvorlig*” og ”*betydelig*” for QWB i henhold til tabell 4.6.

Forskjeller i sosiale preferanser mellom nasjonale populasjoner kan også forklare noe av variasjonen i MAU-instrumentenes resultater (Hawthorne et al. (2000)). Hvis det er kritiske forskjeller som påvirker nytteverdiene, kan de rapporterte forskjellene reflektere dette. Det er systematiske forskjeller i livstilfredsstillelse rundt om i verden, og det vil derfor være overmodig å anta at preferansene er universale. Nedsatt helse eller ulike handikap kan oppleves forskjellig fra land til land. I tillegg varierer behandlings- og rehabiliteringstilbud mellom land, samt at den offentlige støtten er forskjellig⁴⁴. EQ-5D bruker for eksempel britiske vekter, HUI3 kanadiske og 15D finske vekter ved tilskrivelse av en verdi til en helsetilstand. Det er derfor ikke utenkelig at MAU-instrumentene kan få reliabilitetsproblemer dersom de brukes i land hvor preferansene ikke er kartlagt. Så lenge en har slike fundamentale forskjellene mellom instrumentene hvor ingen av de peker seg ut en ”gullstandard”, foreslår de fleste helseøkonomer at en bør ha som en generell regel at en bruker minst to MAU-instrumenter i et studie (Hawthorne et al. (2000), Hurst og Forbes (1999), Suarez-Almazor et al. (2000)).

En bør merke seg at i henhold til tabell 4.6, gir denne støtte til antakelsen om at forskjellene mellom MAU-instrumentene, vist i figur 4.1, kan tilbakeskrives til de fundamentale problemene nevnt ovenfor og til mangelen på en ”gullstandard”. Samtidig gir dette en grunn til å tro at det dermed ikke er feil i selve ”mappingen” som har gjort at det er slike store forskjeller i gjennomsnittlig helse relatert livskvalitet mellom MAU-instrumentene. Resultatene som MAU-instrumentene har kommet frem kan til derfor sees på som ganske

⁴⁴ Norge blir sett på det landet med klart mest universalisme i forhold til befolkningsdekningen. Jo mer like ytelser til alle, dess mer universelle er velferdsordningene (se ref. i Hansen (2000))

valide. Det er også viktig å merke seg at i henhold til Nord sin tabell gir alle MAU-instrumentene våre samme hovedkonklusjon, nemlig at deltakerne har i gjennomsnitt et betydelig helseproblem.

I tillegg til de overnevnte problemer er det blitt avdekket at det også er en stor mangel på empirisk bevis og dekning for det MAU-instrumentene påstår de kan dekke av forskjellige sykdommer/ helsetilstander (Nord (1999b)). Med det menes det at de gjerne ikke er støttet av direkte målinger av preferansene for disse sykdommene/ helsetilstandene. Dermed er det ikke gitt at instrumentene er sensitive nok og “passende” til enkelte sykdommer/ helseproblemer som de påstår å dekke. En slik mangel på empirisk dekning i tillegg til de overnevnte fundamentale problemer, er noe som bør mane til forsiktighet ved tolkningen av tallene og hva de står for.

4.2 Frafallsseleksjon i Bergensprosjektet

4.2.1 Seleksjonsmodellen

Som forklart i kapittel 3.3.2, er et vanlig problem ved sosiale eksperimenter som varer over tid at en del deltakere faller fra underveis. I Bergensprosjektet mangler en oppfølgingsdata på nærmere 1/3 av deltakerne, hvor en stor del er fra kontrollgruppen. Dette skaper et ”missing data” problem, hvor frafallet fører til at en bare har tilgjengelig informasjon på utfall for de som møtte opp til posttest og fylte ut Fellespakken. Hvis frafallet er ikke randomisert og korrelert med utfallsvariabelen/ forklartvariabel (her; MAU-score \approx helse), kan dette påvirke behandlingseffekten siden karakteristika ved de to gruppene ikke nødvendigvis trenger å være like lenger. Dette gjør det nødvendig å undersøke om de observerte forskjellene i utfall mellom behandlings- og kontrollgruppen ved posttest utvalget representerer sann behandlingseffekt, eller om de er generert av frafallsseleksjon. En vanlig metode for å undersøke dette, er å utføre en analyse med en regresjonsmodell som prøver å korrigere for frafallet⁴⁵.

⁴⁵ De to modellene som brukes her er Minste Kvadraters Metode som ikke korrigerer for seleksjon, og Full Maximum Likelihood som korrigerer for seleksjon. Disse vil henholdsvis bli betegnet som MKM og FML utover i analysen.

En vanlig økonometrisk tilnærming for å korrigere for frafallsseleksjon, er å modellere utfallet Y_i (MAU-scoren) for individ i , som en lineær funksjon av det å få behandling D_i , andre observerbare karakteristika X_i (alder, kjønn osv.⁴⁶) og uobserverbare karakteristika U_i der * indikerer den latente variabelen (Y_i^* observeres bare hvis $A_i=1$). Likning (4.1) viser dette hvor \mathbf{b} er en vektor av koeffisienter som måler påvirkningen av X_i på Y_i . \mathbf{d} måler effekten av multidisiplinær behandling ($D_i = 1$ for behandling, 0 ellers), mens U_i er feilleddet. Spørsmålet er om \mathbf{d} klarer konsistent å måle verdien av det å motta multidisiplinær behandling. Det kan være et problem hvis frafallet henger systematisk sammen med forhold målt i MAU-instrumentene, og en ikke prøver å korrigere for dette. Utledningen av seleksjonsmodellen som bygger på Heckman formuleringen, er hentet i fra Heckman (1979) og er omarbeidet litt for å passe inn i her.

$$Y_i^* = \mathbf{b}' X_i + \mathbf{d}D_i + U_i \quad (\text{observeres hvis } A_i=1) \quad (4.1)$$

Man har to former for frafallsseleksjon:

- 1) Seleksjon på observerbare kjennetegn. Dette finner sted hvis frafallet er relatert til observerbare karakteristikk. Nærmere sagt så vil det si at de som møter opp til posttest og de som ikke møter opp, er ulike med hensyn på observerbare kjennetegn (av betydning for MAU-scoren \approx helse) som kjønn, alder osv. Hvis en ikke kontrollerer for dette kan en feilvurdere evalueringresultatet. Ved seleksjon på kun observerbare karakteristikk kan likning, (4.1) konsistent estimeres med minste kvadraters metode (MKM) hvor en korrigerer for denne seleksjonen.
- 2) Seleksjon på uobserverbare kjennetegn. Dette finner sted hvis frafallet er relatert til uobserverte karakteristikk som motivasjon, evne, preferanser osv., som kan være korrelert med utfallsvariabelen. Dermed kan ikke likning (4.1) lenger estimeres konsist med minste kvadraters metode som utledningen nedenfor vil vise.

⁴⁶ Variablene som brukes i X_i for å forklare den helserelaterte livskvaliteten i analysen er: Behandling (Behandlingsgruppe=1, Kontrollgruppe=0), Kjønn (Mann=1, Kvinne=0), Alder, Alder², Mndsyk (Antall måneder syk før pretest) og Inntekt/100000 (Inntekt året før pretest). Disse brukes også til å forklare seleksjon sammen med variabelen Bergen (Bor i Bergen kommune=1, Bor ikke i Bergen kommune=0), som er en identifiserende variabel i seleksjonslikningen.

Frafall grunnet seleksjon på observerbare og uobserverbare kjennetegn, beskrives vanligvis med en latent indeks struktur av typen vist i likning (4.2) og (4.3). Her er Z_i er en vektor med observerbare karakteristikk, og V_i er uobserverte karakteristikk. Seleksjonsmekanismen i likning (4.3) er en dummyvariabel, A_i (er enten lik null eller en), hvor en har at Y_i bare observeres hvis $A_i = 1$ som vist i likning (4.4). Dvs. for at vi skal kunne observere utfallsvariabelen, Y_i som er MAU-scoren ved posttest, må seleksjonsdummyen være lik 1. Med det menes at deltakeren har for eksempel motivasjon, evner osv. for å møte opp til posttest, noe som likningene (4.2) og (4.3) prøver å fange opp.

$$A_i^* = \mathbf{a}' Z_i + V_i \quad (4.2)$$

$$A_i = \begin{cases} 1 & \text{hvis } A_i^* > 0 \\ 0 & \text{hvis } A_i^* \leq 0 \end{cases} = \text{Seleksjonsmekanisme} \quad (4.3)$$

$$Y_i = A_i Y_i^* = A_i (\mathbf{b}' X_i + \mathbf{d} D_i + U_i) \quad (4.4)$$

En trenger videre noen antagelser rundt fordelingene til feilleddene U_i og V_i . Standard tilnærmelsen er å anta at U_i og V_i er bivariat normalfordelt (dvs. trukket fra samme populasjon samtidig), vist ved likning (4.5). Videre antar en og at (U_i, V_i) er uavhengig av (Z_i, X_i) .

$$(U, V) \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{s}_U^2 & \mathbf{r} \mathbf{s}_U \mathbf{s}_V \\ \mathbf{r} \mathbf{s}_U \mathbf{s}_V & \mathbf{s}_V^2 \end{pmatrix} \right] \quad (4.5)$$

Korrelasjonen mellom utfallsvariabelen og de uobserverte kjennetegnene, dvs. korrelasjonen mellom U_i og V_i kan uttrykkes med likning (4.6). Likningen fanger opp korrelasjonen mellom frafallsoppførselen og de endogene variablene i regresjonen. \mathbf{r} (rho) ligger mellom -1 og 1 .

$$\text{Corr}(U_i, V_i) = \mathbf{r} \quad \mathbf{r} \in [-1, 1] \quad (4.6)$$

Hvis det er korrelasjon mellom U_i og V_i , dvs. hvis $\mathbf{r} \neq 0$, gjør dette at minste kvadraters metode blir inkonsistent ved estimering av likning (4.1). Gitt at spesifikasjonen gjort i likningene (4.2) – (4.5) holder for å identifisere mekanismen som avgjør oppførselen for

frafallet, vil en få gjenopprettet representativiteten til behandlings- og kontrollgruppen fra det utvalget de ble trukket fra. Dermed kan en få estimert konsistente estimater av koeffisientene ved hjelp av standard Heckmans Tostegs Estimering⁴⁷, eller Full Maximum likelihood estimering. Dette gjøres ved å (tilnærmet) estimere likning (4.7), hvor \mathbf{r} er korrelasjonskoeffisienten og \mathbf{s}_U er standardfeilen i den latente regresjonslikningen (4.1). Leddet $I_i(\mathbf{g}_V)$ er den inverse millsraten, vist i likning (4.8), hvor $\mathbf{j}(\cdot)$ er standard normalfordeling og $\Phi(\cdot)$ er kumulativ standard normalfordeling.

$$Y_i = \mathbf{b}' X_i + \mathbf{d}D_i + \mathbf{r}\mathbf{s}_U I_i(\mathbf{g}_V) + U_i \quad (4.7)$$

$$I_i(\mathbf{s}_Z) = \frac{\mathbf{j}\left(\frac{\mathbf{g}' Z_i}{\mathbf{s}_V}\right)}{\Phi\left(\frac{\mathbf{g}' Z_i}{\mathbf{s}_V}\right)} \quad (4.8)$$

Implikasjonen av dette er at hvis en estimerer (4.1) uten leddet $\mathbf{r}\mathbf{s}_U I_i(\mathbf{g}_V)$ som i likning (4.7), vil utfallsvariabelen bli forventningsskjev hvis $\mathbf{r} \neq 0$. Leddet $\mathbf{r}\mathbf{s}_U I_i(\mathbf{g}_V)$ er gitt i fra setning 20.5 i Greene (2000) om trunkering i bivariat normalfordeling, som bygger på antagelsen om at (U_i, V_i) er bivariat normalfordelt (Denne antagelsen vil ikke bli utledet videre her, og for detaljer rundt dette henvises det til Greene (2000)). Rent teknisk så er Heckmans tostegs modell mer robust enn Full Maximum Likelihood modellen med hensyn på setning 20.5 i Greene. Dette fordi full maximum likelihood modellen er mer kritisk avhengig av den i sine beregninger (diskuteres ikke videre her). Men gitt at modellen er korrekt spesifisert som ved likningene ovenfor, er Full Maximum Likelihood den modellen som gir de mest effisiente koeffisient estimatene av disse to i henhold til Greene. I denne sammenheng er det derfor valgt å rapportere resultatene i fra Full Maximum Likelihood modellen i analysen.

⁴⁷ Heckmans tostegsmodell: (1) Beregner likning (4.2) først med probit (dvs antar $\mathbf{s}_U^2 = 1$) og regner ut den inverse Mills raten, likning (4.8). (2) Estimerer så $Y_i = \mathbf{b}' X_i + \mathbf{b}_1 I_i(\mathbf{s}_Z) + V_i$ for de som har $A_i = 1$, hvor $\mathbf{b}_1 = \mathbf{r}\mathbf{s}_U$. Hvis $\mathbf{b}_1 = 0$, da er korrelasjonen lik null, dvs $\mathbf{r} = 0$. Dette gir konsistente estimater for \mathbf{b} når korrigeret for standardfeil (Greene (2000)).

4.2.2 Resultater Minste kvadraters metode og Full maximum likelihood

Praktisk bakgrunnsinformasjon

I denne analysen er det brukt fire utfallsmål på det samme helsetiltaket. Det medfører at når en skal undersøke om det er et seleksjonsproblem i datasettet som påvirker utfallsmålene, må FML som ble utledet i seksjon 4.2.1 estimeres en gang for hvert MAU-instrument. EQ-5D ble brukt som et utgangspunkt i letingen etter variabler som kunne forklare "helse" og en eventuell seleksjon. Etersom resultatene varierer etter hvilket instrument sin helserelaterte livskvalitet som er den forklarte variabelen, så vi det å ha en "grunnmodell" å jobbe ut i fra som den mest fornuftige strategien. Forenklet sagt er det et to effekter en er på jakt etter her: 1) Behandlingseffekten av det medisinske tiltaket, 2) Hvorfor enkelte deltakere ikke møtte opp til posttest, hva som karakteriserer de og om dette påvirker behandlingseffekten (Greene (2000)).

Mange variabler ble i denne prosessen prøvd ut i modellen, som for eksempel antall arbeidsår, gift, utdanning osv. Alle disse koeffisientestimatene var i stor grad innsignifikante, og ble derfor ikke tatt med blant variablene i den endelige modellen⁴⁸. Vi fant bare en identifiserende variabel for seleksjon, som er variabelen Bergen som er en dummy for om en bor i Bergen kommune eller ikke⁴⁹. Det er og i denne sammenheng naturlig å sjekke seleksjonsmodellen opp mot en vanlig regresjonsmodell uansett utfall i seleksjonsregresjonen, og bør i alle tilfeller gjøres for å korrigere for observerbare kjennetegn. En vanlig regresjonsmodell er nyttig å ha som et sammenligningsgrunnlag, for å se hva som skjer med variablene når en korrigerer for seleksjon.

Kjøringsresultatene

Tabellene 4.7 til 4.10 viser resultatene fra MKM og FML for de fire instrumentene. Siden det er fire MAU-instrumenter en estimerer for her, vil de forskjellige koeffisientene først bli kommentert hver for seg. Deretter følger en diskusjon rundt spørsmålet om frafallseleksjon påvirker resultatene.

⁴⁸ Det ble sjekket for korrelasjon mellom variablene, hvor vi fant at ingen var i nærheten av å være korrelerte.

⁴⁹ I utgangspunktet var et av inklusjonskriteriene at pasienten skulle bo i Bergen kommune. Men grunnet en rekrutteringssvikt som delvis kunne forklares av en endring i sykemeldingspraksis, hvor legene ble mer restriktive med hensyn til å sykemelde på grunnlag av blant annet de diagnoser som lå til grunn til rekrutteringen til prosjektet, ble prosjektområdet utvidet til å omfatte omegnskommunene (Strand (1998)).

EQ-5D	Minste kvadraters metode		Full maximum likelihood	
Variabler	Koeffisienter	Standardfeil	Koeffisienter	Standardfeil
Behandling	0,0274	0,0201	-0,0323	0,0231
Kjønn	-0,0009	0,0198	0,0132	0,0299
Alder	-0,0098	0,0070	-0,0199*	0,0078
Alder²	0,0001	0,0001	0,0002*	0,0000
Mndsyk	-0,0051	0,0043	-0,0066	0,0047
Inntekt/100000	0,0254*	0,0129	-0,0244*	0,0149
Bergen			0,6127*	0,1361
Behandling			0,5702*	0,1261
Kjønn			-0,0359	0,1170
Alder			0,1011*	0,0393
Alder²			-0,0012*	0,0005
Mndsyk			0,0159	0,0236
Inntekt/100000			-0,0114	0,0823
Justert R²	0,0398			
Rho^a			-0,9047	0,0366
LR-test^b			0,0000*	
# observasjoner	334		334	

^a Rho er korrelasjonskoeffisienten r fra likning (4.6)

^b Estimert rho, chi-kvadrat statistikk test for LR test ($\rho = 0$)

* Signifikant på 10% nivå.

Tabell 4.7 MKM og FML for EQ-5D

QWB	Minste kvadraters metode		Full maximum likelihood	
Variabler	Koeffisienter	Standardfeil	Koeffisienter	Standardfeil
Behandling	0,0334	0,0229	0,0204	0,0256
Kjønn	-0,0153	0,0224	-0,0163	0,0224
Alder	-0,014*	0,0079	-0,0165*	0,0082
Alder²	0,0001	0,0001	0,0002*	0,0001
Mndsyk	-0,0047	0,0048	-0,0049	0,0048
Inntekt/100000	0,0419*	0,0148	0,0432*	0,0148
Bergen			0,6416*	0,1678
Behandling			0,5955*	0,1205
Kjønn			0,0491	0,1317
Alder			0,1246*	0,0413
Alder²			-0,0015*	0,0005
Mndsyk			0,0102	0,0246
Inntekt/100000			-0,0611	0,085
Justert R²	0,0353			
Rho^a			-0,2172	0,2349
LR-test^b			0,3677	
# observasjoner	319		319	

^a Rho er korrelasjonskoeffisienten r fra likning (4.6)

^b Estimert rho, chi-kvadrat statistikk test for LR test ($\rho = 0$)

* Signifikant på 10% nivå.

Tabell 4.8 MKM og FML for QWB

15D	Minste kvadraters metode		Full maximum likelihood	
	Variabler	Koeffisienter	Standardfeil	Koeffisienter
Behandling	0,011	0,0097	-0,0122	0,0112
Kjønn	0,0172*	0,0095	0,0205*	0,0111
Alder	-0,0016	0,0034	-0,0070*	0,0039
Alder²	0,00001	0,0001	0,0001*	0,0001
Mndsyk	-0,0022	0,0019	-0,0045*	0,0022
Inntekt/100000	0,0190*	0,0063	0,0196*	0,0073
Bergen			0,5837*	0,1397
Behandling			0,5153*	0,1174
Kjønn			-0,0168	0,1272
Alder			0,1073*	0,0398
Alder²			-0,0013*	0,0005
Mndsyk			0,0255	0,0239
Inntekt/100000			-0,0875	0,0844
Justert R²	0,0658			
Rho^a			0,872	0,0516
LR-test^b			0,0009*	
# observasjoner	304		304	

^a Rho er korrelasjonskoeffisienten r fra likning (4.6)

^b Estimert rho, chi-kvadrat statistikk test for LR test ($\rho = 0$)

* Signifikant på 10% nivå.

Tabell 4.9 MKM og FML for 15D

HUI3	Minste kvadraters metode		Full maximum likelihood	
	Variabler	Koeffisienter	Standardfeil	Koeffisienter
Behandling	0,0283	0,0207	0,0299	0,0261
Kjønn	0,0748*	0,0208	0,0743*	0,0212
Alder	-0,0034	0,0075	-0,003	0,0083
Alder²	0,0001	0,0001	0,0001	0,0001
Mndsyk	-0,0005	0,0044	-0,0005	0,0044
Inntekt/100000	-0,0172	0,0136	-0,0173	0,0135
Bergen			0,7902*	0,1691
Behandling			0,5616*	0,1225
Kjønn			-0,1349	0,1345
Alder			0,1327*	0,0419
Alder²			-0,0015*	0,0005
Mndsyk			0,0194	0,0249
Inntekt/100000			-0,0437	0,0874
Justert R²	0,0276			
Rho^a			0,0355	0,3492
LR-test^b			0,9197	
# observasjoner	347		347	

^a Rho er korrelasjonskoeffisienten r fra likning (4.6)

^b Estimert rho, chi-kvadrat statistikk test for LR test ($\rho = 0$)

* Signifikant på 10% nivå.

Tabell 4.10 MKM og FML for HUI3

Kommentarer til variablene

Behandling

Er en dummy for om en mottok multidisiplinær behandling (D=1) eller ikke (D=0). Denne koeffisienten gir et estimat på behandlingseffekten til den multidisiplinære behandlingen. Ved bruk av MKM rapporterer alle instrumentene (insignifikante) positive koeffisienter mellom 0.011 til 0.033, noe som tilsier at multidisiplinære behandlingen har en positiv effekt på helsen hvis de skal tolkes i noen retning. Det interessante som er verd å merke seg her, er at i FML blir disse koeffisientene (insignifikante) negative i EQ-5D og 15D. Disse to instrumentene er de eneste som indikerer frafallseleksjon. I de to andre forblir koeffisientene positiv, samt det indikeres ikke for frafallsseleksjon. Denne variabelen gir i FML en sterkt signifikant (p-verdi=0,000) positiv koeffisient i seleksjonsdelen i alle instrumentene, noe som kan tolkes som at det å motta behandling er en viktig determinant for frafallet. Tolkningen her blir således at de som mottok den multidisiplinære behandlingen, har en mye større sannsynlighet for å møte opp til posttest. Dette støttes av funnene til Grasdahl (2001a) og Bratberg et al.(2002).

Kjønn

Dummy for kjønn (Mann=1, Kvinne=0). 15D og HUI3 gir estimater som er signifikante og positive i MKM og FML. Dette kan tolkes i den retning at det å være mann (isolert sett), har en svak positiv påvirkning på helsen. I EQ-5D og QWB er de (insignifikante) negative i MKM og forblir dette i FML. Unntaket er EQ-5D hvor den blir (insignifikant) positiv. I seleksjonslikningen sier variabelen at det er en negativ sammenheng mellom det å være mann og møte opp til posttest i alle instrumentene, unntatt i QWB. Men ingen av disse er skarpt bestemte og legges derfor ingen vekt på. Skal de allikevel tolkes, tyder dette på at kvinnene var mer tilbøyelig til å møte opp til posttest enn mennene.

Alder

Alderskoeffisienten er insignifikant og negativ i alle instrumentene i MKM. Disse forblir dette også i FML, men nå signifikante for alle unntatt for HUI3. Dette kan tolkes dit hen at jo eldre en blir, jo dårligere blir helsen din. Den kvadrerte aldersvariabelen gir en insignifikant positiv koeffisient i alle instrumentene ved MKM. Disse forblir også det samme i FML, men nå signifikant i alle instrumentene unntatt for HUI3. Dette tilsier at helsen ved økt alder har en svak avtagende konkav helning. I FML har økt alder en positiv signifikant påvirkning på

sannsynligheten for å møte opp til posttest i alle instrumentene. Dette kan tyde på at de eldre var mer pliktoppfyllende til å møte opp til posttest enn de unge, noe Bratberg et al. (2002) og fant. Den kvadrerte aldersvariabelen gir en signifikant negativ koeffisient i alle instrumentene. Dette betyr at sannsynligheten for å møte opp til posttest øker i avtagende grad (konkav) med alder.

Mndsyk

Antall måneder en person var syk før pretest, har en negativ (insignifikant) påvirkning på helsen i MKM og FML. Variabelen er en indikator på hvor syk en person var før inntaket. Lengden på sykemeldingsperioden før pretest kan med et romslig kritisk testnivå sies å virke ugunstig på helsen både i MKM og FML. Lengden på sykemeldingsperioden før pretest har og en (insignifikant) positiv sammenheng med det å møte opp til posttest. Dette kan tyde på (isolert sett) at det er de sykeste som forblir, men dette gir ikke noe direkte svar.

Inntekt

Økende inntekt synes å ha en positiv sammenheng med helsen i alle instrumentene, unntatt for HUI3. Dette stemmer overrens med andre studier (se Folland et al. (2001)). Både i MKM og FML er disse estimatene signifikante, unntatt for HUI3. I FML synes økt inntekt å ha en (insignifikant) negativ sammenheng med oppmøte til posttest i alle instrumentene. Kan tyde på at de med høy inntekt uteblir fra posttest.

Bergen

Dummy for om en bor i Bergen kommune ($B=1$) eller utenfor ($B=0$). Er den identifiserende variabelen med hensyn på seleksjon. Bergensvariabelen gir en sterk signifikant (p -verdi=0,000) positiv koeffisient for alle instrumentene i FML. Dette kan tyde på at de som bodde i Bergen, hadde en mye større sannsynlighet for å forbli i forsøket og møte opp til posttest enn de som bodde utenfor.

Seleksjon eller ikke seleksjon?

Resultatene for MAU-instrumentene EQ-5D og 15D tyder på at det er et seleksjonsproblem i våre data, mens resultatene fra de to andre tilsier at det ikke er et seleksjonsproblem til stede. Spørsmålet er da, hvilken av disse skal en tro på? Det er hva vi skal prøve å løse her ved å se nærmere på rho og resultatene i fra Loglikelihood-testen (LR-testen), for hypotesen H^0 : $\rho=0$, for hver av MAU-instrumentene.

Tabell 4.11 viser behandlingskoeffisienten for MKM og FML, samt lambda, rho og i klammer [*] LR-testen for om $\rho=0$ for FML⁵⁰. Rho antar verdier mellom 1 og -1. Hvis rho er statistisk signifikant ulik null er det per definisjon seleksjon tilstede. LR-testen er en test på om frafallsseleksjonen påvirker resultatene eller ikke.

	Effekt (MKM)	Effekt (FML)	Lambda (FML)	Rho (FML)
EQ5D	0,0274 (0,174)	-0,0323 (0,156)	-0,1841	-0,9047 [0,0000]
QWB	0,0334 (0,146)	0,0207 (0,418)	-0,0383	-0,2172 [0,3677]
15D	0,0110 (0,254)	-0,012 (0,279)	-0,8279	-0,8720 [0,0009]
HUI3	0,0283 (0,172)	0,0299 (0,252)	0,0060	0,0355 [0,9197]

- o Effekt = koeffisienten Behandling
 - o Tallene i parentes er p-verdier (bruker signifikans på 10% nivå)
 - o Tallene i klammer [*] viser resultatene for Loglikelihood-testen
- H^0 : $\rho=0$ i FML modellen

Tabell 4.11 MKM og FML modellen, full modell (n=full)

Tabell 4.11 viser at både lambda og korrelasjonskoeffisienten er negativ i alle MAU-instrumentene unntatt for HUI3. En negativ rho betyr at uobserverte faktorer relatert til sannsynligheten for å møte opp til posttest, er negativt korrelert med utfallsmålet (Greene (2000)). LR-testen i EQ-5D og 15D forkaster null hypotesen om at $\rho = 0$ på 10% statistisk signifikans nivå, mens for QWB og HUI3 forkastes den ikke. En mulig tolkning av disse resultatene er at for EQ-5D og 15D uteble de sykeste fra posttest, mens de friskeste møtte opp, samt at denne skjevheten påvirker utfallsmålet. QWB kommer frem til det samme, men at dette har ingen påvirkning på utfallsmålet. HUI3 kommer frem til det motsatte hvor det er de syke som møter opp, men at dette heller ikke har noen innvirkning på utfallsmålet.

Det som er verd å merke seg her, er at i FML er estimatene for behandlingskoeffisientene (insignifikante) negative i EQ-5D og 15D. Disse to er de eneste instrumentene som indikerer

⁵⁰ Observasjonene varierer som sagt mellom instrumentene av nevnte grunner i kapittel 3.4.3. Det ble utført et forsøk hvor vi bare inkluderte de individer som hadde fått utregnet MAU-score i alle instrumentene. Dette var en total på $n=260$ hvor vi kjørte MKM og FML på disse, istedenfor å inkludere alle som hadde i hvert fall score i et instrument, $n=full$. Vi fant ingen signifikante forskjeller mellom $n=full$ og $n=260$ med hensyn til behandlingseffekten. Siden antall observasjoner ikke har noe å si for variasjonen i behandlingseffektene, har vi valgt å vise kjøringene for alle observasjonene, $n=full$.

frafallsseleksjon. Siden FML og MKM gir forskjellige estimat, skal det i teorien være seleksjon til stede (Heckman (1979)). Tilstedeværelsen av signifikante frafallsseleksjonstermer i FML modellen som prøver å korrigere for seleksjon på observerbare, samt uobserverbare individuelle karakteristikk, er ikke en formell test på anvendbarheten til MKM spesifikasjonen som bare korrigerer for frafallsseleksjon på det observerbare (Bratberg et. al. (2002)). Men faktumet at to av instrumentene gir signifikante seleksjonsestimater, styrker troen på at behandlingseffektene estimert av en enkel MKM regresjon av X_i på utfallet til de som møtte opp til posttest er innkorrekt. Dette bildet styrkes enda mer av at EQ-5D og 15D er de to instrumentene som vurderes til best å kunne dekke sykdomsbildet her, og gjengi hva respondentene svarte i Fellespakken. Disse to instrumentene vurderes best til å avspeile behandlingens påvirkning på de to gruppene. En kan dermed under tvil godta det at det er seleksjon i datasettet som plager evalueringresultatene, hvis en velger å tro mest på EQ-5D og 15D. Siden ikke alle instrumentene indikerer seleksjon, kan dette tolkes som et tegn på at datasettet er plaget av frafallsseleksjon som ligger i grenselandet og påvirker utfallsmålene.

Ingen av behandlingskoeffisientene i tabell 4.12 produserer punkttestimater som er statistisk ulik null på 10% nivå. Bare i EQ-5D er en i nærheten, hvor en får punkttestimater statistisk ulik null på 17% nivå. En ting en må unngå ved slike analyser er å gjøre type 2 feil, identifisere signifikante behandlingseffekter, positive eller negative når de ikke er tilstede. Resultatene over medfører at estimatene her blir behandlet som null, og at den samlede effekten av programdeltakelse tolkes som null både i FML og MKM. Dette medfører at vi ender opp med den samme konklusjonen som i kapittel 4.1. Det er ingen nevneverdig behandlingseffekt (positiv eller negativ) av det å motta behandling for et gjennomsnittlig individ med de gitte inklusjonskarakteristikk.

Tilslutt må det bemerkes at estimatene er sensitive til ekskludering av Bergensvariabelen i seleksjonslikningen. Når denne variabelen ekskluderes, estimerer FML et (statistisk insignifkant) punkttestimat for behandlingseffekt i henholdsvis EQ-5D, QWB, 15D og HUI3, på 0.0249, 0.0303, 0.0128 og 0.0284. Estimert rho er på henholdsvis 0.0399, -0.0513, 0.0787 og 0.0023. Men nå kan ikke $H_0: \rho=0$ forkastes lenger i noen av instrumentene. Noe som peker mot at frafallet i dette tilfellet ikke fører til at behandlingseffekten blir forventningsskjev, og modellen dermed kan estimeres uten seleksjonstermene. Dette kan være et resultat av at seleksjonsmodellen ofte blir ustabil når det ikke er noen identifiserende

variabler i seleksjonsdelen til modellen (Greene (2000)). Teoretisk sett trenger en ikke noen identifiserende variabler, men uten slike variabler overlater en til den funksjonelle formen å identifisere modellen. Dermed vil det bli vanskelig å ta slike resultater seriøst siden den funksjonelle form antagelsen ikke har noen stø basis i teorien (Stata (1999)). En LR-test på forklaringskraften til variabelen Bergen forkaster hypotesen, H^0 : Bergen = 0, i alle MAU-instrumentene med god margin. Denne testen sees på som mer nøyaktig enn en vanlig t-test (Greene (2000)) og (Stata (1999)). Det er dermed vanskelig å forkaste hypotesen om at Bergens variabelen ikke skal ha noen forklaringskraft i modellen når to tester sier at den har det. Testene sier at denne variabelen bør være med.

4.2.3 Spesifikasjonstest

Heckman og Hotz (1989) og LaLonde (1986) argumenterer begge for at en god spesifikasjonstest av regresjonsmodellen, er å undersøke om utfallene ved pretest kan forklares med den samme modellen som brukes til å forklare utfallet ved posttest. Hvis dummyen for mottatt behandling, som i dette tilfellet er variabelen "Behandling", gir en programeffekt som er signifikant forskjellig fra null i en modell der utfallsvariabelen ved posttest (posttest MAU-scoren) er byttet ut med den korresponderende pretest (pretest MAU-scoren) utfallsvariabelen, feiler estimatoren testen. I motsatt tilfelle passerer den testen. Tabell 4.12 viser resultatene av denne spesifikasjonstesten.

	EQ-5D	QWB	15D	HUI3
Behandlingskoeffisient	0,022	0,009	-0,003	-0,010
Signifikans (P> Z)	0,264	0,371	0,713	0,613

Tabell 4.12 Pre-program testen

I seksjon 4.1.2 fant vi at det var forskjell i behandlingseffekt mellom behandling og kontrollgruppen ved pretest. Resultatene fra denne testen sier i så fall at denne forskjellen ikke er signifikant forskjellig i fra null, og at gruppene er like. Vi kan ikke forkaste noen av estimatene med denne testen. Ingen av behandlingskoeffisientene er statistisk signifikante på 10% nivå når posttest MAU-scoren byttes ut med pretest MAU-scoren. Dette kan tolkes som et tegn i retning av at resultatene i fra FML modellen er valide og kan stoles på. Vi har ikke nødvendigvis funnet den "sanne" modellen, men har fått utelukket det at modellen er totalt misledende.

4.3 Heterogenitet i behandlingseffekten

Vi har ikke klart å påvise en statistisk signifikant bedring i gjennomsnittlig helserelatert livskvalitet blant alle deltakerne, sett under ett. Dette utelukker derimot ikke muligheten for at behandlingen kan ha hatt en større effekt i enkelte undergrupper, ved at enkelte deltakere kan ha vært mer mottakelig for behandlingen enn andre.

Interaksjonstermer

Det er alltid en mulighet for at behandlingen virker forskjellig i ulike undergrupper. En måte å analysere om virkningen er forskjellig for menn og kvinner, for ulike alders- og inntektsgrupper osv., er å konstruere et sett med interaksjonsvariabler. Dette gjøres ved å lage nye variabler, som er produktet av behandlingsdummyen og andre bakgrunnsvariabler. Dette kan vises med en modell som i likning (4.9), hvor *helse (MAU-scoren)* blir forklart av variablene *behandling (B)*, *kjønn (M)*, *alder (A)*, *behandling * kjønn (B * K)* og *behandling * alder (B * A)*. De to siste termene er interaksjonstermer (Greene (2000)).

$$MAU = \mathbf{b}_0 + \mathbf{b}_1 B + \mathbf{b}_2 M + \mathbf{b}_3 A + \mathbf{b}_4 B * M + \mathbf{b}_5 B * A \quad (4.9)$$

$$\partial MAU / \partial B = \mathbf{b}_1 + \mathbf{b}_4 + \mathbf{b}_5 A \quad (\text{for } M=1) \quad (4.10)$$

$$\partial MAU / \partial B = \mathbf{b}_1 + \mathbf{b}_5 A \quad (\text{for } M=0) \quad (4.11)$$

Deriverer en likning (4.9) med hensyn på behandling, vil en finne behandlingseffekten med hensyn på alder vist i likning (4.10) for menn og (4.11) for kvinner. Likningene impliserer at behandlingseffekten avtar med alderen (hvis en antar \mathbf{b}_5 er negativ) for begge kjønn ($\mathbf{b}_4 = 0$ når $M = 0$, $\mathbf{b}_4 \neq 0$ når $M = 1$). Forskjellen mellom menn og kvinner er gitt ved \mathbf{b}_4 .

Resultat interaksjonsledd

Tabellene 4.14 og 4.15 viser koeffisientene for interaksjonsleddene i MKM og FML for de fire MAU-instrumentene. Interaksjonsleddene gjør det mulig å identifisere i hvilken grad programeffekten er betinget av de oppsummerte bakgrunns karakteristika.

Variabler	EQ-5D		QWB	
	Koeffisienter	Standardfeil	Koeffisienter	Standardfeil
b*mann	-0,0567	0,0441	-0,0446	0,0195
b*alder	0,0030	0,0155	-0,0096	0,0069
b*alder2	-0,0001	0,0002	0,0001	0,0001
b*inntekt	-0,0394	0,0314	-0,0031	0,0138
b*mndsyk	0,0177	0,0109	0,0021	0,0047

Variabler	15D		HUI3	
	Koeffisienter	Standardfeil	Koeffisienter	Standardfeil
b*mann	-0,0118	0,0212	-0,0309	0,0469
b*alder	0,0005	0,0076	-0,0074	0,0163
b*alder2	-0,00002	0,0001	0,0001	0,0002
b*inntekt	-0,0081	0,0171	-0,0148	0,0330
b*mndsyk	0,0025	0,0047	0,0047	0,0107

Tabell 4.13 MKM med interaksjonsvariabler

Variabler	EQ-5D		QWB	
	Koeffisienter	Standardfeil	Koeffisienter	Standardfeil
b*mann	-0,0367	0,03779	-0,087*	0,0498
b*alder	0,0054	0,0130	-0,0192	0,0176
b*alder2	-0,0001	0,0002	0,0002	0,0002
b*inntekt	-0,0435	0,0290	0,0254	0,0351
b*mndsyk	0,0151*	0,0096	0,0052	0,0119
rho	-0,0900	0,0443	-0,3248	0,2944
LR-test	0,0000		0,3481	

Variabler	15D		HUI3	
	Koeffisienter	Standardfeil	Koeffisienter	Standardfeil
b*mann	-0,0106	0,0192	-0,0308	0,0460
b*alder	0,0019	0,0065	-0,0077	0,0162
b*alder2	-0,00002	0,0001	0,0001	0,0002
b*inntekt	-0,0111	0,0156	-0,0144	0,0326
b*mndsyk	0,0033	0,0045	0,0047	0,0105
rho	-0,8744	0,0532	0,0395	0,3718
LR-test	0,0008		0,9165	

Tabell 4.13 FML med interaksjonsvariabler

Det en ser ut i fra tabell 4.13 og 4.14 er at for bakgrunnskjennetegnene som vi har kontrollert for, er de fleste uten signifikant betydning. Sannsynligvis har vi fått disse resultatene på grunn av det ikke eksisterer noen signifikante effekter i de utvalgte undergruppene vi har prøvd ut her. Kjøringene her er forholdsvis like de i kapittel 4.2, slik at en kommer frem til samme konklusjon med hensyn til seleksjon her som der. Dvs. det konkluderes med at datasettet er plaget med frafallseleksjon som ligger i grenseland og påvirker utfallsmålene.

Ved en eventuell tolkning av resultatene for FML, viser alle instrumentene at det å være mann i behandlingsgruppen synes å ha en svak negativ virkning, som nesten opphever den positive virkningen av å være mann i utgangspunktet⁵¹. Dette kan tolkes som at menn har best helse (i gjennomsnitt) i forhold til kvinnene i utgangspunktet, men at menn ikke har noe særlig utbytte av det multidisiplinære behandlingsopplegget. Økende alder i behandlingsgruppen synes å ha en positiv virkning, som har en konkav avtagende helning. Høyere inntekt synes å ha en negativ virkning, mens et økende antall måneder sykemeldt før pretest synes å ha en positiv effekt. Ingen av disse gruppevisse variasjonene i behandlingseffekter er skarpt bestemte programeffekter, og kan derfor ikke legges så altfor stor vekt på.

⁵¹ For EQ-5D: $0,037 - 0,036 = 0,001$

5 Oppsummering

Formålet med denne oppgaven har vært å analysere helseeffekten av rehabiliteringstiltaket Bergensprosjektet: "Tilbake på jobb", i et kostnadsnytte perspektiv. Hensikten med dette har vært å se på om det er mulig å bruke MAU-instrumenter som mål i effektevaluering av helsetiltak, og hvordan dette lar seg implementere i et praktisk eksempel. Det har gjennom oppgaven blitt gjort rede for de vanligste samfunnsøkonomiske lønnsomhetsanalysene for evaluering av helsetiltak. I denne sammenheng er det blitt lagt hovedvekt på CUA og denne metodens grunnleggende prinsipper for å måle livskvalitet, samt ulike metoder og MAU-instrumenter for måling av helse. Dette for å kunne påpeke styrke og svakheter ved denne metoden. Gjennom å overføre kliniske data fra Bergensprosjektet til MAU-instrument, har vi kartlagt gjennomsnittlig helserelatert livskvalitet for deltakerne i forsøket både før og etter tiltak. Resultatene kan brukes til å evaluere rehabiliteringstiltaket og evaluere MAU-instrumentene gjennom praktisk bruk. I tillegg er det gjennomført en regresjonsanalyse for å se om det foreligger en positiv behandlingseffekt av tiltaket, samt det er korrigert for frafallsseleksjon for å vurdere om dette påvirker resultatene. Vi skal her oppsummere noen av resultatene og erfaringene som er gjort underveis.

En av fordelene med CUA som ofte blir fremhevet, er at bruk av en kostnadseffektivitetsrate som kostnad per QALY, muliggjør en sammenligning av kostnadsnytteforholdet mellom ulike behandlinger og tiltak med forskjellige endepunkter og multiple utfall. En del av kritikken mot CUA går på at nytte på kardinalnivå er umulig eller svært vanskelig å måle når det gjelder helse. At det i tillegg foreligger et større antall ulike metoder og MAU-instrument, som kan brukes for å avdekke subjektive så vel som objektive preferanser for et helseutfall, medfører problemer i sammenlikningen av resultater mellom forskjellige tiltak, som dernest vil vanskeliggjøre rangering av tiltak. Dette fordi at resultatene blir til dels usammenlignbare grunnet forskjellene mellom MAU-instrumentene ved måling av helse, og det at verdiene er ilagt forskjellig mening fra instrument til instrument.

De overnevnte problemene gjør det vanskelig å sammenligne resultatene mellom instrumentene i analysen. Vi finner i kapittel 4 at gjennomsnittlig helserelatert livskvalitet varierer fra 0,516 i HUI3 til 0,845 i 15D, som er en differanse i helserelatert livskvalitet på 0,329. Dette er en differanse som de fleste vil karakterisere som betydelig når det gjelder mål

på helse. Et slikt resultat påpeker nødvendigheten av at en standardverdi for nyttemålingene settes, slik at et stort antall ulike metoder ikke undergraver resultatenes verdi og betydning.

Vi har gjennom analysen ikke klart å avdekke noen gjennomsnittlig positiv (eller negativ) signifikant helsegevinst av tiltaket i Bergensprosjektet: ”Tilbake på jobb”, eller for noen undergrupper med hensyn på de bakgrunnskjennetegnene som det ble kontrollert for. Ved korrigerende av frafallseleksjon fikk vi det interessante funnet at de to MAU-instrumentene som gir signifikante seleksjonsestimater, EQ-5D og 15D, er også de to instrumentene som vurderes til best å kunne dekke sykdomsbildet her. Instrumentene vurderes således best til å avspeile behandlingens påvirkning på de to gruppene. Det kan dernest tenkes at disse viser det mest riktige resultatet med hensyn til frafallseleksjon. Men siden ikke alle instrumentene viser dette resultatet i FML, har vi tolket dette som at datasettet muligens er plaget av frafallseleksjon som ligger i grenseland og påvirker utfallsmålene. Men påvirkningen er ikke stor nok til at vi kan si i hvilken retning dette påvirker utfallsmålene.

En kan selvfølgelig stille spørsmål om hvor valid resultatene er, siden modellen vår ikke finner noen ting i retning som kan tyde på en positiv behandlingseffekt. Det ble derfor utført en spesifikasjonstest av regresjonsmodellen i kapittel 4.2.3. Resultatene vi fant der kan tolkes i retning av at modellen som er brukt ikke nødvendigvis er den ”sanne” modellen, men at den heller ikke er total misledende. Dermed konkluderes det med at estimatene våre er til å stole på, og at disse gir et korrekt estimat av behandlingseffekten.

Funnene våre støtter oppunder tidligere funn som tyder på at behandlingsformene gir noenlunde lik behandlingseffekt (Strand (1998) og Bratberg et al. (2002)). Dette kan virke litt merkelig idet at behandlingsgruppen rapporterte bedre subjektiv helse, mindre smerte, osv. ved posttest, faktorer som MAU-instrumentene nettopp er ment å fange opp (Torrance et al. (1995)). Et interessant spørsmål blir således om MAU-instrumentene brukt i analysen er sensitive nok til å registrere små endringer i helsen til deltakerne i Bergensprosjektet. Sammenslåingen av alternativ i Fellespakken kan ha gjort at små endringer i svar i Fellespakken fra pre- til posttest ikke har blitt fanget opp av instrumentene, fordi forandringen blir ”skjult” i sammenslåingen. Et fellestrekk ved overføringen av de kliniske dataene til MAU-instrumentene, er at en som oftest må slå sammen nivå i Fellespakken for å få disse til å passe inn i MAU-instrumentene. Eksempelvis kan dette vises i EQ-5D for mobilitetsdimensjonen som består av tre nivå, mens alternativet fra Fellespakken består av 5

nivå. Det sier seg selv at en mister en del av informasjonsgrunnlaget med hensyn til endring i helse i fra alternativet fra Fellespakken, når en må slå sammen nivå for å få det hele til å passe inn i EQ-5D. Hvis en pasient går i fra alternativ 3 ”sterk smerte” til alternativ 2 ”noe smerte” i Fellespakken med hensyn til mobilitet, vil ikke dette bli fanget opp av EQ-5D siden disse to nivåene er slått sammen inn i nivå 2 ”noen problemer ved å gå” i EQ-5D. Dermed er dette kanskje en medvirkende årsak til vi finner en så liten endring i helsereelatert livskvalitet i fra pre- til posttest. EQ-5D og QWB har bare 3 svaralternativ i hver av dimensjonene sine, mens 15D og HUI3 har 5. Alternativene i Fellespakken er som oftest 5 eller 6. Ut i fra dette skulle en tro at 15D og HUI3 er sensitive nok. Men heller ingen av disse to instrumentene viser noen større endringer i helsereelatert livskvalitet fra pre- til posttest enn de andre. Det er rimelig å tro at HUI3 blir ganske usensitiv, siden det er gjort antagelser på 5 av de 8 dimensjonene. Dette medfører trolig at pasientenes reduserte dimensjonsmål forsvinner blant alle dimensjonene som ikke angår sykdommen, jamfør diskusjonen i kapittel 2.4.2. I så tilfelle virker det som om EQ-5D, QWS og HUI3 er uegnede instrumenter til å fange opp små endringer i helse blant deltakerne i Bergensprosjektet. Men en slik slutning er vanskelig fatte på et slik grunnlag. Et mulig steg videre her, er å se nærmere på endringer innenfor de forskjellige dimensjonene i instrumentene.

Studiet viser at det er fullt mulig å bruke MAU-instrumenter i effektevaluering av helsetiltak, selv om en må i dette tilfellet være forsiktig med tolkningen av resultatene. Å utføre en analyse med MAU-instrument retrospektivt som her, har den ulempen at siden en ikke får de direkte svarene fra pasientene, vil en heller ikke helt få de ”riktige” svarene. I tillegg kan det å utføre CUA retrospektivt medføre upålitelige analyser, fordi forutsetningene i livskvalitetsmålingene kan være ulik forutsetningene i de kliniske forsøkene. Det vil derfor være viktig i tilfeller med CUA å foreta undersøkelser av livskvalitet underveis i prosessen ved utprøving av nye tiltak og behandlinger.

Svakheten ved CUA er metodene for å måle livskvalitet, idet at forskjellige metoder for å måle livskvalitet gir ulike svar. Det som er klart er at CUA og dens metoder for måling av livskvalitet er under utvikling, og foreløpig er dette den beste måten å representere pasientenes preferanser. Nord (1999a) refererer til de som ”syntetiske indikatorer”. Termen er passende fordi den kaster lys over den sanne naturen til indikatorene, dvs. detaljerte konstruksjoner og ikke direkte observasjoner. Termen setter også søkelyset på muligheten for signifikante feil. Ikke desto mindre bør de inkluderes i økonomiske- og programevalueringer,

men med bevissthet om deres begrensninger. I sin nåværende form bidrar CUA uansett til å synliggjøre og systematisere et sett av ulike elementer både på kostnads- og gevinstsiden. Selv om resultatene i fra en CUA analyse ikke kan brukes alene i en beslutningssammenheng, bidrar CUA til at beslutningene til slutt fattes på et bredere informasjonsgrunnlag enn det man ellers ville hatt. Etter hvert som erfaringene og sammenligningsgrunnlaget blir større vil denne metoden sannsynligvis få større legitimitet. Foreløpig må CUA og dennes metode for å måle livskvalitet, sees på som et skritt på veien til noe som kan bli et viktig og nyttig verktøy ved evaluering av helsetiltak. Ettersom flere får øynene opp for potensialet for et slik mål vil nok flere være med på å utvikle denne metoden videre, for som Hurst og Forbes (1999) sier: ”det er bedre og måle imperfekt det som er viktig, enn å måle perfekt det som er uviktig”.

Referanser:

- Bratberg, E., Grasdal, A., og Risa, A.E. (2001): Økonomisk evaluering av rehabiliteringstiltak. I Askildsen, J.E. og Haug, K.Helse (red): Helse, økonomi og politikk, utfordringer for det norske helsevesenet. *Cappelen Akademiske Forlag*.
- Bratberg, E., Grasdal, A., og Risa, A.E. (2002): Evaluating social policy by experimental and noneksperimental methods. *Scandinavian Journal of Economics*, 104, s. 104-171.
- Bratberg, E., Hanssen, A., Risa, A.E., Vaage, K. og Aakvik, A. (1998): Samfunnsøkonomisk evaluering. *Bergensprosjektene 1993-1998. SEFOS*.
- Brooks, R. With the EuroQol Group (1996): EuroQol: The current state of play. *Health Policy* 37, pp. 53-72.
- Bomann-Larsen, P. (1998): Prosessevaluering. *Bergensprosjektene 1993-1998. SEFOS*.
- Carr, A.J. og Higginson, I.J. (2001): Measuring quality of life: Are quality of life measures patient centred? *British Medical Journal*, volume 322.
- Cox, D.R., Fitzpatrick, R., Fletcher, A.E., Gore, S.M., Spiegelhalter, D.J, og Jones, D.R. (1992): Quality of life assessment: Can we keep it simple? *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Volume 155, Issue 3, 353-393.
- Doland, P. og Roberts, J. (2002): To what extent can we explain time trade-offs values from other information about respondents? *Social Science and Medicine* 54, pp. 919-929.
- Doland, P., Gudex, C., Kind, P. og Williams, A. (1996): Valuing health states: A comparison of methods. *Journal of Health Economics*, 209-231.
- Drummond, M., Stoddart, G.L. og Torrance, G.V. (1997): Methods for the economic evaluation of health care programmes. *Oxford Medical Publications*.
- Fitzpatrick, R., Davey, C., Buxton, M.J, og Jones, D.R. (1998): Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment, Vol. 2: No. 14*.
- Feeney, D., Furlong, W., Boyle, M. and Torrance, G.W. (1995): Multi-Attribute Health Status Classification System; Health Utilities Index. *Pharmacoeconomics* 7 (6): 490-502.
- Folland, S., Goodman, A.C. og Stano, M. (2001): The economics of health and health care, 3 utg. *Prentice Hall*.
- Garrat, A., Schmidt, L., Mackintosh, A., Fitzpatrick, R. (2002): Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal*, volume 324.
- Grasdal, A. (2001a): The performance of sample selection estimators to control for attrition bias. *Health Economics*, 10, s. 385-398.

Grasdal, A. (2001b): The empirical importance of randomisation bias in the Bergen experiments. *Working Papers in Economics No.0201. Department of Economics University of Bergen.*

Greene, W.H. (2000): *Econometric analysis*, 4 utg. *Prentice Hall.*

Guzman, J., Esmail, K., Karjalainen, K., Malmivaara, A., Irvin, E. and Bombardier, C. (2001): Multidisciplinary rehabilitation for chronic low back pain: systematic review. *British Medical Journal*, volume 322.

Hansen, T. (2002): Multi Attribute Utility-instrument i økonomiske analyser av legemidler. Hovedfagsoppgave. *Institutt for økonomi Universitetet i Bergen.*

Heckman, J.J (1979): Sample selection bias as a specification error. *Econometrica* 47.

Heckman, J.J. og Hotz, V.J. (1989): Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American Statistical Association*, Volume 84, Issue 408, pp. 862-874.

Heckman, J.J. og Smith, J.A. (1995): Assessing the Case for Social Experiments. *The Journal of Economic Perspectives*. Volume 9, Issue 2.

Hawthorne, G. (2000): Functional status and health-related quality of life assessment in patients with rheumatoid arthritis. *Working Paper 104. Centre for Health Program Evaluation.*

Hawthorne, G., Richardson, J. og Day, N (1999): Using the assessment of quality of life (AQoL) version 1. *Technical report 12. Centre for Health Program Evaluation.*

Hawthorne, G., Richardson, J., Day, N og McNeil, H (2000): Life and death: Theoretical and practical issues in using utility instruments. *Working Paper 102. Centre for Health Program Evaluation.*

Hurst, N.P. and Forbes, J. (1999): Does economic evaluation have anything to offer the rheumatologist?. *Rheumatology*, 38: 2-5.

Håland Haldorsen, E.M. og Ursin, H. (1998): Medisinsk-psykologisk evaluering – resultater etter ett års oppfølging. *Bergensprosjektene 1993-1998. SEFOS.*

Håland Haldorsen, E.M., Brage, S., Strømme Johannesen, T., Tellens, G. og Ursin, H. (1998): Musculoskeletal pain: Concepts of disease, illness, and sickness certification in health professionals in Norway. *Bergensprosjektene 1993-1998. SEFOS.*

Håland Haldorsen, E.M., Grasdal, A.L., Skouen, J.S., Risa, A.E., Kronholm, K. og Ursin, H. (2002): Is there a right treatment for a particular patient group?. *Pain*, 95, 49-63.

Kaplan, R. (1989): Health outcome models for policy analysis. *Health Psychology*, Vol. 8, pp. 723-735.

Kaplan, R.M., Anderson, J.P., Ph.D., Wu, A.W., M.D., Mathews, W.M.C., MD. (1992): The Quality of Well-Being Scale: Applications in Arthritis. *MAPI Research Institute, Quality of Life Newsletter 2*.

Keeney, R. og Raiffa, H. (1976): Decisions with multiple objectives: preferences and value tradeoffs. *Wiley*.

Krabbe, P., Essink-Bot, M.L. og Bonsel, G. (1997): The comparability and reliability of five health-state valuation methods. *Social Science of Medicine, Vol. 45, No.11, pp. 1641-1652*.

Kronholm, K. (1998): Prosjektrapport 1993-96. *Bergensprosjektene 1993-1998. SEFOS*.

LaLonde, R.J. (1986): Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review, Volume 76, Issue 4, pp. 604-620*.

Moffett, J.K., Torgerson, T., Bell-Syer, S., Jackson, D., Llewlyn-Phillips, H., Farrin, A. og Barber, J. (1999): Randomised controlled trials of exercise for low back pain: clinical outcomes, costs and preferences. *British Medical Journal 319*.

Muldoon, M.F. (1998): What are quality of life measurements measuring? *British Medical Journal 316*.

Nord, E. (1999a): Cost-Value analysis in health care. *Cambridge University Press*.

Nord, E. (1999b): Adjusting health state utilities for use in economic evaluation. *MAPI Research Institute, Quality of Life Newsletter 23*.

Nord, E. og Wolfson, M. (1999c): Multi-attribute health state valuations: Ambiguities in meaning. *MAPI Research Institute, Quality of Life Newsletter 21*.

Nord, E., Arnesen, T., Menzel, P. og Pinto, J.L. (2001): Towards a more restricted use of the term 'Quality of life'. *MAPI Research Institute, Quality of Life Newsletter 26*.

Norheim, O.F. og Bringedal, B. (2001): Hvordan bør vi prioritere i helsetjenesten? I Askildsen, J.E. og Haug, K. Helse (red): Helse, økonomi og politikk, utfordringer for det norske helsevesenet. *Cappelen Akademiske Forlag*.

NOU 2000: 27. Et inkluderende arbeidsliv. *Sosial- og helsedepartemanget*.

NOU 1999: 20. Å vite eller ikke vite. *Sosial- og helsedepartemanget*.

NOU 1998: 16. Nyttekostnadsanalyser. *Finansdepartementet*.

NOU 1997: 27. Nytte-kostnadsanalyser. *Finansdepartementet*.

NOU 1997: 18. Prioritering på ny. *Sosial- og helsedepartemanget*.

Rawls, J. (1974): Some reasons for the maximin criterion. *The American Economic Review, Volume 64, Issue 2, 141-146*.

- Richardson, J. (1997): Critique and some recent contributions to the theory of cost utility analysis. *Working Paper 77. Centre for Health Program Evaluation.*
- Roemer, M.I. (1961): Bed supply and hospital utilization: unnatural. *Hospitals, Vol 35.*
- Roland, M. og Torgerson (1998): Understanding controlled trials: What outcomes should be measured? *British Medical Journal, volume 317.*
- Rosen, Harvey S. (2002): Public finance, 6. utgave. *McGraw-Hill Higher Education.*
- O'Connor (1993): Issues in the measurement of health-related quality of life. *Working Paper 30. Centre for Health Program Evaluation.*
- Sanders, C., Egger, M., Donovan, J. Tallon, D. og Frankel, S. (1998): Reporting on quality of life in randomised controlled trials: bibliographic study. *British Medical Journal, volume 317.*
- Sibbald, B. og Roland, M. (1998): Understanding controlled trials: Why are randomised controlled trials important?. *British Medical Journal, volume 316.*
- Sintonen, H. (1994): 15-D Measure of Health Related Quality of Life: Reliability, Validity and Sensitivity of its Health State Descriptive System. *Working Paper 41. Centre for Health Program Evaluation.*
- Sintonen, H. (1995): 15-D Measure of Health Related Quality of Life: Feasibility, Reliability and Validity of its Valuation System. *Working Paper 42. Centre for Health Program Evaluation.*
- Stata Reference Manual (1999): Release 6, Volume 2 H-O. *Stata Press College Station Texas.*
- Strand, L.I. (1998): Fysioterapeutisk evaluering – resultater etter ett års oppfølging. *Bergensprosjektene 1993-1998. SEFOS.*
- Suarez-Almazor, M.E., Kendall, C., Johnson, J.A., Skeith, K. and Vincent, D. (2000): Use of health status measure in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. *Reumatology 39: 783-790.*
- Torgerson, D. and Raftery, J. (1999): Measuring outcomes in economic evaluations. *British Medical Journal, volume 318.*
- Torrance, G.W., Furlong, W., Feeny, D and Boyle, M. (1995): Multi-Attribute Preference Functions; Health Utilities Index. *PharmacoEconomics 7 (6): 503-520.*
- Torrance, G.W., Thomas, W.H. og Sackett, D.L. (1972): A utility maximization model of for evaluation of health care programmes. *Health Service Research, Vol. 7, pp. 118-133.*
- Ubel, P.A. (1999): How stable are people's preferences for giving priority to severely ill patients? *Social Science and Medicine 49, pp. 895-903.*
- Von Neuman, J. and Morgensten, O. (1953): Theory of games and economic behaviour. *3.utgave. Wiley. New York*

Wagstaff, A. (1991): Health care: QALY's and the equity-efficiency tradeoff. *Journal of Health Economics* 10.

Williams, A. (1992): Cost-effectiveness analysis: is it ethical? *Journal of Medical Ethics* 18.

Aakvik, A. (2001): Samfunnsøkonomisk evaluering av legemidler. I Askildsen, J.E. og Haug, K.Helse (red): Helse, økonomi og politikk, utfordringer for det norske helsevesenet. *Cappelen Akademiske Forlag*.

Appendiks:

Dette appendikset viser hvordan dataene i Bergensprosjektet: "Tilbake på jobb", er ført inn i de to MAU-instrumentene QWB og HUI3. For MAU-instrumentet EQ5D henvises det til kapittel 4. "Mappingen" i 15D er utelatt siden dette blir for omfattende å ta med her. For flere detaljer angående MAU-instrumentene og Fellespakken utover det som forklares her, henviser det til henholdsvis Drummond et al. (1997) og Håland Haldorsen og Ursin (1998).

(A1) "Mapping" i QWB

QWB består av de tre dimensjonene, *mobilitet*, *fysisk aktivitet* og *sosial aktivitet*, samt 25 *symptom-/ problemkombinasjoner*. Følgende omforminger er blitt gjort for å få dataene i Fellespakken til å passe inn i QWB.

1) Mobilitet (Mob)

Tabell A1 viser mobilitetsdimensjonen i QWB og dens tre nivåer. Dimensjonen tar for seg hvordan det er for en person å kunne ta offentlig transport og/eller kjøre bil. Eneste alternativet fra Fellespakken er i fra spørsmål 29, tabell 5; "hvor mye smerte er det gjennomsnittlig forbundet med det å kjøre bil?". Tabell A2 og A3 viser henholdsvis "mapping" alternativet fra Fellespakken, og nivå sammenslåingen mellom den og mobilitetsdimensjonen. Hvis et individ svarer alternativ 2; "Noe smerter" i Fellespakken, vil han/hun få nivå 2 på mobilitetsskalaen i QWB med den medfølgende (unytte) vekten: Mob = -0,062.

Mobilitet:

1. Ingen begrensninger av helse grunner. Vekt: -0,000
2. Kjører ikke bil, helse relatert; Kjører ikke bil som vanlig for alderen (yngre enn 15 år), helse relatert, og/eller bruker ikke offentlig transport, helse relatert; eller vil bruke mer hjelp enn normalt for alderen for å kunne bruke offentlig transport. Vekt: -0,062
3. På sykehus, helse relatert. Vekt: -0,090

Tabell A1 Mobilitetsdimensjonen i QWB

Hvor mye smerte er det forbundet med å kjøre bil?	1. Ingen smerte.	2. Noe smerte.	3. Sterk smerte.	4. Nærmest utålelig smerte.	5. Utålelig smerte, uforenelig med aktivitet.
---	------------------	----------------	------------------	-----------------------------	---

Tabell A2 Utdrag fra tabell 5 i Fellespakken

	Nivå 1	Nivå 2		Nivå 3	
Alternativ QWB	1. Ingen...	2. Kjører ikke...		3. På sykehus...	
Alternativ Fellespakken	1. Ingen smerte	2. Noe smerte	3. Sterk smerte	4. Nærmest utålelig smerte	5. Utålelig smerte, uforenelig med aktivitet

Tabell A3 Nivåsammenligningen mellom QWB og Fellespakken i mobilitetsdimensjonen

2) Fysisk aktivitet (Pac)

Fysisk aktivitetsdimensjonen er vist i tabell A4. Dimensjonen tar for seg hvor godt et individ mestrer dagligdagse ting som å gå, bøye seg over noe, løfte ting osv. Fra spørsmål 29, tabell 5 i Fellespakken, er det tatt et utdrag av forskjellige aktiviteter vist i tabell A5. Spørsmålet her tar for seg hvor ”*mye smerte det er forbundet med de forskjellige aktivitetene*”. Snittet i tabell A5 avgjør nivået i fysisk aktivitetsdimensjonen. Nivåsammenligningen er som for mobilitetsdimensjonen, og er vist i tabell A6 sammen med nivåintervallene. I eksempelet i tabell A5, vil individet få regnet ut snittet $2 + 3 + 2 + 4 + 4 = 45/5 = 3$. Dette vil gi vekten: $Pac = -0,060$ i henhold til nivåintervallet i tabell

Fysisk aktivitet:

1. Ingen begrensninger av helse grunner. Vekt: -0,000
2. I rullestol, beveger eller kontrollerer den ved uten hjelp av andre; eller har problemer eller prøvde ikke å løfte, hoppe, bøye seg over, eller bruke trapper eller stige..... Vekt: -0,060
3. I rullestol, beveger eller kontrollerer ikke den ved uten hjelp av noen andre, eller i seng, stol eller sofa hele dagen, helse relatert. Vekt: -0,077

Tabell A4 Fysisk aktivitetsdimensjonen i QWB

	1. Ingen smerte.	2. Noe smerte.	3. Sterk smerte.	4. Nærmest utålelig smerte.	5. Utålelig smerte, uforenelig med aktivitet.
1. Gå		X			
3. Bøye seg			X		
8. Gå i trapp		X			
9. Løfte				X	
10. Bære				X	

Tabell A5 Utdrag fra tabell 5 i Fellespakken

	Nivå 1	Nivå 2		Nivå 3	
Alternativ QWB	1. Ingen...	2. Kjører ikke...		3. På sykehus...	
Alternativ Fellespakken	1. Ingen smerte	2. Noe smerte	3. Sterk smerte	4. Nærmest utålelig smerte	5. Utålelig smerte, uforenelig med aktivitet
Intervall	1 - 1,5	1,51 - 3,5		3,51 - 5	

Tabell A6 Nivåsammenligning og nivåintervallene i fysisk aktivitetsdimensjonen i QWB

3) Sosial aktivitet (Sac)

Den sosiale aktivitetsdimensjonen er vist i tabell A7. Dimensjonen tar for seg hvor godt en person klarer seg i sosiale sammenhenger, og det å ta vare på seg selv. Her er det brukt et snitt av de to spørsmålene "Min evne til å ta vare på meg selv" og "Min sosiale situasjon", for å bestemme nivået i denne dimensjonen. Disse to er hentet fra spørsmål 30, tabell 6 i Fellespakken, og er vist i tabell A8. En person kan være sosialt passiv, men allikevel klare å ta vare på seg selv. Dermed sees det som rimelig å ha med disse to. I eksempelet i tabell A8, vil individet få snittet $3+4 = 7/2 = 3,5$. Dette vil gi vekten: $Sac = -0,061$ i henhold til nivåintervallet i tabell A9.

Sosial aktivitet:

1. Ingen begrensninger av helse grunner. Vekt: -0,000
2. Begrenset i (for eksempel godt humør, humor) rolle aktivitet. Vekt:-0,061
3. Begrenset i det meste av (primær) rolle aktivitet, helse relatert Vekt:-0,061
4. Utfører ikke noen stor rolle aktivitet, helse relatert, men utfører aktiviteter som medfører å ta vare på seg selv. Vekt:-0,061
5. Utfører ikke noen stor rolle aktivitet, helse relatert, og utførte ikke eller trenger mer hjelp enn vanlig i å utføre en eller flere aktiviteter som medfører å ta vare på seg selv, helse relatert. Vekt:-0,106

Tabell A7 Sosial aktivitetsdimensjonen i QWB

	1. Svært utilfredstillende	2. Utilfredsstillende	3. Verken godt eller dårlig	4. Tilfreds	5. Svært tilfreds
2.Min evne til å klare meg selv er				X	
3.Min sosiale situasjon er			X		

Tabell A8 Utdrag fra tabell 6 i Fellespakke

	Nivå 1	Nivå 2		Nivå 3	
Alternativ QWB	1. Ingen...	2, 3 og 4. Begrensninger		5. Utfører ingen ...	
Alternativ Fellespakken	5.Svært tilfreds	4.Tilfreds	3. Verken godt eller dårlig	2. Utilfreds stillende	1. Svært utilfreds stillende
Intervall	5 - 4,51	4,5 - 1,51		1,51 - 1	

Tabell A9 Nivåsammenslåing og nivåintervallene i sosial aktivitetsdimensjonen i QWB

4) Symptom-/ problemkombinasjoner (Cpx)

Videre så inneholder QWB en liste over forskjellige symptom-/ problemkombinasjoner som tap av bevissthet, brannså, blødninger osv. med sine respektive (unytte) vekter. I denne sammenheng ble symptom-/ problemkombinasjon nummer 7: ”Smerte, stivhet, svakhet, lammelse, eller andre former for ubehag i bryst, mage, side, nakke, rygg, hofte eller hender, føtter, armer eller legger” valgt. Denne har vekt: $Cpx = -0,299$. Alle individene i undersøkelsen som har en verdi ulik null i en av dimensjonene, har fått tilskrevet denne symptom-/ problemkombinasjonen. Dette er gjort med tanke på at det er ganske rimelig og tro at disse plagene forholdsvis godt beskriver hvordan personer med belastningslidelser kan ha det.

5) Utregning av helserelatert livskvalitet

Vektene i hver dimensjon i QWB regnes sammen med formel 2.15 som er gjengitt her:

$$U(x)^{QWB} = 1 - [(CPXwt) + (MOBwt) + (PACwt) + (SACwt)] \quad (A1)$$

I eksempelet her vil personen få utregnet følgende helserelatert livskvalitet:

$$U(x)^{QWB} = 1 - [(0,062) + (0,060) + (0,061) + (0,299)] = 0,518 \quad (A2)$$

(A2) “Mapping” i HUI3

HUI3 består av 8 dimensjoner, *syn, hørsel, språk, mobilitet, førlighet, følelser/emosjonalitet, erkjennelse/kognisjon og smerte*. Følgende er blitt gjort for å få dataene i Fellespakken til å passe inn i HUI3.

1) Syn:

Det ble ikke samlet inn data fra deltakerne angående deres syn, samt de med alvorlige problemer i denne dimensjonen ble ekskludert i fra forsøket (Kronholm (1998)). Derfor er nivået i denne dimensjonen satt lik 1 for alle, med den antagelsen at alle deltakerne i forsøket har perfekt syn. Det henviser ellers til kapittel 3.2.3 for kommentarer til dette.

2) Hørsel:

Ser på hvor godt en person hører. Samme antagelser er gjort her som under 1).

3) Språk:

Ser på hvor godt en person kan kommunisere. Samme antagelser er gjort her som under 1).

4) Mobilitet:

I denne dimensjonen ser instrumentet på hvordan det er for en person å bevege seg rundt uten hjelp av andre personer eller andre hjelpemidler. Dimensjonen er vist i tabell A10. Denne dimensjonen er forholdsvis lik mellom MAU-instrumentene. For at bevegelighets-/mobilitetsdimensjonene mellom instrumentene skulle ha noenlunde samsvar, ble det valgt å inkludere de samme alternativene som ble brukt i mobilitetsdimensjonen i 15D og EQ5D. Dette var for å unngå at bevegelighets-/mobilitetsdimensjonen ble bestemt av ett alternativ i et instrument, og så av flere i et annet. Et utvalg av forskjellige aktiviteter, vist i tabell A11, ble valgt. Disse er hentet fra spørsmål 29, tabell 5 i Fellespakken. Spørsmålet her tar for seg ”*hvor mye smerte er det gjennomsnittlig forbundet med de beskrevne aktivitetene*”. Gjennomsnittet i tabell A11 bestemmer nivået i bevegelighetsdimensjonen. I nivåsammeslåingen ble det valgt å droppe alternativ 6 ”*Kan ikke gå i det hele tatt*” i mobilitetsdimensjonen, grunnet det at slike personer ble ekskludert i fra forsøket. I eksempelet i tabell A11 vil personen få $2 + 3 + 2 + 4 + 4 + 3 = 18/6 = 3$ i snitt. I henhold til tabell A12 som viser intervallinndelingen og nivåintervallene, vil personen dermed få tilskrevet nivå 3 og nyttescoren 0,86 i mobilitetsdimensjonen.

Mobilitet:

1. Klarer å gå rundt i nabolaget uten problemer, og uten hjelpemidler.
2. Klarer med små problemer å gå rundt i nabolaget, men trenger ingen hjelpemidler eller hjelp av en annen person.
3. Klarer å gå rundt i nabolaget med hjelpemidler, men trenger ikke hjelp av en annen person.
4. Kan bare gå korte avstander med hjelpemidler, og trenger en rullestol for å kunne komme seg rundt i nabolaget.
5. Klarer ikke å gå alene, selv med hjelpemidler. Klarer å gå korte avstander med ved hjelp av en annen person, og trenger en rullestol for å komme seg rundt i nabolaget.
6. Kan ikke gå i det hele tatt.

Tabell A10. Mobilitetsdimensjonen i HUI3

	1. Ingen smerte.	2. Noe smerte.	3. Sterk smerte.	4. Nærmest utålelig smerte.	5. Utålelig smerte, uforenelig med aktivitet.
1. Gå		X			
2. Sitte			X		
3. Bøye seg		X			
7. Løpe/jogge				X	
8. Gå i trapp				X	
10. Sportsakt			X		

Tabell A11 Utdrag av tabell 5 Fellespakken

	Nivå 1	Nivå 2	Nivå 3	Nivå 4	Nivå 5
Nivå HUI3	1.Klarer å...	2.Klarer å...	3.Klarer å...	4.Kan bare...	5.Klarer ikke.
Alternativ Fellespakken	1. Ingen smerte.	2. Noe smerte.	3. Sterk smerte.	4. Nærmest utålelig smerte.	5. Utålelig smerte, uforenelig med aktivitet.
Nivåintervall	1 – 1,5	1,51 – 2,5	2,51 – 3,51	3,51 – 4,5	4,51 - 5

Tabell A12 Nivåsammenslåing og nivåintervallene i mobilitetsdimensjonen i HUI3

5) Førlighet:

Ser på hvor godt en person kan bruke hendene sine. Samme antagelse er gjort som under 1).

6) Følelser/emosjonalitet:

Her ser HUI3 på en persons livsinnstilling, om personen er normalt glad og interessert i livet. Dette er vist i tabell A13. Fra Fellespakken ble spørsmål 35; "Hvilket trinn passer til livet ditt nå?" valgt. Denne er vist i tabell A14. Spørsmålet passer godt inn med det de spør om i HUI3,

fordi begge steder er de ute etter nøyaktig det samme. Sammenslåingen er enkel og vist i tabell A15. Hvis en person setter livet sitt til trinn 5 i Fellespakken, vil personen få nivå 3 ”*litt trist*” i HUI3 med nyttescoren 0,85.

Følelser/emosjonalitet:

1. Glad og interessert i livet.
2. Noenlunde glad.
3. Litt trist.
4. Veldig trist.
5. Så trist at livet ikke er verd å leve.

Tabell A13 Følelser-/emosjonalitetsdimensjonen i HUI3

Hvilket trinn passer for ditt liv nå?

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

Tabell A14 Spørsmål 35 i Fellespakken

	Nivå 1		Nivå 2		Nivå 3		Nivå 4		Nivå 5	
Alternativ Fellespakken	1	2	3	4	5	6	7	8	9	10
Alternativ HUI3	1		2		3		4		5	

Tabell A15 Nivåsammenslåingen mellom Fellespakken og HUI3 i følelser-/emosjonalitetsdimensjonen

7) Erkjennelse/kognisjon:

HUI3 ser her på hvor godt en person klarer å huske ting, tenke og løse hverdagsproblemer. Samme antagelse som under 1) er gjort her.

8) Smerte:

Denne dimensjonen, vist i tabell A16, ser her på smerte versus aktivitet. Fra Fellespakken ble det valgt å ta utgangspunkt i noen alminnelige helseproblem forbundet med muskel-/skjelettsykdommer. Disse er oppført i tabell A17. Gjennomsnittet der bestemmer nivået i denne dimensjonen. Sammenslåingen av nivå og nivåintervall er vist i tabell A18. Ved sammenslåingen av nivå, ble alternativ/nivå 2 i HUI3 droppet for å få dataene i Fellespakken til å passe inn. Dette ble gjort fordi alternativ 2 og 3 er ganske like. I tillegg har de minst differanse i nyttescore som brukes til å regne ut den helserelevante livskvaliteten. I eksempelet i tabell A17 vil personen få $1 + 2 + 2 + 0 + 1 + 1 = 7/6 = 1,17$ i snitt. I henhold til tabell A18 som viser intervallinndelingen og nivåintervallene, vil personen dermed få tilskrevet nivå 2 og nyttescoren 0,90 i smertedimensjonen.

Smerte:

1. Fri for smerte og ubehag.
2. Mild til moderat smerte som ikke hindrer noen aktiviteter.
3. Moderate smerter som hindrer noen få aktiviteter.
4. Moderat til alvorlige smerter som hindrer en del aktiviteter.
5. Alvorlige smerter som hindrer de fleste aktiviteter.

Tabell A16 Smertedimensjonen i HUI3

Nedenfor nevnes noen alminnelige helseplager	4. Ikke plaget. 5. Litt plaget. 6. En del plaget. 7. Alvorlig plaget.
4. Hodepine	1
5. Nakkesmerter	2
6. Smerter øverst i ryggen	2
7. Smerter i korsryggen	0
8. Smerter i armer	1
9. Smerter i skuldre	1

Tabell A17 Utdrag fra Helseproblem siste 30 døgn i Fellespakken

	Nivå 1	Nivå 2	Nivå 3	Nivå 4
Alternativ Fellespakken	0. Ikke plaget	1. Litt plaget	2. En del plaget	3. Alvorlig plaget
Alternativ HUI 3	1. Fri for...	3. Moderate...	4. Moderat til...	5. Alvorlige...
Intervall	0 – 0,5	0,51 – 1,5	1,51 – 2,5	2,51 - 3

Tabell A18 Nivåsammenslåing og nivåintervallene i smertedimensjonen i HUI3**5) Utregning av helsereelatert livskvalitet:**

Nytten av helsetilstandene i eksempelet regnes sammen med formel (2.20), som er gjengitt her. For flere detaljer angående hvordan dette ser ut og er gjort henvises det til Feeney et al (1995) og Torrance et al. (1995).

$$U(x)^{HUI3} = 1,371 * (1 * 1 * 1 * 0,86 * 1 * 0,85 * 1 * 0,90) - 0,371 = 0,53 \quad (A3)$$

Personen i eksempelet her vil få utregnet en helsereelatert livskvalitet på 0,53.

(A3) Bakgrunnsdata i hver MAU-instrument

Variabler	Behandlingsgruppen (n=297)			Kontrollgruppen (n= 136)		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
Alder	43.066	10.795	20 / 66	42.919	11.008	20 / 67
Kjønn	0,387	0,488	0 / 1	0,375	0,486	0 / 1
Inntekt*	1,881	0,7549	0,217/7,479	1,835	0,727	0,064/4,877
Mndsyk	3,417	2,042	1 / 20	3,544	2,164	1 / 15
Bergen	0,872	0,335	0 / 1	0,875	0,332	0 / 1

Tabell A19 Bakgrunnsdata for EQ-5D ved pretest (*Inntekt/100.000)

Variabler	Behandlingsgruppen (n=249)			Kontrollgruppen (n= 50)		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
Alder	44,056	10,167	24 / 65	42,482	10,150	21 / 64
Kjønn	0,337	0,474	0 / 1	0,4	0,493	0 / 1
Inntekt*	1,864	0,758	0,217/7,480	1,844	0,667	0,482/4,877
Mndsyk	3,533	2,167	1 / 20	3,259	1,816	1 / 12
Bergen	0,904	0,296	0 / 1	0,941	0,237	0 / 1

Tabell A20 Bakgrunnsdata for EQ-5D ved posttest (*Inntekt/100.000)

Variabler	Behandlingsgruppen (n=284)			Kontrollgruppen (n= 134)		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
Alder	42,544	10,598	20 / 66	42,351	10,424	20 / 66
Kjønn	0,430	0,496	0 / 1	0,403	0,492	0 / 1
Inntekt*	1,907	0,758	0,217/7,480	1,90	0,784	0,482/4,877
Mndsyk	3,431	2,021	1 / 20	3,216	1,793	1 / 12
Bergen	0,866	0,341	0 / 1	0,858	0,350	0 / 1

Tabell A21 Bakgrunnsdata for QWB ved pretest (*Inntekt/100.000)

Variabler	Behandlingsgruppen (n=239)			Kontrollgruppen (n= 80)		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
Alder	43,678	9,986	24 / 65	41,713	10,132	21 / 61
Kjønn	0,347	0,477	0 / 1	0,425	0,498	0 / 1
Inntekt*	1,851	0,752	0,217/7,480	1,863	0,691	0,482/4,877
Mndsyk	3,497	2,157	1 / 20	3,333	1,899	1 / 12
Bergen	0,904	0,296	0 / 1	0,925	0,265	0 / 1

Tabell A22 Bakgrunnsdata for QWB ved posttest (*Inntekt/100.000)

Variabler	Behandlingsgruppen (n=212)			Kontrollgruppen (n= 94)		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
Alder	42,729	10952	20 / 66	42,234	10,772	20 / 67
Kjønn	0,396	0,490	0 / 1	0,340	0,476	0 / 1
Inntekt*	1,969	0,777	0,433/7,480	1,819	0,649	0,064/3,539
Mndsyk	3,379	2,213	1 / 20	3,372	2,187	1 / 15
Bergen	0,882	0,323	0 / 1	0,872	0,336	0 / 1

Tabell A23 Bakgrunnsdata for 15D ved pretest (*Inntekt/100.000)

Variabler	Behandlingsgruppen (n=226)			Kontrollgruppen (n= 78)		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
Alder	43,425	9,958	24 / 65	41,436	10,114	21 / 64
Kjønn	0,341	0,475	0 / 1	0,397	0,493	0 / 1
Inntekt*	1,848	0,770	0,217/7,480	1,8233	0,576	0,482/3,593
Mndsyk	3,561	2,229	1 / 20	3,507	2,146	1 / 12
Bergen	0,894	0,309	0 / 1	0,962	0,194	0 / 1

Tabell A24 Bakgrunnsdata for 15D ved posttest (*Inntekt/100.000)

Variabler	Behandlingsgruppen (n=298)			Kontrollgruppen (n= 144)		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
Alder	42,999	10,674	20 / 66	42,715	10,638	20 / 67
Kjønn	0,379	0,486	0 / 1	0,382	0,486	0 / 1
Inntekt*	1,912	0,732	0,414/7,480	1,861	0,753	0,064/4,877
Mndsyk	3,405	2,029	1 / 20	3,188	1,830	1 / 12
Bergen	0,876	0,330	0 / 1	0,875	0,332	0 / 1

Tabell A25 Bakgrunnsdata for HUI3 ved pretest (*Inntekt/100.000)

Variabler	Behandlingsgruppen (n=255)			Kontrollgruppen (n= 92)		
	Gj.sn.	St.avvik	Min / max	Gj.sn.	St.avvik	Min / max
Alder	43,878	10,018	24 / 65	42,804	10,254	21 / 64
Kjønn	0,329	0,471	0 / 1	0,391	0,491	0 / 1
Inntekt*	1,857	0,753	0,217/7,480	1,811	0,679	0,482/4,877
Mndsyk	3,525	2,136	1 / 20	3,359	1,983	1 / 12
Bergen	0,906	0,293	0 / 1	0,946	0,228	0 / 1

Tabell A26 Bakgrunnsdata for HUI3 ved posttest (*Inntekt/100.000)

