

Vurdering af generiske livskvalitetsinstrumenter
- deres kvaliteter i forbindelse med
anvendelse på demensramte

Charlotte Horsted

Health Economics Papers

2004:5

Indhold

1 Baggrund	5
1.1 Formål	7
1.2 Metode og afgrænsning.....	8
2 Patienters vs. proxy respondenters vurdering af livskvalitet	9
2.1 Vurdering af sammenfald mellem patienter og proxyvurderinger.....	10
2.2 Empiriske resultater	11
2.2.1 Overensstemmelsesfaktorer	13
3 Livskvalitetsinstrumenter	15
3.1 Fordele og ulemper ved hhv. generiske og sygdomsspecifikke instrumenter	15
3.2 Indholdsvaliditeten i generiske instrumenter i relation til demensramtes vurdering af relevante attributter	16
4 Generiske instrumenter	20
4.1 Reliabilitet og validitet.....	25
4.1.1 Reliabilitet.....	25
4.1.2 Validitet.....	26
5 Gennemgang af de generiske instrumenter	28
5.1 Duke Health Profile (DUKE).....	28
5.2 EuroQol (EQ-5D).....	32
5.3 Health Utilities Index (HUI)	36
5.4 Sickness Impact Profile (SIP)	40
5.5 15 Dimensioner (15D)	46
5.6 Short Form – 36 (SF-36).....	47
5.8 Quality of Well-Being (QWB).....	56
6 Sammenfatning	61
7 Referenceliste	64
Bilag 1. Oversigt over demensstudier	69

Forord

Projektet ”Metoder og instrumenter til effektvurdering af indsatser over for demente” er finansieret af Servicestyrelsen (Styrelsen for social service, socialministeriet) og er udført i et samarbejde mellem CAST (Center for Anvendt Sundhedstjenesteforskning og Teknologivurdering) og Forskningsenheden for Sundhedsøkonomi, IST (Institut for Sundhedstjenesteforskning), ved Syddansk Universitet, Odense.

Der eksisterer i dag et stort antal metoder og instrumenter med det angivne formål at kunne vurdere og/eller måle effekten af en indsats over for demensramte. Formålet med projektet har været at give potentielle brugere en oversigt derover og i tilknytning hertil at beskrive instrumenternes egenskaber, så det bliver muligt at bedømme instrumenternes egnethed til et givet formål.

Projektet afrapporteres ved 4 publikationer, som alle er tilgængelige på: <http://www.sam.sdu.dk/healtheco/publications/workingpapers.htm>

1) Charlotte Horsted og Terkel Christiansen. Metoder og instrumenter til effektvurdering af indsatsen over for demente – en oversigt. Syddansk Universitet: Health Economics Papers 2004:2

2) Charlotte Horsted og Terkel Christiansen. Metoder og instrumenter til effektvurdering af indsatsen over for demente – psykometriske begreber. Syddansk Universitet: Health Economics Papers 2004:3

3) Charlotte Horsted og Terkel Christiansen. Metoder og instrumenter til effektvurdering af indsatsen over for demente – gennemgang og beskrivelse af instrumenterne. Syddansk Universitet: Health Economics Papers 2004:4

4) Charlotte Horsted. Vurdering af generiske livskvalitetsinstrumenter – deres kvaliteter i forbindelse med anvendelse på demensramte. Syddansk Universitet: Health Economics Papers 2004:5.

Den førstnævnte publikation indeholder en kortfattet oversigt over de metoder og instrumenter, der mere uddybende er gennemgået i den tredje publikation. Formålet med rapporten er at give læseren et hurtigt overblik over instrumenternes egenskaber – dvs. anvendelse, antal items, håndtering, håndteringstid samt vurdering af instrumenternes reliabilitet og validitet.

Den anden publikation indeholder en beskrivelse af nogle af de metoder og kriterier, som anvendes ved kvalitetsvurdering af instrumenter. En vis forståelse herfor kan være væsentlig for læsning af den tredje publikation. Den indeholder desuden en ordliste over de mest anvendte begreber i rapporterne.

Den tredje publikation er en bilagsrapport til den første. Rapporten indeholder en gennemgang og beskrivelse af de udvalgte ca. 70 instrumenter. Fokus er på instrumenternes formål, deres begrebslige grundlag og psykometriske egenskaber (validitet, reliabilitet samt følsomhed over for ændring).

Den fjerde publikation indeholder en beskrivelse af de mest almindelige generiske livskvalitetsinstrumenters egenskaber i forbindelse med anvendelse på demensramte individer. Dette drejer sig om instrumenterne DUKE, EQ5D, HUI, SIP, 15D, SF-36, QWB.

En stor tak til overlæge Kirsten Abelskov, Gerontopsykiatrisk afdeling i Århus Amt og professor, dr. med. Per Bech, Stressklinikken, Hillerød Sygehus, for deres kommentarer til rapporterne. Ligeledes tak til Charlotte Bruun Pedersen og Marie Holmgaard Kristiansen for korrekturlæsning og opsætning. Evt. resterende fejl og mangler er alene forfatterens.

Forskningsassistent, Charlotte Horsted
Professor Terkel Christiansen
Syddansk Universitet

1 Baggrund

I 1995 definerede WHO livskvalitet (QoL) som: ” *individuals’ perceptions of their position in life in the context of the culture and value system in which they live, and in relationship to their goals, expectations, and standards*”(Kuyken et al. 1403-09). Denne definition antager, at individerne har den intellektuelle kapacitet til at foretage komplekse subjektive vurderinger om eget liv. I litteraturen er der ligeledes bred enighed om, at livskvalitet er et multidimensionalt begreb, der indeholder fysisk, psykologisk og social velvære, ligesom der er bred enighed om, at livskvalitet er et subjektivt fænomen (Selai and Trimble 101-11).

Udgangspunktet i forbindelse med vurdering af livskvalitet er, at individer/patienter er den bedste informationskilde til vurderingen af egen livskvalitet. Det er dog ikke altid muligt at opnå de ønskede informationer – specielt i situationer, hvor patienterne er meget syge, skrøbelige, meget gamle eller unge, emotionelt forstyrrede, kognitivt eller mentalt svækkede. Især kognitiv svækkelse udgør et stort problem i forbindelse med vurderingen af sundhedsstatus eller livskvaliteten hos demente patienter. Som (Coucill et al. 760-71) skriver: “*Due to such cognitive dysfunction, patients with dementia may have a loss of insight and capacity for self-observation, and forget recent experiences and feelings...may be unable to recognize disabilities...may lack insight into their illness and psychological mechanisms, such as denial...*”.

Mange har sat spørgsmålstegn ved, om man skal/kan tage højde for demensramtes livskvalitet, idet de ofte ikke selv er i stand til at vurdere/give udtryk for denne. Dette aspekt af sundhedstilstanden har derfor ofte tidligere været udeladt i forbindelse med effektvurderinger af interventioner rettet mod demensramte. Livskvalitet er dog en essentiel komponent i den overordnede vurdering af sundhed og anvendes til at reflektere patientens perspektiv i studier af medicinske

behandlings effekt og udfald. Ligeledes kan livskvalitetsvurderinger anvendes til evaluering af den omsorg/pleje, patienter tilbydes. Da forbedring af livskvalitet/forsinkelse af overgangen til dårligere tilstande er et af de primære mål i forbindelse med omsorgen af demensramte, og da der ikke eksisterer nogen kur til helbredelse af demenssygdommene, er det især væsentligt, at man kan vurdere den omsorg, patienterne tilbydes, for som Brod et al. siger: *”For individuals with dementia... the ability of health care providers to intervene and affect QoL is far greater than to affect the course of the disease”* (Brod et al. 25-35).

Almindeligvis er overvejelser vedrørende demensramtes livskvalitet en væsentlig faktor i forbindelse med:

- Evaluering af nye behandlings- og aflastningsprogrammer
- Effekttestning af nye medicinprodukter
- Udvikling af kliniske guidelines

Effekterne af intellektuel forringelse på patienternes fatteevne, evne til at udtrykke forventninger og ønsker, samt sprog, har ført mange forskere til den konklusion, at demensramte ikke selv er i stand til at vurdere egen livskvalitet (Ankri et al. 1055-63) En indirekte tilgang til vurderingen af livskvalitet kan derfor tages i brug, nemlig at anvende proxy respondenter som alternativ informationskilde. En proxy kan i dette tilfælde enten være ægtefælle, pårørende, ven og/eller en sundhedsyder, fx læge, sygeplejerske, sygehjælper mv. Anvendelsen af proxyvurderinger bygger på antagelsen om, at proxy'en på et givent tidspunkt kan vurdere patientens værdier og præferencer (von Essen 229-34) – ligeledes antages, at proxy'en kan tage højde for, at patientens livssyn, ønsker og behov ændres over tid, hvilket har indflydelse på vurderingen af livskvaliteten.

Brugen af proxy giver potentielt en løsning på informationsproblematikken. Det forudsætter dog, at proxyinformationen stemmer overens med den information, man ville have opnået fra patienten selv. Flere forskere er kritiske overfor

validiteten af denne tilgang. Bl.a. (Weyerer and Schaufele 213-18) udtrykker ulempen ved proxyvurderingen således: ”...*they filter a subjective measure through the opinion of another person. These ratings may be influenced by the proxy’s own expectations and belief system, the prior relationship with the person being rated, and the current burden of care*”. Et potentielt problem med proxymåling er, at data kan være farvet og biased af en anden persons holdning. (Krenz et al. 453-61) giver dog i deres undersøgelse af Sickness Impact Profile (SIP) udtryk for, at familiemedlemmerne giver en mere valid indikator for demente patienters funktionelle status end patienterne selv. Dette gøres ved anvendelse af multiple korrelationer mellem SIP subskalaer (uafhængige variable) og hhv. Mini Mental State Examination (MMSE (Folstein, Folstein, and McHugh 189-98)) og en modificeret Dementia Rating Scale (mDRS (Lawson, Rodenburg, and Dykes 153-59)) (afhængig variabel). Både ujusterede og justerede korrelationer er højere for proxy respondenterne end for patienternes vurderinger. Især de justerede korrelationer (justeret for svind/shrinkage) indikerer at patienternes vurderinger bliver mindre og mindre valide over tid (se også afsnit 5.4).

1.1 Formål

Formålet med dette arbejdsnotat er at få beskrevet de mest almindelige generiske livskvalitetsinstrumenters egenskaber i forbindelse med vurderingen af demensramtes livskvalitet. Intentionen er herigennem at få opbygget viden om de generiske instrumenter og deres (eventuelle) egenskaber i forbindelse med brug på demensramte.

Fokus i arbejdsnotatet er på:

- Proxy-respondent problematikken. Hvor stort et sammenfald er der mellem demensramtes og proxy respondenternes vurdering af livskvalitet?
- Generisk vs. sygdomsspecifikke instrumenter. Hvad er fordelene og ulemperne ved de to instrumenttyper?

- Indholdsvaliditeten i de nyttebaserede generiske livskvalitetsinstrumenter – dækker disse de relevante aspekter, som er væsentlige for demensramtes livskvalitet?
- Hvordan er de generiske instrumenters psykometriske egenskaber (valideringsresultater) i forbindelse med anvendelse på demensramte individer?

I forbindelse med gennemgangen af de generiske instrumenter beskrives kort grundlaget for dannelsen af instrumentet, dvs. formål og begrebsligt grundlag, valideringsresultater, samt eventuelle problemer, der er ved brug af de generiske instrumenter på demensramte (specifikke items og/eller andre problemstillinger, der evt. gøres opmærksom på i fundne artikler).

Der udføres endvidere en oversigt over demensstudier, der har bragt et generisk instrument i anvendelse. Kilde, studietype, kort beskrivelse af populationen samt metodiske overvejelser/håndteringsmetode (interview, selvudfyldelse, proxyvurdering), vil fremstå af oversigten.

1.2 Metode og afgrænsning

Omfanget af generiske instrumenter er begrænset til de mest almindelige/mest anvendte. Fokus er som følge heraf på: EQ-5D, 15D, SF-36, WHOQOL, SIP, QWB. Disse dækker både præference-baserede mål og generelle helbredsprofiler.

Udgangspunktet er artikler fundet ved søgning i hhv. Medline og Embase, der har anvendt det generiske instrument på demensramte. Søgningen er gjort ud fra instrumentets fulde navn og/eller akronym samt 'dementia' og/eller 'Alzheimer*'. Fokus har både været på valideringsstudier, studier/artikler, der har diskuteret proxy respondent problematikken, samt studier, der har anvendt det generiske instrument som outcome mål – fx i kliniske forsøg.

2 Patienters vs. proxy respondenters vurdering af livskvalitet

Det er generelt vedkendt, at proxy målinger *er* en nødvendighed i de senere demensstadier, idet patienterne ikke længere er i stand til at vurdere denne, mens mildt demente – og måske også moderat demente, selv kan vurdere egen livskvalitet. Det er dog påpeget, at patienter i de tidligere stadier sandsynligvis giver en over-optimistisk vurdering af egne funktionelle kapaciteter (Selai and Trimble 101-11), og derved muligvis fremkommer med ikke-valide informationer.

Spørgsmålet er, hvornår patienters og proxy respondenters vurderinger stemmer overens. Generelle studier af proxyudledte data foreslår at:

- *the more objective the question and the more concrete the item in question, the closer the proxy's response will be to the subject's*
- *proxies are poorer reporters for conditions and symptoms that are private and not easily observed*
- *findings regarding proxy-subject agreement for ratings of affective status are inconsistent* (Selai and Trimble 101-11)

Det mest konsistente fund i studierne er, at der er størst overensstemmelse mellem patienter og proxy vurderingerne i forbindelse med items, som spørger om diskrete, observerbare funktionsaspekter så som mobilitet, og at proxy respondenterne har tendens til at overvurdere manglende evner, sammenlignet med patienternes egne vurderinger (Selai and Trimble 101-11;Weyerer and Schaufele 213-18;von Essen 229-34). Dette er en tendens, som oftest er tolket som en underestimering af patienternes livskvalitet (von Essen 229-34;Novella et al. 443-52).

Der eksisterer flere faktorer, der kan forklare denne uoverensstemmelse:

- patientens delvise selvbevidsthed om aftagende funktionelle evner
- tendens til at observatører giver mere vægt til negative end til positive informationer, når der dannes indtryk af andre
- omsorgsbyrden. Bl.a. viser resultaterne i (Novella et al. 443-52), at fænomenet er mindre udpræget når sundhedsprofessionelle agerer som proxy og når patienterne er indlagt på hospital
- proxyperspektivet: skal vurderingen bygge på egen opfattelse af patientens funktion, eller skal vurderingen baseres på hvordan man tror patienten opfatter egen daglig funktion? Sidstnævnte kan føre til større grad af overensstemmelse.

Hvem, der er bedste til at vurdere patientens livskvalitet, afhænger altså af øjet der ser. Et argument i Proxy-respondenternes favør er, at forskellige typer proxy respondenter ofte er mere enige om patientens livskvalitet end patienten selv. Dette *kunne* evt. reflektere, at patienternes respons var af tvivlsom kvalitet (von Essen 229-34). På den anden side viser en analyse af den interne konsistens i (Novella et al. 443-52), at data fremskaffet fra patienter er mere sammenhængende end data fremskaffet fra proxy respondenterne, hvilket taler til fordel for patientvurderingen.

2.1 Vurdering af sammenfald mellem patienter og proxylvurderinger

Spørgsmålet er, hvor lille/stor en forskel mellem proxy- og patientvurderingen kan være, før man kan konkludere, at en overensstemmelse er tilfredsstillende? Der eksisterer forskellige måder, hvormed man kan tolke mål for overensstemmelse. (von Essen 229-34) har opstillet og kort beskrevet de forskellige tilgange:

Statistisk signifikant. Statistisk signifikante forskelle i middelværdierne kan give information om systematisk over- eller underestimering grupperne imellem (gruppeniveau), mens statistisk signifikante associationer tolkes som evidens for systematisk overensstemmelse respondenterne (hhv. proxy og patient) imellem

(individniveau). Anvendeligheden af statistisk signifikans er begrænset af, at der ikke eksisterer nogen klar metode hvormed man kan tolke den kliniske vigtighed af en statistisk signifikant forskel/association. Et andet problem er, at resultatet er afhængigt af stikprøvens størrelse og af at den samme grad af overensstemmelse kan opnås ved tilfældigheder.

Effekt størrelse. En anden måde at vurdere betydningen af en forskel i middelværdien er ved at relatere den til standardafvigelsen. Effektstørrelser tolkes ofte i henhold til Cohens (Cohen) guidelines, hvor 0,20 er lille, 0,50 er moderat og 0,80 er stor forskel.

Guidelines. Styrken af en overensstemmelse tolkes nogle gange mod guidelines. Pearson's r , vægtet Kappa statistik, samt intraklasse korrelationskoefficienter (ICC) kan tolkes som følger: $\leq 0,20 \approx$ dårlig overensstemmelse; $0,21 - 0,40 \approx$ fair overensstemmelse; $0,41 - 0,60 \approx$ moderat overensstemmelse, og $0,81 - 1,00 \approx$ excellent, næsten perfekt overensstemmelse.

Andelen af præcis eller tilnærmelsesvis overensstemmelses konkordans. Selvom der kun er fundet moderate associationer, kan proxy respondenten fremkomme med vurderinger, som i langt de fleste af tilfældene er identiske eller meget lig patientens vurderinger. Det er foreslået, at kategoriforskelle kan tolkes som små, såfremt forskellen kun er en responskategori, og store, hvis forskellen er mere end én responskategori.

2.2 Empiriske resultater

Der er fundet enkelte studier, hvis primære fokus har været på undersøgelsen af overensstemmelsen mellem demensramte og deres proxy respondent.

I (Novella et al. 443-52) blev overensstemmelsen mellem selvhåndtering og proxy rapporteringen (interrater reliabilitet) af helbredsrelateret livskvalitet, i en stikprøve

bestående af 76 patienter med mildt til moderat Alzheimers, undersøgt. De demensramte og proxy respondenterne (hhv. familie og sundhedspersonale) udfyldte '17-item Duke health profile' (jf. også afsnit 4.1), hvor items i proxy respondenternes version var omformulerede. Resultaterne i deres undersøgelse var, at der var dårlige til moderat overensstemmelse mellem de demensramte og deres proxy (præcis overensstemmelse mellem proxy- og patientvurderingen for de 17 items varierede fra 26,3 til 52,6 %). Overensstemmelsen var størst for de observerbare mål for funktion, så som fysisk helbred eller funktionsindskrænkning (eng.: disability). I kontrast hertil var overensstemmelsen dårlig for de mere subjektive subskalaer så som opfattet helbred eller socialt helbred. Der var i studiet generelt højere overensstemmelse mellem demensramt og sundhedspersonaleproxy, end mellem demensramt og familieproxy. Globalt set var børnene dårligere proxy end ægtefællen, ligesom sygehjælper (eng.: nurses aids) var dårligere end andre sundhedsprofessionelle. Generelt set vurderede familie proxyerne fysisk helbred, mentalt helbred, generelt helbred og depression signifikant lavere end de demensramte selv. For sundhedspersonalet gjaldt dette kun for fysisk helbred. Effektstørrelsen var dog relativ lille for alle de signifikante forskelle (Cohens $d < 0,5$). (Coucill et al. 760-71) undersøgte interrater overensstemmelsen mellem 64 demensramte og proxy respondenter i relation til EQ-5D. Her fandtes tilsvarende problemer mht. overensstemmelse mellem patientvurderinger og proxyvurderingerne. Vægtede Kappaværdier mellem de demensramte og proxyen (hhv. omsorgsyder og læge) lå mellem 0,03 (sædvanlige aktiviteter) – 0,4 (mobilitet). Dvs. dårlig til fair overensstemmelse. Mellem omsorgsyderen og lægen lå overensstemmelsen for tre af dimensionerne noget højere (sædvanlige aktiviteter, personlig pleje, mobilitet – vægtede kappa mellem 0,34 – 0,63). Dårligere overensstemmelser fandtes for de to sidste dimensioner (hhv. 0,12 og 0,16), hvor der var højere overensstemmelse mellem den demensramte og proxyen. Der blev ikke fundet bedre overensstemmelsesresultater mellem proxy respondenterne og patienter med mild demens.

2.2.1 Overensstemmelsesfaktorer

Der er flere mulige forklaringer på, hvorfor studierne fremkommer med lave patientproxy overensstemmelsesresultater. Patientproxy overensstemmelse afhænger af flere faktorer, bl.a.:

Metodologiske faktorer: korrelationen mellem to scorere kan aldrig overstige kvadratroden af produktet af scorernes interne konsistens. Den interne konsistens lå for patienterne i (Novella et al. 443-52) mellem 0,17 (social health) og 0,76 (general health); for familieproxy mellem -0,19 (social health) – 0,65 (mental health) og for omsorgsyderproxy mellem -0,06 (social health) – 0,67 (mental health). Den interne konsistens er delvist afhængig af antallet af items og responskategorier såvel som respons variabiliteten.

Stikprøvestørrelsen er en anden faktor, som er relateret til overensstemmelse. Lave overensstemmelsesresultater findes oftest i studier med små stikprøver (ca. 50 patientproxy par eller færre) (von Essen 229-34).

Hvem er proxy?: Man ville måske forvente, at der var større overensstemmelse mellem patienter og omsorgsyderen end mellem patienterne og fx lægen. Dette er dog ikke nødvendigvis tilfældet, indikeret af resultaterne i (Coucill et al. 760-71; Novella et al. 443-52) og generelle fund i andre studier (von Essen 229-34). Hvem, der således er den bedste proxy for patienten, er uklart. Coucill et als ”råd” er, at da forskellige proxy respondentgrupper kan give forskellige resultater, bør man overveje at indsamle data fra både omsorgsydere og læger/professionelle, i stedet for at stole på kun én proxy. Deres resultater viser også, at observerbare og objektive dimensioner ikke nødvendigvis medfører større overensstemmelse mellem patienternes og proxy respondenternes vurderinger – heller ikke, når patienterne er mildt demensramte.

Proxy karakteristik: Generelt forventes der større overensstemmelse mellem patienter og ’signifikante’ andre (primært ægtefællen/børn), der er involveret i

patientens dagligdag. Involvering i patientens daglige liv kan dog forvrænge proxy'ens opfattelse af patienten, hvilket skyldes, at omsorgen for en patient kan være meget byrdefuld. Studier har således vist, at vurderinger af patientens psykosociale funktion er negativt påvirket af omsorgsbyrdeopfattelsen. Byrde kan fremme intolerance, hvilket kan føre til en underestimering af patientens funktionelle kompetence (von Essen 229-34). En proxy, der ikke er involveret i patientens daglige omsorg, men stadig i tæt kontakt med denne, kan i mange situationer derfor være den foretrukne proxy.

Patient karakteristik: Resultater fra visse studier indikerer, at der er større overensstemmelse vedrørende fravær end tilstedeværelsen af et problem. Andre studier understøtter dog ikke dette resultat. Nyere studier foreslår et U-formet forhold mellem patienternes helbredstilstand og niveauet for patientproxy overensstemmelse. Disse fund antyder, at man kan stole på proxy udledte livskvalitets informationer, når behovet for disse er størst (von Essen 229-34).

3 Livskvalitetsinstrumenter

Der eksisterer ingen gylden standard, hvormed man kan måle individers livskvalitet. Der eksisterer ifølge (Thorgrimsen et al. 201-08) på nuværende tidspunkt mere end 1000 generiske og sygdomsspecifikke helbredsrelaterede livskvalitetsinstrumenter. De *generiske instrumenter* dækker et bredt spekter af QOL domæner, og deres primære fordel er, at de kan anvendes til sammenligning af forskellige patientgruppers livskvalitet. Generiske instrumenter anvendes derfor i standard økonomiske evalueringer, der udføres på samfundsniveau. De *sygdomsspecifikke instrumenter* indeholder kun sygdomsrelevante items. Resultaterne for livskvalitet kan derfor ikke direkte sammenlignes med andre sygdomsgruppers indikation af livskvalitet, da fokus ikke er på de samme sygdomsrelaterede problemer.

3.1 Fordele og ulemper ved hhv. generiske og sygdomsspecifikke instrumenter

Der er fordele og ulemper ved begge instrumenttyper: Styrken ved de generiske livskvalitetsinstrumenter er, at de muliggør livskvalitetssammenligninger mellem forskellige patientgrupper, ligeledes mellem raske og syge. At generiske instrumenter skal kunne anvendes på en bred population (unge/gamle, raske/meget syge), er ligeledes denne type instrumenters største ulempe, idet instrumentet skal være følsomt over for selv de mindste helbredsproblemer, hvilket bevirker, at man ofte i forbindelse med brug på meget syge løber ind i problemer med gulveffekter (- eller lofteffekter alt efter hvordan der skaleres), idet en stor procentdel af svarpersonerne opnår den lavest mulige score – en score som er lig den værst tænkelige livskvalitet. Herved får instrumentet problemer med at kunne skelne mellem forskelle i den helbredsrelaterede livskvalitet blandt syge svarpersoner, da de har ramt gulvet inden for det område, hvor skalaen er følsom. Et andet problem er, at items indeholdt i et generisk instrument ikke nødvendigvis er dækkende for livskvaliteten i forhold til en bestemt patientgruppe. Specifikke

sygdomssymptomer som kløe, udslæt, kvalme, kognitiv svækkelse, adfærdsproblemer, mv. er sjældent indeholdt i de generiske instrumenter, selvom de kan spille en stor rolle for livskvaliteten i en bestemt patientgruppe. Generiske instrumenters egenskaber kan derfor være begrænsede for specifikke tilstande og sygdomme. Som Gold et al (1996) siger: *”Any generic health state classification system that is general and simple enough to be applied across all diseases and interventions may lack sensitivity to important differences in health status that are salient for particular diseases or interventions. Such differences, while ostensibly small, may be highly relevant...”* *“Therefore, disease-specific, and condition-specific classification systems have an important role to play in cost effectiveness analysis...”*(Gold et al. 109-10). Alternativet til de generiske instrumenter er derfor de sygdomsspecifikke instrumenter, hvor fokus netop er på de faktorer/symptomer, der spiller stor rolle for den pågældende patientgruppes livskvalitet. Fordelen er, at disse instrumenter som regel er dækkende for livskvaliteten i forhold til den pågældende patientgruppe. Man løber derfor ikke ind i nær så store problemer med gulv og lofteffekter, som man kan gøre ved de generiske instrumenter. Ulempen er, at de ikke kan anvendes til at sammenligne mellem forskellige interventionstyper foretaget på samfundsniveau (dvs. standard økonomiske evalueringer).

3.2 Indholdsvaliditeten i generiske instrumenter i relation til demensramtes vurdering af relevante attributter

Forcen ved de generiske, nyttebaserede livskvalitetsinstrumenter er således i forbindelse med sundhedsøkonomiske analyser, idet de muliggør sammenligning af livskvalitet på tværs af forskellige sygdomstyper. Som et resultat heraf kan det være vanskeligt at finde et generisk instrument, der indeholder alle sygdomsspecifikke livskvalitetsaspekter, da de generiske instrumenter indeholder mere generelle problemstillinger. Et væsentligt spørgsmål er, om instrumenterne opfanger de aspekter, demensramte finder er relevante i forhold til livskvalitet.

(Silberfeld et al. 71-79) har gennem semi-strukturerede interviews med 20 mildt demensramte (AD) og deres omsorgsyder søgt at finde, hvilke livskvalitetsaspekter demensramte og deres omsorgsydere finder er relevante og vigtige. Disse aspekter blev sammenlignet med indholdet i tre generiske nyttebaserede livskvalitetsinstrumenter, nemlig HUI, EQ-5D og QWB. Resultaterne af Silberfeld et als empiriske undersøgelse er gengivet i efterfølgende tabel.

Attributter	EQ-5D	H UI 2	H UI 3	Q W B	Attributter	EQ5 D	HUI 2	HUI 3	QWB
Global impression of quality of life					Frustration/irritability/anger		✓		✓
					Sense of control/autonomy				✓
Life satisfaction		?	?		Fear of embarrassment				
General health	✓				Concerns over taking medication/ side effects/overmedication				
Physical well-being					Physical appearance				
Physical illnesses/ailments/frailty	?	?		✓	Fear of being alone				
Fatigue/lack of energy				✓	Sense of stability				
Sleep disturbances				?	Appreciation of nature				
Hearing problems		✓	✓	✓	Relief from job related stress or responsibilities				
Falls/loss of balance				✓	Functional well-being				
Incontinence				✓	Recreational activities/hobbies/leisure	✓			✓
Loss of appetite				✓	Driving				✓
Vision		✓	✓	✓	Exercise				
All disturbances of consciousness				✓	Self-care activities	✓	✓		✓
Social/family well-being					Instrumental	?			?
Connection to family/family	✓			✓					
SUPPORT Social Interactions	?			✓					

Withdrawn/loneliness/boredom				?	Opportunity to travel				
Loss of competence	✓		✓	✓	Mobility	✓	✓	✓	✓
Access to good medical care					Cognition				
Living arrangements					Memory problems		✓	✓	✓
Financial security					Subjective sense of confusion				
Feeling useful/valuable to others/helpful					Ability to read				✓
Supervision					Disorientation in space/getting lost				
Emotional well-being					Ability to benefit from memory aids				
Burden of memory loss/emotional adjustment to Dx					Communication problems/word-finding difficulties		?	?	?
Personal losses					Disorientation in time				
Lack of interest or motivation					Help with word-finding difficulties				
Religion				✓	Psychomotor difficulties (Apraxia)				
Depression	✓	✓	✓	✓	Disorientation in person/loss of identity				
Appreciation of art					Lack of responsiveness to environment				
Hopes for longevity									
Hopes for a cure or arrest of the progressive cognitive decline									
Restlessness/agitation									

Kilde: Modificeret fra Silberfeld et al., 2002.

Bemærkning: Et ✓ indikerer, at attributten, identificeret i Silberfeld et als undersøgelse, er medtaget i det generiske instrument. Et ? indikerer et begrænset match mellem respondenternes attribut og det generiske instruments attribut.

Der blev gennem interviewene fundet i alt 56 attributter vedrørende livskvalitet i (mildt) demensramte individer. Disse attributter blev af Silberfeld et al., inddelt i 6

brede domæner: (1) globalt livskvalitetsindtryk, (2) fysisk velvære, (3) social/familie velvære, (4) emotionel velvære, (5) funktionel velvære, (6) kognition. En sammenligning med indholdet i de generiske instrumenter fandt, at kun 27 attributter (48 %) kunne inkluderes i mindst et domæne i et af de tre vurderede instrumenter: EQ-5D kunne inkludere 10 af attributterne; HUI3 kunne inkludere 11 og QWB 25.

Resultatet viser, at der, sammenlignet med de respondentgenererede attributter, er problemer med indholdsvaliditeten i de tre nyttebaserede generiske livskvalitetsinstrumenter i relation til demenssygdomme, idet de ikke medtager mange af de livskvalitetsfaktorer, der er vurderet relevante for demensramte patienter. Dette sætter en mulig begrænsning på anvendeligheden af de generiske instrumenter, idet disse måske ikke er i stand til at opfange klinisk væsentlige forskelle mellem behandlinger eller ændringer over tid. Dette er dog ikke nødvendigvis ensbetydende med, at de generiske instrumenter ikke kan anvendes på de demensramte. Som Silberfeld et al. (2002) selv skriver, så er det ”...*still possible that the generic instruments will perform well because they sample those selected areas of quality of life that are most important.*”

Om man vælger et generisk instrument eller et sygdomsspecifikt instrument, afhænger derfor af formålet med studiet. Ofte anbefales det dog både at inkludere et sygdomsspecifikt og et generisk instrument i en undersøgelse (Selai and Trimble 101-11; Drummond et al.).

4 Generiske instrumenter

Der eksisterer to skoler, der begge er enige om, at livskvalitet er et multidimensionalt fænomen – hvad de ikke er ikke enige om, er implikationen heraf. For den ene skole betyder multidimensionaliteten, at målingen af livskvalitet må være multidimensionalt, hvorfor der kun kan dannes profiler, hvor hver profil måler en enkelt livskvalitetsdimension eller domæne. Den anden skole noterer, at idet individer er nødt til at vægte forskellige helbredsattributter for at bestemme, hvilken der synes bedst, bør det være muligt at udlede en overordnet vurdering via nogle passende metoder, som genererer en enkelt indekssværdi for hver helbredstilstand (Williams 199-208)

De generiske instrumenter kan derfor opdeles i to grupper, hhv. i de generelle helbredsprofiler og i de præference-baserede mål¹. SF-36, Nottingham Health Profile², WHOQOL og Sickness Impact Profile (SIP) er eksempler på generelle helbredsprofiler. Der eksisterer fra en (sundheds)økonoms synsvinkel flere problemer med sådanne mål: Da de oftest ikke frembringer et samlet indeks, begrænses deres anvendelighed i forbindelse med sundhedsøkonomiske analyser, idet de har multiple outcome dimensioner.³ Man vil derfor ikke altid få et entydigt svar vedrørende hvilken af fx to behandlinger, der resulterer i det bedste helbred.

En økonom vil derfor oftest foretrække, at de forskellige dimensioner sammenvejes på en veldokumenteret måde – fx i et nytteindeks. Et sådant indeks

¹ Generelle helbredsprofiler har deres force i kliniske forsøg, hvor det kan være nødvendigt med detaljeret information. Præferencebaserede mål har deres force i evaluative studier (CEA/CUA)

² Er ikke medtaget i notatet da den sjældent bringes i anvendelse i dag.

³ Nogle profiler, så som SIP og NHP, kan fremkomme med en summeret score, men er ikke skaleret på et 0 til 1 kontinuum med en klar reference til 'død'. De er derfor ikke velegnet til brug i CEA eller CUA studier. Det kan ligeledes være vanskeligt at opfange det totale kliniske billede ved brug af profiler, da de fleste behandlinger har bivirkninger såvel som benefits (Kaplan et al. 509-20).

bygger ikke (som i psykometriske analyser) på antagelsen om, at man måler en enkelt velafgrænset egenskab. I stedet vægtes forskellige egenskaber (helbredsaspekter) sammen, idet man som udgangspunkt for vægtningen tager folks præferencer for forskellige helbredstilstande.

Typisk vil man bestemme præferencen for en helbredstilstand på en skala, der går fra 1 (fuldstændigt velbefindende og godt helbred) til 0 (død). Der eksisterer forskellige teknikker, hvormed man kan udlede folks præferencer (eller værdisætning af en helbredstilstand): fx time-trade off (hvor man 'ofrer' levetid), standard gamble (øget risiko), eller willingness to pay/accept (indkomst/penge) (Bjørner et al.). Der vil ikke komme nærmere ind på fordele og ulemper ved de forskellige tilgange. I stedet henvises til diverse sundhedsøkonomiske lærebøger for uddybende diskussionen heraf (Gold et al.; Folland, Goodman, and Stano; Drummond et al.).

Kendetegnende for generiske instrumenter er, at der ofte eksisterer mange valideringsstudier af disse. Begrebsvaliditet er fx etableret ved at vise systematisk forhold mellem en variabel og andre mål som formodentlig repræsenterer det samme begreb. Disse associationer definerer begrebets betydning. For nærværende arbejdsrapport er fokus på demensramte, hvorfor en gennemgang af de psykometriske egenskaber, fundet i originalstudier og øvrige valideringsstudier ikke gengives her. Andre har dog beskrevet flere af disse, bl.a. (Hutchinson, Bentzen, and König-Zahn; McDowell and Newell; Petersen and Zachariae). Disse forfattere har ligeledes søgt at give enkelte af de generiske instrumenter "karakter" i henhold til instrumenternes (generelle) psykometriske egenskaber. Disse "karakter" er gengivet i efterfølgende tabel.

Oversigt over generiske instrumenter

Instru- ment	År og kilde	Antal spørgsmål	Håndterings metode nævnt i original artiklen	Afprøvning af pålidelighed		Afprøvning af gyldighed		Dansk oversæt telse	Dansk valide ring	Præfe rence- vægtet?
				Grundig- hed☀	Resultat ♥	Grundig- hed☀	Resultat ♥			
DUKE ☺	1990 (Parkerson, Broadhead, and Tse 1056-72)	17	Selvhånd teret	+++	++	+++	++	Nej	Nej	Nej
EQ- 5D ☺	1990 (Williams 199-208)	5 + 1VAS	Selvhånd teret	++	++	++	+	Ja	Nej (?)	Ja
SF- 36 ☺	1990 (Ware and Sherbourne 473-83)	36	Selvhånd teret <i>eller</i> Interview håndteret	+++	+++	+++	+++	Ja (Bjørner et al.)	Ja (Bjorner et al. 991- 99;Bjorner et al. 1001-11)	Nej (ja♦)
15D	1994 (Sintonen 328-36)	15	Selvhånd teret					Ja	?	Ja

HUI	1982 (HUI1) (Torrance, Boyle, and Horwood 1043-69) 2001 (HUI2+HUI3) (Furlong et al. 375-84)	15 (HUI2 + HUI3)	Selvhånd teret <i>eller</i> Interviewer håndteret					Undervej s	Nej	Ja
WHOQ OL ☺	1995 (Kuyken et al. 1403- 09)	100	Selvhånd teret	+++	++	+++	+++	Ja	Ja (Norholm and Bech 229-35)	Nej
SIP ☺	1976 (Bergner et al. 57-67) 1981 (Bergner et al. 787- 805)	136	Interview håndteret <i>eller</i> Selvhånd teret	+++	++	+++	++	Ja (Folker 175-78)	Ja (delvist) (Folker and Jensen 3347-52)	Nej
QWB ☹	1970♠ ,1973, 1982 (Kaplan and Bush 61-80)	107	Interviewer håndteret	++	+++	+++	++	?	?	Ja

Kilder: ☺ Fischer Petersen og Zachariae. ☹ McDowell og Newell.

☀ Et + betyder, at der kun er rapporteret meget basal information vedrørende instrumentets pålidelighed/gyldighed. To ++ betyder, at der er udført flere forskellige tests til afprøvning af instrumentets pålidelighed/gyldighed. Tre +++ betyder, at alle vigtige former for tests til afprøvning af instrumentets pålidelighed/gyldighed har været udført. ♥ Et + betyder, at resultaterne tyder på svag pålidelighed/gyldighed. To ++ betyder, at resultaterne tyder på acceptabel pålidelighed/gyldighed. Tre +++ betyder, at resultaterne tyder på meget tilfredsstillende pålidelighed/gyldighed.

♦ Der er gjort forsøg på at indeksere SF-36. Denne tilgang kaldes SF-6D og er nærmere beskrevet i (Brazier, Roberts, and Deverill 271-92; Brazier et al. 1115-28)

♠ QWB hed oprindeligt "Health Status Index. Dette blev senere ændret til Index of Well-being og efterfølgende til QWB).

Et generisk instrument kan komme med nok så fine psykometriske egenskaber for den generelle befolkning – det er dog ikke kriterium nok i valget heraf på en specifik befolkningsgruppe, såsom de demensramte. Inden valg af et generisk instrument foretages, er det nødvendigt at undersøge, om instrumentets indhold er passende for formålet; om det før har været anvendt til lignende formål, som det nu tænkes anvendt til; om det tidligere har været benyttet på sygdomsområdet eller patientgruppen.

Oversigt over indholdet i de generiske livskvalitetsinstrumenter

Begreb	DUKE ☺	EQ- 5D ☺	SF- 36 ☺	15D	HUI2	HUI3	WHO QOL	SIP ☺	QWB ☺
Fysisk funktion	X	X	X	X	X	X	X	X	X
Social funktion	X	X	X	X			X	X	X
Rolle funktion	X	X	X		X	X	X	X	X
Psykiske symptomer	X	X	X	X			X	X	
Alment helbred (globalt selvvurderet helbred)	X	X	X				X		
Smerter (fysiske)	X	X	X	X	X	X	X	X	
Energi/træthed	X		X	X			X		X
Psykisk velbefindende	X		X		X	X	X		
Søvn	X			X			X	X	
Kognitiv funktion	X			(X)	X	X	X	X	
Livskvalitet									
Ændringer i alment helbred							X		
Ernæring				X				X	
Miljø							X		
Spirituel							X		
Fertilitet					X				
KARAKTE RISTIKA									
Scoringmetode (L = Likert, T = Thurstone, R = Rasch, U = utility)	L	U	L, R, (U)	U	U	U	L	T	U

Rapportering (P = profil, GK = generelle komponenter, EI = et indeks)	P, EI	EI	P, GK (EI)	P, EI	EI	EI	P, GK	P, GK, EI	EI
---	-------	----	------------------	----------	----	----	-------	-----------------	----

Kilde: ● Delvis efter (Bjørner et al.). Opstillingen er meget subjektiv. Således kan der være forskellige fortolkninger af, hvad de forskellige begreber indeholder.

En kort introduktion/beskrivelse af reliabilitets- og validitetsbegreber gives i efterfølgende afsnit 4.1. Dette efterfølges af en beskrivelse af de mest anvendte generiske livskvalitetsinstrumenter og deres egenskaber. Fokus er primært på baggrunden for udviklingen af de enkelte instrumenter, ligeledes på om instrumenternes egenskaber er vurderet på demensramte, alternativt på andre grupper med mulige kognitive problemer eller ”blot” på ældre, for at vurdere eventuelle problemer med instrumentet på pågældende population (fx problemer med enkelte items eller dimensioner).

4.1 Reliabilitet og validitet

4.1.1 Reliabilitet

Almindeligvis er spørgsmålet, om instrumentet er reliabelt. Dvs. hvor reproducerbare og internt konsistente er de resultater, som instrumentet fremkommer med. Reproducerbarheden er især væsentlig i forbindelse studier, der skal måle effekten af en intervention. Således bør forskellen i pre- og post test situationen ene og alene afspejle interventionen og ikke det anvendte instrument. Der eksisterer forskellige metoder, hvormed man kan vurdere et instruments reliabilitet:

Intern konsistens er et udtryk for graden af skalaitems samvariation. Samvariation mellem items i en skala er nødvendig, da det er udtryk for, at items måler aspekter ved samme dimension. Intern konsistens beregnes ofte ved Cronbach's alpha, der kan variere mellem 0 – 1. Tommelfingerreglen er, at alphaværdien bør ligge mellem 0,70 og 0,90. Den bør ikke overstige 0,90, da dette er tegn på, at der muligvis er overflødige items i skalaen.

Test-retest er et udtryk for graden af stabilitet i besvarelsen af livskvalitetsinstrumentets items over tid. Overensstemmelsen udtrykkes som en korrelationskoefficient, hvor en korrelation større end 0,70 normalt anses for acceptabel. Jo højere en korrelationskoefficient, desto bedre. Dette forudsætter dog, at livskvaliteten ikke har ændret sig i perioden mellem besvarelserne.

Interrater reliabilitet: er udtryk for overensstemmelsen mellem data opnået af forskellige observatører/interviewere, eller patient og proxy respondent overensstemmelse. (jf. også afsnit 2.1).

4.1.2 Validitet

Herudover er spørgsmålet om instrumentet rent faktisk måler, hvad det hævder at skulle måle – et spørgsmål om validitet.

Indholdsvaliditet: er de spørgsmål, der stilles, dækkende for den problemstilling, man ønsker at undersøge? Eller giver item i en skala dækkende og udtømmende informationer om det begreb, som skalaen angiver at måle? Man bør derfor altid vurdere indholdsvaliditetsproblematikken, når der skal vælges spørgeskema i forbindelse med design af en undersøgelse.

Kriterievaliditet: vurderes ud fra uafhængige mål for den egenskab, som skalaen måler. Disse kan være indsamlet på samme tidspunkt som skalaen (samtidig validitet) eller på et senere tidspunkt (prædiktiv validitet). I praksis er tilgængelige kriterier oftest mål for begreber, man formoder har en stærk sammenhæng med den målte egenskab (altså ingen gylden standard), hvorfor grænsen mellem kriterievaliditet og begrebsvaliditet er flydende.

Begrebsvaliditet: begrebsvaliditeten vurderes ved at teste de hypoteser om en skalas struktur og sammenhæng med andre variable, som udspringer af teoretiske overvejelser om det begreb, skalaen måler.

Øversigt over validitetsbegreber

Begreb	Betydning
Indholdsvaliditet	Items repræsenterer på en dækkende måde den egenskab (dvs. univers af mulige items), der ønskes undersøgt.
Umiddelbar validitet	Synes de medtagne items umiddelbart at måle det ønskede begreb?
Kriterie-relateret validitet	Korrelerer et nyt mål højt med det gyldne standardmål for begrebet?
Samtidig	Stemmer testen overens med en anden anerkendt test inden for samme domæne, der håndteres samtidigt?
Prædiktiv	Er testen i stand til at forudse fremtidige præstationer
Begrebsvaliditet	Korrelerer skalaen højt med mål for øvrige variable, som hypoteserne herom har forudsagt?
Konvergent validitet	Korrelerer et mål for fx smerte højt med et mål for effekterne af smerte? - Har et item høj korrelation med summen af de øvrige items i skalaen?
Diskriminant/divergent validitet	Har et mål for fysisk virke en lavere korrelation med et mål for mental sundhed end med et mål for mobilitet? - Korrelation mellem et item og summen af de øvrige items i 'egen skala' er højere end korrelationen mellem itemet og andre skalaer i instrumentet.
Known groups validitet	Er den gennemsnitlige helbredsopfattelse signifikant lavere for en gruppe af patienter end for en generel populationssample?

Andre faktorer, der kan være relevante at kigge på, er fx om:

- Instrumentet er følsomt til at opfange ændringer i sundhedsstatus over tid?
- Scoring – fortolkning og præcision
- Instrumentet bliver accepteret af populationen. En vurdering af afslagsraten, håndteringsmetoden, håndteringstiden, antallet af, samt fordelingen af manglende data, er en anden måde til vurdering af instrumentets anvendelighed på en given population.

5 Gennemgang af de generiske instrumenter

5.1 Duke Health Profile (DUKE)

Formål: Dannelse af et kort, generisk, selv-rapporteret instrument, der vurderer de grundlæggende helbredsmæssige problemstillinger (helbredsstatus).

Tidsreferencerammen for items er 'nu' eller indenfor den forgangne uge, alt efter hvad der skal svares på.

Begrebsligt grundlag: Instrumentet er udledt fra Duke-UNC health Profile (DUHP, (Parkerson et al. 806-28), et instrument med 63 items, som var designet til målingen af outcomes i primære sundhedsomgivelser. DUHP havde flere begrænsninger, hvilket blev søgt løst ved dannelsen af DUKE, bl.a. 1) at alle symptomer blev grupperet i samme symptomstatus dimension, 2) brug af selvværd som eneste indikator for emotionel helbredsfunction, samt 3) afgrænsning af definitionen for social funktion til social rolle udførelse (eng.: role performance). Udviklingen af de vigtigste af DUKEs helbredsmål er baseret på WHO's tre dimensioner (fysisk, mentalt og socialt helbred).

Beskrivelse: Duke Health Profile er et generisk, 17-item instrument, der kan anvendes til at vurdere helbred eller helbredsrelateret livskvalitet. Det består af i alt seks helbredsmål: fysisk helbred (5 items), mentalt helbred (5 items), socialt helbred (5 items), generelt helbred (15 items), opfattet helbred (1 item) og selvværd (5 items), samt fire dysfunktions mål: angst (6 items), depression (5 items), smerte (1 item) og funktionsindskrænkning (eng.: disability) (1 item). De fysiske, mentale, sociale, opfattet helbred og funktionsindskrænkningsskalaerne er uafhængige af hinanden, idet ingen af deres items er indbefattet i de øvrige skalaer, som deler items med hinanden.

Oversigt over items i DUKE

1	I like who I am	10	Sleeping
2	I am not an easy person to get along with	11	Hurting or aching in any part of your body
3	I am basically a healthy person	12	Getting tired easily
4	I give up too easily	13	Feeling depressed or sad
5	I have difficulty concentrating	14	Nervousness
6	I am happy with my family relationships	15	Socialize with other people
7	I am comfortable being around people	16	Take part in social, religious, or recreation activities
8	Walking up a flight of stairs	17	Stay in your home, a nursing home, or hospital because of sickness, injury, or other health problem
9	Running the length of a football field		

Kilde: (Parkerson, Broadhead, and Tse 1056-72)

Hver item vurderes på en tre-punktsskala (0 – 2), og itemrespons i de enkelte mål adderes. Råscorens middelværdi normaliseres til en skala mellem 0 og 100. For fysisk, mental, generel og opfattet sundhed, samt selvværd, går denne fra 0 = værst til 100 = bedst. For angst, depression, smerte og funktionsindskrænkning er skalaen vendt, så 0 = bedst og 100 = værst. Individernes respons præsenteres som 10 profiler, og der udregnes ingen generel score. For yderligere uddybning af metoden henvises til appendikset i (Parkerson, Broadhead, and Tse 1056-72).

Psykometriske egenskaber i forbindelse med demensramte

Der er fundet 2 studier, der vurderer DUKEs egenskaber på demensramte. I (Novella et al. 443-52) gennemgås overensstemmelsen mellem patient og proxy respondent (jf. afsnit 2.2), mens (Novella et al. 158-66) undersøgte instrumentets acceptabilitet, gennemførlighed, reliabilitet og validitet. Dette blev gjort ved håndtering af instrumentet på 148 mildt til svært demensramte – median MMSE scoren for 28 mænd var 12,5 (omfang 2-24). Median MMSE scoren for 120 kvinder var 13 (omfang 0-24). Følgende resultater er taget fra (Novella et al. 158-66).

Reliabilitet: Test-retest reliabilitet vurderet ved intraklasse korrelationskoefficient (ICC): I tilfældig stikprøve (n = 44) var ICC > 0,5 for alle sundhedsmål med undtagelse af opfattet sundhed (0,45) og angst (0,48). I individer med MMSE score ≥ 15 blev der kun fundet en score < 0,5, og det var i målet for selvværd.

Intern konsistens vurderet ved Cronbach's alpha (n = 148): kun tre sundhedsmål (mental sundhed, generel sundhed og depression) nåede en alphaværdi på ca. 0,70. For de resterende sundhedsmål lå alpha mellem 0,23 (selvværd) og 0,58 (fysisk sundhed).

Interrater reliabilitet: patienter og familiær proxy. Overensstemmelsesscoren (ICC) lå mellem 0,03 og 0,57. Kun social sundhed (p = 0,48), smerte (p = 0,74) og disability (p = 0,29) var ikke vurderet signifikant forskelligt. Mellem patienter og sundhedspersonale var overensstemmelsesscoren mellem 0,17 og 0,46. Social sundhed (p = 0,96), smerte (p = 0,92), disability (p = 0,14) og depression (p = 0,14) var ikke vurderet signifikant forskelligt. For de to proxy grupper var den eneste signifikante forskel i vurderingerne i forhold til mental sundhed (0,0009) og depression (p = 0,03). Interrater enigheden for hele stikprøven var generelt dårlig til middelmådig, med undtagelse af funktionsindskrækning, hvor ICC mellem de tre respondentgrupper var 0,55 og mellem patienter og deres familiære proxy'er var 0,57. I en subgruppe af demente (MMSE > 15, n = 54) var overensstemmelsen mellem proxy og patient noget større. ICC lå over 0,6 for 5 af de 10 sundhedsmål (fysisk, mental, generel sundhed, selvværd og funktionsindskrækning)

Validitet: Intern begrebsvaliditet blev vurderet ved principal komponent analyse. Fem faktorer havde egenværdi større end 1 og kunne forklare 59 % af instrumentets totale varians. 1. faktor, 4 mentale sundhedsitems. 2. faktor, 4 angstitems og 3 depressionsitems. 3. faktor, 3 sociale sundhedsitems og 4 selvværdsitems. 4. faktor, 3 fysiske sundhedsitems. 5. faktor, 2 sociale sundhedsitems.

Ekstern begrebsvaliditet: Evne til at diskriminere mellem individer med (n = 28) og uden (n = 80) ekstra kronisk patologi. Blev gjort via en variansanalyse justeret

for demensens sværhedsgrad (målt ved MMSE). Individer uden ekstra kronisk patologi havde bedre livskvalitet i 9 af de 10 mål, med signifikant forskel i 6 af de 10 mål.

Gennemførlighed (feasibility) og acceptabilitet: Instrumentet kunne håndteres på 98 % af respondenterne. 20 % af de demente kunne selv håndtere instrumentet (middel MMSE score var for disse 11,5; $p < 0,01$). Gennemsnitlig håndteringstid var 10,6 minutter (SD = 9 min). Problemer med missing data i 7 – 14 % af instrumentet, højest for de mest demente. Dvs. instrumentets gennemførlighed varierede med sværhedsgraden. Mest problematiske items var item 6 '*I am satisfied with my family relationships*', som ikke var besvaret af 17,5 % af respondenterne, samt item 17 '*During the past week, how often did you stay in your home, a nursing home, or hospital because of sickness, injury, or other health problems*', som 26,6% af respondenterne ikke havde besvaret.

Kommentar.

- Duke Health Profile reducerer ikke helbredsrelateret livskvalitet til en enkelt score, men vurderer i stedet bestemte elementer som indgår i helbredsrelateret livskvalitet.
- Instrumentets acceptabilitet er højt, så længe patienter med svær demens ikke indgår ($MMSE \leq 5$).
- Item 6 involverer en værdivurdering (value judgement) af patientens familie, hvilket ikke er let for patienterne. Problemet med item 17 er, at besvarelsen heraf afhænger af patientens korttidshukommelse.
- Kort håndteringstid – tog i studiet gennemsnitligt ca. 10 minutter.
- Instrumentets reliabilitet (intern konsistens) blev i studiet fundet god for mental sundhed, generel sundhed og depression (Cronbach's alpha $> 0,65$), acceptabel for fysisk sundhed og angst (alpha $> 0,50$), men dårlig for social sundhed og selvværd.

- Der blev fundet en acceptabel intern begrebsvaliditet. Faktoranalysen på de 15 items, der er forbundet med 7 multi-item sundhedsmål, fremkom med 6 særskilte sundhedsmål.
- Instrumentet er utilstrækkeligt hvis man skal studere opfattet sundhed og selvværd blandt demente
- Det blev fundet, at demensens sværhedsgrad kun har en lille indflydelse på reliabiliteten og validiteten. Det påvirker dog dets reproducerbarhed og acceptabilitet.

5.2 EuroQol (EQ-5D)

Formål: Dannelse af et kort, standardiseret, generisk mål, der kan udtrykke helbredsstatus i en enkel indeksscore. Instrumentet er kommet i stand som følge af et tværkulturelt samarbejde mellem Holland, England, Finland, Sverige og Norge, og er som følge heraf dannet, så det kan bringes i anvendelse på tværs af forskellige kulturelle sammenhænge. Instrumentet er tiltænkt anvendt i evaluative studier og policy research, og er designet som et selvhåndterbart spørgeskema, som fx kan bringes i anvendelse i forbindelse med post-udsendte (befolknings-)undersøgelser.

EQ-5D er designet, så det kan supplere andre instrumenter så som SF-36, NHP, SIP eller helbredsspecifikke spørgeskemaer.

Begrebsligt grundlag: EuroQol gruppen blev dannet i 1987 for at beskæftige sig med den voksende variation i mål for livskvalitet. Gruppen bestod af forskere, der havde viden indenfor økonomi, matematik, medicin, pleje, filosofi, psykologi og sociologi. De gjorde opmærksom på, at: *”there are a great number of rival approaches aiming to do very similar things, with little or no attempt at systematic comparison of their respective strengths and weaknesses either at a conceptual or at an empirical level”* (Williams 199-208). Udgangspunktet var derfor at definere en kerne af generisk helbredsrelaterede livskvalitetsitems. Der blev taget

udgangspunkt i allerede eksisterende instrumenter, bl.a. QWB, SIP, Nottingham Health Profile, og Rosser Index.

Initialt bestod instrumentet af seks attributter, nemlig: mobilitet, personlig pleje, sædvanlige aktiviteter, sociale forhold, smerte, og humør, og definerede et teoretisk univers indeholdende 216 mulige helbredstilstande (Williams 199-208).

Beskrivelse: EQ-5D er et spørgeskema, i hvilket respondenterne skal vurdere deres nuværende (i dag) helbredstilstand. Dette gøres på 5 helbredsrelaterede dimensioner: mobilitet, personlig pleje, sædvanlige aktiviteter, smerter/ubehag, angst/depression. Hver dimension vurderes på en 3-pointskala, hvor 1 = 'ingen problem', 2 = 'noget af et problem' og 3 = ekstremt problem. Der kan i alt dannes 243 forskellige helbredsprofiler. Herudover er tilføjet to ekstra helbredstilstande 'bevidstløs' og 'død', så instrumentet i alt kan definere 245 forskellige helbredsprofiler.

Instrumentet indeholder ligeledes en VAS-skala, hvor patienten angiver vurderingen af egen helbredstilstand. Skalaen går fra 0 = værst tænkelige helbredstilstand til 100 = bedst tænkelige helbredstilstand.

Psykometriske egenskaber i forbindelse med demente

Der er fundet 4 studier, der på én eller anden måde har undersøgt EQ-5Ds egenskaber på demensramte. (Silberfeld et al. 71-79) diskuterer indholdsvaliditeten (jf. afsnit 3.2). (Coucill et al. 760-71) fokuserede på instrumentets interrater overensstemmelse (jf. afsnit 2.2). (Thorgrimsen et al. 201-08) har brugte EQ-5D til at undersøge et sygdomsspecifikt instruments kriterievaliditet (Quality of Life-Alzheimer's Disease, QoL-AD), og (Ankri et al. 1055-63) har undersøgt EQ-5Ds acceptabilitet, gennemførlighed, reliabilitet og validitet på 137 mildt til svært demente – omfang af MMSE scoren 0 – 27, middelværdi 12,8. Følgende resultater er taget fra (Ankri et al. 1055-63) med mindre andet er specificeret.

Reliabilitet: Test-retest reliabilitet (n = 47; 3 dage, kappa statistik (% enighed)): der blev fundet moderat overensstemmelse i forbindelse med mobilitet, personlig pleje, sædvanlige aktiviteter og angst/depression: kappaværdier mellem 0,51 - 0,59 (73,3 - 75,6 %) og svag overensstemmelse i forbindelse med smerte/ubehag, kappaværdi = 0,34 (enighed = 62,8 %). Test-retest reliabiliteten for VAS, vurderet ved ICC, viste svag overensstemmelse $r = 0,44$ ($p < 0,04$). Test-retest reliabiliteten for totalscoren for de 5 dimensioner, vurderet ved ICC, var tilfredsstillende, $r = 0,74$ ($p < 0,0001$).

Interrater reliabilitet: kappa koefficienterne mellem patient- og proxyvurderingerne lå for 4 af dimensionerne meget lavt: For patient og familiær proxy var overensstemmelsen mellem 45,3 - 54,6 %, og for patient og institutionel proxy mellem 27,7 - 52,3 %. Kun den mest objektive dimension, mobilitet, opnåede en acceptabel overensstemmelse på hhv. 65,8 % (familiær) og 71,9 % (institutionel). Overensstemmelsen mellem de to proxy respondentgrupper var generelt lidt højere for de enkelte dimensioner – mellem 58,5 - 76,4 %, igen med størst enighed i forbindelse med mobilitetsdimensionen. For VAS var ICC hhv. 0,20 ($p = 0,01$; patient – familiær proxy), 0,23 ($p = 0,006$; patient – institutionel proxy) og 0,40 ($p < 0,0001$; familiær og institutionel proxy). For totalscoren var ICC hhv. 0,41; 0,42 og 0,54.

Validitet: Begrebsvaliditet: De fem items totalscore var signifikant korreleret med VAS scoren ($r = 0,43$; $p = 0,0001$). Forventede korrelationer mellem Katz (Katz et al. 20- &) indikator (ADL) og mobilitet, personlig pleje ($p < 0,0001$), sædvanlige aktiviteter og smerter ($p < 0,002$) viser konvergent validitet. Ingen forhold mellem Katz indikator og angst, hvilket er evidens for diskriminant validitet. Ligeledes ingen forbindelse mellem VAS og ADL scorerne. Der blev heller ikke fundet associationer mellem MMSE, CDR og EQ-5D.

Known group validitet: kvinderne scorede som forventet højere på angst/depression skalaen end mændene. Ligeledes viste resultaterne, at kvinderne

vurderede deres helbredstilstand værende dårligere end mændenes. Dette stemmer overens med fund i generelle populationer.

Kriterievaliditet (samtidig validitet): korrelationen med et sygdomsspecifikt instrument (QOL-AD) var 0,54 for EQ-5D (n = 52; p < 0,0001) og 0,50 med VAS (n = 40; p < 0,01) hvilket indikerer acceptabel kriterievaliditet (Thorgrimsen et al. 201-08).

Gennemførlighed (feasibility) og acceptabilitet: 137 af 142 patienter accepterede at udfylde spørgeskemaet. 14 individer kunne selv udfylde spørgeskemaet (gns. MMSE = 18) og 123 fik hjælp af en interviewer (gns. MMSE = 12,4). Gennemsnitlige håndteringstid var 15,3 min (SD 13,9 min). Varigheden af håndteringen var ikke signifikant korreleret med MMSE-scoren eller med EQ-5D dimensionerne. Responsraten for hver af de 5 items lå mellem 90 og 98 %. Flere end 80 % af individerne (114 ud af 137) besvarede alle 5 items, og kun 7 respondenter havde besvaret færre end 4 items. Jo højere MMSE score, jo bedre responsrate. Der var en ikke-responsrate på 22 % i forbindelse med VAS-vurderingen.

Kommentar:

- Indeholder få items – godt i forbindelse med denne populationstype
- Relativ kort håndteringstid, ca. 15 min., men med stor variabilitet (enkelte tog over 1 time).
- Ikke oplevet problemer med hverken gulv- eller lofteffekter i (Ankri et al. 1055-63). (Coucill et al. 760-71) rapporterer dog mulige problemer med en lofteffekt, idet 48 % af patienterne vurderede, at de ikke havde problemer i relation til dimensionerne. Coucill et al. finder dog, at det højst sandsynligt ikke er en ”sand” loft effekt.
- 59 forskellige profiler ud af de teoretisk 243 mulige profiler blev fundet (n = 114).
- Lav interrater reliabilitet, acceptable test-retest resultater.

- Validitetsresultaterne stemmer godt overens med hypoteserne herom. Resultaterne indikerer, at resultaterne reflekterer patienternes mentale og fysiske helbred.
- Mindre overensstemmelse mellem proxy og patienter med MMSE score \leq 10. Forudsætter ikke nødvendigvis, at patienternes vurderinger er mere tvivlsomme som følge af demensens progression, men kan reflektere at det bliver vanskeligere for proxy respondenterne at afkode patienternes helbredstilstand.
- Sværhedsgraden har kun lille effekt på reliabiliteten og validiteten, men påvirker acceptabiliteten.
- Vanskeligste item at besvare var 'sædvanlige aktiviteter' (non-respons = 10,2 %) samt VAS (non-respons = 21,9 %). I (Coucill et al. 760-71) gøres der ligeledes opmærksom på, at der ved dette item er størst uoverensstemmelse mellem patient- og proxyvurderingen.
- Problem (?): indeholder ingen kognitive aspekter

5.3 Health Utilities Index (HUI)

Formål: HUI er en familie af generiske helbredsstatus- og helbredsrelaterede præferencebaserede livskvalitetsmål. Familien består af Health Utilities Index Mark 1 (HUI1), Health Utilities Index Mark2 (HUI2) og Health Utilities Index Mark3 (HUI3) systemer. Hvert HUI mål indeholder et helbredsstatus klassifikationssystem og en præferencebaseret scoringsformel. Spørgeskemaerne er designet, så de kan fremkalde respons fra et bredt spekter af individer. Der eksisterer standardiserede selvhåndterede og interviewhåndterede versioner i både selvrapporterede og proxyrapporterede versioner, ligesom der er fire standard tidsreference perioder (se evt. www.healthutilities.com).

Begrebslige grundlag: HUI blev udviklet i respons til behovet for et standardiseret system til måling af helbredsstatus og helbredsrelateret livskvalitet for at beskrive 1) *the experience of patients undergoing therapy*; 2) *longterm outcomes associated with disease og therapy*; 3) *the efficacy, effectiveness and efficiency of healthcare interventions*; and 4) *the health status of general populations* (Horsman et al. 54).

HUI1 blev udviklet til vurderingen af outcome for spædbørn med meget lav fødselsvægt. HUI2 blev initialt udviklet til vurdering af outcome i individer, der havde overlevet kræft i barndommen. Da ufrugtbarhed kan være følge af nogle kræftformer i barndommen, blev en 'fertilite'ts' attribut medtaget. HUI2 blev senere modificeret til brug på voksne. HUI3 blev oprindeligt udviklet for 'the 1990 Statistics Canada Ontario Health Survey'. Den blev udviklet til at tage hånd om nogle af de problemer, der var omkring definitionen af HUI2. HUI3 skulle være anvendelig i både kliniske og generelle populationsstudier.

HUI bliver på nuværende tidspunkt defineret som indeholdende både HUI2 og HUI3, hvorfor nuværende HUI spørgeskemaer dækker begge systemer. Det vigtigste kriterium for udvælgelse af attributter i HUI systemet var den betydning, medlemmer af den generelle offentlighed tillagde hver attribut.

Beskrivelse: HUI2 består af 7 helbredsattributter⁴ med tre, fire el. fem niveauer pr. attribut. Niveauerne varierer fra 'meget begrænset' til 'normal'. Attributterne er: sanser (eng.: sensation – se, høre og tale), mobilitet, emotion (kvaler og angst), kognition (indlæring), personlig pleje, smerte (frekvens og kontroltype) og fertilitet (kan udelades). HUI3 består af 8 attributter med fem eller seks niveauer pr. attribut. Attributterne er: syn, hørelse, tale, begrænset bevægelighed (eng.: ambulation), fingerfærdighed (eng.: dexterity), emotion (glæde versus depression), kognition (evne til at løse dag til dag problemer) og smerte (sværhedsgraden

⁴ HUI anvender begrebet attribut i stedet for dimension eller domæne for at understrege instrumentets relation til 'multiattribute utility theory' (MAUT) (Furlong et al. 375-84), som er det, nyttescoringsfunktionerne er baseret på.

heraf). HUI2 og HUI3s helbredsstatus klassifikationssystemer er komplementære. Sammen yder de deskriptive mål for evner eller begrænsninger for helbredsstilstands attributter og beskrivelser af omfattende helbredsstatus.

For en uddybning af HUIs scoringsfunktioner henvises til (Furlong et al. 375-84). HUI data indsamles ved anvendelse af et eller flere spørgeskemaer i de formater, der udvælges til at kunne matche det specifikke studiedesign. Det kan fx dreje sig om håndtering (selv eller interview), vurderingssynspunkt (selv eller proxy) eller vurderingsperiode (1, 2, 4 uger eller 'normal').

Psykometriske egenskaber i forbindelse med demente

Der er ikke fundet deciderede valideringsstudier, der vurderer brugen af systemerne på demensramte. Umiddelbart dækker HUI2 og HUI3 mange af de områder, som er væsentlige for demensramte, hvilket højner systemets umiddelbare validitet. Instrumentets indholdsvaliditet er diskuteret i (Silberfeld et al. 71-79) (jf. afsnit 3.2), mens (Neumann et al. 413-22) i deres studie sammenligner HUI2s og HUI3s nyttescorer. Det er fra Neumann et al. følgende 'validitetsresultater' er taget.

Validitet: Prædiktiv validitet: multipel lineær regressionsanalyse viser, at de forskellige AD-stadier (målt ved CDR og korrigeret for socioøkonomiske faktorer) er en negativ og signifikant prædikator for patientens overordnede nyttescore. Sammenlignet med mild AD er en forøgelse i sværhedsgraden associeret med et fald på mellem 0,10 – 0,15 i nyttescoren. Forklaringsgraden for de to systemer: $R^2 = 0,51$ for HUI2 og $R^2 = 0,44$ for HUI3.

Konvergent validitet. Sammenholdt med graden af demens, målt ved CDR, blev der fundet substantielle, signifikante variationer i scorene på HUI2 og HUI3. De mest demente havde lavere score end de moderat eller mildt demente.

Known group validitet. Den gennemsnitlige globale nyttescore for omsorgsyderne er stort set ens uanset anvendelsen af HUI2 (0,87; sd. 0,18) eller HUI3 (0,87; sd.

0,14). Den gennemsnitlige nyttescore for AD-patienterne ved brug af HUI3 er 0,22 (sd 0,26) og for HUI2 er denne 0,53 (sd 0,21).

Kommentar:

- Kan anvendes til sundhedsøkonomiske analyser - både CEA og CUA
- HUI består pt. af to systemer HUI2 og HUI3, der tilsammen kan beskrive næsten 1.000.000 unikke helbredstilstande. (HUI2 definerer 24.000 og HUI3 972.000)
- Kan bruges hver for sig eller sammen
- HUI2 er at foretrække i situationer hvor generel information omkring personlig pleje er relevant
- HUIs nytter bliver målt via anvendelse af standard gamble metoden
- I (Neumann et al. 413-22) blev der fundet en multiattribut nyttescore for HUI3 på -0,36 og for HUI2 på -0,03 (skalaerne ankeret på 1 for perfekt helbred og 0 for død) – dvs. begge fremkommer med score for værre situationer end død
- HUI2's globale nyttescore varierede fra 0,73 (sd.0,15) i patienter med mulig AD til 0,14 (sd. 0,07) i patienter med terminal AD. De samme resultater for HUI3 var 0,47 (0,024) til -0,23 (0,08)
- HUI3 kan bedre opfange ændringer i svært svækkede end HUI2. Større sandsynlighed for at få problemer med gulveffekt med HUI2 end med HUI3
- Er et dyrt system at anvende. Minimum betaling er 3000\$ for brug af én HUI spørgeskemaversion
- Let at anvende og håndtere
- Indeholder spørgsmål vedrørende kognitiv funktion til forskel fra fx EQ-5D
- Få spørgsmål, hvilket er godt i forbindelse med denne population
- Der blev ikke fundet evidens for, at omsorgsydernes egne nytter varierede med patienternes sygdomsgrad (Neumann et al. 413-22). Dette kan dog reflektere ”...the inability of the HUI3 to detect subtle but important changes

in caregiver health-related quality of life, which may in turn be attributable to the heavy influence of attributes of physical health in the HUI3”
(Neumann et al. 413-22)

Yderligere information om HUITH; ansøgning om brug af HUI spørgeskemaer, algoritmer og ekstra materiale samt prisliste kan fås ved henvendelse til:

William Furlong
Health Utilities Inc.
88 Sydenham Street
Dundas, Ontario L9H 2V3 Canada
Telefon: (905) 525-9140, lokalnummer 22389; FAX: (905) 627-7914; E-Mail:
furlongb@mcmaster.ca

5.4 *Sickness Impact Profile (SIP)*

Formål: Dannelse af et mål for opfattet helbredsstatus, som er følsomt nok til at opfange ændringer eller forskelle i helbredsstatus som indtræffer over tid eller mellem grupper. Det er et instrument, der har til hensigt at fremkomme med mål for effekterne eller outcome af sundhedspleje (eng.: health care) som kan anvendes i forbindelse med evalueringer, programplanlægning og policy formuleringer (Bergner et al. 787-805). Tidsreferencerammen er her-og-nu vurderinger af helbredsstilstanden. Håndteringstiden er i Bergner (1981) angivet til at ligge mellem 20 – 30 min. ved interviewhåndtering.

Begrebslige grundlag: Det initiale arbejde med dannelsen af SIP begyndte i 1972. Bevæggrunden for dannelse af SIP var en observation af, at der var sket et skift i de udviklede lande: i stedet for kun at kigge på helbredelsen af en sygdom var fokus i stedet på en minimering af sygdommens indflydelse på de daglige aktiviteter. Det var derfor ikke nok at vurdere en behandlings virkningsfuldhed og effektivitet ud fra morbiditets- eller mortalitetsrater.

Ønsket med dannelsen af SIP er et instrumentet, der kan måle patientens helbredsstatus ved at vurdere, hvordan den pågældende sygdom ændrer patientens

daglige aktiviteter og adfærd. Fokus er på observerbare sygdomsrelaterede adfærdsændringer, og disse er indsamlet fra patienter, fra primære omsorgsydere og professionelle.

Beskrivelse: SIP er et funktionelt statusinstrument, der består af 136 udsagn, som beskriver sygdoms-relaterede adfærds dysfunktioner. Disse udsagn kan inddeles i 12 subskalaer; tre af disse subskalaer kan ligeledes danne en fysisk dimension, fire andre en psykosocial dimension.

Items fremstår således som en samling udsagn om ændret adfærd, hvor der for undersøgelsespersonens vedkommende skal tages stilling til, om det enkelte udsagn er rigtigt eller forkert i relation til deres helbred. Hvert item har en vægt som reflekterer den relative begrænsningssværhedsgrad, der er forbundet med itemet (Pollard and Johnston 921-34).

SIP totalscoren fremkommer ved at summere skalaværdien af alle de afkrydsede items, dividere denne sum med summen af værdien for alle items i SIP og gange den opnåede kvotient med 100. Scoren for hver kategori/dimension kan beregnes på tilsvarende måde. Jo højere score, der opnås, jo højere dysfunktion (Bergner et al. 787-805).

Oversigt over indhold i Sickness Impact Profile

Dimension	Kategori	Betegnelser	Antal udsagn	Max mulige score	Eksempel på udsagn
Fysisk	BCM	Body care and movement	23	2003	Du skal have hjælp til at kunne stå oprejst
	M	Mobility	10	719	Du ligger i sengen det meste af tiden
	A	Ambulation	12	842	Du kan ikke gå på trapper
Psykosocial	EB	Emotional behaviour	9	705	Du sukker og stønner ofte på grund af smerter eller ubehag
	SI	Social	20	1450	Du tager mindre på

		interaction			
Uafhængige kategorier	AB	Alertness behaviour	10	777	besøg hos venner og bekendte Du kan ikke koncentrere dig i længere tid om noget
	C	Comunication	9	725	Du er kun i stand til at skrive din underskrift
	SR	Sleep and rest	7	499	Du sover eller døser det meste af tiden
	HM	Home management	10	668	Du gør intet for at vedligeholde eller reparere noget
	W	Work	9	515	Du arbejder på deltid på grund af din helbredstilstand
	RP	Recreation and pastimes	8	422	Du bruger mindre tid til hobby og fornøjelser
	E	Eating	9	705	Du skal have hjælp til at spise

Kilde: (Folker 175-78) og (Pollard and Johnston 921-34)

Der er identificeret seks typer af problemer ved SIP: (a) *illogical scoring*, (b) *the nature and meaning of overall scores*, (c) *ambiguity of items*, (d) *the order of items*, (e) *the length of questionnaire* and (f) *the work category* (Pollard and Johnston 921-34). Flere af disse problemer er inter-relaterede.

Bl.a. har der været en del kritik af SIPs mest anvendte scoringsmetode, idet den fører til inkonsistente og ulogiske resultater (Pollard and Johnston 921-34). Et eksempel, gengivet i (Pollard and Johnston 921-34), er i forbindelse med hhv. en person, som overhovedet ikke kan gå (lam) og en person som kan gå med besvær (gigtplaget). Problemet er, at en person, som slet ikke kan gå vil være mindre begrænset i kategorien ”gåfunktion”, end en person, som kan gå med besvær. Det skyldes, at den lammede person kun kan bekræfte og opnå score på følgende items: *I do not walk at all; I get around in a wheelchair; I do not use stairs at all*, mens den gigtplagede person fx kan score på følgende items: *I walk shorter distances...*;

I walk more slowly; I walk by myself but with some difficulty...; I do not walk up or down hills; I go up or down stairs more slowly...; I use stairs only with mechanical support. Sammenlagt opnår den lammede patient en begrænsning i forhold til den bedst tænkelige funktion på 34 % mens den gigtplagede patient opnår en begrænsning på 37 % ved brug af de vægte, der er tillagt items. Ligeledes består de forskellige kategorier af ulige antal items. Dette betyder, at kategorierne ikke får samme vægt i totalscoren.

På baggrund af de problemer, der har været i forbindelse med den oprindelige scoringsteknik, har (Pollard and Johnston 921-34) foreslået en ny scoringsmetode som skulle tage højde for flere af disse forhold.

Psykometriske egenskaber i forbindelse med demente

Der er kun fundet ét studie (Krenz et al. 453-61), der meget basalt undersøger SIPs egenskaber i forbindelse med brug på demensramte. Data i (Krenz et al. 453-61) stammer fra ambulante patienter og deres familie. De inkluderede patienter har haft symptomer på kognitiv svækkelse i mindst 3 måneder. MMSEs middelværdien for 88 demensramte = $17,1 \pm 0,8$ og for 16 ikke-demensramte = $26,6 \pm 1,2$.

Validitet: Samtidig validitet for patienternes SIPs og familiemedlemmernes vurdering af patienternes SIPs. SIP scorerne blev sammenlignet med scorerne opnået på to validerede instrumenter, der vurderede demensens sværhedsgrad, nemlig MMSE (Folstein, Folstein, and McHugh 189-98) og modificeret Dementia Rating Scale (mDRS (Lawson, Rodenburg, and Dykes 153-59)). Variansen i mDRS og MMSE scorerne som kunne forklares af SIP subskalaerne anvendtes som et relativt mål for samtidig validitet. SIP subskalaerne var de uafhængige variable og hhv. mDRS eller MMSE scoren den afhængige variabel (multipel regression).

Afhængige variabel = mDRS. Kilde = patienterne, ikke-justeret korrelation: Initial vurdering: $r = 0,55$ ($p = 0,08$; $n = 57$), follow-up vurdering $r = 0,65$ (1 år; $p = 0,36$; $n = 30$). Kilde = omsorgsyderne: initial vurdering: $r = 0,87$ ($p < 0,001$; $n = 41$),

follow-up vurdering $r = 0,84$ ($p < 0,001$, $n = 48$). Afhængige variable = MMSE. Kilde = patienterne, ikke-justeret korrelation: Initial vurdering: $r = 0,58$ ($p = 0,04$; $n = 57$), follow-up vurdering $r = 0,63$ (1 år; $p = 0,31$; $n = 33$). Kilde = omsorgsyderne: initial vurdering: $r = 0,82$ ($p < 0,001$; $n = 41$), follow-up vurdering $r = 0,72$ ($p < 0,001$, $n = 46$). De justerede 'shrinkage'⁵ korrelationer var noget lavere og ikke-signifikante for patienterne, mens de kun var lidt lavere for omsorgsyderne og signifikante med undtagelse af follow-up med MMSE som afhængig variabel. De justerede værdier giver et relativt unbiased estimat af de multiple korrelationskoefficienter i populationen. De fremkomne resultater indikerer, at proxy respondenterne fremkommer med validt mål for patienternes funktionelle status mens patienterne ikke selv gør.

Known group validitet: SIP scorerne (subskala scorerne, dimensionsscorerne og totalscorerne) for hhv. AD og ikke-demensramte patienter. Forskellen mellem grupperne blev vurderet ved kvadrerede subskala koefficienter. Disse kan give indikation af hvilke subskalaer, der bedst diskriminerer mellem demente og ikke-demente. Ved initialvurderingen var der to subskalaer, 'Alertness Behavior' og 'Communication', der viste væsentlige forskelle mellem de to grupper. Ved follow-up vurderingen 1 år efter viste især 'Home Management' og 'Mobilitet' subskalaerne forskelle mellem de to grupper. Mindre forskelle på subskalaerne 'Recreation', 'Social Interaction' og 'Body Care and Movement', dvs. både fysiske og psykosociale dimensioner diskriminerede mellem demente og ikke-demente som følge af sygdommens progression.

Følsomhed over for ændringer: SIP scorerne (subskala scorerne, dimensionsscorerne og totalscorerne) for AD (baseline $n = 28$).

⁵ Shrinkage is a term that refers to the expected attenuation of sample multiple correlation coefficients relative to their population parameters. Since chance-level associations within a sample inflate the relationship between predictors and criterion, the magnitude of the original multiple correlation is expected to be larger than its value in the population. Shrinkage is especially likely to occur if the original multiple correlation is based on a large number of correlated predictors and a relatively small sample (Krenz et al. 453-61)

Konfidensintervaller for forskelle mellem initial og follow-up SIP score viser, at subskalaerne: Body Care and Movement, Mobility, Ambulation, og Home Management viser signifikant svækkelse over tid for de demente.

(Sano et al. 1137-41) har i en undersøgelse af levocecarpine ikke fundet evidens for, at SIP kan måle forbedringer i livskvaliteten i medicinske interventionsstudier.

Kommentar

- 'Work' subskalaen er upassende for geriatriske (Krenz et al. 453-61) og dermed også demente patienter. Problem, da ingen items vil blive afkrydset i denne. Hermed opnås en score lig nul, hvilket indikerer, at individerne ingen begrænsninger har i denne kategori (Pollard and Johnston 921-34)
- Er et langt instrument. En besvarelse af de mange spørgsmål medfører øget risiko for upålidelige besvarelser/missing data
- Items er primært rettet imod observerbare adfærdsændringer – øger pålideligheden for proxyvurderingen
- Proxy-respondenters (familie) vurdering er mere valid end patienternes egen vurdering – især over tid (Krenz et al. 453-61)
- Registrerer om en sygdom har medført ændringer i patientens adfærd – ikke om patienten føler sig generet af de ændringer, som sygdommen har medført
- Ulogisk scoringsteknik såfremt man ønsker dannelse af indeks
- Er en amerikansk spørgeskemametode baseret på et empirisk udviklet helbredsbegreb, som er i overensstemmelse med WHO's sundhedsdefinition af optimal fysisk, psykisk og social funktion og velfærd
- Er dog ikke decideret et livskvalitetsinstrument, da det ikke er patienternes subjektive oplevelse af deres fysiske, psykiske og sociale trivsel, der måles. I stedet er det det fysiske, psykiske og sociale funktionsniveau, der måles (Folker et al. 2352-54)
- Der eksisterer en version, dannet til plejehjemsbeboere (Gerety et al. M2-M8), der indeholder 66 items. Løser dog ikke problemet med ønsket om et generisk instrument, der kan anvendes på demente

- <http://www.outcomes-trust.org/instruments.htm#SIP>

5.5 15 Dimensioner (15D)

Formål: 15D er et generisk, multidimensionalt, standardiseret, selvhåndteret helbredsrelateret livskvalitetsinstrument, som beskriver individers helbredsstatus.

Begrebslige grundlag: 15D er et finsk udviklet helbredsrelateret livskvalitetsinstrument, der holder sig til WHO's helbredsdefinition. Som følge heraf er 15D sammensat af fysisk (funktion af det psyko-fysiske system), mental og social velvære (social funktion i sædvanlige roller og opgaver).

Beskrivelse: De 15 dimensioner, der er indeholdt i 15D, er: bevægelighed, syn, hørelse, vejrtrækning, søvn, spisning, tale, vandladning/afføring, sædvanlige aktiviteter, mentale funktion, ubehag og symptomer, depression, stress, livskraft, seksuel aktivitet. Hver dimension er inddelt i en 5-punktsskala, der afspejler alvorligheden af dimensionen. Det er et instrument, der både kan danne en profil eller give en enkelt indeksscore. Indeksscoren opnås ved at inkorporere populationsbaserede præferencevægte til dimensionerne. 15D indlemmer med de 15 dimensioner og 5 udfald over 30 mio. mulige forskellige helbredsstilstande, hvorfor der teoretisk er tale om et meget følsomt instrument.

Psykometriske egenskaber i forbindelse med demenspatienter:

Der er ikke fundet studier, der har anvendt 15D i rene demensstudier. Instrumentet er dog brugt og fundet validt i forbindelse med livskvalitetsvurderingen af 260 ældre (gns. 66,5 år gamle; SD 10,6) patienter med Parkinsons syge, en sygdom, der for mange medfører kognitive dysfunktioner, således var 25 af patienterne ligeledes demente. Responsraten var høj (99,2 %), og 84,9% havde besvaret alle 15 items. Patienter, der ikke havde udfyldt hele spørgeskemaet var ældre (gns. alder 73,5 versus 65,4 år), mens sværhedsgraden af Parkinson eller demens ikke havde

nogen påvirkning på responsraten. I alle tilfælde med manglende data var item 15, vedrørende seksuel aktivitet, ikke besvaret.

Known group validitet: Demens resulterer i dårligere HRQoL scorer: gennemsnitlig 15D score for ikke-demensramte var 0,78 (0,1) og for demensramte 0,65 (0,1); $p < 0,001$.

Kommentar:

- Item 15 vedrørende seksuel aktivitet er vanskelig at besvare, da mange ældre er enlige og/eller ikke længere seksuelt aktive.
- Alder korrelerer signifikant med 15D scoren

5.6 Short Form – 36 (SF-36)

Formål: Dannelse af en standardiseret, generisk indikator for helbredsstatus, der kan anvendes i klinisk praksis, i forskning, i sundhedspolitiske evalueringer og i generelle populationsundersøgelser (Ware and Sherbourne 473-83). SF-36 kan enten håndteres af patienten selv, via interview (ansigt til ansigt eller via telefon), eller computer (Ware 2062). Instrumentet kan håndteres på individer ældre end 14 år.

Begrebslige grundlag: SF-36 blev dannet, så den kunne tilfredsstille mindstekravet for de psykometriske standarder, der er nødvendige til gruppesammenligninger. De 8 helbredsbegreber, som SF-36 består af, blev valgt ud fra de 40 begreber (og 151 items), som var indeholdt i 'The Medical Outcomes Study (MOS)', der havde sit udgangspunkt i 'The Health Insurance Experiment' (Brook et al. 1-&). De fleste items i SF-36 har derfor rødder i instrumenter, som har været i brug siden 1970'erne og 80'erne.

Beskrivelse: SF-36 består af 36 items, der er inddelt i otte separate multiitem-skalaer og en enkelt 'Ændringer i alment helbred (HT)' spørgsmål. De 8 multiitem-

skalaer er: fysisk funktion (PF), begrænsninger – fysisk betinget (RP), fysisk smerte (BP), alment helbred (GH), energi (VT), social funktion (SF), begrænsninger – psykisk betinget (RE), psykisk velbefindende (MH). Fem skalaer (PF, RP, RE, SF, BP) måler grad af dysfunktion, mens de tre øvrige skalaer (MH, VT, GH) omfatter hele spektret fra negative til positive helbredstilstande (Bjørner et al.). Alle skalaerne scores, så en høj score indikerer en positiv helbredsstatus. Råscorene i hvert domæne transformeres til en score, som repræsenterer respondentens relative position på et kontinuum mellem 0 og 100 %.

Formlen for transformationen af den rå sumscore er:

$$\text{Transformeret skala} = \left[\frac{(\text{Aktuelle råscore} - \text{lavest mulige råscore})}{(\text{Højest mulige råscore} - \text{lavest mulige råscore})} \right] \times 100$$

Ufuldstændige domæner kan beregnes, givet at mindst 50 % af items i skalaen er tilstedeværende. I disse tilfælde kan en aritmetisk middelværdi udregnes fra værdier af de items, som er vurderet af patienten, og anvendes som en tilskrevet værdi for de manglende items. For en yderligere beskrivelse af scoringsmetoden henvises til den danske manual (Bjørner et al.).

En række nye metoder til scoring af SF-36 er udviklet. Fx har det vist sig, at der er en høj grad af samvariation mellem skalaerne. Dette gælder fire skalaer, som primært måler fysisk helbred (PF, RP, BP og GH) og fire skalaer som primært måler psykisk helbred (VT, SF, RE og MH). Således kan 82 % af den reliable variation i de otte skalaer forklares af de to komponenter. Disse resultater har ledt frem til en alternativ måde at præsentere data på – hhv. et generelt mål for den fysiske helbreds-komponent (PCS, Physical Component Summary) og et generelt mål for den psykiske helbreds-komponent (Mental Component Summary). For yderligere uddybning se (Ware, Kosinske, and Keller; Ware et al. AS264-AS279; Bjørner et al.) Denne tilgang bør dog ikke anvendes, hvis man har missing data.

En anden metode er anvendelse af item respons modeller til estimering af helbredsstatus. I stedet for almindelig Likertskalering, foreslås i stedet Rasch score (en-dimensionalitet). Se (Haley, Mchorney, and Ware 671-84).

Som følge af sundhedsøkonomers ønske om nytteindeks, har man ligeledes forsøgt at udlede et præferencebaseret indeks baseret på SF-36, se (Brazier et al. 1115-28; Brazier, Roberts, and Deverill 271-92). Denne version kaldes SF-6D og kan definere 18.000 helbredsstilstande. SF-6D består af seks dimensioner: fysisk funktion, rolle begrænsning, social funktion, smerte, mentalt helbred, og vitalitet, og kan anvende data fremskaffet ved SF-36.

Psykometriske egenskaber: Der er fundet et enkelt studie (Novella et al. 362-69), der har undersøgt SF-36s egenskaber på 138 demensramte patienter (forskellige demensformer og demensgrad), rekrutteret fra 16 centre i hhv. Frankrig, Schweiz og Belgien.

Reliabilitet: Intern konsistens, målt ved Cronbach's alpha. For PF, RP, RE, BP og GH blev der fundet høje alpha koefficienter: alpha lå mellem 0,74 og 0,92 (PF). Alpha koefficienten for hhv. MH, SF og VT var lavere men acceptable – hhv. 0,69, 0,63 og 59.

Test-retest reliabiliteten blev vurderet på data fra 39 tilfældigt udvalgte demensramte patienter. ICC var dårligst for RE, hvor denne var lig 0,17. For MH og BP var ICC hhv. 0,45 og 0,49. De resterende skalaer fremkom med acceptable ICC scorer mellem 0,51 (HT) og 0,81 (PF). Hvis fokus blev begrænset til de patienter, der havde MMSE score ≥ 15 (n = 18), blev reproducerbarheden acceptabel til god for alle 8 subskalaer (omfang 0,53 – 0,90).

Interrater reliabilitet, vurderet ved overensstemmelsesgraden (ICC, n = 125) viste følgende: Mellem patienter og familiær proxy lå koefficienterne mellem 0,10 (SF) – 0,38 (PF). Mellem patienter og professionel omsorgsyder lå koefficienterne mellem 0,00 (SF) – 0,39 (PF). Mellem familiær proxy of professionelle omsorgsydere lå koefficienterne mellem 0,22 (VT) – 0,66 (PF). Begrænset til

patienter med MMSE score ≥ 15 ($n = 52$) var overensstemmelseskoefficienterne generelt større for alle subskalaerne.

Validitet: Den interne begrebsvaliditet blev vurderet ved principalkomponent analyse på data fra alle 125 demensramte. To komponenter kunne forklare 66% af den totalt målte varians. Man fandt ikke samme faktor struktur, som teoretiseret af Ware i tidligere studier (hhv. fysisk og mental sundhedsdimensioner).

Ekstern begrebsvaliditet: SF-36's evne til at sondre mellem individer med og uden en ekstra kronisk sygdom ($n = 36$ uden ekstra kronisk sygdom; $n = 89$ med ekstra kronisk sygdom) var signifikant for 5 af de 8 subskalaer (PF, RP, BP, RE, SF) ($p = 0,003 - p = 0,0004$).

Gennemførlighed (feasibility) og acceptabilitet: 125 ud af i alt 138 demensramte individer indvilligede i at udfylde spørgeskemaet. 27,2 % af de demensramte kunne selv håndtere dette. Den gennemsnitlige MMSE score for disse var 20, ligesom disse havde flere års skolegang. Ellers var hjælp fra interviewer påkrævet i 72,8 % af tilfældene.

Den gennemsnitlige håndteringstid var 18,8 min (SD 12). Problemer med missing data fra individerne varierede fra 1,8 – 19 %. Hvis individer med MMSE score < 15 blev ekskluderet, varierede non-responsraten for items mellem 0 – 13 %. Non-responsraten for subskalaerne lå mellem 1,6 – 8,8 % for alle patienter. Hvis patienterne med MMSE score < 15 blev ekskluderet, lå non-responsraten for subskalaerne mellem 0 og 3,8 %.

Kommentar:

- Relativt kort instrument, der har udvist gode egenskaber (god intern konsistens, test-retest reliabilitet med undtagelse af RE, gennemførlighed og acceptabilitet)
- Kan bringes i anvendelse på mildt til moderat, men ikke svært demensramte patienter

- Responsraten er relateret til alder, aktivitetsbegrænsning og kognitiv svækkelse. Interview håndtering øger responsraten (Parker et al. 13-18; Novella et al. 362-69)
- Tre spørgsmål, der repræsenterede 11 skalaitems stod i (Parker et al. 13-18) for 41 % af de manglende data blandt ældre patienter (og ikke kun kognitivt svækkede). Det drejede sig om spørgsmål 4 (14 %), som omhandler problemer med arbejde eller andre daglige aktiviteter som følge af fysisk helbred; spørgsmål 5 (14 %), som omhandler problemer med arbejde eller andre daglige aktiviteter som følge af emotionelle problemer; og spørgsmål 11 (14 %), som søger opfattelsen af helbredsstatus
- Problem med manglende data er i Danmark særlig udtalt for skalaen for alment helbred blandt personer over 66 år, hvor der kun har kunnet beregnes en skalaværdi for 79 % af de postomdelte spørgeskemaer
- Har scoringsrutine, som tillader op til 50 % manglende data
- Spørgsmål 2 forventer, at respondenterne skal kunne huske hvordan helbredet var for 1 år siden – kan være svært med pålidelig respons fra dement
- Tidsreferencerammen er de foregående 4 uger – igen kan der være problemer med hukommelsen
- Håndteringstiden for ældre er ca. 15 min. (McDowell and Newell), for demente lidt længere
- Et problem kunne være at kognitive funktionsitems ikke medtaget
- Problemer med gulv- og lofteffekter i visse subskalaer. Procentvis andel af svarpersonerne med score 100 er i 'Social funktion' = 63,8 % og i 'Begrænsninger – psykisk betinget' = 57,5 %. I 'Begrænsninger – fysisk betinget' er den procentvise andel af svarpersoner med score 0 (gulveffekt) = 33,4, og med score 100 = 37 %. For skalaen 'fysisk smerte' er lofteffekten = 39,1 (Bjørner et al.). (Novella et al. 362-69) finder ligeledes problemer med gulv og loft-effekter i enkelte af subskalaerne. For

gulveffekter drejer dette sig om RE, RF og PF og for lofteffekter gælder dette subskalaerne RE, RF, SF og BP

- Gulv- og lofteffekter begrænser SF-36 anvendelse i interventionsstudier
- Forekomsten af logiske fejl er lav (Bjørner et al.), hvilket generelt tyder på letforståelige spørgsmål. Ophobning af fejl, der involverer items om fysisk funktion
- Er ikke oprindeligt designet til brug i økonomiske evalueringer eller til udledningen af QALY

Man kan få det danske SF-36 spørgeskema samt tilladelse til at anvende skemaet ved henvendelse til en af nedenstående:

Jakob Bjørner
Afdeling for Sundhedstjenesteforskning
Institut for Folkesundhedsvidenskab
Panum, Blegdamsvej 3
DK-2200 København N
Telefon 35 32 79 80
Email: J.B.Bjorner@socmed.ku.dk

Per Bech
Forskningsenheden
Psykiatrisk Sygehus
Dyrehavevej 48
DK-3400 Hillerød
Telefon: 48 29 32 53
Email: p.bech@vip.cypercitiy.dk

5.7 WHOQOL

Formål: Dannelse af et internationalt, selvhåndteret, helbredsrelateret, kumulativt livskvalitetsmåleinstrument, der gør det muligt at udføre livskvalitetsstudier i forskellige kulturelle omgivelser samtidigt.

Begrebslige grundlag: WHO definerer helbred som “*A state of complete physical, mental, and social well-being not merely the absence of disease...*”. Det følger heraf, at målingen af helbred og effekterne af pleje ikke blot må indeholde en indikation af ændringer i frekvensen og sygdommens sværhedsgrad, men også en estimering af velvære, hvilket kan vurderes ved måling af de forbedringer, der er i livskvaliteten relateret til plejen.⁶ WHO's definition for livskvalitet er: “*individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns*”(Kuyken et al. 1403-09). Det er ud fra denne livskvalitetsdefinition, items i WHOQOL er vurderet.

WHOQOL er resultatet af et tværkulturelt samarbejde, der havde til formål at udvikle et helbredsrelateret livskvalitetsinstrument, der kunne anvendes globalt. Der var 15 centre involveret i udviklingsprocessen. Disse jævnt fordelt over hele kloden. Items blev genereret ud fra eksisterende viden om livskvalitet, samt de synspunkter, praktiserende sundhedsprofessionelle, patienter (indlagte og ambulante) samt raske individer gav udtryk for. En item-pool fra de 15 centre bestod af ca. 1800 items, som dannede udgangspunktet for dannelsen af WHOQOL.

Beskrivelse: WHOQOL består af 100 items, der tilsammen dækker 6 livskvalitetsdimensioner: en fysisk dimension, en psykologisk dimension, en uafhængighedsdimension, en social dimension, en miljømæssig dimension og en spirituel dimension. Hver dimension er inddelt i underskalaer, der kaldes 'facetter'

⁶ http://depts.washington.edu/yqol/docs/WHOQOL_Info.pdf

(Kuyken et al. 1403-09). I alt er der 24 facetter, som hver indeholder 4 items. Herudover indeholder instrumentet også 4 generelle items. Hvert item vurderes på en 5-punktsskala, og hver facet beregnes ved at summere itemscorene. Scorene for facetterne varierer derfor mellem 4 – 20, hvor højere score betegner højere livskvalitet. Dette gælder dog ikke for facetterne 1, 8 og 11, der har omvendt scoring (Power et al. 1569-85).

Oversigt over WHOQOLs facetter

Facet 1	Pain and Discomfort	Facet 12	Working capacity
Facet 2	Energy and Fatigue	Facet 13	Personal relationships
Facet 3	Sleep and Rest	Facet 14	Social support
Facet 4	Positive feelings	Facet 15	Sexual activity
Facet 5	Thinking, learning, memory and concentration	Facet 16	Physical safety and security
Facet 6	Self-esteem	Facet 17	Home environment
Facet 7	Bodily image and appearance	Facet 18	Financial resources
Facet 8	Negative feelings	Facet 19	Health and social care: availability and quality
Facet 9	Mobility	Facet 20	Opportunities for acquiring new information and skills
Facet 10	Activities of daily living	Facet 21	Participation in and new opportunities for recreation/leisure
Facet 11	Dependence on medication or treatments	Facet 22	Physical environment: (pollution/noise/traffic/climate)
		Facet 23	Transport
		Facet 24	Spirituality/religion/personal beliefs

Kilde: (Power et al. 1569-85)

Psykometriske egenskaber: Der er fundet et enkelt studie, (Struttmann et al. 273-79), som i et pilotstudie har undersøgt om WHOQOL-100 kan anvendes til at vurdere hhv. demente (n = 27 moderat demente, MMSE > 15) og kræftframtes (n = 30) livskvalitet – om der dannes divergente profiler for de to grupper.

Reliabilitet: Intern konsistens: alphaværdierne for de enkelte facetter lå mellem 0,65 – 0,93. Laveste værdier for facetterne 13 (alpha = 0,68) og 22 (alpha = 0,65), der lå lige under grænsen på 0,70 for acceptabel alfaværdi; højeste værdier for facetterne 11 (alpha = 0,91) og 12 (alpha = 0,93), der lå lige over grænsen på 0,90 (tegn på overflødige items).

Test-retest: (2 uger, patientvurdering, Pearson's r): Rimelig stabilitet i dimensionerne psykologi (r = 0,725) og autonomi (r = 0,698) for AD patienternes vurderinger. Lave korrelationer i fysisk (r = 0,095) og social interaktions (r = -0,196) dimensionerne (Struttmann et al. 273-79).

Validitet: Known group: de to grupper (demente og kræftsyge) scorede signifikant forskelligt på flere af facetterne og spørgsmålene (Struttmann et al. 273-79).

Kommentar:

- Der er flere facetter, som er vurderet ubrugbare på ældre. WHO har derfor gang i dannelsen af en version (WHOQOL-old), der netop kan bringes i anvendelse på ældre. Instrumentet består fortsat af 100 items, men de problematiske omskrives, så de passer bedre overens med denne aldersgruppe
- Der eksisterer kun begrænset evidens for instrumentets anvendelighed på demente
- Der eksisterer en kortere, generisk version af WHOQOL, kaldet WHOQOL-BREEF. Har dog de samme problemer som WHOQOL i forbindelse med brug på ældre
- <http://www.acpmh.unimelb.edu.au/whoqol/instruments.html>

5.8 *Quality of Well-Being (QWB)*

Formål: Dannelse af et generelt, helbredsrelateret, interviewerhåndteret livskvalitetsspørgeskema, der kan tilvejebringe et estimat for værdien af helbredsstatus på et bestemt tidspunkt, som er nødvendig i cost-utility analyser. Kan ligeledes anvendes i kliniske forsøg og populationsstudier (Kerner et al. 44-61). Tidsreferencen er normalt de seneste 6 dage.

Begrebslige grundlag: QWB er del af en generel helbreds policy model, der er fremkommet ud fra selvstændige teorier i hhv. økonomi, psykologi, medicin, og almen helbred (eng.: public health) (Kaplan et al. 509-20). Modellen for helbredsstatus indeholder komponenter for mortalitet (død), morbiditet (helbredsrelateret livskvalitet) og tid. Rationalet for modellen er, at sygdomme og manglende evner er vigtige af to årsager: 1) sygdom kan medføre, at den forventede levetid bliver afkortet; 2) sygdom kan gøre livet mindre ønskeligt. QWB er metoden, der anvendes til målingen af livskvalitet til beregning i modellen.

Beskrivelse: QWB er et præference-vægtet mål der kombinerer tre funktionsskalaer (hhv. mobilitet, fysisk aktivitet, og social aktivitet) med et mål for symptomer og problemer (valg mellem 27 items) til fremskaffelsen af et point-in-time udtryk for trivsel, som løber fra 0 (død) til 1 (for asymptomatisk fuld funktion). Modellen separerer aspekter for hhv. helbredsstatus og livskvalitet ind i adskilte komponenter. Disse er forventet levetid (mortalitet), funktion og symptomer (morbiditet), præferencer for observerede funktionstilstande (nytter) og varigheden af helbredstilstanden (progression). Udover klassificeringen i observerbare funktionsniveauer, bliver individer også klassificeret ved symptomer eller problemer (Kaplan et al. 509-20).

Quality of Well-Being classification system

PART 1. Quality of Well-Being/general health policy model: function scales with step definitions		Part 2. Quality of Well-Being/general health policy model: symptom/problem complexes (CPX)	
Step no.	Step definition	CPX no.	
	Mobility Scale (MOB)	1	Death (not on respondent's card)
5	No limitations for health reasons	2	Loss of consciousness such as seizure (fits), fainting, or coma (out cold or knocked out)
4	Did not drive a car, health related; did not ride in a car as usual for age (younger than 15yr), health related, and/or did not use public transportation, health related; or had or would have used more help than usual for age to use public transportation, health related	3	Burn over large areas of face, body, arms, or legs
		4	Pain, bleeding, itching, or discharge (drainage) from sexual organs – does not include normal menstrual (monthly) bleeding
		5	Trouble learning, remembering, or thinking clearly
2	In hospital, health related	6	Any combination of one or more hands, feet, arms, or legs either missing, deformed (crooked), paralyzed (unable to move), or broken – includes wearing artificial limbs or braces
	Physical Activity (PAC)	7	Pain, stiffness, weakness, numbness or other discomfort in chest, stomach (including hernia or rupture), side, neck, back, hips, or any joints or hands, feet, arms, or legs
4	No limitations for health reasons	8	Pain, burning, bleeding, itching, or other difficulty with rectum, bowel movements, or urination (passing water)
3	In wheelchair, moved or controlled movement of wheelchair without help from someone else; or had trouble or did not try to lift, stoop, bend over, or use stairs or inclines, health related; and/or limped, used a cane, crutches, or walker, health related; and/or had any other physical limitation in walking, or did not try to walk as far or as fast as other the same age are able to do, health related	9	Sick or upset stomach, vomiting or loose bowel movement, with or without chills, or aching all over
		10	General tiredness, weakness, or weight loss
		11	Cough, wheezing, or shortness of breath, with or without fever, chills, or aching all over
		12	Spells of feeling, upset, being depressed, or of crying
		13	Headache, or dizziness, or ringing in ears, or spells of feeling hot, nervous, or shaky

1	In wheelchair, did not move or control the movement of wheelchair without help from someone else, or in bed, chair, or couch for most or all of the day, health related	14	Burning or itching rash on large areas of face, body, arms, or legs
		15	Trouble talking, such as lisp, stuttering, hoarseness, or being unable to speak
		16	Pain or discomfort in one or both eyes (such as burning or itching) or any trouble seeing after correction
	Social Activity Scale (SAC)	17	Overweight for age and height or skin defect of face, body, arms, or legs, such as scars, pimples, warts, bruises, or changes in colour
5	No limitations for health reasons		
4	Limited in other (e.g., recreational) role activity, health related	18	Pain in ear, tooth, jaw, throat, lips, tongue, several missing or crooked permanent teeth – includes wearing bridges or false teeth, stuffy, runny nose; or any trouble hearing – includes wearing a hearing aid
3	Limited in major (primary role activity, health related		
2	Performed no major role activity, health related, but did perform selfcare activities	19	Taking medication or staying on a prescribed diet – for health reasons
		20	Wore eyeglasses or contact lenses
		21	Breathing smog or unpleasant air
1	Performed no major role activity, health related, and did not perform or had more help than usual in performance of one or more selfcare activities, health related	22	No symptoms or problem (not on respondent's card)
		23	Standard symptom/problem
		24	Trouble sleeping
		25	Intoxication
		26	Problems with sexual interest or performance
		27	Excessive worry or anxiety

Kilde: Gengivet med tilpasning fra (Drummond et al.)

Psykometriske egenskaber i forbindelse med demente:

Der er fundet to studier, der fokuserer på validitetsaspekter ved QWB i forbindelse med anvendelse på demensramte individer. I (Silberfeld et al. 71-79) vurderes indholdsvaliditeten (jf. afsnit, 3.2), og i (Kerner et al. 44-61) vurderes begrebsvaliditeten. Kerner et al. har anvendt QWB på 211 patient-ægtefælle par og kontrol par. 159 var AD-patienter (diagnosticeret mulig eller med sikkerhed) og 52 var kontrolgruppe. Følgende validitetsresultater stammer derfra.

Validitet: Evidens for QWBs begrebsvaliditet blev vurderet via korrelationer mod forskellige mål for demens, hukommelses- og adfærdsproblemer, samt psykiatriske symptomer [Mattis Dementia Rating Scale' (MDRS: undersøgelse af kognitiv funktion, (Coblentz et al. 299-308), 'Memory and Behavior Problem Checklist (MBPC: består af 29 ofte sete problemer, (Zarit, Reeve, and Bach-Peterson 655), Brief Symptom Inventory (BSI: psykiatriske symptomkvaler (DEROGATILR et al. 1-15)]. Patienter med dårligere kognitiv funktion, målt med MDRS, havde lavere QWB score ($r = 0,52$, $p < 0,01$), ligesom lavere QWB scorer var forbundet med større adfærdssvækkelse, målt ved MBPC ($r = 0,64$, $p < 0,01$). QWB scorerne var også signifikant forbundet med psykiatriske kvaler ($r = -0,26$, $p < 0,01$).

Kommentar:

- Kritikken af QWB har primært været, at det er et dyrt og vanskeligere instrument at håndtere end de fleste andre HRQOL instrumenter (kræver trænet interviewer), ligesom det er et langt og komplekst spørgeskema. Dette har ført til dannelse af en selvhåndterbar version, der ligeledes kan anvendes til frembringelse af QALY (QWB-SA). Indeholder flere items, der vurderer mentalt helbred (kritiseret mangel ved QWB). Tidsreferencen er de foregående 3 dage. Denne version kan ligeledes håndteres ved interview (face-to-face eller telefon), psykometriske egenskaber er dog ikke vurderet specifikt ved disse tilgange. Normal håndteringstid: ca. 10 min.
- Proxy vurdering anbefales ikke ved brug af QWB-SA, da mange af de vurderede symptomer er subjektive og personlige af natur. Forfatterne erkender dog, at det kan være nødvendigt, hvis instrumentet skal anvendes på hjerneskadede eller demente
- Brugen af QWB-SA er gratis for non-profit organisationer (se brugermanual), og er valideret til brug på ældre
- (Kerner et al. 44-61) finder i deres studie, at i kvalitetsjusteret overlevelsesanalyse, så ligger AD patienter på ca. 0,5QALY per overlevet år

- Generelt ikke oplevet problemer med gulv- eller lofteffekter (Kaplan et al. 509-20)
- <http://www.outcomes-trust.org/instruments.htm#QWB>
- Kan ligeledes fremkomme med profiler, men er knap så ”god” som fx SF-36 i denne henseende (Kaplan et al. 509-20)

6 Sammenfatning

De fleste er enige om, at mild til moderat demente patienter selv kan vurdere egen livskvalitet – og bør gøre dette. Som konsekvens af de variable evner demenspatienter besidder, kan det dog være vanskeligt at beslutte, hvornår det vil være behørigt at indsamle proxyinformationer om patientens tilstand frem for indsamle information hos patienten selv.

Hvem, der er den bedste proxy for patienten – og om proxyvurderinger er valide, er der ligeledes ikke enighed om. Validering af proxyvurderingen af ældre svækkede individer er især vanskelig, da der ikke eksisterer nogen klar standard, som man kan anvende til at validere proxyvurderinger imod. Ligeledes vanskeliggøres validitetsstudier af, at validitetskriteriet typisk er patientvurderinger. Idet patienterne er vurderet værende ikke-reliable, pga. deres kognitive tilstand, bliver vurderingen af validiteten meget vanskelig. Generelle fund i studier, der har gjort et forsøg på at vurdere overensstemmelsen mellem patient og proxyvurderinger finder, at proxyvurderingerne er mere pessimistiske end individernes selvurderinger. Således undervurderer de fleste proxy vurderinger patienternes egen vurdering af fysisk og psykisk livskvalitet.

Hvem der agerer proxy for patienten, har oftest også indflydelse på overensstemmelsen af patientens og tredjepartsvurdering. Således fremkommer der i visse studier resultater, der peger på at sundhedspersonalet er bedre end en familiær proxy, mens andre studier fremkommer med det modsatte resultat. Der er således ikke nogen entydige indikator for, hvilken type proxy respondent man bør anvende. Man finder dog, at man ikke bør blande data: hvis man anvender proxy respondenter i forbindelse med indsamlingen, skal dette gøres på alle patienter og ikke kun for de mest svækkede.

Præferencebaserede generiske instrumenter anvendes ofte til vurdering af interventioners indflydelse på helbredsrelateret livskvalitet og anvendes ofte i økonomiske evalueringer. Disse redskaber er baseret på et multidimensionalt helbredstilstandsklassifikationssystem, hver med et sæt af præferencevægte. De genererer en enkelt indeksscore som kan anvendes som et mål for præferencebaseret helbredsrelateret livskvalitet, eller kan kombineres med leveår til brug i økonomiske analyser (Conner-Spady and Suarez-Almazor 791-801).

Generelt set måler de forskellige instrumenter det samme underliggende begreb, indikeret ved moderat til høje korrelationer mellem instrumenterne. Dette er dog ikke ensbetydende med, at de fremkommer med samme resultater og/eller er lige anvendelige på alle populationstyper. Bl.a. kan forskellige indeksscorer potentielt fremkomme med forskellige cost-utility ratioer. Dette vises bl.a. i (Conner-Spady and Suarez-Almazor 791-801), som har anvendt hhv. EQ-5D, SF-6D og HUI3 i et længerevarende studie af en gruppe primært gigtplagede patienter. Resultaterne viser, at indeksscorerne ikke kan udveksles i beregningen af longitudinal-baserede QALYs, da de ikke har ækvivalente middelværdier og varians, ligesom de viser forskellig ændring over tid. Ligeledes er den begrebslige forståelse af helbredsrelateret livskvalitet forskellig.

Herudover viser resultaterne i et studie af Neumann et al (Neumann et al. 413-22), at selv beslægtede instrumenter som HUI2 og HUI3, der indeholder mange af de samme kvaliteter, fremkommer med forskellige nyttescore. I dette tilfælde skyldes forskellene i scorerne primært hvordan systemerne behandler tilstande værre end død, hvilket i visse situationer kan få indflydelse på den inkrementielle cost-effectiveness ratio: *While it does not immediately follow that incremental cost-effectiveness ratios will differ in all cases, there is the potential for that to happen. Thus, if a scoring system has a wider range of negative scores [som fx HUI3], and if the disease or problem involves a lot of morbidity, the cost per QALY will probably differ, whereas if the disease or condition involves only minor morbidity, the results would not differ very much. The differential scoring illustrates another*

complication for comparing results and for constructing league tables (Neumann et al. 413-22).

Der er derfor flere overvejelser, som må foretages inden valg af instrument træffes. Det primære må dog bero på, at instrumentet begrebsligt indeholder komponenter for livskvalitet, som er relevante for den population, instrumentet skal bringes i anvendelse på.

7 Referenceliste

- Ankri, J. et al. "Use of the EQ-5D among patients suffering from dementia." Journal of Clinical Epidemiology 56.11 (2003): 1055-63.
- Bergner, M. et al. "The Sickness Impact Profile - Development and Final Revision of A Health-Status Measure." Medical Care 19.8 (1981): 787-805.
- Bergner, M. et al. "Sickness Impact Profile - Validation of A Health Status Measure." Medical Care 14.1 (1976): 57-67.
- Bjorner, J. B. et al. "Tests of data quality, scaling assumptions, and reliability of the Danish SF-36." Journal of Clinical Epidemiology 51.11 (1998): 1001-11.
- Bjorner, J. B. et al. "The Danish SF-36 Health Survey: Translation and preliminary validity studies." Journal of Clinical Epidemiology 51.11 (1998): 991-99.
- Bjørner, JB et al. Dansk Manual Til SF-36. Et Spørgeskema Om Helbredsstatus. Lif Lægemedelindustriforeningen, 1997.
- Brazier, J., J. Roberts, and M. Deverill. "The estimation of a preference-based measure of health from the SF-36." Journal of Health Economics 21.2 (2002): 271-92.
- Brazier, J. et al. "Deriving a preference-based single index from the UK SF-36 Health Survey." Journal of Clinical Epidemiology 51.11 (1998): 1115-28.
- Brod, M. et al. "Conceptualization and measurement of quality of life in dementia: The dementia quality of life instrument (DQoL)." Gerontologist 39.1 (1999): 25-35.
- Brook, R. H. et al. "Overview of Adult Health-Status Measures Fielded in Rands Health-Insurance Study." Medical Care 17.7 (1979): 1-&.
- Coblentz, J. M. et al. "Presenile Dementia - Clinical Aspects and Evaluation of Cerebrospinal-Fluid Dynamics." Archives of Neurology 29.5 (1973): 299-308.
- Cohen, J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum, 1988.
- Conner-Spady, B. and M. E. Suarez-Almazor. "Variation in the estimation of quality-adjusted life-years by different preference-based instruments." Medical Care 41.7 (2003): 791-801.
- Coucill, W. et al. "EQ-5D in patients with dementia - An investigation of inter-rater agreement." Medical Care 39.8 (2001): 760-71.

- DEROGATI.LR et al. "Hopkins Symptom Checklist (Hscl) - Self-Report Symptom Inventory." Behavioral Science 19.1 (1974): 1-15.
- Drummond, MF et al. Methods for the Economic Evaluation of Health Care Programmes. 2.nd ed. Oxford University Press, 1997.
- Folker, H. "Sickness impact profile - SIP. En metode til vurdering af helbred." Ugeskrift for læger 149 (1987): 175-78.
- Folker, H et al. "Undersøgelse af målekvalitet af sickness impact profile." Ugeskrift for læger 152 (1990): 2352-54.
- Folker, H. and B. M. Jensen. "Study of selected methods of self-evaluation of health, the quality of life, and satisfaction with treatment: use in patients four weeks after discharge from a psychiatric ward." Ugeskrift for Laeger 163.24 (2001): 3347-52.
- Folland, S, AC Goodman, and M Stano. The Economics of Health and Health Care. 3ed ed. ed. Upper Saddle River: Prentice-Hall, 2001.
- Folstein, MF, SE Folstein, and PR McHugh. "Mini-Mental State: a practical method for grading the cognitive state of patients for the clinicial." Journal of Psychosomatic Research 12 (1975): 189-98.
- Furlong, W. J. et al. "The Health Utilities Index (HUI (R)) system for assessing health-related quality of life in clinical studies." Annals of Medicine 33.5 (2001): 375-84.
- Gerety, M. B. et al. "The Sickness Impact Profile for Nursing-Homes (Sip-Nh)." Journals of Gerontology 49.1 (1994): M2-M8.
- Gold, MR et al. " Identifying and Valuing Outcomes." Ed. MR Gold et al. New York: Oxford University Press, 1996. 109-10.
- Gold, MR et al. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
- Haley, S. M., C. A. Mchorney, and J. E. Ware. "Evaluation of the Mos Sf-36 Physical Functioning Scale (Pf-10) .1. Unidimensionality and Reproducibility of the Rasch Item Scale." Journal of Clinical Epidemiology 47.6 (1994): 671-84.
- Horsman, J et al. "The Health Utilities Index (HUI): concepts, measurement properties and applications." Health Qual Life Outcomes 1.1 (2003): 54.
- Hutchinson, A, N Bentzen, and C König-Zahn. Cross Cultural Health Outcome Assessment; a User's Guide. ERGHO, 1996.
- Kaplan, R. M. et al. "The Quality of Well-Being Scale: critical similarities and differences with SF-36." International journal for Quality in Health Care 10.6 (1998): 509-20.
- Kaplan, RM. and JW Bush. "Health-related quality of life measurement for evaluation research and policy analysis." Health Psychol 1 (1982): 61-80.

- Katz, S. et al. "Progress in Development of Index of Adl." Gerontologist 10.1 (1970): 20-&.
- Kerner, D. N. et al. "Validity of the quality of well-being scale for patients with Alzheimer's disease." Journal of Aging and Health 10.1 (1998): 44-61.
- Krenz, C. et al. "Characterizing Patient Dysfunction in Alzheimers-Type Dementia." Medical Care 26.5 (1988): 453-61.
- Kuyken, W. et al. "The World-Health-Organization Quality-Of-Life Assessment (Whoqol) - Position Paper from the World-Health-Organization." Social Science & Medicine 41.10 (1995): 1403-09.
- Lawson, JS, M Rodenburg, and JA Dykes. " A Dementia Rating Scale for use with psychogeriatric patients." Journal of Gerontology 32 (1977): 153-59.
- McDowell, I and C Newell. Measuring Health. A Guide to Rating Scales and Questionnaires. 2nd ed. Oxford University Press, 1996.
- Neumann, P. J. et al. "A comparison of HUI2 and HUI3 utility scores in Alzheimer's disease." Medical Decision Making 20.4 (2000): 413-22.
- Norholm, V. and P. Bech. "The WHO Quality of Life (WHOQOL) Questionnaire: Danish validation study." Nordic Journal of Psychiatry 55.4 (2001): 229-35.
- Novella, J. L. et al. "Evaluation of the quality of life in dementia with a generic quality of life questionnaire: The Duke Health Profile." Dementia and Geriatric Cognitive Disorders 12.2 (2001): 158-66.
- Novella, J. L. et al. "Measuring general health status in dementia: Practical and methodological issues in using the." Aging Clinical and Experimental Research 13.5 (2001): 362-69.
- Novella, J. L. et al. "Agreement between patients' and proxies' reports of quality of life in Alzheimer's disease." Quality of Life Research 10.5 (2001): 443-52.
- Parker, S. G. et al. "Measuring health status in elder patients. The SF-36 in practice." Age and Ageing 27.1 (1998): 13-18.
- Parkerson, G. R., W. E. Broadhead, and C. K. J. Tse. "The Duke Health Profile - A 17-Item Measure of Health and Dysfunction." Medical Care 28.11 (1990): 1056-72.
- Parkerson, G. R. et al. "The Duke-Unc Health Profile - An Adult Health-Status Instrument for Primary Care." Medical Care 19.8 (1981): 806-28.
- Petersen, AF and Zachariae, B. Livskvalitetsmåling i sundhedsvæsnet - en introduktion. 2003. Aarhus Amts Trykkeri, Aarhus Amt. De mellem menneskelige relationer.
- Pollard, B. and M. Johnston. "Problems with the Sickness Impact Profile: a theoretically based analysis and a proposal for a new method of implementation and scoring." Social Science & Medicine 52.6 (2001): 921-34.

- Power, M. et al. "The World Health Organization Quality of Life assessment (WHOQOL): Development and general psychometric properties." Social Science & Medicine 46.12 (1998): 1569-85.
- Sano, M. et al. "Double-Blind Parallel Design Pilot-Study of Acetyl Levocarnitine in Patients with Alzheimers-Disease." Archives of Neurology 49.11 (1992): 1137-41.
- Selai, C. and M. R. Trimble. "Assessing quality of life in dementia." Aging & Mental Health 3.2 (1999): 101-11.
- Silberfeld, M. et al. "Content validity for dementia of three generic preference based health related quality of life instruments." Quality of Life Research 11.1 (2002): 71-79.
- Sintonen, H. The 15D-Measure of Health-Related Quality of Life. I. Reliability, Validity, and Sensitivity of Its Health State Descriptive System. Melbourne: National Center for Health Program Evaluation, 1994.
- Sintonen, H. The 15D-Measure of Health-Related Quality of Life. II. Feasibility, Reliability and Validity of Its Valuation System. Melbourne: National Center for Health Program Evaluation, 1995.
- Sintonen, H. "The 15D instrument of health-related quality of life: properties and applications." Annals of Medicine 33.5 (2001): 328-36.
- Struttman, T et al. "Quality-of-Life Assessment in the Old Using the WHOQOL 100: Differences Between Patients With Senile Dementia and Patients With Cancer." International Psychogeriatrics 11.3 (1999): 273-79.
- Thorgrimsen, L. et al. "Whose quality of life is it anyway? The validity and reliability of the quality of life-Alzheimers disease (QoL-AD) scale." Alzheimer Disease & Associated Disorders 17.4 (2003): 201-08.
- Torrance, GW, MH Boyle, and SP Horwood. "Application of multi-attribute utility theory to measure social preferences for health states." Operations Research 30 (1982): 1043-69.
- von Essen, L. "Proxy ratings of patient quality of life - Factors related to patient-proxy agreement." Acta Oncologica 43.3 (2004): 229-34.
- Ware, J. E. "SF-36 health survey update (vol 25, pg 3130, 2000)." Spine 26.18 (2001): 2062.
- Ware, J. E. et al. "Comparison of Methods for the Scoring and Statistical-Analysis of Sf-36 Health Profile and Summary Measures - Summary of Results from the Medical Outcomes Study." Medical Care 33.4 (1995): AS264-AS279.
- Ware, J. E. and C. D. Sherbourne. "The Mos 36-Item Short-Form Health Survey (Sf-36) .1. Conceptual-Framework and Item Selection." Medical Care 30.6 (1992): 473-83.

Ware, JE, M Kosinske, and SD Keller. SF-36 Physical and Mental Health Summary Scales - A User's Manual. Boston: The Health Institute, 1994.

Weyerer, S and M Schaufele. "The assessment of quality of life in dementia." Int Psychogeriatr 15.3 (2003): 213-18.

Williams, A. "Euroqol - A New Facility for the Measurement of Health-Related Quality-Of-Life." Health Policy 16.3 (1990): 199-208.

Zarit, SH, KE Reever, and J Bach-Peterson. "Relatives of the impaired elderly: Correlates of feelings of burden." The Gerontologist 20.649 (1980): 655.

Bilag 1. Oversigt over demensstudier

Instrument	Kilde	Studietype	Beskrivelse	Håndteringsmetode
WHOQOL	(Struttmann et al. 273-79)	Pilotstudie: Undersøgelse af livskvaliteten hos de gamle ved brug af WHOQOL – forskel mellem kræftsyge og dementes profiler?	N = 57; 27 moderat demente (MMSE > 15) og 30 kræftsyge	Selvhåndteret eller interview af patient
EQ-5D	(Ankri et al. 1055-63)	Valideringsstudie af den franske version af EQ-5D til måling af HRQoL i demensramte	N = 137 ambulante eller indlagte demensramte individer. MMSE score i omfang 2 - 27	Interviewhåndteret Selvudfyldelse (mildt demenramte)
	(Coucill et al. 760-71)	Undersøgelse af interrater overensstemmelse	43 patienter vurderet i kliniske omgivelser, 21 patienter vurderet i eget hjem	Proxy-vurdering af hhv. familie og sundhedspersonale Interviewhåndteret Selvhåndteret
	(Thorgrimsen et al. 201-08)	EQ-5D anvendt til vurdering af QoL-AD's samtidige validitet	Meget mild til moderat demensramte (CDR ½ - 2) N = 261 MMSE score i omfang 3 - 28	Proxy-vurderet af hhv. omsorgsyder (familie) og læge Interviewhåndteret
15D				

Ingen fundne demensstudier/artikler

HUI	(Neumann et al. 1138-45)	Modelstudie (Markov), der ønsker at undersøge den økonomiske virkning	Data stammer fra 528 AD patienter (201 mildt, 175	HUI2: Proxy-vurdering, gjort af omsorgsyderne
-----	--------------------------	---	---	---

	(CUA) af donepezil i behandlingen af AD.	moderat, 142 svært dementtramte), der enten har været hjemmeboende eller bosat på plejehjem.	
(Neumann et al. 27-32)	Cross-sectional studie: Teste muligheden for at måle helbredsnytte i AD med generisk præference-vægtet instrument	N = 679 mildt til svært demente AD patienter fra forskellige omgivelser.	HUI2 håndteret af professionel interviewer på proxy respondenter (omsorgs-ydere) via telefon
(McMahon et al. 515-22)	Modelstudie: CEA af PET (positron emission tomography) til demens diagnose	'Mild til moderat hjemmeboende demensramte'	HUI3 Udledt fra eksisterende data – fra (Neumann et al. 413-22)
(Neumann et al. 413-22)	Sammenligning af HUI2 og HUI3 nyttescore i AD.	N = 679 AD patient/omsorgsyder par (mulig – terminalt demensramte) i forskellige omgivelser.	HUI2 og HUI3 håndteret af professionel interviewer på proxy respondenter (omsorgsyder – formel eller uformel) via telefon
DUKE			
(Novella et al. 443-52)	Undersøgelse af interrater overensstemmelse	N = 76 AD patienter	Selv håndteret Proxyvurderet af hhv. familie og sundhedspersonale
(Novella et al. 158-66)	Valideringsstudie	N = 148 patienter med mental sygdom Kliniske omgivelser	Selv håndteret Interview håndteret Proxy-vurderet af hhv. omsorgsyder (familie) og læge
SF-36			
(Teri et al. 2015-22)	Randomiseret kontrolleret klinisk forsøg: kan motion reducere funktionel afhængighed og udskyde institutionaliseringstidspunktet for patienter med AD?	N = 153 AD patienter Moderat til alvorlig demente. Gns. MMSE = 16,8 (7,1)	Proxy-håndteret (familiær) To subskalaer: physical functioning and physical role functioning

(Novella et al. 362-69)	Valideringsstudie	N = 138 demensramte patienter (forskellige demensformer). Gns. MMSE 13 (variation 0 – 27)	Selvhåndteret Interviewhåndteret (interview af proxy hhv. patient)
(Yeardley et al. 385)	Krydssektionelt studie som forsøger at beskrive forholdet mellem kognitive, funktionelle og adfærdsmål for sværhedsgrad og overordnet sundhedsstatus i AD patienter.	N = 128	Omsorgsydervurderet
SIP			
(Sano et al. 1137-41)	Dobbelt-blind, parallel design, placebo-kontrolleret pilotstudie af acetyl levocarnitine i AD patienter.	N = 30 mild til moderat hjemmeboende demensramte patienter	Proxy-vurderet af omsorgsyderen
(Sano et al.)	Dobbelt-blind placebo kontrolleret forsøg	N = 20 Parkinson syge patienter med intellektuel svækkelse eller demens	Håndteret og vurderet af proxy (familiemedlem)
(Teri et al. 2015-22)	Randomiseret kontrolleret klinisk forsøg: kan motion reducere funktionel afhængighed og udskyde institutionaliseringstidspunktet for patienter med AD?	N = 153 AD patienter Moderat til alvorligt demensramte. Gns. MMSE = 16,8 (7,1)	Proxyhåndteret/vurderet ud fra kendskab til patienten Subskalaerne 'Body care and movement', 'mobility', og 'home management'
(Krenz et al. 453-61)	Bestemmelse af samtidig validitet af familiemedlemmers vurdering af AD patienters funktionelle status	Forholdsvist mildt demente	Interview af hhv. proxy (familie) og patient 'Work' subskalaen ikke anvendt på populationen
(Morishita et al. 680-83)	Studie, der undersøger validiteten af en telefonhåndtering af SIP	N = 31 geriatriske patienter Middel MMSE score = 24,8 (4,4) omfang 10 - 30	Hhv. telefon-interview + alm. interview af de ældre Interview af proxy for patienter, der udviste tegn på kognitiv svækkelse. Kun fokus på de tre subskalaer, der indgår i fysisk funktions dimensionen

QWB

(Kerner et al. 44-61)

Undersøgelse af QWBs egenskaber i forbindelse med AD patienter - vurdering af QWBs betydning i studier af AD.

N = 211 patient-ægtefælle par, $\frac{3}{4}$ havde mulig eller med sikkerhed AD, $\frac{1}{4}$ var kontrolgruppe

Interview af omsorgsyder proxy

- Ankri, J. et al. "Use of the EQ-5D among patients suffering from dementia." Journal of Clinical Epidemiology 56.11 (2003): 1055-63.
- Coucill, W. et al. "EQ-5D in patients with dementia - An investigation of inter-rater agreement." Medical Care 39.8 (2001): 760-71.
- Kerner, D. N. et al. "Validity of the quality of well-being scale for patients with Alzheimer's disease." Journal of Aging and Health 10.1 (1998): 44-61.
- Krenz, C. et al. "Characterizing Patient Dysfunction in Alzheimers-Type Dementia." Medical Care 26.5 (1988): 453-61.
- McMahon, PM et al. "Cost-effectiveness of PET in the diagnosis of Alzheimer disease." Radiology 228.2 (2003): 515-22.
- Morishita, L. et al. "Concurrent Validity of Administering the Geriatric Depression Scale and the Physical Functioning Dimension of the Sip by Telephone." Journal of the American Geriatrics Society 43.6 (1995): 680-83.
- Neumann, P. J. et al. "Cost-effectiveness of donepezil in the treatment of mild or moderate Alzheimer's disease." Neurology 52.6 (1999): 1138-45.
- Neumann, P. J. et al. "Health utilities in Alzheimer's disease - A cross-sectional study of patients and caregivers." Medical Care 37.1 (1999): 27-32.
- Neumann, P. J. et al. "A comparison of HUI2 and HUI3 utility scores in Alzheimer's disease." Medical Decision Making 20.4 (2000): 413-22.
- Neumann, P. J. et al. "A comparison of HUI2 and HUI3 utility scores in Alzheimer's disease." Medical Decision Making 20.4 (2000): 413-22.
- Novella, J. L. et al. "Evaluation of the quality of life in dementia with a generic quality of life questionnaire: The Duke Health Profile." Dementia and Geriatric Cognitive Disorders 12.2 (2001): 158-66.
- Novella, J. L. et al. "Measuring general health status in dementia: Practical and methodological issues in using the." Aging Clinical and Experimental Research 13.5 (2001): 362-69.
- Novella, J. L. et al. "Agreement between patients' and proxies' reports of quality of life in Alzheimer's disease." Quality of Life Research 10.5 (2001): 443-52.
- Sano, M et al. "A controlled trial of piracetam in intellectually impaired patients with Parkinson's disease." Mov Disord (1990).
- Sano, M. et al. "Double-Blind Parallel Design Pilot-Study of Acetyl Levocarnitine in Patients with Alzheimers-Disease." Archives of Neurology 49.11 (1992): 1137-41.

- Struttmann, T et al. "Quality-of-Life Assessment in the Old Using the WHOQOL 100: Differences Between Patients With Senile Dementia and Patients With Cancer." International Psychogeriatrics 11.3 (1999): 273-79.
- Teri, L. et al. "Exercise plus behavioral management in patients with Alzheimer disease - A randomized controlled trial." Jama-Journal of the American Medical Association 290.15 (2003): 2015-22.
- Thorgrimsen, L. et al. "Whose quality of life is it anyway? The validity and reliability of the quality of life-Alzheimers disease (QoL-AD) scale." Alzheimer Disease & Associated Disorders 17.4 (2003): 201-08.
- Yeardley, H. L. et al. "Behaviour, functional capacity and overall health status all decline with the increasing severity of Alzheimer's disease." European Neuropsychopharmacology 5.3 (1995): 385.