Statistisk analyse af kræftpatienters kropslige potentiale

Jacob Pedersen c948533

LYNGBY 2003
EKSAMENSPROJEKT
NR. 10/03

IMM

Trykt af IMM, DTU

Forord

Denne rapport er blevet til i samarbejde med Universitetshospitalernes Center for Sygepleje- og Omsorgsforskning (UCSF), Center for Muskelforskning, Onkologisk og hæmatologisk klinik på Rigshospitale samt Institut for Informatik og Matematisk Modellering (IMM) ved Danmarks Tekniske Universitet (DTU) i Lyngby.

I forbindelse med tilblivelsen af rapporten vil undertegnet sige specielt tak til mine vejledere: Poul Thyregod (IMM), Judith L. Jacobsen (Novonordisk) og Helle Rootzen (IMM). Samt til: Lis Adamsen (UCSF), Julie Midtgaard Rasmussen (UCSF) og Lasse Knutsen (UCSF). Alle for deres behjælpelighed ved spørgsmål og kommentarer samt deres engagerede samarbejde under hele projektforløbet.

Lyngby Februar 2003

Jacob Pedersen
c948533

Abstract

This report is a statistical analysis of health data from 23 cancer patients in Chemo therapy, from the phase 1 of the project "Krop & Kræft" (Body and Cancer), at the University Hospitals Centre for Nursing and Care Research (UCSF), Copenhagen Muscle Research Centre, and the clinic for Oncology & haematology at Copenhagen University Hospital. The report uses data from three questionnaires (SF-36, EORTC QLQ-C30 and HAD/HADS) and data from six objective physical measuring methods of the physical health of the patients, included measurements of strength and VO_2max.

The statistical analysis concerns longitudinal data, and is carried out by using several statistical approaches, for example the Proc Mixed analysis from the SAS program. This report focuses on a statistical approach to the data, and on being a guide/introduction, on how to approach similar data from future phases of the project.

All the plottings are made in Matlab and Staroffice Spreadsheed except for the pictures which have been created/edited in Adope PhotoShop. The statistical analysis is made with the use of SAS and Matlab.

Keywords: Longitudional, SAS Proc Mixed, Krop & Kræft, Cancer, UCSF, SF-36, EORTC QLQ-C30, HAD.

Resumé

Denne rapport er en statistisk analyse af helbredsdata fra 23 kræftpatienter i kemoterapi, fra første fase af projekt "Krop & Kræft", ved Universitetshospitalernes Center for Sygepleje- og Omsorgsforskning (UCSF), Center for Muskelforskning samt Onkologisk og Hæmatologisk klinik på Rigshospitalet. Rapporten bruger data fra tre spørgeskemaer (SF-36, EORTC QLQ-C30 og HAD/HADS) og data fra seks objektive fysiske målemetoder af patientens fysiske helbred, bl.a. målinger af styrke og VO_2max.

Den statistiske analyse vedrører longitudionale data, og er foretaget ved at benytte flere forskellige statistiske metoder, fx Proc Mixed analysen fra pc-programmet SAS. Denne rapport fokuserer på den statistiske tilgangsvinkel til dataene, samt på at være en guide/introduktion, til at behandle tilsvarende data fra fremtidige faser af projekt "Krop & Kræft".

Alle plots er lavet vha. Matlab og Staroffice Spreadsheed med undtagelse af billeder, som er bearbejdet i Adope PhotoShop. Den statistiske analyse er lavet vha. SAS og Matlab.

Indhold

Kapitel 1

Indledning

På Universitetshospitalernes Center for Sygepleje- og Omsorgsforskning (UCSF), Center for Muskelforskning samt Onkologisk og hæmatologisk klinik på Rigshospitalet, er i foråret 2001 startet et projekt kaldet "Krop og Kræft"[1]. Projektet har henblik på at forbedre Kræftpatienters kropslige potentiale, og dermed medvirke til at øge deres kræfter og ressourcer til at blive bedre fysisk samt psykisk rustet i den kemobehandling, de alle gennemgår.

Projektet "Krop og Kræft" foregår ved at patienterne gennemgår et 6 ugers træningsforløb, der indeholder fysisk træning, afspænding, undervisning i kropsbevidsthed samt individuel massage. Ideen til projektet er ifølge projektets leder, sociolog, Ph.d. Lis Adamsen, udsprunget af et ønske om at forbedre forholdene for kræftpatienter generelt. Holdet bag projekt "Krop & Kræft"[2] følte, igennem sit daglige arbejde med kræftpatienter, at der manglede nye impulser i patienternes hverdag, som ikke mindede dem om

[1] En kropsorienteret indsats til kræftpatienter i kemoterapi - et klinisk kontrolleret forsøg.

[2] Universitetshospitalernes Center for Sygepleje- og omsorgsforskning: Forskningsleder, sociolog, Ph.d, sygeplejerske, Lis Adamsen, Ph.d. studerende, cand.psych. Julie Midtgaard Rasmussen, Projektfysioterapeut, Morten Quist, Forskningssygeplejerske, Christina Andersen (Onkologisk Klinik), Forskningssygeplejerske, Tom Møller (Hæmatologisk Klinik). Klinisk ansvarlige Finsencentret: Professor, overlæge, dr.med. Mikael Rørth (Onkologisk Klinik), professor, overlæge, dr.med Niels Borregaard (Hæmatologisk Klinik). Samarbejdspartnere: Center for Muskelforskning, (CMRC), Rigshospitalet: Professor, dr.m. Bengt Saltin, Cand.scient. Morten Zacho.

deres sygdom, eller i hvert fald kunne give dem mod på at tage kampen op med samme.

Projekt Krop & Kræft er tilrettelagt som et klinisk kontrolleret forsøg i tre faser: (1) en pilotfase, hvor 25 patienter har indgået: (2) en fase, som afprøver interventionen overfor udvalgte diagnosegrupper, og hvor der indgår 100 patienter: Og endelig (3) en randomiseringsfase, hvor 200 patienter fordeles i en kontrolgruppe og en interventionsgruppe.

Denne rapport beskæftiger sig udelukkende med data fra 1. fase af projektet.

1.1 Formål

Rapport er begrundet af et ønske om en statistisk analyse af data fra første fase af projekt Krop & Kræft, som ud fra et statistisk synspunkt, kan vurdere samt påpege eventuelle effekter af interventionen. Rapporten sigter yderligere på at være en introduktion til ligende behandlinger af data fra efterfølgende faser af projektet.

1.2 Udvælgelse af brugbare data

I Projekt Krop & Kræft er, som beskrevet i indledningen, foretaget en række indsamlinger af data fra patienterne før, under og efter det egentlige trænings- og terapiforløb. Dette er foregået igennem forskellige former for spørgeskemaundersøgelser, patientdagbøger, interviews samt fysiske målinger på patienterne[3].

I forbindelse med en statistisk analyse er det vurderet af UCSF's personale samt undertegnede, udelukkende at anvende data fra spørgeskemaundersøgelserne samt de fysiske målinger af patienterne. Denne beslutning er taget ud fra hvilke data det syntes relevant at basere en sådan undersøgelse på, her under: Dataenes anvendelighed, deres sammenlignelighed, om de respektive mål var validerbare samt deres generalitet overfor projektet. På baggrund af ovenstående er data fra patientdagbøger og interviews valgt fra, bl.a. fordi de på mange punkter er svære at sammenligne og validere.

[3] Kvantitative, kvalitative samt objektive og subjektive metoder.

1.3 Opbygning af rapport

I første del af rapporten beskrives de tre anvendte spørgeskemaer fra spørgeskemaundersøgelserne. Dette er valgt ud fra at disse spørgeskemaer består af mange enkeltkomponenter, der hver især bidrager til en yderst kompleks beskrivelse af patientens tilstand/ helbred. Derefter følger beskrivelse af spørgeskemadataene, grafiske fremstillinger, og statistiske behandling af dataene vha. test og analyser.

Anden del af rapporten beskæftiger sig med de objektive data, der i modsætning til data fra spørgeskemaundersøgelserne, består af mere afgrænsede komponenter, der i form af gentagende fysiske målinger, objektivt beskriver patientens fysiske mål, som fx patientens styrke. De objektive data er først beskrevet og dernæst gennemgået vha. grafiske fremstillinger, samt statistiske test og analyser.

Slutligt er resultater fra de forrige kapitlers behandling af spørgeskema- samt objektivedata sammenlignet og diskuteret.

Kapitel 2

Indledende litteraturgennemgang

For at kunne anvende de indsamlede patientdata, er det nødvendigt at vide hvad de enkelte data giver af information om patienten. I denne forbindelse er især spørgeskemadataene svære at forstå hvis man ikke på forhånd er bekendt med disse. Derfor er dette kapitel tilegnet en grundig gennemgang af, hvordan det enkelte spørgeskema er opbygget samt hvordan det anvendes.

2.1 Anvendte spørgeskemaer

Der er i projektet brugt følgende 3 forskellige spørgeskemaer til at indsamle data om patienterne:

- SF-36 - Et spørgeskema om helbredsstatus
- EORTC QLQ-C30 - (Quality of Life Questionnaire)
- HAD - Hospitals Angst og Depression

Fælles for de tre spørgeskemaer er at de alle kan bruges til at give en overordnet vurdering af helbredet hos en patientgruppe, og er dermed interessante, i forhold til projekt Krop & Kræft, mht. det formål at måle

ændringer som følge af en intervention (i det efterfølgende kaldet projektforløb).

De tre spørgeskemaer vil i det efterfølgende blive gennemgået enkeltvis med henblik på at give et overblik, med vægt på det overordnede, og dermed skabe en forståelse for hver af disse og deres anvendelse i projektet.

2.1.1 SF-36 - Spørgeskema om helbredsstatus

Af hensyn til rettighedsindehaverne af den danske version af SF-36 manualen og spørgeskema, citeres hermed som anbefalet i beskrivelsen af rettigheder og brug af manualen til videnskabeligt arbejde[1]: Jakob Bue Bjørner, Mogens Trap Damsgaard, Torquil Watt, Per Beck, Niels Kristian Rasmussen, Tage Søndergaard Kristensen, Jens Modvig, Kate Thunedborg.

Teknisk gennemgang/overblik

Opdeling af afsnit:

- Hovedformål, samt begrænsninger
- Praktisk anvendelse
- Udformning
- Brug af SF-36 i forskellige henseende
- Undersøgelsmæssig styrke og publikationsstørrelser
- Tildeling af score
- 0-100 skalering af scores
- Missing, behandling af ikke besvarede spørgsmål/items
- Reliabilitet, Floor, Ceiling og Validitets begreber
- Variable for fysisk og psykisk helbred

Hovedformål, samt begrænsninger

I manualen for den danske SF-36 står følgende:

> "SF-36 er udarbejdet som et generelt helbredsmål med bred anvendelse, Ved hjælp af spørgeskemaet kan man vurdere effekten på den sygdomsrelaterede livskvalitet af f.eks. sygdom eller behandling." [SF-36 manual, 1997][2]

[1] [SF-36 manual, 1997]
[2] Side 8, første afsnit.

Hovedformål: Skemaets hovedformål er at vurdere den generelle helbredstilstand hos den adspurgte dels via en sammenligning med en gruppe af andre adspurgte, samt gentagende målinger. Med gruppe menes at et skema ikke kan stå/vurderes alene, men er beregnet til at indgå i en undersøgelse af flere personer.

Begrænsninger: Skemaet er ikke udformet til at bestemme resultater af en snæver og specifik behandling, bivirkninger eller bestemte sygdomme. Skemaet kan bruges til at give et overordnet mål for den almene helbredstilstand patienten befinder sig i, hvilket manualen betegner "funktionsbegrænsninger/velbefindende"[3]. Dette betyder samtidigt at der ligger et stort vurderings arbejde bag brugen af SF-36 i forskellige undersøgelser, da man i hver enkelt undersøgelse er nødsaget til at vurdere om de data man får ud, giver et brugbart resultat og om resultatet har relevans overfor det ønskede formål.

Praktisk anvendelse

SF-36 benytter spørgsmål med 8 underliggende helbredsbegreber/kategorier. Disse helbredsbegreber indeholder hver spørgsmål, der primært retter sig mod undesøgelse af den enkelte helbredskategori, og sekundært indeholder information, som senere kan bruges til at betegne patientens primære fysiske helbred og psykiske helbred. Det skal dog understreges at ikke alle helbredsbegreber omhandler såvidt fysisk som psykisk helbred, men er begrænset til at omhandle den ene af disse to.

Svarene på et spørgeskema bruges til at placere personens helbred (indenfor hver helbredskategori og fysisk/psykisk helbred), på en skala (0-100) og kan derved sammenlignes med andre spørgeskemaer eller undersøgelser.

I undersøgelser er det ofte nødvendigt at bruge flere spørgeskemaer end SF-36. Til dette brug er bl.a. SF-36 et vældigt benyttet værktøj, men det tilrådes fra SF-36'manualens side, at der laves et pilotprojekt hvor det afprøves på nogle relevante personer, hvorledes udfyldelsen af skemaerne forløber og modtages af de adspurgte. Nogle vigtige aspekter at afprøve via et pilotprojekt, er fx om skemaerne er "for lange" og man derved opnår svar, der ikke er velovervejet af den adspurgte. Det kan også tjekkes om spørgsmålene er forståelige, og ikke mindst læselige. Med læselige menes ikke kun visuelt men også i høj grad også om de adspurgte kan forstå

[3][SF-36 manual, 1997](side 13, andet afsnit).

hvad der menes i hvert enkelt spørgsmål. Af samme grund anbefales det at skemaet (SF-36) ikke anvendes på personer under 14 år, mens der til svagtseende (fx ældre mennesker) er lavet specielle skemaer, disse skemaer er visuelt "nemmere" at forstå, dvs. mere overskuelige.

SF-36 spørgeskemaet er beregnet til at blive udfyldt af éen person af gangen, og uden indblanding fra anden person. Skemaerne kan også udfyldes via interview vha. fx telefon, men i sådan situationer er det pålagt intervieweren at være så upåvirkelig overfor den interviewede som muligt. Foretages undersøgelserne over tid, som fx i nærværende projekt, tilrådes det at der går tilstrækkeligt med tid mellem hvert spørgeskema patienten skal udfylde, til at patienten ikke umiddelbart kan erindre hvad denne svarede i forrige skema, i manualen er nævnt min. 14 dage.

I forbindelse med validering af data er der til dette formål, lavet meget store undersøgelser af bl.a. den danske samt den amerikanske befolkning. Disse undersøgelser kan bruges hvis man ønsker at sammenligne med ens egne data. Dataene fra den danske undersøgelse kaldes også for "Danske normdata for SF-36"[4], og bygger på undersøgelse af 6000 ikke-institutionaliserede danskere over 14 år[5].

Udformning

SF-36 er et standardiseret amerikansk spørgeskema oversat til dansk. Skemaet består af 8 hovedkategorier (helbredsbegreber), der hver er givet ved en række spørgsmål kaldet "items". Hvert spørgsmål besvares via nogle valgmuligheder, der enten givet via skalaer (fx 1 - 4) eller Ja/Nej spørgsmål. Hvor der er brugt skalaer er det klart defineret, hvad hver enkelt skalatrin står for.

[4][SF-36 manual, 1997] (side 69).
[5]For yderligere detaljer se Danske manual til SF-36: [SF-36 manual, 1997], kapitel 10.

De 8 Helbredskategorier/begreber er:

Helbredskategori	Items	skala
Fysisk funktion (Physical function - PF)	10	1-3
Begrænsninger - fysisk betinget (Role physical - RP)	4	ja/ nej
Fysisk smerte (Bodily pain - BP)	2	1-5 samt 1-6
Alment helbred (General health perceptions - GH)	5	1-5
Energi (vitality - VT)	4	1-6
Social funktion (Social function - SF)	2	1-5
Begrænsninger - psykisk betinget (Role emotional - RE)	3	ja/nej
Psykisk velbefindende (Mental health - MH)	5	1-6
SUM	35	

Tabel 2.1: Oversigt for de 8 helbredskategorier, givet i SF-36'skemaet.

Der findes i et SF-36 spørgeskema, desuden et ekstra spørgsmål om ændringer i den adspurgtes almene helbred, men dette spørgsmål indgår ikke i nogen skalaberegning[6].

Helbredskategori	Items	skala
Ændringer i alment helbred	1	1-5

Tabel 2.2: Ekstra spørgsmål givet i SF-36'skemaet.

Dvs. sammenlagt er der (35 + 1)= 36 spørgsmål.

Brug af SF-36 i forskellige henseende.

Når man benytter SF-36 til at undersøge helbredsstatus før og efter et projektforløb/intervention, eller af personers/gruppers udvikling over tid, er der nogle generelle antagelser man må være bekendt med, disse er gengivet

[6] Se efterfølgende underafsnit: 0-100 skalering af scores.

i dette samt efterfølgende afsnit: (Undersøgelsesmæssig styrke, og populationsstørrelse).

En vigtig antagelse i startfasen af en undersøgelse, er at tage højde for den måleusikkerhed der er ved at bruge forudbestemte/afgrænsede skalaer og standardiserede svar. Med dette menes at det er oplagt at, hvis en person kun oplever ganske små ændringer i en helbredstilstand, vil dette ikke kunne måles i et spørgeskema hvor man fx kan vælge mellem (fx; meget, mindre eller lidt ond i knæet). Samtidigt er det vigtigt at tage højde for forskellige menneskers præferencer, idet nogle fx har lavere smertegrænser end andre, og derfor vil have vidt forskellige udgangspunkter at vælge ud fra.

Ovenstående er en konsekvens af bl.a. en SF-36 spørgeskemaundersøgelse, og betyder at det ikke er muligt, fx at bestemme hvor meget smerte en person døjer med. I tilfældet med smerte, vil det <u>kun</u> være muligt at bedømme hvorvidt der i forhold til forrige målinger er sket; en forværing, det modsatte eller ingen ting. Dermed understreges det at SF-36 bruges til at finde personers generelle helbredsstatus og at dette udelukkende er vurderet ud fra hvad personen selv oplever, og <u>ikke</u> hvad man fx kan måle eksakt med et givent instrument.

Undersøgelsmæssig styrke, og populationsstørrelse

I ovenstående afsnit, blev nævnt at det er vigtigt at vurdere den måleusikkerhed der er forbundet med den undersøgelse man skal i gang med. Dermed er det relevant at kunne bestemme, hvor store ændringer man kan acceptere som signifikante. En hjælp til dette er givet i den danske SF-36 manual i form af bl.a. nogle tabeller[7]. Tabellerne bruges til at angive de minimum- pupulationsstørrelser/antal-patienter, som er nødvendige hvis man ønsker at kunne bestemme ændringer i en skala, som signifikante ($\alpha = 5\%$). I understående tabel er vist et udpluk fra tabel (5.4)[8] der angiver minimumpopulationsstørrelser for undersøgelser foretaget over tid, og iøvrigt er aktuel overfor projektet "krop og kræft".

[7] [SF-36 manual, 1997], tabel 5.2 - 5.5
[8] [SF-36 manual, 1997], side 20

Skala		Antal points forskel 2	20
Fysisk funktion	PF	852	9
Begrænsninger - fysisk betinget	RP	1826	19
Fysisk smerte	BP	883	10
Alment helbred	GH	654	7
Energi	VT	693	8
Social funktion	SF	809	9
Begrænsninger - psykisk betinget	RE	1721	18
Psykisk velbefindende	MH	515	6

Note: baseret på følgende forudsætninger: $\alpha = 0.05$, tosidet t-test, power = 80% og en korrelation på 0.60 mellem målinger (Cohen, 1988; Ware, Jr. et al., 1993)

Tabel 2.3: Nødvendige populationsstørrelser, for identifikation af signifikante ($\alpha = 5\%$) pointforskelle over tid.

Ovenstående tabel angiver, at for at en forskel på 2 skalapoint er signifikant ($\alpha = 5\%$) for en måling på Fysisk smerte (BP), er det nødvendigt at undersøgelsesgruppen består af minimum 883 personer. Har man derimod en undersøgelsesgruppe på 10 personer, vil det for samme parameter (Fysisk smerte) gælde, at der skal være tale om en forskel på minimum 20 point, før man kan antage ændringer for signifikante.

Tildeling af score

Når et skema er udfyldt, bruges en fastlagt metode til at give de forskellige svar en score. Disse scores bruges til at anbringe de enkelte helbredskategorier på en skala fra 0-100, der overordnet indikerer følgende[9]:

- En lav værdi (min. 0) indikerer et dårligt helbred.
- En høj værdi (max. 100) indikerer et godt helbred.

Eksempel på score: I spørgsmål med svarmulighederne Ja/nej, gives et “Ja” score 1 og svaret “Nej” score 2.

[9] Se tabel: (2.6), for yderligere forklaring.

I de resterende spørgsmål gives scores ud fra understående tabel[10]:

H	Item	Svarmuligheder		O-kode	E-kode
PF	3a-j	Ja, meget begrænset		1	1
		Ja, lidt begrænset		2	2
		Nej, slet ikke begrænset		3	3
RP	4a-d	Ja		1	1
		Nej		2	2
BP	7	Ingen smerte		1	6
		Meget lette smerter		2	5.4
		Lette smerter		3	4.2
		Middelstærke smerter		4	3.1
		Stærke smerter		5	2.2
		Meget stærke smerter		6	1
	8 (+7)*		(8)	(7)	
		Slet ikke	1	1	6
		Slet ikke	1	2-6	5
		Lidt	2	1-6	4
		Noget	3	1-6	3
		En hel del	4	1-6	2
		Virkelig meget	5	1-6	1
	8 (−7)	Slet ikke	1		6
		Lidt	2		4.75
		Noget	3		3.5
		En hel del	4		2.25
		Virkelig meget	5		1

Note: Hele tabellen kan ses i Bilag A, tabel: (A.2).

Tabel 2.4: Udpluk af scoretildeling for SF-36'skemaet.

[10] De originale tabeller kan ses i manualen kapitel 7 side 33-41 [SF-36 manual, 1997]. For at se SF-36'spørgsmålene til tabel: (2.4), henvises til selve SF-36 skemaet.

H	= Helbredskategori.
O-kode	= Oprindelig kode.
E-kode	= Endelig kode.

* = koden for spørgsmål 8 og 7, som begge er besvaret

Tabel 2.5: Tegnforklaring til tabel: (2.4).

Det ses af tabel (2.4, samt A.2 bilag A) at der er ingen endelig kode for helbredskategorien HT, hvilket skyldes at denne helbredskategori ikke bruges i nogen skala. Det ses desuden at for flere af spørgsmålene/ items vendes skalaen i forhold til den endelige kode/score.

For andre items ses det at den endelige score tildeles en bestemt vægtning, dette gøres for at alle summationer (se næste afsnit) stemmer overens og "tæller" lige meget.

Specielt for spørgsmål 7 og 8 foretages først følgende konstatering:

- Hvis de begge er besvaret, foretages vægtningen betegnet 8 $(+7)$.
- Hvis kun det ene af de to items (7 eller 8) er besvaret foretages vægtningen hhv. 7 eller 8 (-7).

0-100 skalering af scores.

Når hvert besvaret spørgsmål/item har fået tildelt en score (råscore), summerer man råscorene fra samme helbredskategori (aktuelle råscore), og omregner for hver helbredskategori disse til skalaværdier (transformeret skala) vha. følgende formel:

$$Transformeret\ Skala = \left[\frac{(Aktuelle\ raascore - lavest\ mulige\ raascore)}{(Hoejest\ mulige\ raascore - lavest\ mulige\ raascore)}\right] \times 100$$

$$Note : (Raascore = tildelt\ score)$$

Man behøver ikke at benytte omregningen til den transformerede skala, idet man kan nøjes med at sammenligne de utransformerede scores eller summationer. Dette betyder imidlertid at man ikke umiddelbart kan sammenligne sine helbredskategorier med andre slags spørgeskemaundersøgelser, hvorfor det anbefales at man transformerer.

Generelt vurderes de enkelte helbredskategorier på samme måde i den ovenfor nævnte skala, om denne skala er transformeret eller utransformeret. En høj score betegner et godt helbred mens en lav score et dårligt helbred. For den enkelte helbredskategori har høj og lav score følgende betydning:

	Laveste score (0)	Højeste score (100)
PF	Meget begrænset i udførelsen af alle typer fysiske aktivitere	Udfører alle typer aktivitete
RP	Problemer med daglige aktiviter p.g.a. fysisk helbred	Ingen problemer med daglige aktiviteter p.g.a. fysisk helbred
BP	Meget stærke smerter	Ingen smerter
GH	Vurderer eget helbred som dårligt, og forventer forværring i fremtiden	Vurderer eget helbred som fremragende
VT	Føler sig træt og udslidt	Føler sig fuld af liv og energi hele tiden
SF	Begænsninger i normale sociale aktiviteter p.g.a. helbredsproblemer	Ingen begrænsninger i sociale aktiviteter p.g.a. helbredsproblemer
RP	Problemer med daglige aktiviter p.g.a. psykisk helbred	Ingen problemer med daglige aktiviteter p.g.a. psykisk helbred
MH	Nervøs og trist til mode hele tiden	Rolig, glad og tilfreds hele tiden

Note: Tabellen er et uddrag af: Tabel (9.2)[SF-36 manual, 1997].

Tabel 2.6: Betydning af max. og min. skalatrin for de 8 SF-36 helbredskategorier.

Missing, behandling af ikke besvarede spørgsmål/items.

I SF-36 behandles tilfælde af missing (spørgsmål/items der ikke er besvaret) på følgende måder:

- Hvis mindre end $(<)$ 50% af en helbredskategoris spørgsmål er besvaret, skal helbredskategorien ikke regnes med for dette spørgeskema.

- Hvis mere end (≥) 50% (samt 50%) af en helbredskategoris spørgsmål er besvaret, laves et afrundet gennemsnit (til nærmeste skalatrin) for de besvarede spørgsmål, og de ikke-besvarede spørgsmål (i) tildeles denne værdi.

Reliabilitet, Floor, Ceiling og Validitets begreber

I udfærdigelsen af manualen for SF-36, har man forsøgt at tage højde for så mange usikkerhedsmomenter og problemstillinger som muligt. Dette betyder at man har lavet en del analyser af besvarede spørgeskemaer, ud fra hvilke man har undersøgt beregninger for forskellige "fejl". Disse beregninger kan i visse tilfælde benyttes til at korrigere ens egne undersøgelser, udover at de selvfølgelig gør opmærksom på de forskellige problemstillinger man kan komme ud i.

Der opereres med følgende "fejl" i SF-36 spørgeundersøgelser:

Reliabilitet: I SF-36's helbredskategorier er der, med undtagelse af HT, flere end et spørgsmål for hver helbredskategori. Dette skyldes at man ønsker at øge præcisionen og dermed reliabiliteten af den givne skala. For at kunne angive et mål for hvor gode de anvendte skalaer er til at "måle den sande værdi" af en given helbredstilstand hos en patient, har man brugt Crombach's koefficient alpha[11]. Cronbach's koefficient alpha benyttes ved, på en tilfældig måde, at splitte spørgsmål/items, som afdækker den samme helbredskategori, op i to halvdele "split-half metoden". Derefter beregnes korrelationen mellem de to opsplittede halvdele.

I SF-36 manualen[12] er angivet estimater for Cronbach's alpha, baseret på en SF-36 stikprøveundersøgelse af den danske befolkning. Disse estimater viser at SF-36 på nær følgende 3 tilfælde: SF 16-24 år (0,66), RE 16-24 år (0,62) og RE 25-44 år (0,69)), generelt ligger med en korrelation mellem items, for forskellige opdelinger af befolkningen[13], på over 0.70. En korrelation mellem items på min. 0.70, er påkrævet for spørgeskemaskalaer til sammenligning af grupper (Nunnally & Bernstein, 1994[14]). Det angiver samtidigt at SF-36 har svært ved at opnå korrelationer over 0.90, der ligeledes er påkrævet hvis spørgeskemaskalaerne skal kunne anvendes til at vurdere enkelt patienter[15].

[11] [SF-36 manual, 1997] Side 48, 1 afsnit.
[12] [SF-36 manual, 1997], side 49 tabel 8.1.
[13] Mand/kvinde, alder, uddannelse, kronisk/ingen kronisk sygdom samt total.
[14] [SF-36 manual, 1997] side 48, 3 afsnit.
[15] [SF-36 manual, 1997] side 48, 3 afsnit.

Floor = Skalaen er ikke tilstrækkelig til at måle ændringer i helbred for den nedre del af skalaen.

Ceiling = Skalaen er ikke tilstrækkelig til at måle ændringer i helbred i den øvre del af skalaen (modsatte af Floor effekt).

Betegnelserne (Ceiling eller Floor) bruges til at udtrykke følgende situationer:

- Hvis de adspurgte ikke kan udtrykke ændringer via den givne skala, fx når en forringelse af et fysisk helbred, ikke kan udtrykkes i bunden af skalaen, og den adspurgte kun kan udtrykke sit eget niveau, ved at benytte den samme laveste værdi, denne brugte ved forrige undersøgelse (Floor).
- I undersøgelser af flere personer med fx et generelt meget dårligt helbred, hvor det ikke vil være muligt at opnå en tilstrækkelig differentiering imellem de adspurgte (Floor).

Hvis lignende situationer gør sig gældende i den øvre del af skalaen, bruges betegnelsen Ceiling istedet for Floor.

For validitesbegreber er følgende beskrevet:

- Indholdsvaliditet.
- Kriterievaliditet.
- Begrebsvaliditet: Konvergens validitet, diskriminativ validitet.

Her af følger en kort gennemgang.

Indholdsvaliditet = "Er de spørgsmål, man stiller, dækkende for den problemstilling, man ønsker at undersøge?", og "giver items i en skala dækkende og udtømmende informationer om det begreb, som skalaen angiver at måle?". [16]

Kriterievaliditet: Overgangen mellem kriterievaliditet og begrebsvaliditet er flydende og derfor er her valg kun at medtage begrebsvaliditet, da dette er bedst forklarede, og det umiddelbart bedst dækkende for begge begreber.

Begrebsvaliditet = For dette validitetsbegreb stilles følgende spørgsmål til selve spørgeskemaet og sammenhænge imellem spørgsmål/items;

[16] [SF-36 manual, 1997] Side 54 4 afsnit.

- "Er inddelingen af items i bestemte SF-36 skalaer fornuftig ?". Her under:
  1. "Har et item en forventet høj korrelation med summen af de øvrige items i skalaen (konvergens validitet)?"
  2. "Er korrelationen mellem et item og dets egen skala (summen af de øvrige items) tilstrækkelig højere end korrelationen mellem dette item og nogen anden skala (diskriminativ validitet)?"

Sammenhængen mellem SF-36 og begreberne konvergens/- diskriminativ validitet, kan umiddelbart tolkes på følgende måde:

Konvergens validitet: Er besvarelsen af spørgsmål logisk sammenhængende, for spørgsmål der måles i samme skala. Dette vil fx sige, om spørgsmål fra samme helbredsbegreb generelt besvares så de udtrykker en entydig sammenhæng, eller kan man omvendt, ikke kan drage nogle fornuftige konklusioner af besvarelserne.

Diskriminativ validitet: Måles det stillede spørgsmål på en fornuftig skala, og er der utiltænkt sammenhæng mellem et spørgsmål og et andet spørgsmål stillet i en helt anden sammenhæng. Hvis det sidste tilfælde gør sig gældende, kan konsekvensen være at man måler samme begreb på to forskellige skalaer og derved opnår forvirring om begge spørgsmål. En sådan forvirring kan have stor betydning for om de konklusioner man kunne tænkes at drage i denne forbindelse giver mening.

Variable for fysisk og psykisk helbred

De fundne værdier for de 8 helbredskategorier (PF, RP, BP, GH, VT, SF, RE, MH), kan bruges til at finde to værdier for den adspurgtes generelle hhv. fysiske (PCS Physical Component Summary) og psykiske (MCS Mental Component Summary) helbred. Dette gøres ved at vægte de standardiserede 8 helbredskategorier, udfra nedenstående vægtningstabel[17] og derefter reskalere de to opnåede værdier til værdier for PCS og MCS.

[17] [SF-36 manual, 1997] tabel 11.1 side 79: Konstanter som indgår i scoring af PCS og MCS. Baseret på undersøgelser af den amerikanske befolkning (Ware, Jr. et al. 1994b).

	Trin 1		Trin 2	
Skala	Middelværdi	Standardafv.	PCS	MCS
PF	84,52	22,89	0,42	-0,23
RP	81,20	33,80	0,35	-0,12
BP	75,49	23,56	0,32	-0,10
GH	72,21	20,17	0,25	-0,02
VT	61,05	20,87	0,03	0,24
SF	83,60	22,38	-0,01	0,27
RE	81,29	33,03	-0,19	0,43
MH	74,84	18,01	-0,22	0,49

Tabel 2.7: Konstanter til beregning af PCS og MCS værdier.

Værdier for PCS og MCS, opnåes via følgende 3 trin:

Trin 1: Standardisering:

$$Standardisering = \frac{Skala - Middelvaerdi}{Spredning}$$

Trin 2: Vægtning og summering:

De standardiserede helbredskategorier multipliceres med de i ovenstående tabel angivne tal for hhv. PCS og MCS og der fremkommer to nye komponenter kaldet: PCS-trin(2) og MCS-trin(2).

Trin 3: Reskalering:

$$(middelværdi(USA) = 50,\ spredning(USA) = 10)$$

$$PCS = PCS(trin\ 2) \times spredning(USA) + middelvaerdi(USA)$$
$$MCS = MCS(trin\ 2) \times spredning(USA) + middelvaerdi(USA)$$

De angivne værdier for middelværdi(USA) og spredning(USA), er standardværdier lavet på baggrund af undersøgelser i USA. Der er ifølge den danske manual til SF-36, endnu ikke lignende tal for Danmark, og derfor bruges de ovennævnte tal fra USA indtil videre.

2.1.2 HAD - Hospitalsangst og depression

Opdeling af afsnit:

- Hovedformål, samt begrænsninger
- Anvendelse og udformning
- Tildeling af score
- Reliabilitet og Validitets begreber

Hovedformål, samt begrænsninger

The Hospital anxiety and depression scale (HADS/HAD)[18] anvendes til at bestemme angst og depression i kliniske undersøgelser i forbindelse med hospitalsbehandlinger på somatiske sygdomme[19].

Der tages udgangspunkt i mentale symptomer på fysiske sygdomme[20] (somatiske sygdomme), specielt er der fokus på livstruende sygdomme hvor der bl.a. er foretaget undersøgelser vha. HAD på 568 kræftpatienter på et hospital i Norge [Skarstein, 2000].

Der beregnes to variable for hhv. angst og depression, disse to variable er defineret via følgende tilfælde:

Angst: "En tilstand af generel angst, ikke nødvendigvis fokuseret på en speciel situation" her under "state of anxious mood, restlessness and anxious thoughts"[Snaith & Zigmond, 1994][21].

Depression: Er en generaliseret depression, hvor der er tegn på "en tilstand af mangel på interesse" og "diminished pleasure response", desuden henvises til "The lowering of hedonic tone" hvilket bruges som guide til genkendelse af "humør svingninger af biologisk oprindelse"[22].

Begrænsninger:

Den anvendte engelsk manual for HAD er langt fra så dybdegående, som manualen for SF-36 skemaet. Dette har betydet at det har været nødvendigt

[18][Snaith & Zigmond, 1994] side 1.
[19]Legemlig, organisk sygdom (mods. psykisk).
[20][Johnston, 2000] side 579
[21]Side 2 midt.
[22][Snaith & Zigmond, 1994] side 2 sidst.

at hente information om skemaet fra flere forskellige kilder, hvilket har gjort det svært at få et entydigt billede af HAD og dets anvendelse.

En vigtig mangel for den anvendte litteratur er beskrivelsen af hvordan man behandler spørgsmål med tilfælde af missing, dvs. ubesvarede spørgsmål.

Der findes heller ikke beskrivelse af Ceiling og Floor effekter, hvilket ligesom for SF-36 skemaet er yderst relevant.

Der er ingen tabel eller beregninger for statistisk styrke for HAD, hvilket også betyder at der heller ikke er beregninger/tabeller for pupulationsstørrelser, som det var gældende i SF-36 skemaet.

Ud over de nævnte begrænsninger i beskrivelsen af HAD, er det vigtigt at man ikke benytter HAD til at stille diagnoser [Brendan, 1993][23], da HAD langt fra er et tilstrækkeligt værktøj til dette. Det rådes derfor til at man udelukkende anvender HAD til at se tendenser til angst og depression hos patienter, samt udvikling over tid.

Anvendelse og udformning

Skemaet er et udfyldningsskema, hvor der er givet en række spørgsmål (14 stk.) og for hvert spørgsmål er givet en skala (1-4) hvor på man kan afkrydse sit svar.

Skemaet er relativt kort, dvs. skemaet kan udfyldes af en patient umiddelbart mens denne venter på at komme til ved en behandling/undersøgelse. Det har yderligere været hensigten i forbindelse med udformningen af skemaet, at skemaet udfyldes ved undersøgelsstedet, hvilket i denne hensigt er et hospital (her af: Hospital Anxiety Depression Skale HADS/HAD).

De 14 spørgsmål er opdelt så der er 7 spørgsmål om angst, og ligeledes 7 om depression. Der er ikke spørgsmål der omhandler begge kategorier (angst og depression).

Skalaerne hvorpå man kan afgive sit svar, går alle fra 1-4 (4 svarmuligheder) men er forskellige i udformningen. Et eksempel på spørgsmål og svar er givet i følgende tabel, her ses desuden den oprindelige engelske formulering overfor den danske formulering anvendt i projektet "Krop og Kræft".

[23]Side 73, 1 spalte nederst

Svar nr.	Engelske formulering	Danske formulering
	I feel tense or 'wound up':	**Jeg føler mig anspændt:**
1	Most of the time	Næsten hele tiden
2	A lot of the time	Meget af tiden
3	From time to time, occasionally	Engang imellem
4	Not at all	Slet ikke

Tabel 2.8: Spørgsmål og svarmuligheder for HAD-spørgeskemaets spm. 1.

De forskellige skalatrin (1,2,3,4) har ikke altid svar 4 som det "normale" svar, for en patient med hverken angst eller depression, hvilket betyder at skalaen vendes. Dette kompenseres der imidlertid for i den efterfølgende behandling af data[24].

Tildeling af score

Når et spørgeskema er udfyldt helt, registeres de angivne svar (1-4), og der tildeles en score for hvert spørgsmål. Der scores i to skalaer hhv. angst og depression for hvilke der senere udregnes en værdi. I nedenforstående tabel, er vist et eksempel på score i A-skalaen, dvs. angst skalaen (D-skala = depression skala):

Spørgsmål nr. 1	Svar	Svar nr.	Score
Jeg føler mig anspændt:			
	Næsten hele tiden	1	3
	Meget af tiden	2	2
	Engang imellem	3	1
	Slet ikke	4	0

Tabel 2.9: Scoretabel for spm. 1, HAD-skemaet.

Tabel: (A.1, bilag A)[25] angiver sammenhængen mellem svarangivelse for HAD's 14 spørgsmål og tildeling af score. Opdelingen i A og D angiver hvilken variabel/helbredkategori (Angst eller Depression) hvor til den respektive score tildeles, som vist i understående.

[24] Se næste afsnit.
[25] [Zigmond & Snaith, 1983] side 368-370.

Der udregnes to værdier for hhv. angst og depression (A_{sum},D_{sum}), dette gøres ved at summere de tildelte scores for hver af de to skalaer (A og D), dvs. for skalaens (A/D) respektive tilhørende spørgsmål (1-14):

$$\sum_{i=1}^{13} A_i = A_{sum} \quad i \in \{1,3,5,7,9,11,13\}$$
$$\sum_{i=2}^{14} D_i = D_{sum} \quad i \in \{2,4,6,8,10,12,14\}$$

Der findes i litteraturen om HAD, to forskellige mål for anvendelsen af de summerede scores, i manualen (1994) til HAD bruges følgende[26]:

Angst / Depression	
A_{sum}/D_{sum}	Betegnelse:
0-7	Normal
8-10	Mild
11-14	Moderat
15-21	Svær

Tabel 2.10: Skalaopdeling for HAD, 1994.

I den tidligere (1983) beskrivelse af HAD [Zigmond & Snaith, 1983], bruges følgende opdeling:

Angst / Depression	
A_{sum}/D_{sum}	Betegnelse:
0-7	ikke tilfælde
8-10	måske tilfælde
11-21	tilfælde

Tabel 2.11: Skalaopdeling for HAD, 1982.

Hvilken en opdeling man vælger at bruge, er op til den enkelte undersøgelse.

Reliabilitet og Validitet

Reliabilitet: I Zigmond og Snaith fra 1983[27], er foretaget en undersøgelse af HAD skemaets effekt til at finde depression samt angst hos 50 udvalgte

[26] [Snaith & Zigmond, 1994]
[27] [Zigmond & Snaith, 1983]

patienter. Undersøgelsen blev foretaget ved at patienterne først udfyldte et HAD skema og derefter blev interviewet med henblik på at vurdere om patienten led af angst og/eller depression. Derefter blev HAD skemaet og interviewene sammenlignet og i understående tabel ses de korrelationer, som blev fundet at være mellem de to slags undersøgelser[28].

	Interview rating	
Patient rating	Anxiety	Depression
Anxiety	+ 0,54*	+ 0,19
Depression	+ 0,08	+ 0,79**

* P < 0,05

** P < 0,01

Tabel 2.12: Korrelation mellem interview- og HAD undersøgelse.

Af ovenstående tabel ses at HAD har høj korrelation ved vurderingen af depressioner, og dermed stemmer overens med interview delen, idet det antages at interview delen udtrykker den "sande" værdi. Ses på korrelationen mellem angst målt ved hhv. HAD og interview, er det straks sværere at sige om HAD kan anvendes til at bestemme tilfælde af angst, idet korrelationen her er betragtelig lavere ($+0,54\ ved\ P < 0,05$).

Manualen for HAD[29], anvender en norsk undersøgelsen af 568 kræftpatienter[30], til at understrege at begge variable angst og depression er signifikante i HAD undersøgelser. I Moorey undersøgelsen bruges Cronbach's alpha[31], og der er fundet et alpha for A-skalaen på 0,93 og et alpha for D-skalaen på 0,90. Det tyder derfor på, i modsætning til Zigmond and Snaith undersøgelsen i ovenstående tabel, at man godt kan anvende HAD til at bestemme både angst og depression hos patienter.

Validitet: Der skelnes i manualen for HAD ikke mellem forskellige former for validitet hvilket er en tydelig svaghed for vurderingen af HAD. Der vil ikke blive gået nærmere i detaljer med de anvendte undersøgelser mht. validitet, men nøjes med samme konstatering, som er foretaget i manualen til HAD, at der synes at være en rimelig tilkendegivelse af validiteten

[28] [Zigmond & Snaith, 1983] side 364, tabel 2.
[29] [Snaith & Zigmond, 1994]
[30] Moorey et al. (1991).
[31] Se beskrivelse af SF-36'skemaet.

for samtlige 14 spørgsmål, og opdelingen i 7 spørgsmål til hhv. angst og depression.

Måleusikkerhed: Der er ikke angivet beregninger, opstillinger for måleusikkerheder eller "fejl" i forbindelse med en HAD måling (målingens usikkerhed). Manualen[32] taget heller ikke stilling til hvorledes man bør reagere i forhold til forskellige problemstillinger der kunne tænkes at opstå i forbindelse med en HAD undersøgelse, dette er derfor et indlysende emne at tage op i egne undersøgelser, samt en indikation af at man skal passe på ikke at drage snævre og specifikke antagelser ud fra de målinger man får.

For andre undersøgelser af HAD er her nogle nøglekonklusioner der er relevante i forbindelse med problemstillinger omhandlende måleusikkerhed:

- "Self-reported distress is not always correlated with the presence of a given disorder"[33].
- "Mild (or neurotic) psychiatric disorder cannot be considered to be either present or absent since the degree of distress is continuously distributed in the population"[34].
- Mere relevante fortolkning af scores: "(0-7) Probable absence, (8-10) possible presence, and (11-21) probable presence of disorder"[35].
- Hvis undersøgelsen retter sig mod at finde patienter med høj sandsynlighed for "mood disorder", bør kun scores over 10/11 bruges.
- Hvis undersøgelsen retter sig mod at finde alle mulige tilfælde af angst og depression, bør kun scores over 8/9 bruges[36].

[32] [Snaith & Zigmond, 1994]
[33] [Zigmond & Snaith, 1983] side 72, 1 afsnit.
[34] [Zigmond & Snaith, 1983] side 364-365.
[35] [Zigmond & Snaith, 1983] side 365 linje 8-9.
[36] [Zigmond & Snaith, 1983] side 365, 1 spalte midt for.

2.1.3 EORTC QLQ C-30 - Livskvalitetsbestemmelse

QLQ-C30 (Quality of Life Questionnaire) version 3.0 (2001), er udarbejdet af The European Organization for Research and Treatment of Cancer (EORTC). Version 1.0 var på dets udgivelsesår (1993), testet på lungekræft patienter i 13 forskellige lande, mens EORTC QLQ-C30 version 3.0 i 2001 var registreret i mere end 2200 studier/undersøgelser.

(Ref.: [EORTC scoring manual, 2001] samt [EORTC QLQ-C30, 1993]).

Opdeling af afsnit:

- Formål og begrænsninger
- Anvendelse og udformning
- Scoring
- Skala/data anvendelse
- Missing og bias
- Validitet og reliabilitet

Formål

QLQ-C30 er udarbejdet med henblik på at vurdere den helbredsrelaterede livkvalitet, hos især kræftpatienter i forbindelse med kliniske undersøgelser.

Anvendelse og udformning

Udarbejdelsen af QLQ-C30 er siden 1993 foregået i foreløbig 3 faser, hhv. 1,2 og 3 version. Dette har betydet at spørgeskemaet har været igennem en række af evalueringsfaser der hver har været med til at tilpasse og optimere skemaet.

Til skemaet findes idag en del specielle moduler (bl.a. QLQ-BR23 (Breast cancer module), QLQ-LC13 (Lung cancer module)), der kan anvendes i tilfælde hvor man ønsker at undersøge patienter med en bestemt kræftsygdom.

For at give en forståelse af QLQ-C30 skemaet ide, er her en oversigt over de emneområder der behandles:

Fem funktions skalaer:	Tre symptom skalaer:
- Physical	- Fatigue
- Role	- Pain
- Cognitive	- Nausea and vomiting
- Emotional	
- Social	
En global helbredsstatus skala:	**Enkelt spørgsmål*:**
- Global helbredsstatus/QoL	- Dyspnoea
	- Loss of appetite
	- Insomnia
	- Constipation
	- Diarrhoea
	- Financial impact**

**Omhandlende symptomer ofte rapporteret hos kræftpatienter.*
***Perceived financial impact of the disease.*

Tabel 2.13: Emneområder for EORTC QLQ-C30, version 3.0.

De forskelige emneområder behandles vha. nogle spørgsmål der hver har et antal mulige svar på en skala. I det nedenfor stående skema kan ses spørgsmålenes fordeling på de forskellige emneområder/ helbredskategorier.

	Skala	Antal spm.	spm. nr.	Function skala
Global helbreds status/Qol:				
Global helbreds status/Qol	QL2	6	29,30	
Functional scales:				
Physical functioning	PF2	5	1 to 5	F
Role functioning	RF2	2	6,7	F
Emotional functioning	EF	4	21 to 24	F
Cognitive functioning	CF	2	20, 25	F
Social functioning	SF	2	26 ,27	F
Symptom scales/items:				
Fatigue	FA	3	10,12,18	
Nausea and vomiting	NV	2	14,15	
Pain	PA	2	9,19	
Dyspnoea	DY	1	8	
Insomnia	SL	1	11	
Appetite loss	AP	1	13	
Constipation	CO	1	16	
Diarrhoea	DI	1	17	
Financial difficulties	FI	1	28	

Tabel 2.14: Helbredskategorioversigt for EORTC QLQ-C30.

Spørgsmålene 29 og 30 har begge skalaer der går fra 1 til og med 7:

(1 = Meget dårligt, 7 = Særdeles godt).

De øvrige spørgsmål (1-28) har alle skalaer der går fra 1-4:

(hhv, Slet ikke (1), lidt (2). En del (3). Meget (4)).

Patienten udfylder selv skemaet eller i form af interview, hvilket stiller store krav til skemaets opbygning og ordvalg. I denne forbindelse er QLQ-C30 udarbejdet med henblik på at kunne bruges i mange forskellige kultur sammenhænge, hvilket bl.a. giver sig udslag i at skemaet er oversat til en del forskellige sprog[37].

Den nyeste (2001) version hedder 3.0, hvilket er den version som EORTC

[37] EORTCmanual side 31.

anbefaler at man anvender. Man kan på følgende internet adresse finde de nyeste oplysninger omkring QLQ-C30 skemaet (www.eortc.be/home/qol/).

Scoring

For hvert spørgsmål der besvares er angivet en værdi, denne værdi angiver scoren for det enkelte spørgsmål. For at beregne en råscore for et emneområde udregnes et gennemsnit at de besvarede spørgsmål der tilhører det givne emneområde (se tabel: (2.14))[38].

Man kan vælge at bruge andre måder at udregne sin råscore på, men ifølge manualen anbefales det at man fortsat benytter gennemsnitsberegningerne. Ifølge manualen arbejdes der fortsat på alternative måder at beregne råscore, samt lave skalatransformationen på [EORTC scoring manual, 2001].

Til at anbringe de beregnede råscore på en skala, som går fra 0-100, anvendes to forskellige omregningsformler/skalaberegninger(S).

Råscore(RS):

$$RS = (I_1 + I_2 + \ldots + I_n)/n$$

Funktional skala:

$$S = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100)$$

Symptom samt globale helbreds skalaer:[39]

$$S = \left\{\frac{(RS-1)}{range}\right\} \times 100$$

Range er forskellen mellem den største og mindste svarmulighed for hvert spørgsmål, dvs. at hvis man for et spørgsmål kan svare lavt(1), og højt(6) er range = 5 (6-1).

Grunden til at der anvendes to forskellige formler til skalaberegning, er at funktional skalaen vendes, så den er sammenlignelig med den globale helbredsstatus, samt resultater af andre undersøgelser.

En høj værdi på den anvendte skala (0-100) repræsenterer på:

[38] For tilfælde af "ikke" besvaret spm. se underafsnit: Missing og bias.
[39] [EORTC scoring manual, 2001] side 7.

- Funktionsskalaen - en sund tilstand.
- Den Globale helbredsskalaen - en høj livskvalitet (quality of life QoL).

Men: På Symptom skalaen angiver en høj værdi - mange symptomer.

Det bemærkes at symptom skalaen skiller sig ud, da det her ikke kan tilstræbes at gå efter høje værdier. Der nævnes ikke noget om i manualen, at man også kan vende skalaen for symptomer, hvilket umiddelbart betyder at emneområder i denne skala, ikke kan sammenlignes med de øvrige emneområder.

Skala/data anvendelse

For de forskellige emneområder, anvendes skalaværdierne generelt selvstændigt indenfor de 4 områder de er opdelt i. Ønsker man et generelt mål for patienter/patientens helbred, udfra QLQ-C30 skemaet, fraråder manualen stærkt at man forsøger sig med andet end den dertil beregnede Global health status/QoL variabel[40].

Styrkeforhold i undersøgelser: Der findes ikke nogle beregninger på styrkeforhold (statistisk styrke) for QLQ-C30 skemaer i manualen. Vil man derfor undersøge styrkeforhold, fx i forbindelse med publikationsstørrelser, er det nødvendigt at indhente information fra andre undersøgelser og artikler der benytter sig af samme skema. Til dette formål findes en del materiale at bruge da QLQ-C30 skemaet, som nævn tidligere, er behandlet/anvendt i mere end 2200 studier (2001).

Data referencer: En anden måde at anvende sine data på er at sammenligne dem med andre relevante data. Til dette findes en referencemanual (EORTC QLQ-C30 Reference Values manual (Fayers et al., 1998))[41], hvor data er inddelt i tilstande af sygdomme.

Anker sammenligning: Anker baserede sammenligninger er en anden måde at anvende sine data på. Dette er særligt relevant hvis man ønsker at bestemme en udvikling over tid. Anker sammenligning baserer sig på Subjective Significance Questionnaire (SSQ) (Osoba et al., 1998)[42], hvor man lader patienter udfylde QLQ-C30 skemaer af fx to omgange og den anden gang også beder dem udfylde et SSQ skema. Denne skitserede fremgangsmåde for anvendelsen af et SSQ skema, blev i undersøgelsen (Osoba

[40][EORTC scoring manual, 2001] side 13 nederst.
[41]Ref. i: [EORTC scoring manual, 2001].
[42]Ref. i: [EORTC scoring manual, 2001].

et al., 1998)[43] anvendt til at vurdere hvorvidt ændringer, i henhold til to QLQ-C30 skemaer, kunne betegnes som hhv. lille, moderat eller meget store. Udfra denne undersøgelse kom man frem til følgende antagelser om interval inddeling af ændringer i skala (se understående tabel)[44].

Ændring:	
Kategori	Interval
Lille	5-10
Moderat	10-20
Meget	>20

Tabel 2.15: SSQ-kategorisering af ændringer i helbred/livskvalitet.

Ovenstående tabel kan anvendes i egne undersøgelser, til at kategorisere de patienter som oplever ændringer i deres helbred/livskvalitet.

I forhold til at bestemme signifikans af ændringer for score over tid eller mellem grupper, findes i manualen ikke nogen umiddelbar hjælp til dette formål. Manualen påpeger dog følgende[45]:

> "The fact that a change is statistically significant does not necessarily imply that it also has clinical significance."
>
> (Lydick and Epstein, 1993)

Ovenstående citat antyder en vigtig problemstilling, der ligeledes er relevant for spørgeskemaerne hhv. SF-36 og HAD.

Missing og bias

Missing: Proceduren for manglende besvarelse af spørgsmål, er for QLQ-C30 som følgende:

- $> 50\%$: Hvis mere en 50% af spørgsmålene for et emneområde ikke er besvaret (missing) udelades dette emneområde i senere beregninger.

[43]Ref. i: [EORTC scoring manual, 2001].
[44][EORTC scoring manual, 2001] side 14, 3 afsnit nedefra.
[45]Ref. i: [EORTC scoring manual, 2001] side 14 midt.

- $\leq 50\%$: Hvis mindre end eller lig med 50% er spørgesmålene er besvaret, laves en score ved at tage et gennemsnit at de besvarede spørgsmål, og efterfølgende lave skalatransformationer udfra dette "mindre" gennemsnit.

Bias: Mht. bias i data, der kan skyldes forskellige udgangspunkter/niveauer for patienterne, for hvilke de foretager deres valg i spørgeskemaet, nævnes i manualen nogle metoder til at kunne identificere bias samt korrigere sine data. Af disse metoder er: Kovarians-, regression- og faktoranalyser (Fayers and Machin, 2000)[46].

Validitet og reliabilitet

I manualen for QLQ-C30 foreligger ikke umiddelbart undersøgelser, der i samme grad som manualen for SF-36 skemaet, beskæftiger sig med validitet og reliabilitet. Der findes imidlertid andre artikler, der beskæftiger sig med validitet og reliabilitet for QLQ-C30[47].

Reliabilitet: Den anvendte undersøgelse (fra ovenstående artikel) behandler reliabilitets begrebet mht. den tiltænkte multisammenlignelige skalering af spørgsmålene i QLQ-C30. Der blev lavet tests vha. Cronbach's alpha[48] på udfyldte QLQ-C30 skemaer, og fundet en intern sammenhørrighed mellem skalaerne på alpha=0.70 og højere, hvilket ifølge (Nunnally & Bernstein, 1994)[49] er tilstrækkeligt til at sammenligne grupper.

Validitet: I den nævnte ovenstående artikel, er testet for 3 forskellige former for validitet. Hver test er foretaget ved at lade en gruppe patienter udfylde QLQ-C30 skemaer hhv. før og under en behandling:

- **Skala validitet:** Denne test undersøgte om der var en forventet høj korrelation (numerisk > 0.40 (korrelationen er tilladt at være negativ)) mellem sammenhørrige skalaer (fx fysisk funktion og træthed (fatigue)), og en tilsvarende (numerisk) lav korrelation mellem ikke sammenhørrige skalaer (numerisk < 0.40) (fx cognitive functioning and nausea and vomiting). I undersøgelsen fandt man at der var tilstrækkelig (hhv. høj og lav) korrelation mellem de forventede skalaer,

[46] Ref. i: [EORTC scoring manual, 2001] side 11, nederst.
[47] [N.K. Aaronson, 1993] A Quality-of-Life Instrument for Use in International Clinical Trials in Oncology.
[48] Behandlet i afsnit: (2.1.1) SF-36 - Spørgeskema om helbredsstatus.
[49] Ref. i: [N.K. Aaronson, 1993]

desuden var der en høj korrelation mellem den generelle helbred status/QoL og samtlige øvrige skalaer, hvilket bekræfter at denne skala kan bruges som et overordnet mål for patientens helbred.

- **Klinisk validitet:** Er test for om QLQ-C30 skemaet kan bruges til at udskille kliniske undergrupper, hvor patienter udskiller sig fx mht. klinisk status af sygdom. Testen blev udført ved på forhånd at kende til de undergrupper man derefter søgte efter i QLQ-C30 resultaterne. Vha. ensidet variance analyse blev fundet at der ikke var tilstrækkelig signifikante forskelle i QLQ-C30 mht. detektion af klinisk status af sygdom hos de adspurgte patienter (med en enkelt undtagelse ved emotional functioning). I tilfælde af undergrupper med andre kliniske karakteristika[50], blev i nogle tilfælde fundet en tilstrækkelig signifikant udskillelse, men disse resultater har ikke ikke umiddelbart givet anledning til generelle antagelser at gå ud fra.

- **Ændring i helbredsstatus over tid:** Er test for om QLQ-C30 kan anvendes til at bestemme helbredsstatus ændringer over tid. Testen blev udført ved at holde QLQ-C30 resultater (ved start og midt i en behandling) op imod observerede ændringer i helbred. Resultaterne af undersøgelsen viste, at ved på forhånd at opdele i grupper af hhv. forbedret, uforandret og forværret helbredsstatus, var der signifikante tendenser til stigende funktionsniveau og et generelt fald i symptomer hos patienter med et forbedret helbred, mens det modsatte var tilfældet hos patienter med forværret helbred og ingen ændringer var hos patienter med et uforandret helbred. Med denne undersøgelse bekræftes at QLQ-C30 skemaet godt kan bruges til at se ændringer over tid for patienter der oplever forandringer i deres helbred, men med det forbehold at man som minimum, på forhånd bør udskille patienterne i grupper af hhv. forbedret, uforandret og forværret helbred, og derved kan QLQ-C30 skemaet antageligt kun bruges til at bekræfte en sådan forhåndsudskillelse.

Af alternative undersøgelser, som i validitets sammenhæng kan nævnes er Cox et al. (1992), Olschewski and Schumacher (1990)[51], der beskæftiger sig med alternative måder at behandle skalerings beregningerne på.

[50] For specifikke data se undersøgelse: [N.K. Aaronson, 1993].
[51] Ref i: [EORTC scoring manual, 2001] side 13, 3 afsnit nedefra.

I det følgende er medtaget en problemstilling der nævnes i manualen. Problemstillingen, der omhandler vægtning af de enkelte spørgsmål i forhold til gennemsnitsberegningerne, er medtaget for at beskrive den stadige evaluering QLQ-C30 skemaet befinder sig i, mht. vurdering af skemaet og fortolkning af data.

- Problemstillingen går ud på følgende: Skal alle spørgsmål vægtes lige ? (ved nuværende gennemsnitsberegning foretages ingen vægtning). Der kan fx sættes tvivl om det enkelte spørgsmål skal korrigeres i forhold til den senere råscore beregning, altså om hvert enkelt spørgsmål skal "tælle" lige meget i gennemsnitsberegningen.

I manualen[52] nævnes at man stadig anbefaler at bruge den simple summering og gennemsnitsberegning (Likert Metoden), men at man er i gang med at undersøge om alternative måder at beregne råscore og skalaer på er mere relevante.

For yderligere validitets og reliabilitets begreber henvises til underafsnit: "Reliabilitet, Floor, Ceiling og Validitets begreber" af afsnit: (2.1.1) "SF-36 - Spørgeskema om helbredsstatus", da samme betingelser og begreber synes gældende for QLQ-C30.

2.1.4 Evaluering af generelle helbredsmål PCS og MCS

I SF-36 skemaet er det muligt at udregne 2 generelle mål/ estimater for hhv. patientens fysiske som psykiske helbred (PCS og MCS)[53], disse mål udregnes vha. følgende tre trin:

- Trin 1: Transformering af de endelige scores vha. amerikanske middelværdier og spredninger (SF-36 data).

- Trin 2: Vægtning og summering af de transformerede scores, vægtningerne er estimeret vha. "principal komponent analyser af data fra den amerikanske befolkning".

- Trin 3: Tilbagetransformering vha. middelværdi og spredningsestimater, ligeledes fra SF-36 data af den amerikanske befolkning.

[52] [EORTC scoring manual, 2001]

[53] Afsnit: (2.1.1) SF-36 - Spørgeskema om helbredsstatus, underafsnit: Variable for fysisk og psykisk helbred.

De tre ovenfor skitserede trin giver ikke anledning til nogle problemer som sådan. Problemerne opstår som følge af den skala hvor de to mål (PCS og MCS) befinder sig på, da denne ikke er sammenlignelig med den tidligere anvendte 0-100 skala. Dette giver anledning til nogle problemer omkring PCS og MCS målenes definitionen og anvendelsen.

I denne forbindelse er det aktuelt at se på de vægtninger, der anvendes i trin 2, vægtningerne er gengivet i følgende tabel.

Skala	Vægtning	
	PCS	MCS
PF	0,42	-0,23
RP	0,35	-0,12
BP	0,32	-0,10
GH	0,25	-0,02
VT	0,03	0,24
SF	-0,01	0,27
RE	-0,19	0,43
MH	-0,22	0,49

Tabel 2.16: Vægtningstabel for PCS og MCS estimering (SF-36).

Det ses af ovenstående tabel, at alle 8 Skalaer/helbredskategorier (PF, RP, BP, GH, VT, SF, RE og MH) indgår i vægtningsmålet for PCS og MCS, samt at nogle af vægtningerne er negative. Denne kombination af positive og især negative vægtninger, giver en meget abstrakt fortolkning af de generelle mål PCS og MCS, som i nedenstående er skitseret via nogle eksempler:

- 1. En patient, der opnår maksimal score i alle 8 helbredskategorier, scorer ifølge SF-36 manualen skalaværdien 100 i samtlige helbredskategorier. Dette kan umiddelbart tolkes, som den bedste mulige sundhedsmæssige tilstand, og giver derfor maksimale værdier for hhv. PCS og MCS.

  De negative vægtninger betyder imidlertid at en kombination med lavere scores kan giver en endnu højere skalaværdi for PCS eller MCS, end forrige eksempel. For PCS vil en sådan kombination fx være givet ved, at score maksimalt på de 5 første helbredskategorier og minimalt på de sidste 3 (SF, RE og MH).

- 2. En patient kan ligeledes opnå en højere værdi for dennes generelle psykiske helbreds (MCS), ved at have mindsket score i de 4 første helbredskategorier og maksimal score (100) i de øvrige 4.

Ovenstående eksempler giver som tidligere nævnt en meget abstrakt opfattelse af de generelle helbreds komponenter (PCS og MCS), hvilket fx giver sig udslag i følgende hændelige situationer.

- 1. En patient opnår maksimal score i samtlige 8 helbredskategorier, derved opnår personen følgende værdier:

  $$\boxed{\begin{array}{c}\text{PCS} = 57{,}87\\ \text{MCS} = 62{,}14\end{array}}$$

  Den samme patient oplever efterfølgende store fysiske smerter, derfor giver patienten udtryk for disse smerter i næste SF-36 undersøgelse. Dette sker ved at patienten udfylder det nye SF-36 skema som ved forrige test, med undtagelse af spørgsmål, som vedrører fysiske smerter. Derved opnår patienten fx en score for BP-Bodily pain (Fysisk smerte) på 20, hvilket er 80 point lavere end den forrige score på 100.

  Ved udregning af PCS og MCS for det nye SF-36 skema, giver den mindre score i BP imidlertid en højere værdi for MCS på 65,44 mens PCS falder til 47,09. Forøgelsen i MCS kan derfor tolkes således: Patienten får et generelt bedre psykisk helbred, ved at have fysiske smerter.

  Note: Ovenstående situation bliver yderligere forstærket hvis patienten ligeledes oplever: Dårligere fysisk funktion, fysisk betinget begrænsninger og nedgang i alment helbred, hvorved MCS stiger yderligere.

- 2. En Patient, der har scoret 0 i samtlige 8 helbredskategorier og umiddelbart har det værst tænkeligt (ud fra SF-36), opnår en generel psykisk helbreds værdi (MCS) på 17,34. Den samme patient kan imidlertid opnå en endnu lavere værdi for MCS, ved at score højt på de 4 første helbredskategorier (PF, RP, BP og GH). Dette vil i praksis betyde at patienten der psykisk har det meget dårligt, får det psykisk endnu dårligere ved samtidig at have: En fysisk god funktion, ingen fysisk betinget begrænsninger, ingen fysiske smerte og et alment godt helbred (mindst værdi for MCS er -1,26).

Problemet med tolkningen af PCS og MCS skalaen, kan imidlertid anskues fra en anden vinkel:

- **MCS**: En patient, der har scoret 100 i samtlige 8 helbredskategorier opnår et MCS på den umiddelbare grænse (62,14). Da samme patient i et senere spørgeskema giver udtryk for fx smerte, kan den derved øgede MCS værdi forklares ud fra følgende vurdering: Patientens generelle mentale helbred er blevet forbedret, da patienten scorer højt på de psykiske helbredsskalaer, selvom patienten nu også døjer med smerter. Dette betyder at den fysiske smerte hos patienten sammen med den helbredsmæssige gode psykiske tilstand, er grundlag for at de psykiske helbredskategorier vurderes højere.

En anden måde at anskue problemet, kan ligeledes anvendes i den nedre del af MCS skalaen:

- **MCS**: En patient, som har et generelt godt fysisk helbred og som stadig scorer generelt lavt mht. de psykiske helbredskategorier, kan med en hvis rimelighed, vurderes at have en forværret psykisk helbreds tilstand. Dette skyldes at man må formode at et forbedret fysisk helbred ligeledes burde give anledning til et løft i den psykiske tilstand, men da dette ikke indtræffer kan en lavere MCS-værdi være hensigtsmæssig.

Mht. det generelle fysiske helbred (PCS), bliver ovenstående betragtninger en anelse sværere at gennemskue.

- **PCS**: Er en patient fx fysisk i topform men har psykiske problemer, skal patienten ifølge de anvendte vægtninger have en højere værdi for PCS end hvis patienten ikke havde psykiske problemer. Umiddelbart synes dette ulogisk men kan ligeledes tolkes som: At patienten må være "i ekstra god form" hvis personens psykiske helbred ikke har indflydelse på det fysiske helbred.

- **PCS**: I den nedre del af PCS skalaen kan et lavere PCS betyde at hvis en patient har det psykisk bedre men har et uforandret lavt fysisk helbred, betyder dette at patienten vurderes til at have et generelt "lavere" fysisk helbred, da det forbedret psykiske helbred ikke har givet anledning til forbedringer fysisk.

De anvendte vægtninger har den konsekvens, at kender man ikke den "score-mæssige" baggrund for de estimeret værdier af PCS og MCS, er det forbundet med stor usikkerhed at antage noget om patientens helbred baggrund af disse skalaer. Dette skyldes hovedsageligt at forskelle eller ændringer i PCS og MCS værdier kan have uigennemsigtige årsager, og ikke nødvendigvis være grundet af reelle fysiske eller mentale helbredsændringer.

I forbindelse med fortolkning af helbredsværdierne, er det ligeledes vigtigt at man ikke adskiller PCS og MCS målene estimeret på den samme patient (ved samme tidspunkt/måling). Dette skyldes at en forbedring i fx MCS vil kunne ses som en forværring af det andet mål PCS (og omvendt), disse modsatrettede ændringer indholder vigtig information, om tolkning af patientens egentlige helbredsmæssige tilstand.

På baggrund af ovenstående, er det vurderet at PCS og MCS estimaterne ikke er tilstrækkelige entydige mht. skalering og fortolkning. Derfor er det valgt at undgå disse generelle helbredsestimater fra SF-36 spørgeskemaet i denne rapport.

Kapitel 3

Behandling af spørgeskemadata, fase 1

Kapitlet behandler data fra første fase af projektet mht. de tre spørgeskemaer (SF-36, HAD og EORTC QLQ-C30). Kapitlet er opbygget som følgende:

- Den indledende databehandling af spørgeskemaer, behandler projektets forløb og dataindsamling mht. spørgeskemaerne. Det er beskrevet med hvilke variable det er muligt at inddele dataene samt hvilke variable, der er benyttet i den videre databehandling.
- I den grafiske præsentation af spørgeskema data, er vha. forskellige former for plots givet en overordnet gennemgang af det indsamlede datamateriale. De enkelte plots er sammenlignet og diskuteret mht. generelle træk og tendenser, og relevante problemstillinger er søgt fremhævet.
- I den efterfølgende analyse og tests af spørgeskema data, er tendenser fra den grafiske fremstilling af dataene efterprøvet, og der er givet en kort beskrivelse af teorien bag de forskellige tests og analyse metoder.

3.1 Indledende databehandling af spørgeskemaer

Den indledende databehandling består af følgende afsnit:

- Beskrivelse af data fra spørgeskemaer, fase 1
- Udvælgelse af patienter til fase 1
- Indsamling af spørgeskema data
- Relevante variable
- Opdeling af patienter

3.1.1 Beskrivelse af spørgeskemadata fra fase 1

Projektet Krop og Kræft er inddelt i forskellige faser, hvor foreløbig 1. fase er fuldt gennemført[1]. Fase 1 af projektet er tiltænkt som en slags indkøringsfase, og er gennemført med 25 patienter (se nedenstående note), mens efterfølgende faser (2 og 3) forventes gennemført med et øget antal patienter. Hver fase gennemføres af éen omgang[2].

Note: 2 af de 25 patienter er pga. frafald i projektetfasen, taget ud af datasættet, og der skal i det følgende ikke regnes med 25 patienter men udelukkende 23 patienter (udvælgelsen af de 2 patienter, er foretaget af UCSF).

3.1.2 Udvælgelse af patienter til fase 1

Patienterne fra fase 1 har primært selv henvendt sig til projektet, og er derefter blevet udvalgt af medarbejderne i projektet. Denne udvælgelse er udført på baggrund af en række betingelser som beskrevet i det følgende:

Rekrutteringen af patienter er foretaget[3]: "Ud fra et ønske om en bred differentiering på køn, alder og sygdomsgrupper".

Af forskellige sygdomsgrupper, repræsenterer de 23 patienter i fase 1 følgende:

[1] Fase 2, er ikke fuldt gennemmført mens denne rapport skrives
[2] Se evt.: Indledning.
[3] Projekt Krop & Kræft. Resultater for fase 1: Rekruttering.

Antal patienter	Kræftdiagnose/kræfttype
6	Breast-cancer (Brystkræft)
6	Colon-cancer (Tarmkræft)
2	Testicular-cancer (Testikelkræft)
2	Ovarie-cancer (Æggestokkræft)
4	Haematologic-cancer (Blodkræft)
3	Others (Uteri, lung, ect.) (Livmoder, lunge osv.)

Tabel 3.1: Fordeling af kræftdiagnose for de 23 patienter.

I forhold til sygdomsstatus repræsenterer de 23 patienter følgende:

- 10 patienter var sygdomsfri (complete remission).
- 7 patienter var delvis sygdomsfri (partial remission).
- 6 patienter havde klinisk restsygdom (progressive).

3.1.3 Indsamling af spørgeskemadata

Fase 1 har følgende overordnet forløb mht. indsamling af data til spørgeskemaer:

- De 23 patienter er inddelt i tre hold af 7-10 personer.
- Hvert hold har gennemløbet en 6 ugers projektfase (intervention).
- Når et hold er færdigt med en 6 ugers projektfase, er et nyt hold startet umiddelbart derefter.

Udfyldelsen af skemaer er for hvert hold foregået som følgende:

- Ved projektetfasens start, der betegnes med uge 1.
- Ved projektetfasens afslutning, dvs. 6 uger fra uge 1.
- Tre måned fra projektetfasens afslutning hvilket er 18 uger fra uge 1.

Patienterne har selv stået for udfyldelsen af de forskellige spørgeskemaer, hvilket bl.a. har betydet at datoerne for udfyldelserne ofte er forskellige og ikke nødvendigvis er den samme for hvert hold. Med enkelte undtagelser er et holds udfyldelse af skemaer, sket indenfor nogle få dage så derfor er de anførte datoer ikke medregnet i behandlingen af data, men udelukkende adskilte ud fra hvilket "nummer" af test de tilhører:

$$\begin{aligned} \text{Uge}(1) &= \quad \text{testnr. 1} \\ \text{Uge}(6) &= \quad \text{testnr. 2.} \\ \text{Uge}(18) &= \quad \text{testnr. 3} \end{aligned}$$

3.1.4 Relevante variable

I de efterfølgende 3 tabeller er kort opsummeret betydningen af de forskellige relevante helbredskategorier, fra hver af de tre spørgeskemaer. For yderligere information om disse variable henvises til afsnit for det respektive spørgeskema, samt til den anvendte litteratur.

SF-36

Variabel	Variabel forklaring
PF	- Fysisk funktion
RP	- Begrænsninger - fysisk betinget
BP	- Fysisk smerte
GH	- Alment helbred
VT	- Energi
SF	- Social funktion
RE	- Begrænsninger - psykisk betinget
MH	- Psykisk velbefindende
PCS	- Generel fysisk helbred
MCS	- Generel psykisk helbred

Tabel 3.2: Helbredskategorier SF-36.

HAD

Variabel	Variabel forklaring
ANGST	- Angst skala
DEPRESS	- Depressions skala

Tabel 3.3: Helbredskategorier HAD.

EORTC QLQ-C30

Variabel	Variabel forklaring
Global helbreds status/Qol:	
QL2	- Global helbreds status/Qol
Funktions skala:	
PF2	- Fysisk
RF2	- Begrænsninger - fysisk
EF	- Følelsesmæssig
CF	- Koncentration
SF	- Social
Symptom scales/items:	
FA	- Træthed udmattelse
NV	- Kvalme og opkastning
PA	- Smerte
DY	- Åndenød
SL	- Søvnløshed
AP	- Manglende appetit
CO	- Forstoppelse
DI	- Diarre
FI	- Økonomiske vanskeligheder

Tabel 3.4: Helbredskategorier EORTC QLQ-C30.

Note: Til patientidentifikation haves patientens navn og fulde personnummer. Disse oplysninger er dog i henhold til aftale med UCSF fortrolige, og vil udelukkende blive anvendt til at inddele patienterne i undergrupper (se efterfølgende afsnit: Opdeling af patienter).

3.1.5 Opdeling af patienter

I dette afsnit vil blive gennemgået en overordnet opdeling af patienterne samt en diskussion af hver enkelt opdelings betydning for den videre databehandling.

I forbindelse med bearbejdningen af data fra spørgeskemaerne er det interessant at undersøge i hvor stor grad effekter af forskellige inddelinger eventuelt har på udfaldet af spørgeskema undersøgelserne, inddelingerne er følgende: Hold/team, køn, alder, fysisk aktivitet ved baseline, civilstatus,

sygdomsstatus ved inklusion (optagelse i projektet) samt den tidsmæssige opdeling af spørgeskemaundersøgelserne i tre tests. Opdelingerne er primært lavet ud fra information givet af patienterne til UCSF, og er fundet interessant at undersøge i forbindelse med vurdering af projekt Krop & Kræft.

I det følgende er kort skitseret ideen med de forskellige opdelinger:

- **Hold/team:** I forhold til køn- og aldersopdelinger er hold opdelingen allerede foretaget fra projektets start, i form af hold af 7 til 10 patienter.

  Hold 1.
  Hold 2.
  Hold 3.

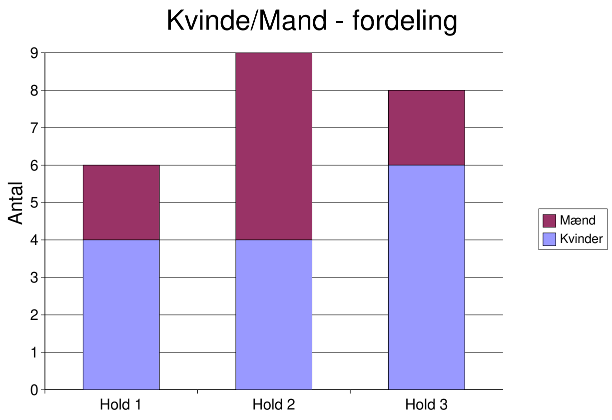
  Det interessante i denne sammenhæng er om patienter på samme hold har haft en indvirkning på hinanden, så det har betydet en effekt på den enkelte patients besvarelse af spørgeskemaerne. Et tænkt eksempel i denne forbindelse, kunne fx være at et særligt godt holdsammenhold har givet deltagerne ekstra selvtillid og derved påvirket dem til at svare mere positivt på enkelte spørgsmål. I en efterfølgende behandling af data vil en forskel mellem hold kunne medføre, at der skal medtages en "holdfaktor" i analyse af spørgeskemaresultater.

- **Køn:** En opdeling af patienterne i køn, sker ud fra den antagelse at der er forskel på kvinder og mænds præferencer. Fx er en mulig præferenceforskel at mænd har en lavere smertetærskel end kvinder og derved kan have tendenser til at score generelt lavere på skalaer omhandlende smerte (fx helbredskategorien BP (fysisk smerte) i EORTC skemaet).

  Mand/kvinde fordeling:

  M: Mand.
  K: Kvinde.

  I nedenstående graf ses fordelingen af kvinder og mænd i de tre hold.

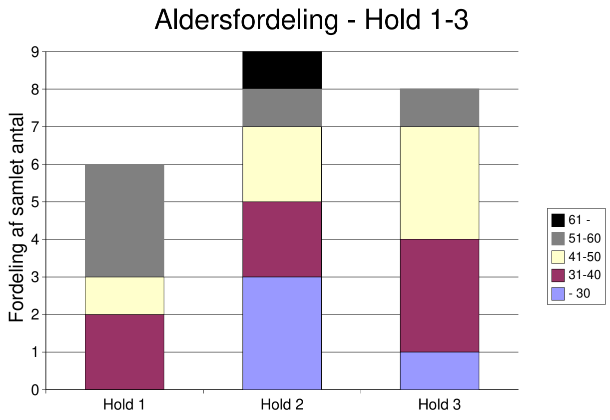


Figur 3.1: Fordelingen af kvinder og mænd pr. hold.

Af ovenstående fordelingen ses, at der for to af holdene er en lille overvægt af kvinder (med undtagelse af hold 2). Til dette skal indskydes at det ikke er de faktiske hold, som ses i ovenstående graf, men udelukkende de 23 patienter der er modtaget data fra.

Det ses yderligere af kvinde/mand fordelingen, at der datamæssigt er 6 patienter i hold 1, 9 patienter i hold 2 samt 8 patienter i hold 3.

- **Alder:** I nedenstående graf er vist aldersfordelingen mellem patienterne i de 3 hold. Patienterne er blevet inddelt i 5 grupper mht. alder. Denne opdeling er valgt for at gøre grafen mere overskuelig.



Figur 3.2: Fordelingen af alder pr. hold.

Det ses af ovenstående aldersfordeling, at patienterne aldersmæssigt er fordelt umiddelbart jævnt mellem de tre hold.

Den store spredning i alder mellem patienterne (18 til 63 år), kan i forbindelse med spørgeskemaudfyldelse, antageligt give anledning til præferenceforskelle mellem de forskellige alderstrin. Denne mulige præferenceforskel har derfor betydet at det er valgt at opdele patienterne mht. alder. Opdelingen er foretaget ud fra følgende vurderinger af data, samt af den ovenstående graf:

- Medianalderen for de 23 patienter der gennemførte fase 1 er 40 år.
- Da antallet af patienter er forholdsvist lavt (23 personer.), betyder en grov opdeling, frem for en finere opdeling, at der vil være et større datagrundlag at foretage beregninger ud fra. Den grovere opdeling medfører desuden tab af information og dermed en øget støj.

Der er valgt følgende opdeling/gruppering af patienterne mht. alder:

A: Født 1962 og efter ($\leq$ 40 år).
B: Født før 1962 ($>$ 40 år).

- **Aktivitet baseline:** Henviser til, hvor aktiv den enkelte patient er i sin fritid før han/hun starter i projektet, og dækker over følgende kriterier (baseline = første projektdag):

  I: Stillesiddende (bevæger sig stort set ikke).
  II: Cykler og går ture af lyst (dvs. ikke planlagt - tilfældigt).
  III: Regelmæssigt fysisk aktiv svarende til mindst 3 timer om ugen.
  IV: (Hård) fysisk træning i over 4 timer om ugen.

  Denne opdeling tager udgangspunkt i om patienter, med fx et i forvejen højt fysisk aktivitetsniveau, får mere ud af projektforløbet ved fx at opnå bedre resultater end patienter der på forhånd har et meget lavt fysisk aktivitetsniveau.

- **Sygdomsstatus ved inklusion:** Der er tre inddelinger af patienterne, som hver betegner den enkelte patients sygdomsstatus ved indgang til projektet (inklusionstidspunktet er der, hvor patienten underskriver et samtykke om projektet). Inddelingerne ser ud som følgende:

  ET: Complete remission (sygdomsfri i forebyggende behandling).
  TO: Partiel remission (delvis sygdomsfri).
  TRE: No change (fortsat kræftsygdom).

  Formålet med denne opdeling er at have mulighed for at teste om fx patienter med en sygdomsstatus på "komplet remission", adskiller sig fra de øvrige patienter i den anvendte behandlingen.

- **Civilstatus:** Hvorvidt patienten bor alene, er samboende med anden person (fx kæreste) eller er gift og derved bor med sin ægtefælle:

  Alene.
  Samboende.
  Gift.

  Denne opdeling er i forhold til de foregående af en mere social karakter. Dens formål er at teste om sociale forskelle påvirker de resultater patienten opnår i forbindelse med projektforløbet. Det kunne fx tænkes at patienter der bor alene er mere tilbøjelige til at fokusere på sin kræftsygdom og derved have større tilbøjelighed til angst eller depression, end en patient der i det daglige modtager personlig støtte fra en ægtefælde.

- **Testnr:** Udfyldelsen af de forskellige spørgeskemaer er for hvert hold af patienter foretaget over tre omgange (hhv. 1, 6 og 18 uger, hvilke er beskrevet i det foregående afsnit: Indsamling af spørgeskema data). Disse tre omgange vil i det efterfølgende blive behandlet som tre opdelinger af fase 1 (hhv. test nummer 1, 2 og 3), hvilket betyder at test nummereringerne foretages på tværs af holdopdelingen, selvom holdene i virkeligheden ikke har udfyldt spørgeskemaerne i de samme tidsintervaller.

  Testnr 1.
  Testnr 2.
  Testnr 3.

  Opdelingen i de tre testnumre betyder at dataene kan betragtes som longitudionale data, med en observation/måling pr. patient pr. helbredskategori pr. projektfase.

Det er yderligere fundet hensigtsmæssigt at opdele patienterne i hhv. patienter med komplette data, samt patienter med manglende data (missing) hhv.:

KOMP.: Komplet data.
MISS.: Missing data.

Patienterne er dermed opdelt i patienter der har udfyldt og afleveret samtlige spørgeskemaundersøgelser (KOMP.), samt patienter der kun har udfyldt og afleveret spørgeskemaer i enkelte spørgeskemaundersøgelser (MISS.).

3.2 Grafisk præsentation af spørgeskemadata

I dette afsnit er vha. forskellige plots vist et udvalg af data fra spørgeskemaerne af første fase af projektet. Formålet er at give et overblik over patienternes generelle helbred udtrykt i form af helbredsvariablene, samt fremhæve interessante mønstre og tendenser mht. dataenes opførsel.

Den grafisk fremstilling af de longitudionale data, er lavet ud fra følgende fire hovedregler, beskrevet af Peter J. Diggle[4]:

- 1: Vis så meget af det relevante rådata som muligt, istedet for kun at vise det summariske (summaries).
- 2: Fremhæv udvalgte mønstre af potentiel videnskabelig interesse.
- 3: Udpeg både mønster på tværs af sektioner samt longitudionale mønstre.
- 4: Gør det let at identificere usædvanlige individer eller usædvanlige observationer.

Outliers: Punkt 4 i ovenstående punkter drejer sig bl.a. om så kaldte "outliers", hvilke er observationer der udelukkes af datamængden, da de fx vurderes at være væsentlig afvigende fra de resterende data. Da datamaterialet i første fase er relativt lille, vil det være forbundet med stor usikkerhed i denne henseende at udelukke enkelte afvigende observationer/outliers. Dette skyldes at der ikke er tilstrækkelig med information i det begrænsede datamateriale til at kunne bestemme om en måling er tilstrækkeligt afvigende fra de øvrige, samt at skalaerne hvor-på målinger opgives, i sig selv virker afgrænsende.

Udvalgte helbredskategorier: Af pladshensyn er det vurderet at det ikke er hensigtsmæssigt at vise data fra samtlige helbredskategorier. Derfor består de efterfølgende plots af følgende udvalgte helbredskategorier, fra hver af de tre spørgeskemadata:

SF-36	**HAD**	**EORTC QLQ-C30**
VT	ANGST	QL2
GH	DEPRESS	
SF		

Tabel 3.5: Udvalgte helbredskategorier.

[4][Diggle, 1994] side 33, andet afsnit.

De udvalgte helbredskategorier i ovenstående tabel, er valgt ud fra følgende vurderinger:

- EORTC QLQ-C30: I dette skema er valgt den generelle helbredskategori QL2 der måler livskvalitet hos patienten, og er den eneste overordnet helbredskategori det ifølge skemaet er forsvarligt at anvende.

- SF-36: Ud fra SF-36'manualen[5], afspejler alment helbred (GH), energi/træthed (VT) og social funktion (SF), patientens overordnet helbred. Dette gør disse helbredskategorier fordelagtige at anvende som generelle helbredsmål, i forhold til de resterende helbredskategorier, der er mere specifikke mål for enten fysisk eller mentalt helbred.[6]

  (Note: De overordnede helbredskategorier PCS og MCS er valgt fra)[7].

- HAD: For dette skema findes udelukkende 2 helbredskategorier hhv. ANGST og DEPRESS (depression).

De øvrige helbredskategorier vil blive behandlet i den efterfølgende parret t-test, og for disse henvises til det omtalte kapitel af samme navn.

Missing data: Datamaterialet består af en del missing data. Dette skyldes primært at 11 af de 23 patienter der er med i første fase af projektet, ikke har besvaret alle tre spørgeskemaer mht. de tre testtidspunkter. De missing data har bevirket at for de 11 patienters vedkommende, har det ikke været muligt at estimere deres scores for alle tre tests (testnr. 1, 2 og 3), og derved vise disse i de efterfølgende plots. I nedenforstående tabel er angivet hvor mange patienter der har besvaret de enkelte spørgeskemaer.

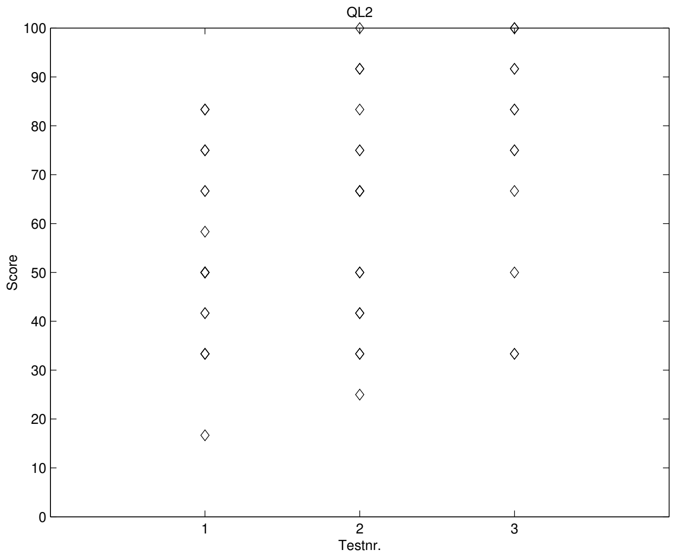
Antal patienter pr. testnr. pr. spørgeskema:			
Testnr.	SF-36	HAD	EORTC QLQ-C30
1	23	23	23
2	22	23	22
3	13	13	13

Tabel 3.6: Antal besvarede spørgeskemaer.

[5][SF-36 manual, 1997] Side 14, afsnit 5.
[6][SF-36 manual, 1997]
[7]Se evt. afsnit: (2.1.4) Evaluering af generelle helbredsmål PCS og MCS (SF-36).

Tabel note: De patientantal der er angivet i ovenforstående tabel, er for testnr. 1, 2 og 3 med få undtagelser, udtryk for de samme patienter. Af de 23 patienter har reelt kun 12 patienter besvaret alle tre tests og alle tre af spørgeskemaer SF-36, HAD og EORTC.

3.2.1 EORTC QLQ-C30: QL2



Figur 3.3: Scatterplot af QL2-score (*test 1 - 3 mod scores*).

I ovenstående scatterplot er vist patienternes score af den overordnede helbredskategori QL2 (quality of life), som funktion af tiden/tests (testnr.). Scatterplottets primære formål er at vise patienternes QL2-score og gengive dem longitudionalt på den anvendte skala.

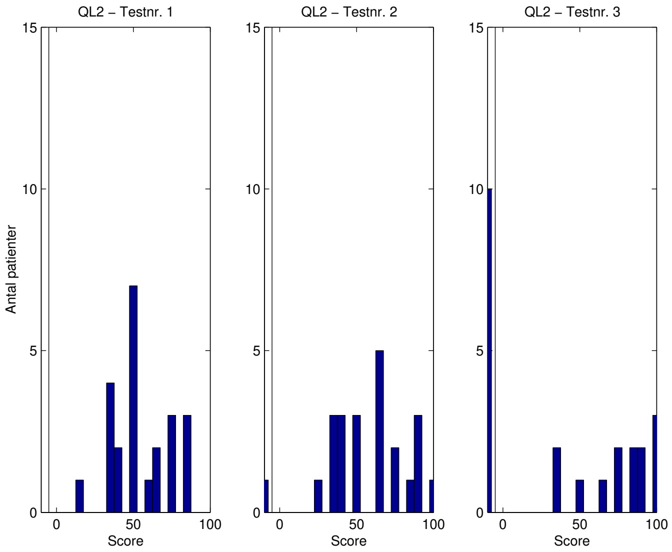
Mht. spredning af scores for de enkelte tests, synes scorene for test 1 at være jævnt fordelte imellem ca. 15 og ca. 85. Spredningen synes svagt øget i test 2, men fra test 2 til 3 synes spredningen ikke yderligere øget, snarer tvært imod.

I forbindelse med visning af scores i et scatterplot, har følgende problemstilling en vigtig betydning:

- Diskret udfaldsrum: Da scoren for helbredskategorierne beregnes ud fra diskrete svarmuligheder (fx 1, 2, 3 eller 4), er scoren for hver helbredskategori ligeledes begrænset til et diskret udfaldsrum. Dette diskrete udfaldsrum er givet ved et endeligt sæt af værdier på skalaen (0-100), og betyder at når disse scores vises vha. et scatterplot, vil nogle punkter i plottet kunne dække over flere observationer/punkter på samme tid.

Det "diskrete udfaldsrum" i ovenstående scatterplot, medvirker i stor grad til at information om fordelingen af scores, i de enkelte tests, ikke kan ses. Denne information kan i stedet ses i nedenstående histogrammer, hvor hver test gengives mht. fordeling af scores.

I histogrammerne findes en lodret linie, denne linie adskiller de missing patientdata, fra patientdata der er beregnet scores på. Dvs. at højre side af histogrammerne (adskilt af linien), viser fordelingen af scores mellem patienter der har besvaret de pågældende spørgeskema, mens venstre side viser hvor mange patienter der ikke har besvaret spørgeskemaet.

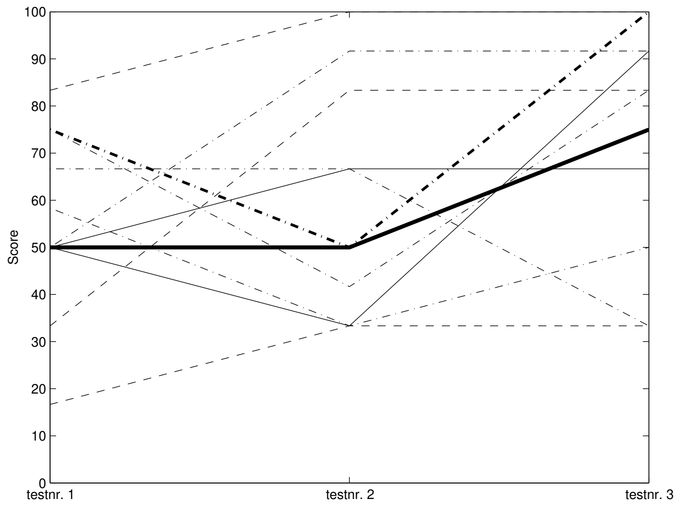


Figur 3.4: Histogrammer over fordeling af QL2-scores (*score mod antal*).

I ovenstående histogrammer ses at det er svært at sige noget konkret om spredningen, for de tre tests.

Ses de både på histogrammerne og scatterplottet, synes der at være en tendens til at det generelle QL2-niveau stiger 6 uger efter behandlingsstart (testnr. 2). En ligeledes tendens til generel stigning, synes at gøre sig gældende 18 uger efter projektets start (testnr. 3: dvs. 12 uger efter projektets slutning).

Det ses yderligere at antallet af patientdata falder markant fra test 2 til 3 med 10 patienter. Disse manglende data betyder at eventuelle tendenser, der synes gældende mellem test 1 til 2 er yderst svære at føre videre til test 2 til 3. Denne problemstilling vil ligeledes gøre sig gældende i de øvrige plots.



Figur 3.5: Spaghettiplot af QL2-score (*test 1 - 3 mod score*).

I ovenstående plot ses sammenhængen mellem 12 udvalgte patienters score på QL2-skalaen. Punkterne viser hvilke QL2-scores den enkelte patient har opnået i de enkelte testnumre, og stregerne viser sammenhænge mellem punkterne. De 12 patienter er dem som har besvaret EORTC-skemaet i alle tre tests, og derved kan vises som ubrudte linier[8].

Det ses af spaghettiplottet at tendenserne fra scatterplottet (QL2), om stigende score for QL2 for stigende tests, i nogen grad bekræftes i et spaghettiplot. Dette vurderes ud fra følgende iagttagelser fra plottet:

Testnr. 1 - 2:

- 5 patienter oplever en stigende QL2-score.
- 4 patienter oplever et fald i QL2-score.
- De resterende 3 patienter opnår samme score som ved forrige test.

[8]Linien som har følgende scores (50 - 50 - 75) mht. testnr 1 til 3, viser scores for to patienter.

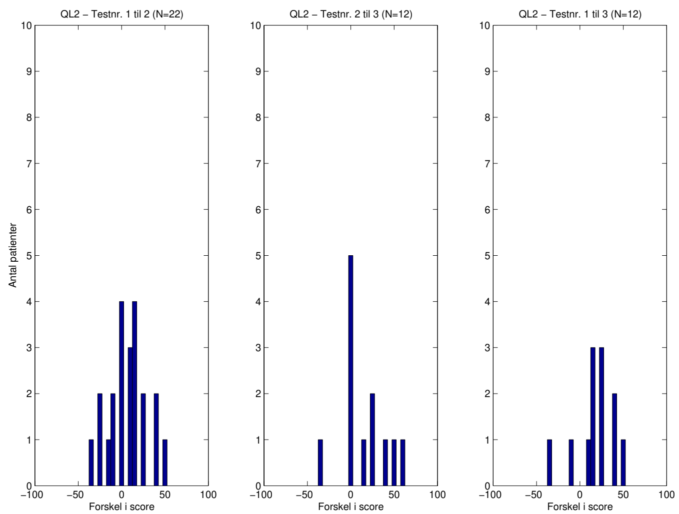
Testnr. 2 - 3, samt 1 - 3:

- Af de 5 patienter der oplevede et øget QL2 imellem testnr. 1 - 2, oplever kun 1 patient en videre forøgelse. De resterende bliver på forrige niveau.
- Af de 4 patienter der oplevede et fald i QL2 imellem testnr. 1 - 2, oplever de alle nu en forøgelse af QL2, men den ene patient kommer ikke op over sit startniveau.
- Af de 3 patienter der oplevede samme niveau i testnr. 2, som ved testnr. 1 oplever 2 et øget QL2, mens en enkelt patient oplever et fald i QL2.

Ovenstående iagttagelser giver fundament for følgende betragtninger:

- Den stigende tendens fra QL2-scatterplottet mellem testnr. 1 til 2 synes ikke at være gældende, idet kun lidt over halvdelen af de i alt 12 patienter reelt oplever en stigning.
- Den stigende tendens fra testnr 1 til 3, synes gældende idet 10 patienter ud af 12 reelt oplever en forøgelse af deres QL2.

I nedenstående histogram, vises fordelingen af ændringer i scores for de patienter der har besvaret de respektive spørgeskemaer. Dvs. at reelt 22 patienter har besvaret tests 1 og 2, mens kun 12 patienter af disse har besvaret test 3.



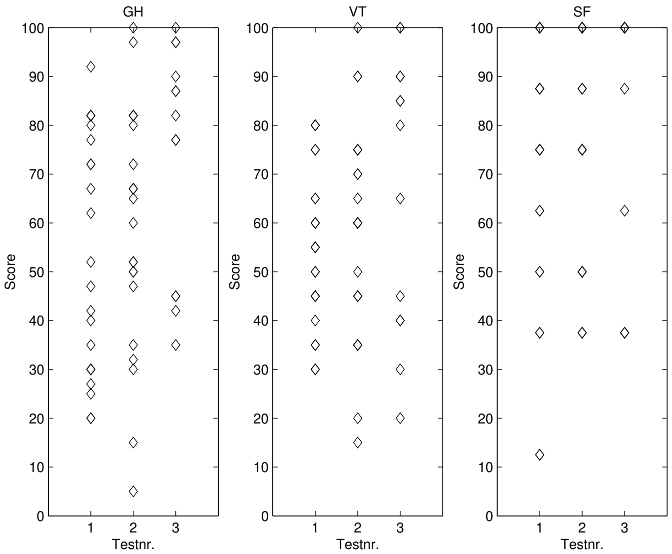
Figur 3.6: Histogrammer for forskelle i QL2-scores (*forskelle i scores mod antal*).

I ovenstående histogrammer, ses at der er en lille overvægt af patienter der har oplevet en stigning i deres QL2-score fra test 1 til 2. Ligeledes ses i det midterste histogram, at 5 patienter ikke oplever en yderligere stigning mellem test 2 til 3, og at een patient oplever et svækket QL2 (de resterende 6 patienter oplever et øget QL2).

I histogrammet yderst til venstre, ses udviklingen af 12 patienters scores fra test 1 til 3, og det ses at reelt 10 patienter oplever et øget QL2, mens 2 patienter oplever et fald på hhv. 10 og 40 point. Af de patienter der oplevede et øget QL2, oplever en enkelt en forøgelse på 50 point, mens hovedvægten af de resterende ligger på cirka 10 til 20 point.

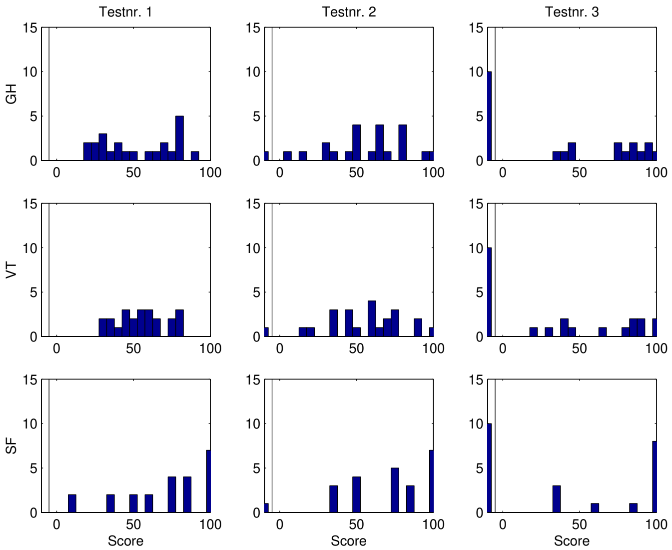
Ud fra de viste plots (scatterplot, spaghettiplots og histogrammer), synes der at være en tendens til en forøgelse af størstedelen af patienternes QL2-score.

3.2.2 SF-36: VT, GH og SF



Figur 3.7: Scatterplots af GH, VT og SF-scores (*test 1 - 3 mod scores*).

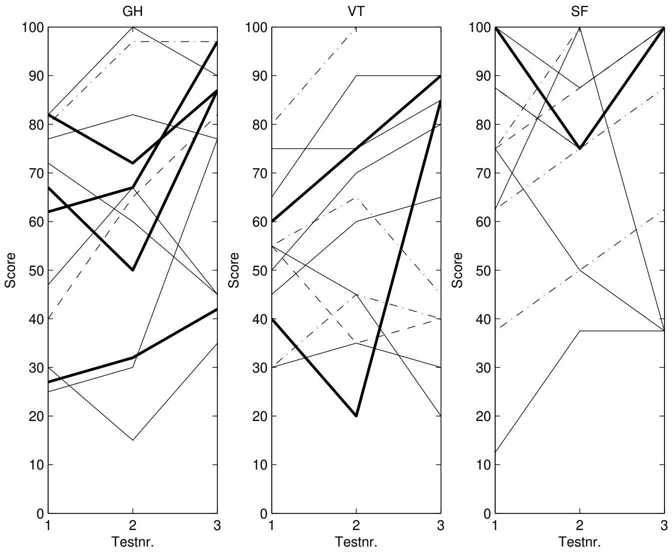
I ovenforstående scatterplots ses scores for helbredskategorien GH, VT og SF (hhv. generelt helbred, energi og social funktion), fra SF-36 spørgeskemaet. Ligesom i foregående behandling af helbredskategorien QL2, er der ligeledes for ovenstående scatterplots gældende en problemstilling om "diskret udfaldsrum". Denne problemstilling fremtræder endnu tydeligere i nedenstående histogrammer, der viser fordelingerne af scores mellem patienter for hver test. Det diskrete udfaldsrum ses af den meget opdelte fordeling af scores, især for helbredskategorien SF er dette tydeligt.



Figur 3.8: Histogrammer over fordeling GH, VT og SF-scores (*Score mod antal*).

Ses der både på scatterplottene og ovenstående histogrammer, synes der ikke at være andre tendenser end at spredningen for scorene øges for hver test (1, 2 og 3), samt at scorene generelt synes at fordele sig i den øvre halvdel af skalaerne (med undtagelse af testnr. 1 for kategorien GH).

Antagelsen om stigende spredning bekræftes af nedenstående spaghettiplot, der viser udviklingen af scores for hver af de tre helbredskategorier.



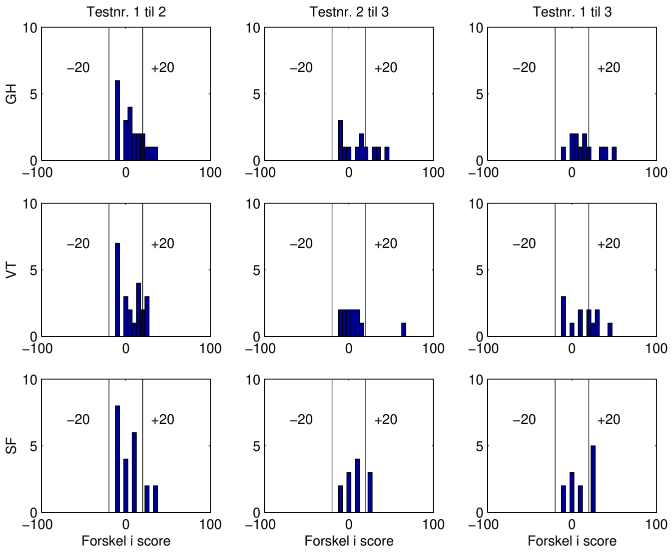
Figur 3.9: Spaghettiplots af GH, VT og SF-scores (*test 1 - 3 mod score*).

I ovenstående spaghettiplot ses en meget forvirrende udvikling af alle tre helbredskategorier, og der synes ikke umiddelbart at være nogle tendenser at se.

I nedenstående ni histogrammer ses fordelingen af ændringer i patienternes scores, mellem de respektive tests. De lodrette linier indikerer signifikante forskelle på plus/minus 20 point eller mere. Linerne er indsat i histogrammerne ud fra følgende forudsætninger [9]:

($\alpha = 0.5$, tosidede t-test, power = 80% og korrelation på 0.60 mellem gentagende målinger, samt for tests af et minimum af patienter for GH, VT og SF på hhv. 7, 8 og 9 personer).

[9] Tal taget fra tabel 5.4 side 20 i SF-36'manualen [SF-36 manual, 1997].



Figur 3.10: Histogrammer for forskelle i GH, VT og SF-scores (*forskelle i scores mod antal*).

Af ovenstående histogrammer ses følgende:

GH:

- Hos 4 patienter ses signifikante forøgelser af deres generelle helbred (GH), med 20 point eller mere, over test 1 til 2, samt 2 til 3.

VT:

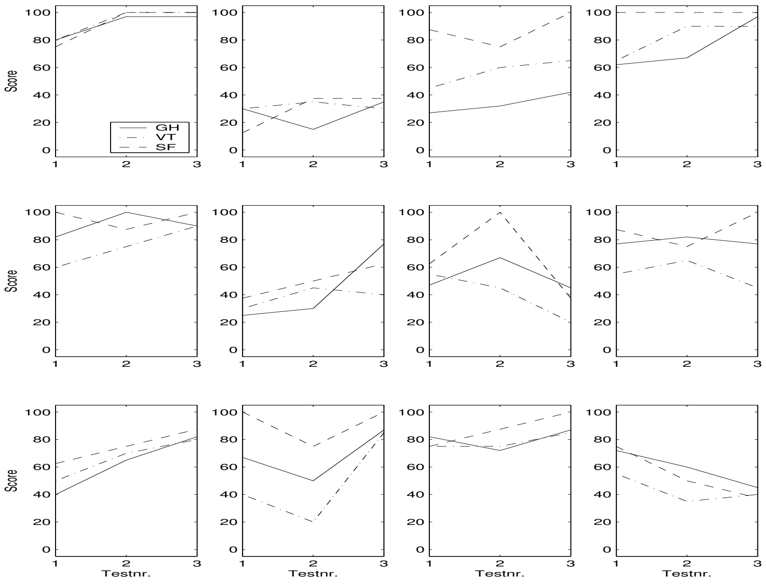
- Fra test 1 til 2 oplever cirka 5 patienter en signifikant forøgelse af VT-score. Mellem test 2 til 3 sker der næsten ikke nogle ændringer, men overordnet for test 1 til 3 har 6 patienter oplevet et signifikant forøgelse (af de 12 patienter der har besvaret samtlige skemaer).

SF:

- For test 1 til 2 oplever 4 patienter en signifikant forøgelse af SF-score. For test 2 til 3 oplever 3 patienter ligeledes en signifikant forøgelse, og mellem test 1 til 3 har i alt 5 patienter oplevet en signifikant øget social funktion (SF).

Da antallet af patienter falder i test 3, kan det ikke vises hvor mange patienter, af de oprindelige 23 stk. der reelt oplever signifikante ændringer i de respektive helbredskategorier. For alle tre helbredskategorier, synes der alligevel at være en generel forøget score, og mellem 4-6 af disse forøgelser er signifikante. Dermed er der en øget indikation, på at der antageligt sker en generel forøgelse af scores for helbredskategorierne GH, VT og SF hen over patientens tre tests.

Ovenstående indikation forstærkes yderligere i nedenstående 12 plots, hvor der for hver af de 12 patienter, i samme plot er vist udviklingen af de tre helbredskategorier.

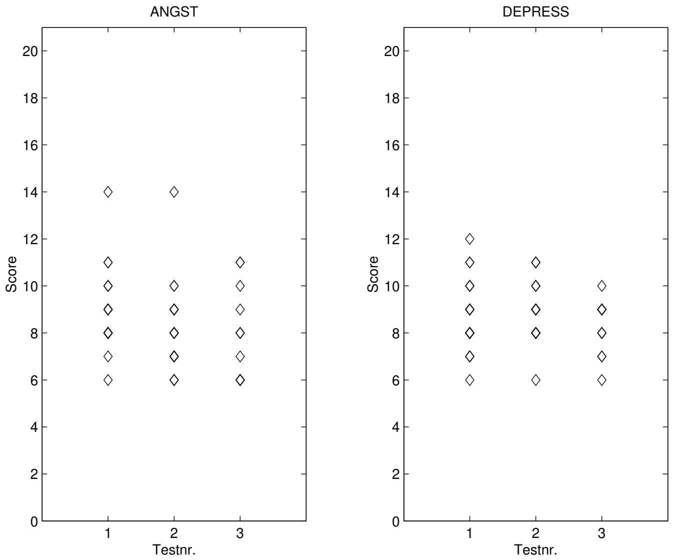


Figur 3.11: Linieplots af 12 patienters GH, VT samt SF-scores (*test 1 - 3 mod score*).

Det ses af ovenstående 12 plots, at scorene for de tre helbredskategorier, for de fleste af de 12 patienters vedkommende, har tendens til at følge hinanden parallet, hvilket kan skyldes en indbyrdes afhængighed, og derved muligvis at de forstærker hinanden. Ovenstående indikation om generel forøgelse af alle tre helbredskategorier, synes derved at være yderligere sandsynlig.

3.2.3 HAD: ANGST og DEPRESS

Efterfølgende scatterplot viser scoren for HAD-skemaets to variable Angst og Depression. Skalaen for HAD-skemaet er ikke den samme 0-100 skala, som for de to øvrige spørgeskemaer, hvilket betyder at HAD-skemaet ikke umiddelbart kan sammenlignes med disse. Ligeledes er HAD-skalene udelukkende diskret mht. at scoren for angst og depressionsvariablene, kan antage heltalsværdier mellem 0 og 21 (0 og 21 inc.). Det skal yderligere bemærkes at en mindsket score på skalaen for dette skema, modsat de øvrige to skemaer, skal ses som en fordel for patientens helbred.

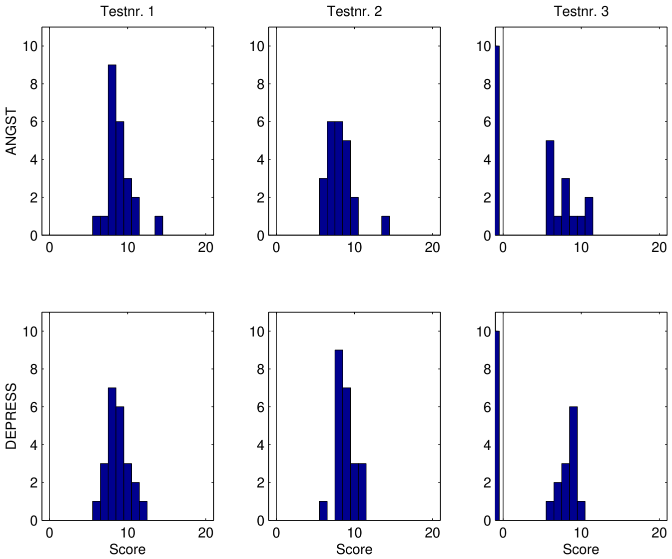


Figur 3.12: Scatterplots af ANGST og DEPRESS variable (*test 1 - 3 mod score*).

I ovenstående scatterplots ses at punkterne for de forskellige tests, ligger meget tæt samt i nogenlunde samme niveau. Dette betyder at scoren for ANGST og DEPRESS variablenes vedkommende, ikke umiddelbart forandrer sig mellem de tre tests, men ligger forholdsvist stabilt mellem en score på 6 og 11 point (med enkelte undtagelser).

En vigtig problemstilling for ovenstående iagttagelse, er ligesom for de to øvrige spørgeskemaers vedkommende, indflydelsen af den mindskede datamængde (missing data) for test nummer 3. En mindre datamængde kan bl.a. tænkes at have indflydelse på at spredningen synes en smule reduceret i test 3, og derved give anledning til forkerte tolkninger.

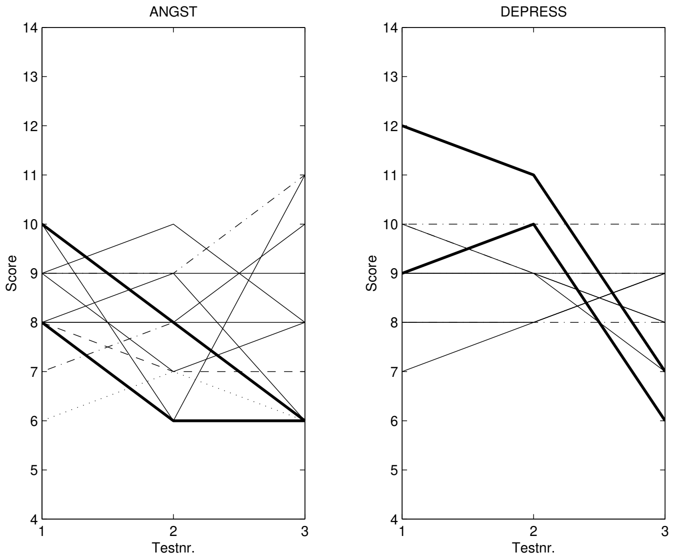
I nedenstående histogrammer ses fordelingen af scores for angst og Depression hos patienterne.



Figur 3.13: Histogrammer over fordeling af ANGST og DEPRESS-scores (*score mod antal*).

Betragtes ovenstående histogrammer, samt foregående scatterplot for samme, synes der ikke at være tegn på der sker nogle betydelige ændringer for Angst eller Depression, igennem de tre tests.

I det nedenstående spaghettiplot, ligeledes af Angst og Depressions variablene, synes der heller ikke at være nogle yderligere tendenser, idet kun meget få patienter oplever ændringer i scores. Bemærk at den lodrette akse for score er mindsket, så de enkelte ændringer i linierne træder tydeligere frem.



Figur 3.14: Spaghettiplots for ANGST og DEPRESS-scores (*test mod score*).

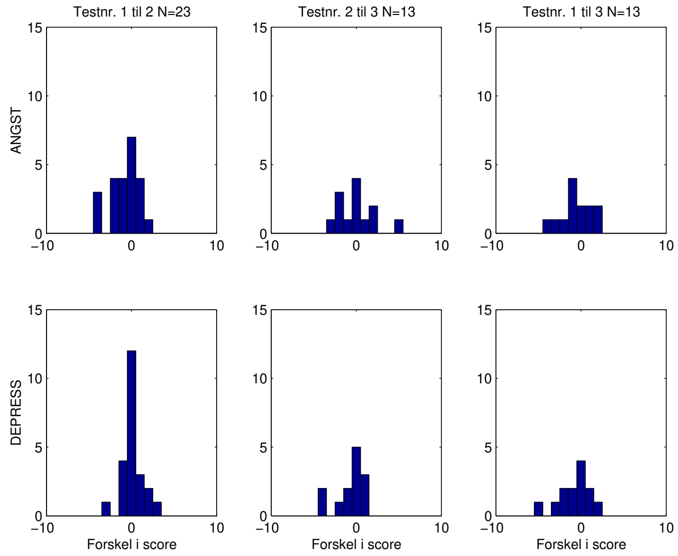
I behandlingen af HAD-skemaet, blev det (i afsnit: (2.1.2) "HAD - Hospitalsangst og depression", underafsnit: "Tildeling af score"), vist hvorledes skalaen for Angst og Depressions variablene kunne inddeles[10]:

- Ikke tilfælde (0-7 point)
- Måske tilfælde (8-10 point)
- Tilfælde (11-21point)

For Angst variablen ligger stort set alle patienter, i kategorien "måske tilfælde" eller der under (med undtagelse af 2 patienter i test nummer 3).

For Depressions variablen ligger, med undtagelse af een patient i test 1, ligeledes samtlige patienter i kategorien "måske tilfælde" eller der under.

[10][Zigmond & Snaith, 1983]



Figur 3.15: Histogrammer for forskelle i ANGST og DEPRESS-scores (*forskelle i scores mod antal*).

Ovenstående seks histogrammer viser patienternes ændringer af scores for de tre tests. Histogrammerne forstærker indikationen fra de forrige Angst og Depressionsplots, at der ikke synes at ske nævneværdige ændringer for patienterne mht. disse to variable.

3.3 Analyser og tests af spørgeskemadata

For de anførte analyser og tests i dette afsnit gælder følgende:

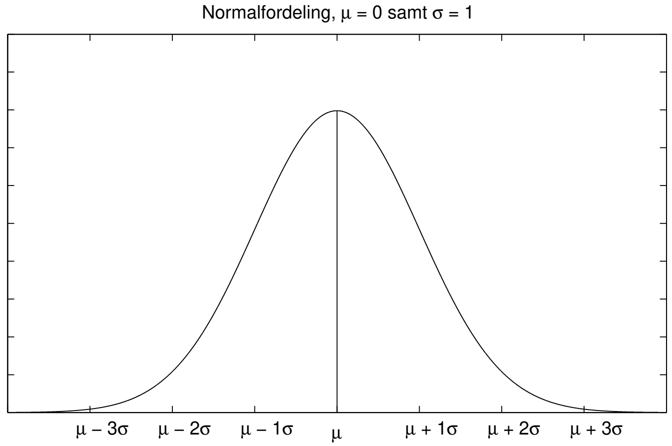
Det lave antal af patienter der har medvirket i fase 1 af projektet, samt det meget lille antal af besvarede spørgeskemaer for alle tre test tidspunkter (12 stk. for SF-36 og EORTC samt 13 stk. for HAD), betyder at resultater af analyser og tests, vil være foretaget på et meget lille datagrundlag. Derfor kan resultater af analyser og tests i denne forbindelse, udelukkende give anledning til gisninger om hvordan datamaterialet opfører sig. De anførte analyser og tests er derfor udelukkende tænkt som en guide til hvilke tests der synes rimelige og relevante at foretage på et senere og større datagrundlag, i form af efterfølgende faser af projekt “Krop & Kræft”.

På baggrund af ovenstående, er der i det efterfølgende anført og beskrevet en række analyser og tests, der vil blive anvendt på det af fase 1 givet datamateriale fra spørgeskemaerne hhv. SF-36, HAD og EORTC QLQ-C30. Det skal understreges at beskrivelser af analyser og tests ikke er udførlige statistiske indføringer i den bagvedliggende teori, men skal forstås som en overordnet forklaring, til hjælp for læsere der ikke på forhånd er bekendt med den anvendte statistik.

3.3.1 Populationsantagelser

Fordelingen af scorene for helbredskategorierne, har stor betydning for hvilke metoder der anvendes under de forskellige tests og analyser. Derfor er en fordelingsantagelse nødvendig og beskrevet i det følgende.

Det antages at scorene følger en normalfordeling, da det synes rimeligt at dataene vil tage en såkaldt klokkeformet fordeling, herved vælges det at overse følgende to betragtninger:



Figur 3.16: Normalfordeling.

- Værdierne for de enkelte helbredsvariable er givet ved et diskret udfaldsrum[11].
- Værdierne af helbredskategorierne er afgrænset af 0-100 skalaen for SF-36 og EORTC spørgeskemaet, og 0-21 skala for HAD skemaets vedkommende.

Ovenstående betragtninger er ikke i overensstemmelse med definitionen af en normalfordeling, men vil i fordelingsmæssig sammenhæng kunne ses bort fra da følgende antagelser gælder:

- Eftersom det diskrete udfaldsrum mindskes vil fordelingen af disse tilnærmer sig en kontinuert normalfordeling.
- De skala afgrænsede værdier, kan forklares af Floor og Ceiling effekter[12], og derved ses skalaerne som et afgrænset billede af den faktiske fordeling af scorene.

3.3.2 Parret T-test

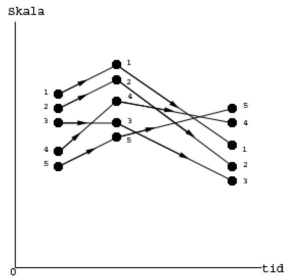
Et t-test går i hovedtræk ud på at teste om en middelværdi er lig med eller forskellig fra nul. Denne måde at lave tests på er fx fordelagtig i tilfælde

[11] Se afsnit: (3.2.1):"Diskret udfaldsrum", hhv. EORTC QLQ-C30: QL2.
[12] Beskrevet i afsnit: (2.1.1): SF-36 - Spørgeskema om helbredsstatus.

hvor man ønsker at bestemme, om der er forskel på middelværdier i to uafhængige normalfordelte undersøgelser.

I tilfælde hvor der ønskes undersøgt om der er sket en ændring i middelværdier for to undersøgelser foretaget over tid, er der imidlertid ikke længere tale om to uafhængige undersøgelser, da der her er afhængighed mellem målinger foretaget på den samme patient. I sådanne tilfælde anvendes i stedet en såkaldt parret t-test hvor man benytter sig af "ændring i målinger pr. patient" (i det følgende defineret vha. $\Delta_{scores(i)}$, hvor (i) er den enkelte patients nummer).

I nedenforstående illustration er vist resultaterne af tre fiktive undersøgelser, foretaget over en tidsperiode. Undersøgelserne er foretaget på fem patienter og linierne mellem observationerne illustrerer overgangene mellem de tre undersøgelser.



Figur 3.17: Parret t-test illustration.

Det ses af ovenstående illustration, at ikke alle linier er parallelle og der synes derfor ikke umiddelbart at vise sig entydige tendenser for målingerne igennem de tre undersøgelser. I ovenstående illustration kan det derfor være nyttigt at anvende en parret t-test.

For hver helbredskategori laves en parret t-test via følgende formel[13]:

$$t_0 = \frac{\overline{d}}{S_d/\sqrt{n}} \tag{3.1}$$

[13] [Douglas C. Montgomery, 2001]

$\overline{d}$: Middelværdien af $\Delta_{scores(i)}$ for $i = 1, 2, ..., n$.
S_p: Estimat for spredningen af $\Delta_{scores(i)}$ for $i = 1, 2, ..., n$.
n: Antallet af patienter.
t_0: Teststørrelse.

Teststørrelsen t_0 sammenlignes med en t-test størrelse fra en t-fordeling ($t_{\alpha/2,n-1}$), med (n-1) frihedsgrader og et givent signifikansniveau ($\alpha/2$).

Er $|t_0| > t_{\alpha/2,n-1}$, antages middelværdien $\overline{d}$ at være forskellig fra 0, hvilket oftest betyder at man på et givent signifikans niveau $(100 - \alpha)$ kan antage, at der er sket en ændring af middelværdier i forhold til de to testede undersøgelser.

Afhængighed mellem $\Delta_{scores(i)}$: Der antages at være afhængighed mellem patientens scores, for hver helbredskategori. Denne afhængighed kan beskrives i form af en tidsmæssig korrelation mellem patientens svar på ens spørgsmål (Longitudional korrelation)[14]. Korrelation betyder at der ligeledes kan antages at være afhængighed mellem ændringer i patienternes scores fx givet ved:

$$\Delta_{scores(i)} : Test\ 2\ til\ 3, \text{ er afhængig af } \Delta_{scores(i)} : Test\ 1\ til\ 2.$$

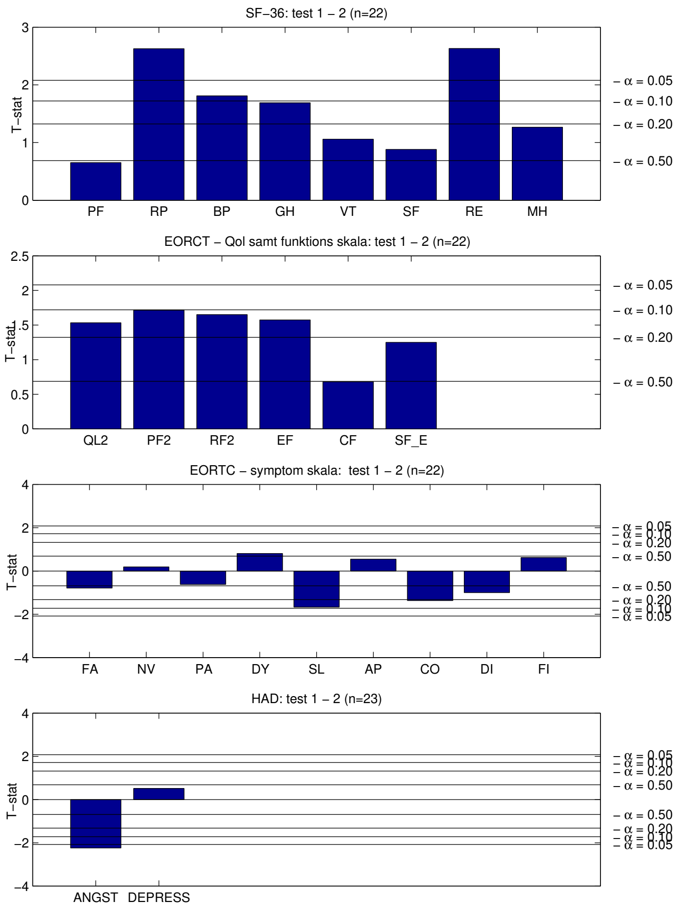
Ovenstående kan udtrykkes i form af afhængighed mellem $\Delta_{scores(i)}$. Afhængigheden mellem $\Delta_{scores(i)}$ betyder at en parret t-test mellem test 2 og 3 er afhængig med en parret t-test mellem test 1 og 2, og at disse derfor ikke uden videre kan sammenlignes.

Da afhængighed mellem tests ikke kan ikke gå forud i tid, fx ved at $\Delta_{scores(i)}$: Test 1 til 2, er afhængig af $\Delta_{scores(i)}$: Test 2 til 3, har afhængigheden mellem $\Delta_{scores(i)}$, derfor ikke indflydelse på en parret t-test mellem test 1 og 2. For parret t-test mellem de øvrige tests (2 og 3, samt 1 og 3), er i det følgende ikke taget højde for afhængigheder mellem $\Delta_{scores(i)}$, og de nævnte tendenser som tolkes ud fra disse testresultater, skal derfor vurderes med forbehold for ovenfornævnte.

I nedenforstående grafer er vist t-stat estimater (t_0) for samtlige helbredsvariable. De vandrette linier viser $t_{\alpha/2,n-1}$-værdierne for forskellige α'er

[14] Er forklaret nærmere i efterfølgende afsnit: (3.3.3) Korrelation mellem variable.

(hhv. 0.05, 0.10, 0.20 og 0.50). Negative t-stat værdier, betyder at den estimerede middelværdi $\overline{d}$ er negativ, og derved at den gennemsnitlige difference/tendens er negativ/faldende (modsat en øget tendens for positiv $\overline{d}$ og uændret for $\overline{d} = 0$).



Figur 3.18: Parret t-test af testnr. 1 mod testnr. 2 (n=22 samt 23).

I ovenstående fire grafer ses at for teststørrelserne for helbredskategorierne mellem test 1 og 2, er især helbredskategorierne RP og RE i SF-36 skemaet signifikante for meget små værdier af α. Dernæst kommer helbredskategorierne BP og GH, mens de øvrige ikke synes at være specielt signifikante.

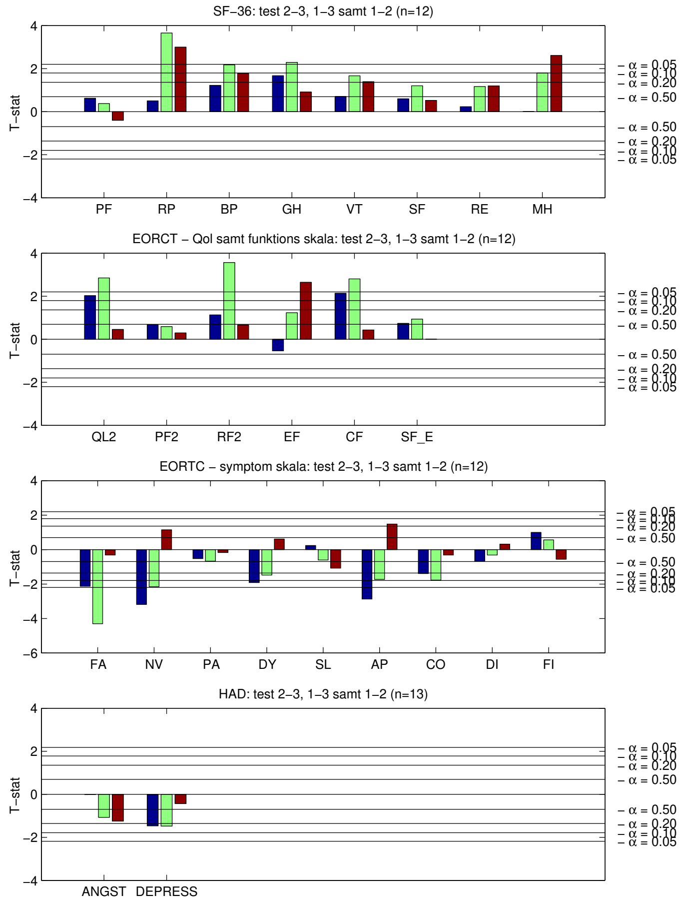
For EORTC-skemaet sker der ikke umiddelbart signifikante ændringer for symptom skalaen, da overordnet kun SL-symptomer synes at mindskes. For funktionsskalaen er der ligeledes tale om små differencer mellem test 1 til 2 helbredskategorierne PF2, RF2 har de største ændringer ($\alpha = 10\%$), og der efter følger EF samt QL2.

I HAD-skemaet synes der umiddelbart at være sket signifikante ændring ($\alpha = 5\%$) for ANGST-variablens vedkommende. Dette skyldes imidlertid at en enkelt patient oplever en relativ stor ændring i forhold til de øvrige patienter, der oplever meget små ændringer[15]. Den "store ændring" medfører en signifikant betydning for den samlede parret t-test, hvilket påpeger en risiko ved at anvende resultater af parret t-test, uden at sammenligne med grafiske fremstillinger af det testede datamateriale. Ændringen i ANGST-variablen vurderes på baggrund af ovenstående, at være uden væsentlig betydning. For DERESS-variablen synes der ikke at være tegn på ændringer i scores.

I nedenstående fire grafer er for hver helbredskategori vist tre parret t-test for de 12 patienter (13 for HAD), der har udfyldt spørgeskemaer ved alle tre tests. Det skal understreges at der ikke er taget højde for afhængighed mellem $\Delta_{scores(i)}$ for de tre tests[16].

[15] Se figur: (3.14).

[16] Se "Afhængighed mellem $\Delta_{scores(i)}$" (dette afsnit).



Figur 3.19: Parret t-test af testnr. 2 mod 3, 1 mod 3 samt 1 mod 2 (n=12 samt 13).

Sammenlignes SF-36 helbredskategorierne, for denne test med forrige test af 22 patienter, ses at umiddelbart følger ændringerne i SF-36 helbredskategorierne ikke samme mønstre. Af særlige tilfælde er helbredskategorien RE, hvor der synes at være væsentlige lavere signifikante ændringer $(1-\alpha)$. Dette kan umiddelbart forklares ved at de patienter, der oplever en ændring mellem test 1 til 2, ikke har lavet test 3 og derved ikke medregnes i hele denne test. Dette bekræfter nødvendigheden af at se på en større datamængde, for at udpege mere konsistente tendenser i materialet.

For EORTC-skemaet ses, for de enkelte helbredskategorier, en større disharmoni mellem test af 22 og 12 patienter end det var tilfældet for SF-36 skemaet. Især helbredskategorierne QL2, FA, NV og AP har overraskende høje signifikante ændringer mellem test 2 til 3, hvilket ligeledes gør sig gældende for helbredskategorierne QL2, RF2, CF, FA samt NV, ved test 1 til 3. Disse ændringer skal her ses i forhold til parret t-test med 22 patienter, og betyder at for disse helbredskategoriers vedkommende, er der umiddelbart tendenser til at der sker signifikante ændringer mellem test 2 til 3, som eventuelt kan eftervises ved at teste på et større datamateriale.

Overordnet for alle helbredskategorierne (for SF-36 og EORTC), synes der at være en tendens til at teststørrelserne for de enkelte helbredskategorier mellem test 1 til 2 (22 patienter), er generelt mere signifikante end teststørrelserne mellem test 2 til 3 (12 patienter). Ses på teststørrelserne mellem test 1 til 3, ud fra den synsvinkel at denne teststørrelse repræsentere en overordnet vurdering af alle tre tests, synes der samlet set at være tegn på signifikante teststørrelser ($\alpha \leq 10\%$), for cirka halvdelen af helbredskategorierne, hvilket er til patienternes fordel. Umiddelbart kan ovenstående tolkes som at patienterne indenfor de pågældende helbredskategorier har fået det signifikant[17] bedre.

For symptomskalaen ses yderligere en tendens til at der sker ændringer mellem test 2 til 3 mens der ikke i første omgang sker ændringer mellem test 1 til 2. Dette er ikke umiddelbart til at forklare, men kan skyldes at observationer fra de resterende 10 patienter (af de 22 patienter) mangler for at mindske differensgennemsnittet.

HAD-skemaet giver ikke umiddelbart grund til positive eller negative tendenser.

3.3.3 Korrelation mellem variable

Der behandles to former for korrelations begreber hhv.:

- **Longitudional korrelation**
- **Sektionel korrelation**

[17] 10% signifikansniveau.

Longitudional korrelation: Når der foretages ens målinger på den samme patient over tid, optræder der ofte en afhængighed mellem disse målinger. Denne afhængighed skyldes at det er den samme patient der måles på, og jo kortere tid der er mellem to målinger jo mere er målingerne afhængige af hinanden.

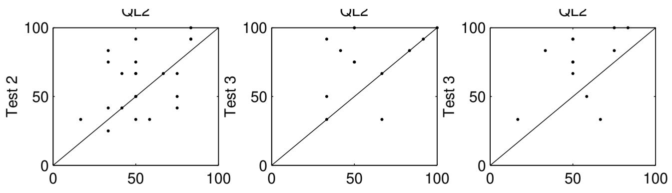
I et tilfælde hvor der fx måles hvilepuls på den samme person over de 5 hverdage i ugen, betyder afhængigheden mellem målingerne at hvilepulsen forventes at ligge i omtrent samme niveau (lineær afhængighed), jo tættere dagene ligger op ad hinanden. Sker der i uges løb ændringer i patientens hvilepuls, vil det være nødvendigt at trække afhængighedsdelen (korrelationen) ud af den respektive måling, for derefter fx at vurdere om ændringen er statistisk signifikant.

I forbindelse med spørgeskemaundersøgelser, hvor patienten for hver undersøgelse udfylder et specifikt spørgeskema, vil afhængighed mellem patientes svar yderligere være begrundet med at patienten vil kunne erindre hvad han/hun svarede i forrige undersøgelse (hvis undersøgelserne foretages indenfor en rimelig tid, fx hvert halve år).

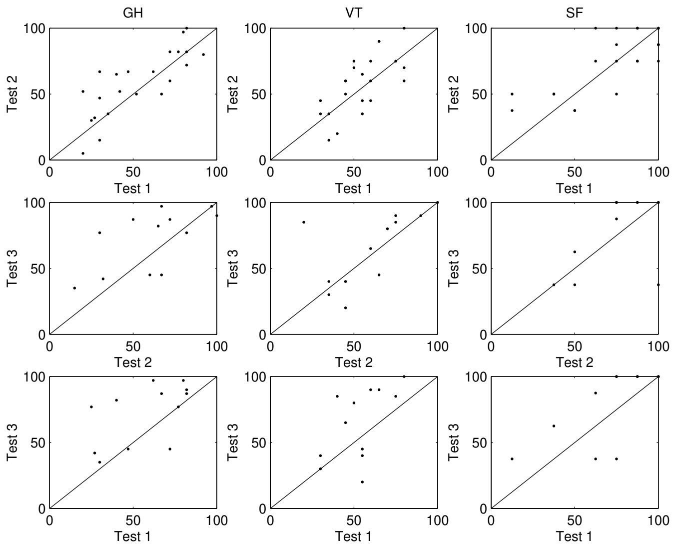
Sektionel korrelation: Denne afhængighed opstår hvis der optræder korrelation mellem patientmålingerne for det samme tidspunkt. Et eksempel på denne korrelation er hvis en bestemt gruppe/ undergruppe af patienter, på en eller anden måde skiller sig ud fra de øvrige patienter i de målinger der foretages. Populært set kan sektionel korrelation ses som afhængighed mellem ens målinger på tværs af tiden, men longitudional korrelation ses på langs af tiden.

Forekommer der en afhængighed, som nævnt i ovenstående eksempel, vil dette betyde at man med fordel ville kunne benytte en yderligere opdeling af patienterne (fx i alder eller køn) til at forklare en del af den variation der optræder mellem målingerne.

I understående grafer ses nogle overordnede helbredskategorier plottet mod sig selv til forskellig tidspunkter (tests). Dette skal illustrere den longitudionale korrelation.



Figur 3.20: Korrelationsplots af QL2.



Figur 3.21: Korrelationsplots af GH, VT og SF.

Det antages at der er uafhængighed mellem patienterne i de enkelte tests, derved er den sektionelle korrelation ikke relevant, så længe en opdeling af patienterne i fx hold eller civilstatus[18], ikke viser sig at være signifikant mht. de i proc Mixed testede modeller[19].

[18]Se afsnit: (3.1.5) Opdeling af patienter.
[19]Se evt. efterfølgende afsnit: (3.3.4) SAS proc Mixed.

Korrelation mellem helbredsvariable

Samtlige af helbredsvariable fra SF-36 og EORTC QLQ-C30 spørgeskemaerne er testet mod hinanden, for at vurdere hvorvidt disse korrelerer[20]. Resultaterne af de mest korrelerede (p-value $< 0.10\%$) variable ses i understående tabel[21]:

Variable		Korrelation	p-value (%)
EORTC	SF-36	Test 1	
PF2	PF	0.77	<0.01
SF_E	SF	0.75	<0.01
CF	SF	0.70	0.02
EF	MH	0.70	0.02
RF2	SF	0.69	0.03
EF	SF	0.68	0.04
FA	SF	-0.67	0.04
		Test 2	
SF_E	SF	0.85	<0.01
EF	MH	0.78	<0.01
QL2	GH	0.78	<0.01
SF_E	GH	0.77	<0.01
PA	BP	-0.77	<0.01
QL2	VT	0.74	<0.01
PF2	PF	0.74	<0.01
QL2	SF	0.71	0.02
FA	GH	-0.68	0.05
SF_E	VT	0.66	0.08

Tabel 3.7: Mest signifikante korrelationer mellem helbredskategorier fra test 1 og 2.

Note: Det er valgt ikke at anvende test 3, da antallet af observationer er meget lille i forhold til test 1 og 2 (12 stk. mod 22 stk.). Dette skyldes at et eventuelt estimat for korrelationen mellem helbredsvariable i test 3, med større sandsynlighed kan være præget af tilfældige faktorer.

I ovenstående tabel ses for de positive korrelationer:

[20]Testen er foretaget vha. Multivariate analyse i SAS INSIGT.
[21]Negativ korrelation betyder at helbredsvariablene afhænger omvendt proportionalt med hinanden.

- Af de helbredskategorier der er størst korreleret i både test 1 og 2, findes helbredskategorier der pr. definition ligner hinanden fx PF2 og PF samt SF_E og SF.
- For SF_E samt SF helbredskategorierne, synes der at være en del af de øvrige helbredskategorier disse er korrelere med, fx RF, EF samt VT. Dette skyldes antageligt at den Sociale funktion har relationer til mange forskellige fysiske som psykiske faktorer, fx er det sandsynligt at patientens sociale velvære afhænger af dennes mentale sundhed.
- For test 2 ses at den overordnet helbredskategori QL2 korrelere med GH, VT samt SF fra SF-36 skemaet. Dette understøtter udvælgelsen af disse helbredskategorier i afsnit: (3.2) “Grafisk præsentation af spørgeskemadata”.

For de negative korrelationer ses følgende:

- De negative korrelationer ses udelukkende med symptom kategorierne fra EORTC-skemaet. Dette skyldes hovedsageligt at en forbedret symptomscore i modsætning til de øvrige kategorier er negativ.
- FA-symptomer (træthed, udmattelse) synes at påvirke den Sociale Funktion.
- Spørgsmål om fysisk smerte synes at korrelere indenfor begge spørgeskemaer (PA og BP).

Ovenstående korrelationer er et nyttigt værktøj til at sammenligne to spørgeskemaer med hinanden, fx med hvor vidt forskellige former at udforme spørgsmål på, kan antages at “måle” det samme. Værktøjet kan ligeledes benyttes til at teste om to helbredskategorier, der pr. definition skulle måle det samme hos patienten, tilsyneladende viser sig ikke at korrelere med hinanden og derfor bør give grund til genovervejelse.

Endnu en anvendelse af ovenstående korrelationer mellem helbredskategorier, kunne være at finde frem til patienter der fx kan have misforstået enkelte spørgsmål eller generalt viser sig at være “særlige tilfælde”. Sådanne patienter kan det være hensigtsmæssigt at behandle særskilt/specielt, og vil kunne udpeges hvis fx deres resultater/scores korrelerer væsentligt fra de øvrige adspurgte patienters scores.

3.3.4 SAS proc Mixed

I forbindelse med en parret t-test, fjernes den ovenfornævnte longitudionale korrelation/afhængighed ved at trække to målinger fra hinanden, ud fra den

antagelse at ændringerne i målingerne ses som uafhængige størrelser (der ses bort fra afhængighed mellem $\Delta_{scores(i)}$). Hvis der er flere målinger er en mulighed ligeledes at lave parret t-test for samtlige kombinationer at test tidspunkter. Dette er imidlertid en uhensigtsmæssig metode at anvende, da mulige kombinationer stiger betydeligt jo flere test tidspunkter der er. Selv om antallet af test i dette projekt er relativt få (tre), er det valgt at benytte proc Mixed, idet der dels vil kunne sammenlignes med resultater af de Parret t-test der er foretaget, og desuden vil der kunne indrages flere faktorer i den enkelte tests, som er beskrevet i det følgende.

Proc Mixed[22] går i hovedtræk ud på at tilnærme en model til sine data, og efterprøve hvorvidt denne model er hensigtsmæssig. I modellen er det forholdsvist nemt at inddrage flere forskellige faktorer, og derefter teste disse på de egentlige data. I det efterfølgende er det valgt at benytte to forskellige modeller, disse er valgt ud fra hvad der umiddelbart syntes logisk i det pågældende tilfælde (primært vurderet ud fra de viste grafer[23]), samt forudgående antagelser for hvordan dataene kunne tænkes at opfører sig[24].

Testens primærer formål er at teste i hvilket omfang det kan antages at projekt Krop & Kræft, har haft en virkning på patienternes scores i de respektive spørgeskemaundersøgelser. Overordnet set gøres dette ved at adskille effekter af korrelationer i dataene (tilfældige komponenter), og dernæst give et kvalificeret bud på de resterende effekters betydning (systematiske komponenter), hvis testresultaterne viser at det kan antages at der er sådanne effekter.

> ”Proc Mixed er direkte rettet mod analyse af "blandede normalfordelingsmodeller", dvs lineære modeller, der både indeholder systematiske og tilfældige komponenter.”
>
> [P. Thyregod, 1998][25]

Den generelle model der anvendes i procMixed ser ud som følgende:

$$y_i = X_i\beta + Z_i\gamma_i + \epsilon_i \tag{3.2}$$

(Note: hhv. et systematisk led, et tilfældigt led samt fejl).

[22] www.sas.com
[23] 3.2 Grafisk præsentation af spørgeskemadata.
[24] Se afsnit: (3.3.3) Korrelation mellem helbredsvariable.
[25] Side 521, bind 3C.

$$i = 1, 2, 3, ..., n \quad for\ n = antal\ patienter \tag{3.3}$$

y_i: Data vektor for den afhængige variabel (se understående).

$$y_i = \begin{pmatrix} y_{test_1} \\ y_{test_2} \\ y_{test_3} \end{pmatrix} \tag{3.4}$$

X_i: Kendt design matrice.
β_i: Ukendt systematisk effekt (skalar eller vektor).
Z_i: Kendt design matrice.
γ_i: Ukendt tilfældig effekt (skalar eller vektor).
ϵ_i: Residual fejl.

γ og ϵ er uafhængige normalfordelte størrelser med middelværdi (0) samt varians og covariansmatricer givet ved hhv. G og R_i[Wolfinger, 1997].

Den forventede værdi af den afhængige variabel, er givet ved følgende:

$$E(y_i) = X_i\beta \tag{3.5}$$

Variansen på den afhængige variabel, er givet ved følgende:

$$V(y_i) = Z_iGZ_i' + R_i = V(\gamma_i) + V(\epsilon_i) \tag{3.6}$$

Ovenstående formel, viser at variansen for den afhængige variabel består af et bidrag fra variansen for den tilfældige effekt (γ_i), samt et bidrag fra variansen fra residual fejlen (ϵ_i). Dette betyder at den longitudionale korrelationen, er afhængige af hvordan ovenstående to variable defineres, ud fra deres respektive varians samt covarians parametre.

Missing data: I forbindelse med missing data, er der med én enkelt undtagelse tale om at udeblivende data, sker efter den sidste måling på den pågældende patient. Dette betyder at understående standard indstilling i proc MIXED gør sig gældende:

> "By default, a dropout mechanism is assumed to be responsible of missing data by MIXED: All missing data are considered to occur after the last observed measurement." [Thiébaut, 2002][26]

Fordi Proc MIXED tager højde for ovenstående situation, vil det ikke være nødvendigt at foretage yderligere hensyntagen til situationer med missing data i de efterfølgende modeller.

Maksimum likelihood estimatorer: Det mål der anvendes til at bestemme hvor "godt" en given model passer til dataene, er i denne forbindelse REML-estimatet. Dette estimat går i hovedtræk ud på at beregne afstanden fra de analyserede data, og hen til den estimerede model (dvs. et residualestimatet). Jo mindre dette afstandsestimat er, jo bedre antages modellen at beskrive de analyserede data.

For at kunne sammenligne de to anvendte modeller, er det nødvendigt at korrigere REML-estimatet for at tage højde for eksempel vist det forskellige antal af frihedsgrader i de to modeller. Til dette er valgt følgende to informations kriterier:

AIC: Akaike Information Criterion (Akaike, 1974).
BIC: Sawa's Bayesian Information Criterion[27](Schwarz, 1978).

AIC-kriteriet korrigere med dimensionen af den anvendte model, mens BIC-kriteriet desuden korrigerer med antallet af frihedsgrader i modellen. For begge kriterier gælder at et mindre estimat foretrækkes frem for et større (smaller is better).

I det følgende er beskrev de to anvendte modeller.

Model 1: Compound Symmetry

For denne model gælder at variansen for hver test (1, 2 og 3) er givet ved en varians indenfor test (σ_0^2) samt en covarians mellem tests (σ_1^2). Dette

[26] Side 252, 1 spalte nederst.

betyder at der er samme covarians mellem samtlige tests, samt at der er samme varians indenfor tests.

Designmatricen for den systematiske effekt ser ud som følgende, for hver patient:

$$X_i = I_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \tag{3.7}$$

Den systematiske effekt er givet ved følgende:

$$\beta = skalar\ (\beta),\ eller\ vektor \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \tag{3.8}$$

Designmatricen for den tilfældige effekt ser ud som følgende, for hver patient:

$$Z_i = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{3.9}$$

Både γ og ϵ er normalfordelte, og uafhængige størrelser, og er givet ved følgende:

$$\gamma_i = N(0, G) \tag{3.10}$$

$$\epsilon_i \in N(0, R_i) \tag{3.11}$$

Opbygningen af G og R_i ser ud som følgende for hver patient:

G: Udtrykker det <u>ens</u> variantions bidrag mellem tests.

$$G = \sigma_1^2 \tag{3.12}$$

R_i: Udtrykker variationsbidraget indenfor tests, hvilket ligeledes er ens for hver test.

$$R_i = \sigma_0^2 I_3 = \begin{bmatrix} \sigma_0^2 & & \\ & \sigma_0^2 & \\ & & \sigma_0^2 \end{bmatrix},\ (Blankt\ felt = 0) \tag{3.13}$$

Den forventede værdi af den afhængige variabel er givet ved formel 3.5, og kan i det aktuelle tilfælde skrives som følgende:

$$\overline{y_i} = E[y_i] = X_i\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \tag{3.14}$$

Note: De tre β'er er ikke nødvendigvis forskellige.

Variansen for y_i er givet ved formel 3.6, og kan i denne model yderligere udtrykkes via følgende:

$$\begin{aligned} \sigma_{y_i}^2 &= V(y_i) \\ &= Z_i G Z_i' + R_i \\ &= \begin{bmatrix} \sigma_1^2 + \sigma_0^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma_0^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma_0^2 \end{bmatrix} \end{aligned} \tag{3.15}$$

idet det gælder:

$$Z_i G Z_i' = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \sigma_1^2 (1\ 1\ 1) = \begin{bmatrix} \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \end{bmatrix} \tag{3.16}$$

Ovenstående betyder at for <u>hver</u> patient, ser model 1 ud som følgende:

$X_i =$ I_3 matrice.
$\beta_i =$ systematisk effekt (vektor eller skalar) af projekt Krop & Kræft.
$Z_i =$ rækkevektor af 1'taller.
$\gamma_i =$ tilfældig effekt med følgende fordeling N(0,$\sigma_{y_i}^2$).

Modellens opbygning i ord:

Hver patient har samme covarians mellem alle tre tests på (σ_1^2). Variationen for den enkelte test består af et bidrag fra covariansen mellem tests på (σ_1^2) plus et bidrag fra variansen indenfor tests (σ_0^2).

Model 2: Unstructured

Parametre og matricer ser ud som i forrige model, med følgende undtagelser.

G og R_i er i denne model bygget op som følgende:

G er et ens variations bidrag for alle tests samt mellem tests.

$$G = \sigma_0^2 \tag{3.17}$$

R_i er en matrice med forskellig variationsbidrag mellem samt indenfor tests.

$$R_i = \begin{bmatrix} \sigma_{1'}^2 & \sigma_{2'}^2 & \sigma_{4'}^2 \\ \sigma_{2'}^2 & \sigma_{3'}^2 & \sigma_{5'}^2 \\ \sigma_{4'}^2 & \sigma_{5'}^2 & \sigma_{6'}^2 \end{bmatrix} \tag{3.18}$$

Den forventede værdi af den afhængige variabel y_i er givet ved formel 3.5.

Variansen for y_i kan skrives som følgende:

$$\begin{aligned} \sigma_{y_i}^2 &= Z_i G Z_i' + R_i \\ &= \begin{bmatrix} \sigma_0^2 & \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 & \sigma_0^2 \end{bmatrix} + \begin{bmatrix} \sigma_{1'}^2 & \sigma_{2'}^2 & \sigma_{4'}^2 \\ \sigma_{2'}^2 & \sigma_{3'}^2 & \sigma_{5'}^2 \\ \sigma_{4'}^2 & \sigma_{5'}^2 & \sigma_{6'}^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \sigma_4^2 \\ \sigma_2^2 & \sigma_3^2 & \sigma_5^2 \\ \sigma_4^2 & \sigma_5^2 & \sigma_6^2 \end{bmatrix} \end{aligned} \tag{3.19}$$

Ovenstående betyder at for hver patient, ser proc Mixed modellen 3.2 ud som følgende:

$X_i =$ I_3 matrice.
$\beta_i =$ systematisk effekt af projekt Krop & Kræft.
$Z_i =$ rækkevektor af 1'taller.
$\gamma_i =$ tilfældig effekt med følgende fordeling N(0,$\sigma_{y_i}^2$).

Modellens opbygning:

Covariansen mellem test er givet ved følgende:

Cov(test 1-2): σ_2^2
Cov(test 2-3): σ_5^2
Cov(test 1-3): σ_4^2

Variansen for de enkelte test er givet ved følgende:

V(test 1): σ_1^2
V(test 2): σ_3^2
V(test 3): σ_6^2

3.3.5 Resultater af proc Mixed

Bearbejdning af resultater består af en række proc Mixed analyser samt en række udvælgelses trin (selektioner). Selektionerne har til formål at udvælge de variable der ud fra en proc Mixed analyse bedst muligt forklarer variationen i datamaterialet, hvilket vil sige forklarer variationen i patienternes score for de enkelte helbredskategorier.

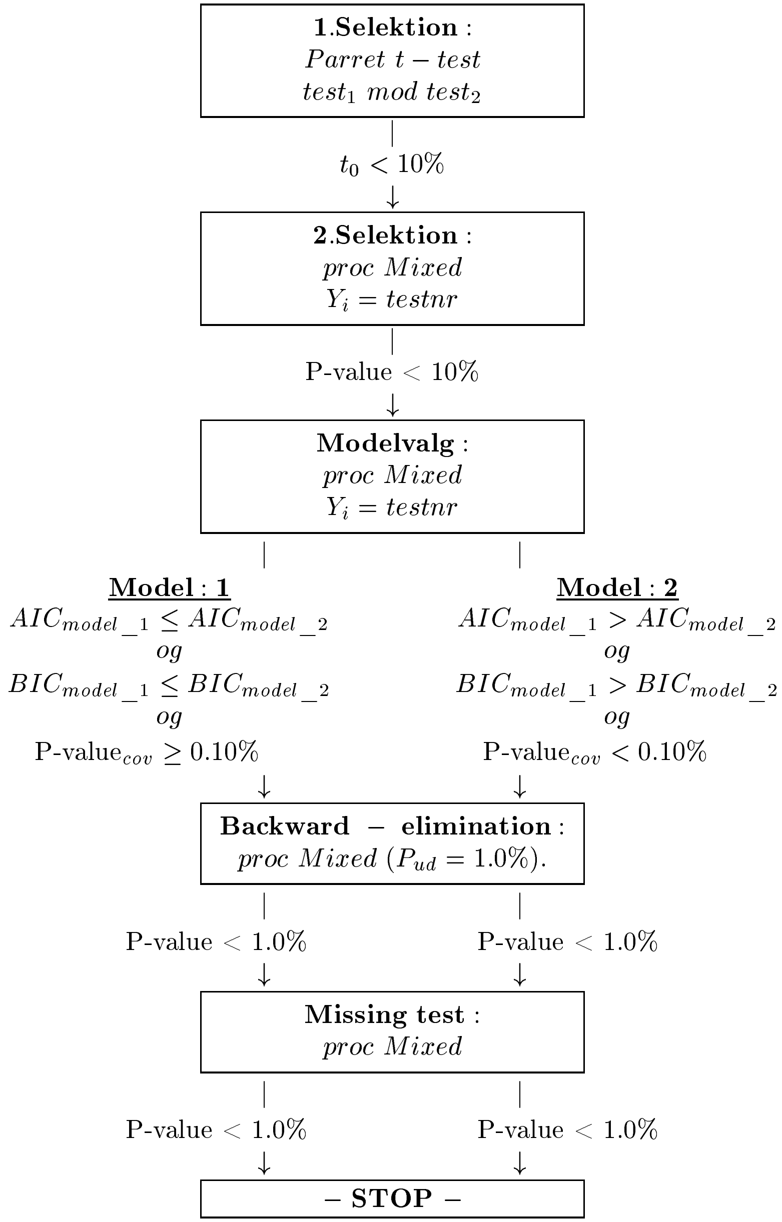
I nedenstående diagram (figur: 3.22), er vist hvilke trin der er valgt for at selektere i de testede helbredskategorier. Diagrammet består bl.a. af forskellige "grænseværdier" i form af p-værdier der bruges for hver selektion. Disse grænseværdier er fastsat ud fra den prioritering, at de skal øges for hvert selektionstrin. Prioritering er anvendt for at udvælge de variable der har den største mulige signifikans, i forhold til at forklare variationen i det testede datamateriale. De forklarende variable selekteres ud fra følgende hypoteser mht. P-værdierne[28]

- Er p-value $> p\!-\!value_{graensevrdi}$: Er der ikke nogen signifikant forskel mellem den pågældende gruppe/opdeling (fx cicilstatus og Køn).
- Er p-value $\leq p - value_{graensevrdi}$: Er der signifikant forskel mellem grupper i den pågældende opdeling, som forklares af forskellen mellem grupper, udtrykt igennem variationen mellem grupper.

P-værdierne kan ændres hvis man ønsker at øge antallet af de forklarende variable, og dermed åbne op for en bredere forklaring af variationen. Men da datamaterialet i fase 1 er forholdsvist "lille", vil en analyse let kunne

[28]Se: [K. Conradsen, 1995] side 393, 1B.

påvirkes af fx enkeltobservationer/målinger, som afviger selv begrænset fra de øvrige målinger selvom observationen ikke kan betegnes som ekstrem. Denne følsomhed søges så vidt muligt af undgået ved at anvende forholdsvis skrappere grænseværdier.



Figur 3.22: Selektionsdiagram.

I det efterfølgende er selektionsdiagrammet gennemgået trin for trin, med henblik på at give læseren et indblik i diagrammets anvendelse.

1. Selektion:

Det er valgt at analysere på de helbredskategorier der i den parret t-test (se afsnit: (3.3.2)), viste teststørrelser med signifikansniveau på minimum 10%[29].

- Valget af den parret t-test mellem test 1 og 2 udelukker forbehold for afhængighed mellem $\Delta_{scores(i)}$, for ændringer i scores pr. helbredskategori[30].

Signifikansniveaugrænsen på minimum 20%, betyder imidlertid at samtlige af EORTC- skemaets helbredskategorier udelukkes. For at kompensere for dette er medtaget de fire første helbredskategorier fra dette spørgeskema. HAD-skemaet er fravalgt da det i afsnittet (Parret t-test) blev vurderet at den signifikante teststørrelse for angst-variablen, skyldtes en enkelt patients ændring i score.

Udvalgte helbredskategorier	
SF-36:	RP, BP, RE
EORTC QLQ-C30:	QL2, PF2, RF2, EF
HAD :	-

Tabel 3.8: Helbredskategorier efter 1. selektion.

2. Selektion:

I denne selektion udvælges de helbredskategorier der kan antages at have en signifikant forskel mellem de tre tests, foretaget igennem projekt Krop & Kræft. Dette gøres for ikke at beskæftige sig med helbredskategorier, der ikke kan antages[31] at være påvirkede af ovenstående projekt, og derved ikke er interessante i forhold til fx at bestemme hvilke effekter det statistisk kan vises at være i forbindelse med et givent projektforløb. Selve analysen laves

[29] ($\alpha = 0.10$) Test 1 til 2 med n=22 patienter, figur: (3.18).
[30] Se afsnit: (3.3.2) Parret t-test.
[31] For et givent signifikansniveau.

vha. begge modeller (hhv. Compound Symmetry (model 1) og Unstructured (model 2)), for den mindste testmodel (hhv. "model Yi = testnr".)[32].

I understående to tabeller er vist syntaksen for anvendelse af model 1 og 2 i en proc Mixed analyse.

Model - 1: Compound Symmetry
proc mixed data=data_input;
class testnr ptid;
model Yi = testnr/ outpm=data_output_model1;
random intercept / subject=ptid;
id testnr ptid Yi;
run;

Model - 2: Unstructured
proc mixed data=data_input;
class testnr ptid;
model Yi = testnr/ outpm=data_output_model2;
repeated testnr / type=un subject=ptid;
id testnr ptid Yi;
run;

Tabel 3.9: Proc Mixed'syntaks for model - 1 og 2.

Ovenstående syntaks består i hovedtræk af enkeltlinier, der hver især er beskrevet i understående skema.

[32] Se understående syntaks for model 1 og 2.

Linie nr.	Forklaring
1.	Definerer en indput fil, bestående af de relevante variable.
2.	Klassificering af de kvalitative variable.
3.	Angiver den testede model, samt en output fil.
4.	Angiver den ønskede model (Compound Symmetry eller Unstructured (un)), samt tildele R_i til den enkelte patient, og angiver uafhængighed mellem patienterne.
5.	Angiver output variable.
6.	Kører samt afslutter programmet.

Tabel 3.10: Proc Mixed - syntaksforklaring.

Nedenstående tabel viser relevante resultater af ovenstående proc Mixed analyse.

	Testet model: $y_i = testnr$			
Variabel (y_i)	Model - 1		Model - 2	
RP	P-value:	0.0009	P-value:	0.0034
	AIC:	545.4	AIC:	536.7
	BIC:	547.7	BIC:	543.5
			P-value$_{Cov}$:	0.0006
BP	P-value:	0.0673	P-value:	0.1857
	AIC:	499.7	AIC:	502.7
	BIC:	501.9	BIC:	509.5
			P-value$_{Cov}$:	0.2466
RE	P-value:	0.0127	P-value:	0.0179
	AIC:	562.3	AIC:	567.4
	BIC:	564.6	BIC:	574.2
			P-value$_{Cov}$:	0.3186
QL2	P-value:	0.0029	P-value:	0.0049
	AIC:	497.1	AIC:	503.5
	BIC:	499.4	BIC:	510.4
			P-value$_{Cov}$:	0.0809
PF2	P-value:	0.2067	P-value:	0.1873
	AIC:	463.7	AIC:	457.3
	BIC:	465.9	BIC:	464.1
			P-value$_{Cov}$:	0.0002
RF2	P-value:	0.0319	P-value:	0.0034
	AIC:	534.1	AIC:	535.3
	BIC:	536.4	BIC:	542.1
			P-value$_{Cov}$:	0.0167
EF	P-value:	0.1927	P-value:	0.2240
	AIC :	481.6	AIC:	484.4
	BIC:	483.9	BIC:	491.2
			P-value$_{Cov}$:	0.0012

Tabel 3.11: Resultater fra proc Mixed analyser af helbredskategorier fra 1 selektion.

Ud fra 2. selektions kriterierne[33], udvælges følgende helbredskategorier:

2. selektion	
SF-36:	RP, BP, RE
EORTC QLQ-C30:	QL2, RF2

Tabel 3.12: Helbredskategorier efter 2. selektion.

Modelvalg:

Udvælgelsen af model foregår via fremgangsmåden beskrevet i understående skema. Fremgangsmåden gennemløbes for hver af de helbredskategorier der blev udvalgt i 2. selektion. Baggrunden for den valgte prioritering mellem modellerne vha. nedenstående fremgangsmåde, er udelukkende at anvende model 2. hvis det kan antages at model 2, som er en mere kompleks model end model 1, er et bedre bud på hvordan dataene opfører sig (AIC samt BIC er mindre for model 2 end for model 1)[34]. Desuden vurderes modelvalget ud fra om korrelationsestimaterne mellem tests, specielt for model 2 (hhv. $\sigma_1^2, \sigma_2^2, ..., \sigma_6^2$), er signifikante forskellige (P-value$_{cov}$), da dette er en væsentlig begrundelse for at vælge model 1 fra, da korrelationsestimatet mellem tests for model 1 er ens (σ_0^2).

Hvis	AIC_model2 < AIC_model1
	og BIC_model2 < BIC_model1
	og P-value$_{cov}$ < 0.10%.
	vælg → model - 2.
Ellers	
	vælg → model - 1.

Tabel 3.13: Fremgangsmåde for modelvalg (hhv. model - 1 eller 2).

Resultatet af ovenstående syntaks ses i understående tabel, der viser hvilken model, der er udvalgt til den videre analyse af de respektive helbredskategorier.

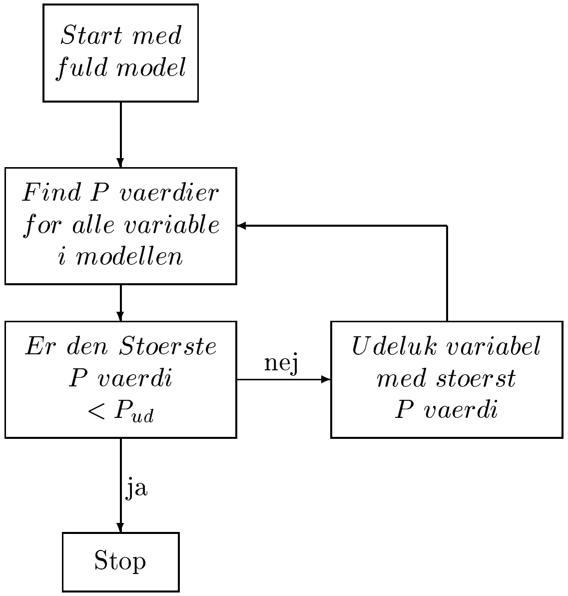
[33] Se selektionsdiagram, figur: (3.22).
[34] Det antages at en simpel model, altid er at foretrække frem for en mere kompleks, da en øget kompleksitet af en model i dette tilfælde, ligeledes øger antallet af parameterestimater, og derved øget antal af usikkerhedsmomenter for hvert parameterestimat.

Variable	Modelvalg:	
	Model - 1	Model - 2
RP		X
BP	X	
RE	X	
QL2	X	
RF2	X	

Tabel 3.14: Resultater af modelvalg.

Backward - elimination:

Den baglæns elimination (Backward - elimination) foregår som vist i nedenstående diagram[35].



Figur 3.23: Backward elimination.

I nedenstående skema er vist den fulde model, som gradvist reduceres for hver proc Mixed analyse, det er nødvendigt af foretage før de resterende forklarende variable overholder grænseværdien i ovenstående diagram (figur: 3.23):

QL2 =	testnr Team Age_40 ... sex Sygdomsstatus_ved_inklusion ... Aktivitet_baseline Civilstatus

Tabel 3.15: Fulde model.

[35][H. Rootzén, 1999]

I tabel: (B.1) i bilag B, er vist resultaterne af en backward elimination af ovenstående forklarende variable, foretaget vha. de respektive udvalgte modeller. Resultaterne er ordnet i den rækkefølge som de elimineres fra den fulde model. For helbredskategorier der har syv elimineringer[36] forsvinder den respektive helbredskategori ud af testen, da ingen af de forklarende variable er signifikante nok til, i denne sammenhæng, at kunne forklare den respektive variation i datamaterialet.

Af tabel: (B.1) i bilag B, ses at tre helbredskategorier ender ud med signifikante forklarende variable, dette drejer sig om følgende helbredskategorier:

Yi =	Forklarende variable
RP =	Testnr
RE =	Team
QL2 =	Testnr

Tabel 3.16: Resultater af "Backward elimination".

Variationen for de tre helbredskategorierne i ovenstående tabel, kan for RP, samt QL2 forklares af en signifikant p-værdi $< 1.0\%$ forskel mellem de tre tests, mens variationen for scores i RE-kategorien kan forklares vha. en ligeledes signifikant forskel mellem opdelingen af patienterne i tre teams[37].

Krydsning af variable: Det er undladt at benytte krydsede variable i den fulde model, da disse ville dele det testede datamaterialet op i meget små grupperinger. Fx ville en relativ lille krydsning af hhv. testnr, Age_40 og sex (testnr*Age_40*sex), lave 4 grupperinger pr. testnr (1, 2 og 3):

Mandlige patienter over 40 år.
Kvindelige patienter over 40 år.
Mandlige patienter under ($\leq$) 40 år.
Kvindelige patienter under ($\leq$) 40 år.

Ovenstående krydsning betyder at i praksis vil hver gruppe bestå af 2 til 6 patienter pr. test, og medfører at parametre i en proc Mixed analyse, estimeres ud fra meget få observationer hvilket vil gøre dem særdeles påvirkelige af tilfældige faktorer og usikkerhedsmomenter.

[36]Dvs. at samtlige forklarende variable fjernes fra modellen.
[37]Se evt. tabel: (B.1) bilag B.

En yderligere årsag til ikke at benytte krydsninger, er den mindsket betydning for projektet, i at tests om variationen i det givne datamateriale, skyldes at der er forskel på om patienterne er fx kvinder eller mænd samt over eller under 40 år, for hver test (se ovenstående eksempel). Det synes i denne sammenhæng mere relevant at teste, om variationen i højere grad eksempelvis skyldes en opdeling mellem kvinder og mænd (sex), end den skyldes den tidsmæssig effekt (testnr.), der kan realiteters til en effekt af projekt Krop & Kræft.

Missing test:

Denne tests er medtaget for at vurdere, om det er af signifikant betydning for variationen i ovenstående modeller, at der forekommer missing data. Testen er forbundet med stor usikkerhed, da en opdeling af patienterne i hhv. dem med missing data og dem med komplette data[38], udelukkende vil bestå af gruppen af patienter med komplette datasæt i sidste test (nr. 3). Af samme grund er der kun foretaget en sådan opdeling af patienterne i test 1 og 2.

Testen for evt. effekt af "missing data" er foretaget vha. følgende modeller[39]:

Yi =	Forklarende variable
RP =	Missing
QL2 =	Missing

Tabel 3.17: Anvendte modeller til Missing test.

Af testresultaterne i understående tabel, ses at der ikke synes at være signifikant forskel på gruppen af patienter med missing data, frem for patienter komplette data (udelukkende ved at kigge på test 1 og 2).

Yi	Model/variable	P-value
RP	Testnr Missing	0.7121
QL2	Testnr Missing	0.6197

Tabel 3.18: Resultater af Missing test.

[38] Se afsnit: (3.1.5) Opdeling af patienter, under "Missing".
[39] Gennemført med en proc Mixed analyse ud fra model 1.

Parameterestimater af Proc Mixed

I understående to tabeller (hhv. tabel: 3.19 og 3.20), er vist parameterestimater for de to helbredskategorier hvis variation kunne forklares af en forskel mellem tests.

- $E(y_i)$: Angiver prædikteret middelværdi for de respektive testnr, med hhv. estimeret spredning ($\hat{S_d}$) og (95%) konfidensinterval.
- $\Delta E(y_i)$: Angiver den prædikteret forskel i scores mellem de respektive tre tests.

Prædikteret parameterestimater for helbredskategorien RP (Fysisk betinget begrænsninger):

Model - 2	RP = testnr		
	$E(y_i)$	$\hat{S_d}$	95% konfidensinterval
$test_1$	12.27	3.93	12.27 ± 7.88
$test_2$	33.41	8.92	33.41 ± 17.89
$test_3$	49.65	11.11	49.65 ± 22.27
Δ_{scores}	$\Delta E(y_i)$		
$test_{1-2}$	21.13		
$test_{2-3}$	16.25		
$test_{1-3}$	37.38		

Tabel 3.19: Parameterestimater og proc Mixed resulater for helbredskategorien RP.

Prædikterede parameterestimater for helbredskategorien QL2 (Global helbredsstatus (Quality of Life 2)):

Model - 1	QL2 = testnr		
	$E(y_i)$	$\hat{S}_d$	95% konfidensinterval
$test_1$	54.35	4.45	54.35 ± 9.05
$test_2$	61.50	4.52	61.50 ± 9.20
$test_3$	75.74	5.56	75.74 ± 11.30
Δ_{scores}	$\Delta E(y_i)$		
$test_{1-2}$	7.15		
$test_{2-3}$	14.34		
$test_{1-3}$	21.49		

Tabel 3.20: Parameterestimater og proc Mixed resulater for helbredskategorien QL2.

Af ovenstående to tabeller kan følgende konstateres:

- Patienternes score for helbredskategorien RP stiger i gennemsnit med 21.1 point $\Delta E(y_i)$, mellem første og sidste uge i projektfasen (hhv. test 1 og 2).

- For samme periode, stiger helbredskategorien QL2 knapt så markant, da den gennemsnitligt forøgelse i score her udgør 7.2 point.

- Ses på forskellen mellem uge 1 og uge 18 i projektetforløbet (hhv. test 1 og 3), stiger helbredskategorien RP gennemsnitligt med 37.4 point, mens QL2 stiger med 21.5 point.

Kapitel 4

Objektive data fra fase 1

4.0.6 Beskrivelse af objektive data

De objektive data består hovedsageligt af målinger af patienternes fysiske kapacitet. Målingerne er foretaget af personalet fra projekt Krop & Kræft, og den enkelte patient har derved ikke haft indflydelse på disse målinger, hvilket var tilfældet hvor patienten selv skulle udfylde et spørgeskema. De objektive data er i det følgende givet ud fra konkrete "objektive" mål, foretaget på hver patient:

Objektive data	
Blodtryk	(I hvile)
Puls	(Hvilepuls)
Temperatur	(Kropstemperatur)
Vægt	(Kropsvægt)
Styrke	($\sum(leqpress, Benchpress\ og\ Pulldown/lat.machine)$)
VO_2max	(Maksimale iltoptagelse)

Tabel 4.1: Objektive data (patientmål).

Det skal understreges at der ikke er foretaget tilsvarende målinger på en kontrolgruppe af nogen art, og det derfor ikke er umiddelbart muligt at

foretage en sammenligning med en sådan. Dette betyder at når der i det efterfølgende beskrives ændringer af nogen art begrundes dette udelukkende på forskelle[1] ud fra det opgivne datamateriale og ikke sammenligninger med eventuelle lignende udefra kommende faktorer.

Variabel forklaring

I nedenstående er givet en forklaring på hvad de enkelte objektive målinger, måler hos den enkelte patient.

Blodtryk:

> "Hjertet er en muskel som pumper blodet rundt i kroppen. Når hjertet trækker sig sammen (hjertets systole) pumpes der blod mættet med ilt ud i legemspulsåren og videre ud i pulsårerne til kroppens organer. Når organerne har brugt ilten føres blodet tilbage via blodkar til hjertet, som fyldes med blod (hjertets diastole). Blodet føres videre ud til lungerne for at blive iltet. Ved blodtrykket forstås det tryk, som blodet udøver på indersiden af blodkarrenes vægge. Blodtrykket er størst i legemspulsåren og aftager jo længere man kommer væk fra hjertet. Det der i almindelighed omtales som blodtrykket, er trykket i overarmens store blodkar.
>
> - Det høje blodtryk (systoliske blodtryk), er det tryk der udøves på blodkarrenes vægge, når hjertet pumper blodet ud i legemspulsåren (hjertets systole)
> - Det lave blodtryk (diastolisk blodtryk), er det tryk der udøves på blodkarrenes vægge, når hjertet fyldes med blod (hjertets diastole).
>
> Blodtrykket består således af to tal, det høje og det lave blodtryk, og det måles i millimeter kviksølv (mmHg), f.eks. 130/70 mmHg". (Citat fra: www.netpatient.dk/hypertension.htm)

Puls: Puls er det antal slag hjertet slår pr. minut, og hvilepuls er det antal slag hjertet slår pr. minut mens kroppen er i hvile.

Temperatur: Patientens kropstemperatur i grader celsius.

[1] Mellem målinger indenfor samme patientmål/gruppe.

Vægt: Patientens kropsvægt målt i kg.

Styrke(Σ): Måles i tre forskellige styrke maskiner, hvilke ses af underståénde tre billeder. Øverst til venstre ses chest/benchpress, øverst til højre ses pulldown/lat. machine, og nederste billede viser en legpress øvelse. Styrken er summen(Σ) af de antal kg, patienten maksimalt kan trække, skubbe eller hive i de respektive tre maskiner.

Figur 4.1: Styrkeøvelser: (øverst tv.) chest/benchpress, pulldown/lat. machine og nederst legpress.

$\mathbf{VO_2max}$: Er målinger af patientens iltoptagelse (målt i ml pr kg pr minut) også kaldet “Den maksimale iltoptagelse” eller “Den Aerobiske kapacitet (Aerobic capacity)”. VO_2max udtrykker følgende:

> “The magnitude of VO_2max depends on the capacity of the lungs to exchange oxygen between the air and the blood in lung capillaries, the capacity of the cardiovascular system to transport oxygen to the muscles, and the muscles's capacity to use oxygen.” [Cureton & Plowman, 2002]

Mere specifikt kan målinger af VO_2max vise følgende hos patienten:

> “The rate of aerobic metabolism and oxygen uptake increases as intensity of exercise increases up to the point at which the aerobic capacity is reached. At this point, even though the exercise intensity can be increased, the oxygen uptake no longer increases proportionally, and there is a plateau in the relation of rate of oxygen uptake to work rate (exercise intensity). The rate of oxygen uptake at the plateau is aerobic capacity.” [Cureton & Plowman, 2002]

Direkte målinger af VO_2max er teknisk krævende, da der kræves avanceret udstyr til at måle iltoptagelsen hos patienten mens denne udfører en fysisk øvelse, som fx et løbebånd eller en kondicykel. I projekt Krop & Kræft er benyttet en indirekte VO_2max-test, der udføres på ergonometercykel (Monark ergomedic 839E). Første del af testen er en steady state test i 8 minutter og anden del er en såkaldt trappetest[2]:

- Steady state testen: Er med en belastning på 67 watt og varer i 8 minutter. Testen starter med en belastning på 67 watt og deltagernes puls i det sidste minut af testen noteres. Testen er et udtryk (målt i puls) for hvor meget kredsløbet skal arbejde ved en given fast belastning i 8 minutter.
- Trappetesten: Starter med en belastning på 67 watt. Belastningen stiger for hvert minut med 20 watt. Testen stoppes når deltageren ikke kan eller ønsker at cykle mere. Herefter noteres puls, belastning opnået i watt samt den samlede tid der er cyklet i. VO_2max. Udregnes. Hver deltager får udregnet deres kondital (ml/kg/min). Testen er et udtryk (målt i l/min) for hvor mange liter ilt blodet maksimalt kan indeholde ved start og slut af de 6 ugers træning.

[2] Steady state, samt trappetesten er begge beskrevet af projekt Krop & Kræfts personale.

En indirekte VO_2max-måling er illusteret i understående billede, som er fra projekt Krop & Kræft:

Figur 4.2: Indirekte VO_2max-måling vha. en ergonometercykel (Monark ergomedic 839E).

Fremgang ses ved en øget VO_2max-måling mellem test (hhv. uge 1 og 6 (18)), hvilket vil sige at blodet kan indeholde en øget mængde ilt. Tilbagegang ses ved en mindsket VO_2max.

4.0.7 Indsamling af objektive data

De enkelte objektive data er indsamlet som vist i nedenstående to skemaer:

		Interventionsperiode		
Krops- og træningsjournal:	Inden	Dagligt	Ugentligt	6. uge
Blodtryk/temperatur		x		
Hvilepuls		x		
Konditionstest (VO_2max)	x			x
Styrketest	x			x
Vægt	x		x	

	Opfølgning*				
Krops- og træningsjournal:	1 måned	3 måned	1 år	2 år	3 år
Blodtryk/temperatur					
Hvilepuls					
Konditionstest (VO_2max)		x	x		
Styrketest		x	x		
Vægt					

**Se understående note: Opfølgningstidspunkterne.*

Udpluk af skema: Forskningsmetoder og tidspunkt for dataindsamling (Interventionsgruppen), fra UCSF Metodeskema Dec 2002.

Tabel 4.2: Plan for indsamling af objektive data, i fase 1.

Noter til skemaer (tabel: 4.2):

- **Opfølgningstidspunkterne:** Er sket efter projektets afslutning, dvs. efter de første 6 uger, eks.: 3 måneder skal derfor læses som 12 uger efter projektafslutning dvs. 3 måneder og 6 uger fra uge 1 (dvs. uge 18).
- **Fridage:** Der er reelt ikke foretaget målinger samtlige dage i ugerne i projektfasen (uge 1 til 6). Dette betyder at tirsdag, torsdag, lørdag samt søndag, skal opfattes som fridage i projektet.
- **Blodtryk i hvile, hvilepuls og kropstemperatur:** I de 6 uger projektet varer, er der inden patienten er startet på nogen form for træning, blevet målt patientens blodtryk, puls samt kropstemperatur.
- **Vægt:** Patienten er éen gang om ugen blevet vejet, hvilket som regel er sket hver mandag.
- **VO_2max samt Styrke:** Den Aerobiske kapacitet samt patientes styrke, er blevet målt to gange hhv. ved projektets start (uge 1.), samt ved projektets afslutning (uge 6.).

Yderligere noter om indsamling af objektive data:

- For patienter der ikke er mødt op til de enkelte målinger, foreligger der missing data.
- Data fra opfølgningsskemaet er ikke medtaget for fase 1, da disse ikke er modtaget fra projekt Krop & Kræft.

4.1 Grafisk præsentation af objektive data

De objektive data er i det følgende primært[3] vist i form af forskellige scatterplots. For at gøre det lettere for læseren af gennemskue graferne, er det valgt at vise projektforløbet i form af 42 dage (mandag (uge 1) til søndag (uge 6)), selvom der i projektet ikke er foretaget målinger mandage, torsdage samt lørdage og søndage[4]. Dette sker for at give en realistisk visning af afstande mellem de reelle måletidspunkter.

Deltaplots istedet for spaghettiplots: Det er valgt ikke at vise spaghettiplots for data med flere end 2 målinger pr. patient. Årsagen til dette valg

[3]Styrke samt VO_2max er vist i form af spaghettiplots.
[4]Se "fridage" i afsnit: (4.0.7) Indsamling af objektive data.

er illustreret i figur 4.9 samt bilag C. Her ses spaghettiplots for de objektive data, og det ses at disse plots meget hurtigt bliver uoverskueligt, hvis der optræder mange målinger. Yderligere forvirres plottene af missing data, som i disse målinger forekommer mere spredt, fx midt i projektforløbet. Dette betyder at det er svært at sammenligne kurver da det ikke umiddelbart er logisk hvilke linier der hænger sammen for den enkelte patient (se bilag C).

Da spaghettiplots bliver meget uoverskuelige, er det istedet valgt at illustrere ændringerne i patientmålingerne, i form af såkaldte deltaplots. Disse plots er lavet for følgende patientmål: Højt og lavt blodtryk, hvilepuls samt kropstemperatur, og viser sammenhænge mellem ændringer i patientmålene ($\Delta\ Test\ nr.\ x =$ test nr. x til test nr. x+1), som er forbundet vha. af linier for hver patient (hvis der er målinger på de pågældende testnumre).

Glidende gennemsnit: I scatterplottene er, med undtagelse af plot for Styrke og VO_2max, indlagt et glidende gennemsnit, dette gennemsnit er beregnet af observationer for det givne tidspunkt og glider/ løber indtil der kan beregnes et nyt gennemsnit, hvilket betyder næste tidspunkt med målinger. De glidende gennemsnit viser derved det generelle niveau for de målinger der er foretaget til samme tidspunkt, samt indikerer, hvordan dette niveau opfører sig igennem projektfasen.

Outliers: Der er ikke fjernet nogle former for outliers for de pågældende scatterplots, med undtagelse af plottet for hvilepuls. I hvilepulsplottet er fjernet målinger med puls $\leq$ 10 her under puls 0, disse observationer er fjernet ud fra den vurdering at der er tale om decideret målefejl/tastefejl[5]. Begrundelsen for at der ikke er fjernet outliers fra de øvrige scatterplots samt datasæt, er at det er vurderet at det ikke er hensigtsmæssigt at fjerne målinger der ikke i sig selv er ekstreme eller abnorme, selv om de afviger fra de øvrige målinger. Afvigende målinger er istedet vurderet til at være del af en naturlig spredning i forhold til at patienterne er fysisk vidt forskellige.

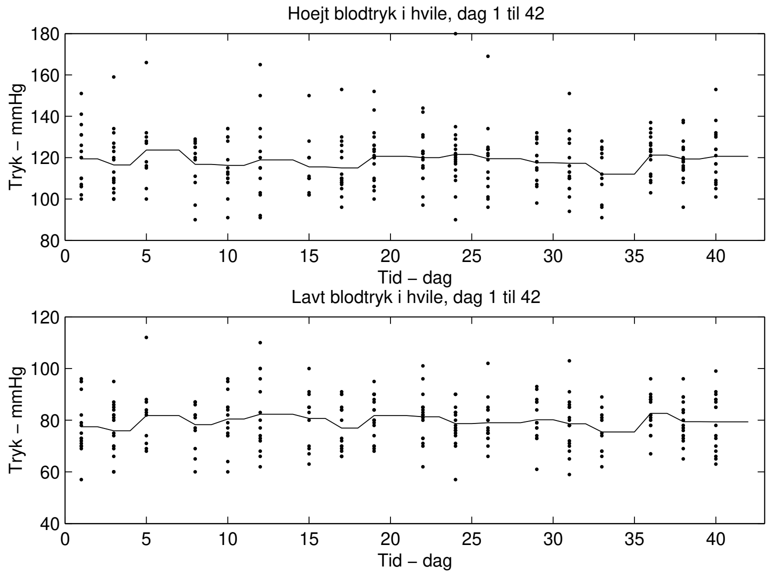
Måleusikkerhed: Det skal i det efterfølgende plots af data, tages højde for en vis måleusikkerhed, som må påregnes da der er tale om målinger med eksempelvis termometre, der i sig selv vil have en vis unøjagtighed. Desuden vil der med stor sandsynlighed være tilfælde af fx aflæsningsfejl fra personen, som foretager den pågældende måling. Da der ikke er opgivet

[5] Puls 0 er det samme som patienten er død, og puls 10 er ekstremt lavt og urealistisk i forhold til at der ifølge UCSF ikke er opstået uventede/utilsigtede komplikationer under projektfasen/træningen (patienter i overhængende livsfare).

størrelser for eventuelle måleusikkerheder, som beskrevet i ovenstående, antages det at måleusikkerhederne er normalfordelte med en middelværdi på nul(0), samt en ukendt spredning ($N(0, \sigma_x),\ \sigma_x = ukendt$).

4.1.1 Højt og Lavt hvileblodtryk

Understående to scatterplots viser hhv. det høje og det lave blodtryk for patienterne igennem projektfasen.



Figur 4.3: Scatterplot af højt og lavt blodtryk (mmHg).

Af scatterplottene for højt og lavt blodtryk, ses i hovedpunkter følgende:

Højt blodtryk:

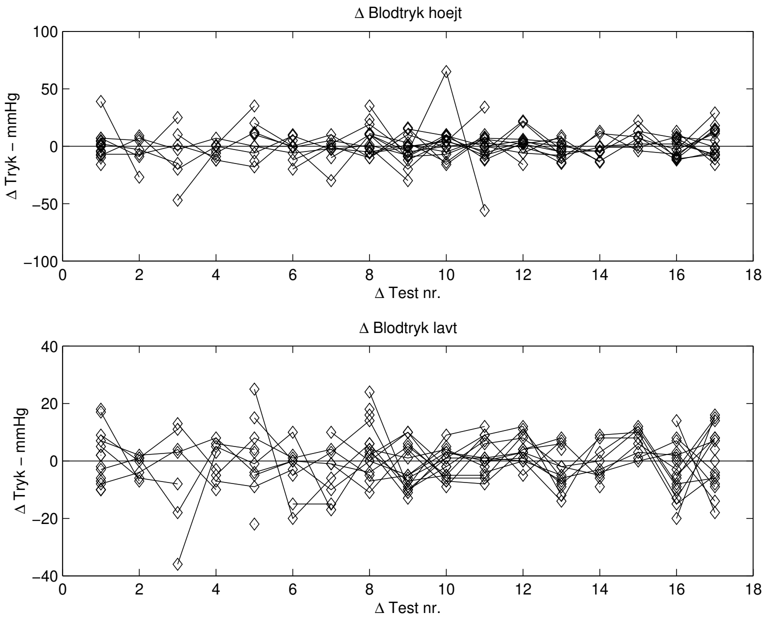
- For det høje blodtryk findes der en forskel på cirka 90 mmHg mellem højeste og laveste måling (180 mod 90), men flertallet af målinger befinder sig indenfor max. 140 til min. 95 mmHg.
- Mht. forholdet mellem måletidspunkterne longitudionalt, synes der ikke at være en entydige tendenser til forandringer, denne vurdering

understøttes ligeledes af det glidende gennemsnit (linien), der momentant stiger og falder mellem 125 og 110 mmHg, men tilsyneladende falder tilbage til det et generelt niveau (120 mmHg) efter relativt kort tid.

Lavt blodtryk:

- Det lave blodtryk synes at følge det høje blodtryk nøje, med den forskel at det ovenfornævnte generalle niveauet, for det lave blodtryk ligger på cirka 80 mmHg.

I understående to deltaplots ses ændringer af patienternes blodtryk igennem de i alt 18 tests der er foretaget på patienterne[6].



Figur 4.4: Deltaplots af højt og lavt blodtryk (mmHg).

Note: Det ses at deltaplottene ud af den vandrette akse, ikke udtrykker de faktiske tidsforskelle mellem tests, hvilket var tilfældet med scatterplottene. Dette betyder, at der for flere af de beregnede ændringer mellem tests (Δ Test nr.), er op til 3 dage mellem de faktiske målinger[7].

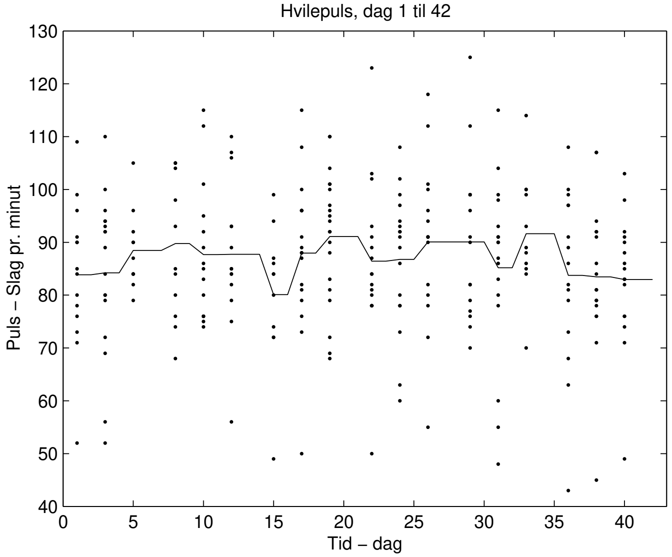
[6] $\Delta Test\ nr. = 17stk.\ (18-1)$.

Af ovenstående deltaplots ses at ændringer i blodtryk (højt og lavt), for flertallet af patienterne er relative små (Δ *blodtryk* ≈ 0), kun tre ændringer er på mere end ± 40 mmHg for det høje blodtryksmålinger, og disse efterfølges umiddelbart af en modsvarende ændring. Ændringerne i de lave blodtryksmålinger ligger hovedsageligt indenfor en grænse på plus/minus ca. 25 mmHg, og viser en umiddelbar stabilitet.

Det skal understreges at deltaplots anvendes til fx at vise om patienter oplever eventuelle parallelle ændringer, men ved små positive og negative ændringer kan have svært ved at vise om patientmålingerne ændrer sig på længere sigt. Derfor kan deltaplots ikke umiddelbart anvendes hvis enkelte patienter oplever mindre vedvarende ændringer, da fx mange små positive ændringer kan få generel betydning for patientens start- og slutniveau. For at tage højde for sådanne ændringer betragtes istedet det glidendegennemsnit i scatterplottene, samt spaghettiplottene i bilag C (behandles kort i efterfølgende afsnit om “Kropstemperatur” under: Generelle tendenser).

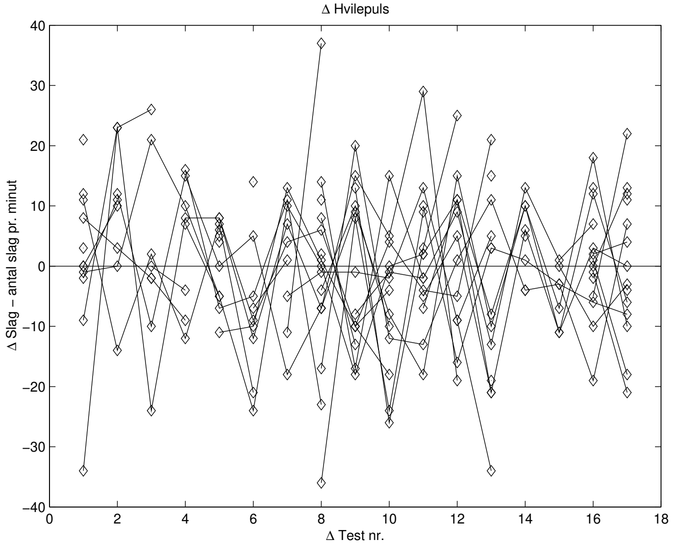
[7] Fra lørdag til og med mandag.

4.1.2 Hvilepuls



Figur 4.5: Scatterplot af hvilepuls (slag pr. minut).

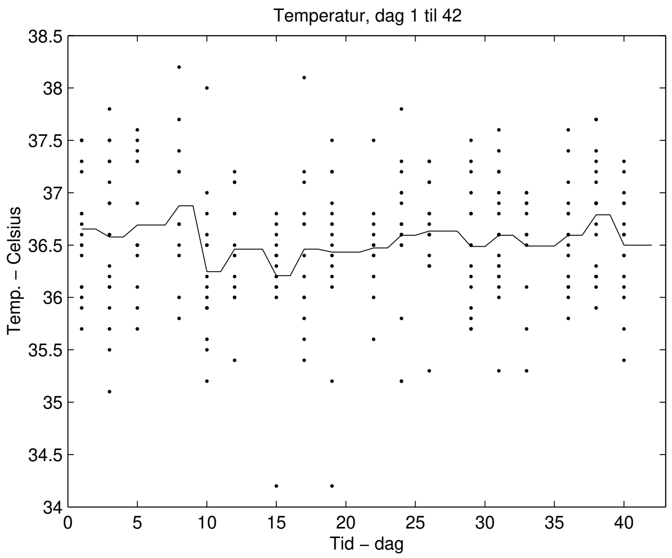
I ovenstående scatterplot ses hvilepulsmålingerne af patienterne i projektfasen. Det ses at der heller ikke her er tydelige tendenser på ændringer i den generelle hvilepuls (longitudionalt). Dette vurderes ud fra at det glidendegennemsnit svinger umiddelbart ligeligt hhv. over og under et generelt niveau på ca. 85 pulsslag pr. minut, under hele projektfasen (max. 92 til min. 78 pulsslag pr. minut), samt at de enkelte målinger er fordelt relativt jævnt på begge sider af de samme generelle niveau (85 pulsslag pr. minut).



Figur 4.6: Deltaplot af hvilepuls.

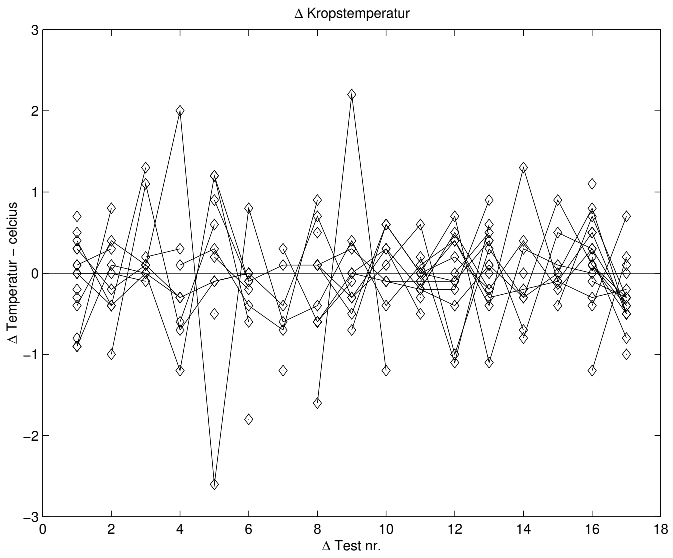
Af ovenstående deltaplot forstærkes iagtagelserne fra scatterplottet, om at det er svært at se tydelige tendenser i punkterne. Dette ses af de mange spredte punkter hhv. over og under nul-linien, for hvilke der ikke synes at være tegn på noget generelt mønster.

4.1.3 Kropstemperatur



Figur 4.7: Scatterplot af kropstemperatur (celsius).

Kropstemperaturen synes ud fra ovenstående scatterplot samt det glidende-gennemsnit, umiddelbart at opfører sig stabilt med en lille variation for det glidendegennemsnit på omkring 1 grad, mens variationen for målingerne er på ca. 2 grader. Det glidendegennemsnit fordeler sig over tid relativt jævnt omkring 36,5 grader, hvilket også er omtrent samme slutværdi som startværdi. Der synes ikke at være nogle umiddelbare tendenser at se. I det understående deltaplot, bekræftes ovenstående.



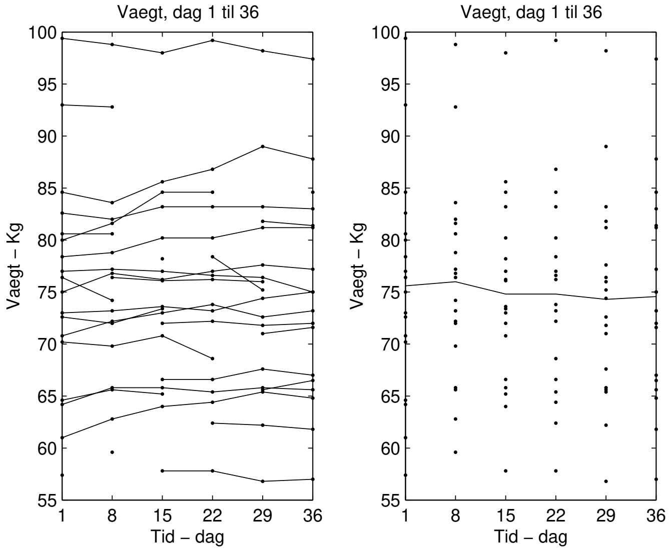
Figur 4.8: Deltaplot af kropstemperatur.

I ovenstående deltaplot ses at langt hovedparten af ændringerne ligger indenfor plus/minus 1 grad (celsius), og er jævnt fordelt omkring nul-linien. De enkelte store udsving (±2 grader) synes at være kortvarige, og modsvares umiddelbart efter af lignende udsving, til den modsatte side af nul-linien, hvorefter de stabilisere sig om nul-linien igen.

Generelle tendenser: For deltaplottene af hhv. højt og lavt blodtryk, hvilepuls samt kropstemperatur, ses en generel tendens til at en positiv ændring umiddelbart efterfølges af omtrent lige så stor negativ ændring (og omvendt). Denne tendens bekræftes af de glidendegennemsnit i de respektive scatterplots, der momentant har begrænsede udsving, og der efter synes at falde tilbage til et generelt niveau for det pågældende patientmål.

Betragtes spaghettiplottene vist i bilag C, giver heller ikke disse anledning til andre antagelser end de ovenfor viste scatterplots og deltaplots. Dette vurderes ud fra at der ikke synes at være nogen ensartet tendens, da linierne (for hvert plot) umiddelbart krydser hinanden vilkårligt.

4.1.4 Vægt



Figur 4.9: Spaghettiplot samt scatterplot af vægt (kg).

Note: Da der i modsætning til blodtryk, hvilepuls samt kropstemperatur, udelukkende er foretaget vejning af patienterne for hver mandag i projektfasen vises vægtdataene med kun en uges mellemrum, hvilket ses af ovenstående to plots.

Spaghettiplottet i ovenstående figur (venstre plot), illustrerer problematikken med at gengive patienternes udvikling longitudionalt, ved at forbinde deres observationer med linier, og derved lave et såkaldt spaghettiplot. Selv for vægtdataene, der er relativt begrænset i mængde i forhold til fx hvilepulsmålinger, forvirres ovenstående plot af de mange data med enkeltpunkter og brudte linier igennem tidsforløbet (udgør omkring 16 stk.), ligesom det er svært at sammenligne linier når der er mange af dem og de ikke viser et gennemskueligt mønster.

For de hele linier[8], synes der ikke umiddelbart at være tendenser til generelle ændringer i patienternes vægt, selvom to patienter umiddelbart tager et par

[8]Patienter med målinger for alle 6 tidspunkter.

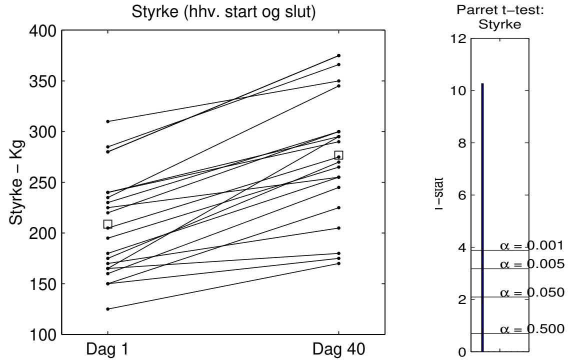
kilo på (linierne startende ved hhv. 61 og 84,5 kg). De udeblivende tendenser bekræftes ligeledes af scatterplottet (figuren til højre), hvor det glidende gennemsnit svinger yderst minimalt omkring 75 kg.

4.1.5 Styrke(Σ) og VO$_2$max

For målingerne af patienternes styrke samt VO_2max er der foretaget målinger hhv. ved projektets start samt slutning. Dette samt den faktor at der er tale om en relativt begrænset mængde patienter hhv. 21 for styrke og 20 for VO_2max, gør disse målinger fordelagtige at vise i form af spaghettiplots frem for scatterplots[9].

Styrke(Σ):

For styrkemålingernes vedkommende, er disse givet ved summen af tre forskellige styrkeprøver[10].



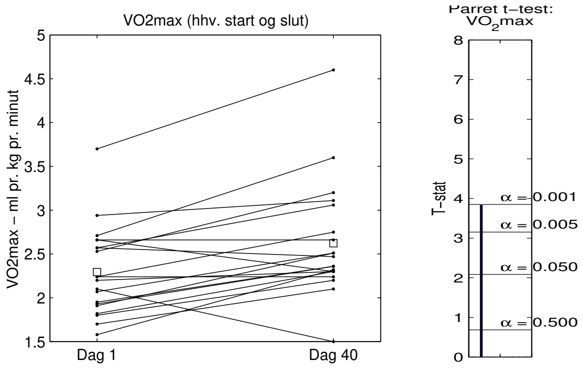
Figur 4.10: Spaghettiplot samt parret t-test af styrke(Σ) (kg).

I ovenstående figur ses patienternes styrke (grafen til venstre). Der synes i dette tilfælde at være tale om en tydelig tendens til en forøgelse af samtlige

[9]Styrke(Σ) og VO_2max antages at være normalfordelte størrelser, med en middelværdi på μ og en spredning på σ.
[10]Beskrevet i afsnit: (4.0.6) Variabel forklaring.

patienters styrke(Σ), selvom hver patient synes at starte fra et meget individuelt udgangspunkt (læs start). Den stigende tendens er yderligere bekræftet af grafen til højre, som viser resultatet af en parret t-test[11] mellem hhv. start og slut målingerne (hhv. dag 1 og 40). De vandrette linier viser signifikansniveauerne 50 til 0.1 procent, og testresultatet viser en yderst signifikant ændring ($t_0 = 10.2$) af styrken(Σ) hos patienterne, hvor den gennemsnitlige forøgelse af den summeret styrke(Σ) pr. patient er 67.9 kg ($N(\hat{\mu}_d, \hat{\sigma}_d) = (67.90, 29.58)\ kg$).

VO_2max:



Figur 4.11: Spaghettiplot samt parret t-test af VO_2max (ml pr. kg. pr. minut).

I ovenstående figur for VO_2max'målingerne, synes der at være tendens til en generel forøgelse i VO_2max, selvom enkelte patienter oplever en forværring. I den parret t-test for målinger af VO_2max (plottet til højre), ses en signifikant testsstørrelse ($t_0 = 3.83$), der i forhold til teststørrelsen for styrke(Σ), er mindre signifikant da dets grænseniveau er på 0.5 procent. Den gennemsnitlige forøgelse i VO_2max mellem hhv. start og slut målingerne af projektfasen, er estimeret til 0.33 ml pr. kg. pr. minut ($N(\hat{\mu}_d, \hat{\sigma}_d) = (0.33, 0.39)\ ml\ pr.\ kg\ pr.\ minut$).

[11] Se afsnit: (3.3.2) Parret t-test.

4.2 Proc Mixed analyser af objektive data

Det er valgt at anvende en proc Mixed analyse på følgende patientmål hhv. højt og lavt blodtryk, hvilepuls, kropstemperatur samt vægt. Dette er gjort ud fra følgende betragtninger:

- Proc Mixed kan analysere testdagene gruppevist, samt med forbehold for longitudional korrelation[12], hvilket kan anvendes til at undersøge om der er signifikante forskelle mellem de i projektfasen foretaget tests (test_dag).
- De grafiske præsentationer af samme udvalgte patientmål, var svære at gennemskue p.g.a. mange observationer.

4.2.1 Beskrivelse af testparametre.

Proc Mixed analysen af objektive data, er til forskel for analyserne af spørgeskemadataene, kendetegnet ved at der er flere teststidspunkter, samt at afstandene mellem testtidspunkterne varierer ud fra hvilke dage i ugen testene/ målingerne er udført. Desuden er målingerne kontinuert fordelt og kun afgrænset af en fysisk betinget måleskala (fx puls 0 eller 300 for hvilepuls).

Til at beskrive målingerne, er det valgt at anvende Compound Symmetry modellen (model 1), der er beskrevet i afsnit:(3.3.4) SAS proc Mixed. Model 1 er valgt ud fra følgende vurderinger:

- Model 1 vurderes ud fra bl.a. antagelser om longitudional korrelation, at være et godt bud på hvordan dataene opfører sig.
- Det vurderes at være unødvendigt at anvende en mere kompliceret model, da fx model 2 ville betyde estimering af væsentligt flere parametre, og at disse ikke nødvendigvis, ville beskrive dataene bedre.
- En simplere model vurderes i denne henseende utilstrækkelig, da der skal tages højde for bl.a. longitudional korrelation i dataene.

Syntaksen der er anvendt ses i nedenstående skema. Den største forskel fra den lignende anvendelse af samme model for spørgeskemadataene, er at variablen testnr for de objektive data er erstattet af test_dag. Dette kan gøres da testdagene kan opfattes som en testnummerering, mens test_uge

[12] Se afsnit: (3.3.3) Korrelation mellem variable.

er indført som testnummerering for vægtdataene. De resterende linier er beskrevet i afsnit: (3.3.5) Resultater af ProcMixed, underafsnit: (2. Selektion).

Model - 1: Compound Symmetry
proc mixed data=data_input;
class test_dag ptid;
model Yi = test_dag / outpm=data_output_model1;
random intercept / subject=ptid;
id test_dag ptid Yi;
run;

Tabel 4.3: Proc Mixed'syntaks for analyse af objektive data.

I ovenstående syntaks er den afhængige variabel (Yi), for hver analyse givet ved hhv. højt og lavt blodtryk, hvilepuls, kropstemperatur samt vægt.

Fordelingsantagelse: Det antages at målingerne af hhv. højt og lavt blodtryk, hvilepuls, kropstemperatur og vægt hver især er normalfordelte størrelser, med en middelværdi på μ og en spredning på σ.

Målingens usikkerhed: Et væsentligt usikkerhedsmoment i forbindelse med analyse af objektive data, er den usikkerhed som hver af de respektive målinger er forbundet med. Da der ikke er opgive nogle estimater for nogle af disse usikkerheder, er der i stedet antaget at disse er normalfordelte med en middelværdi på 0, samt en ukendt spredning. Dette betyder at sådanne usikkerheder er inkluderet i parameteren R_i for model 1 i proc Mixed analysen[13].

4.2.2 Resultater af proc Mixed analyse

I understående tabel ses resultaterne af proc Mixed analyse på de respektive objektive data.

[13] Se afsnit: (3.3.4).

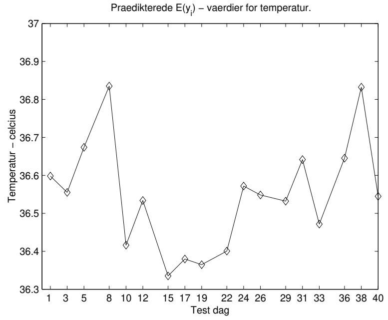
Variabel (y_i)	P-value
Blodtryk - højt:	0.4035
Blodtryk - lavt:	0.2425
Hvilepuls:	0.5235
Temperatur:	0.0882
Vægt:	0,00004

Tabel 4.4: Resultater fra proc Mixed analyse af objektive data.

Af analyseresultaterne i ovenstående tabel vurderes følgende:

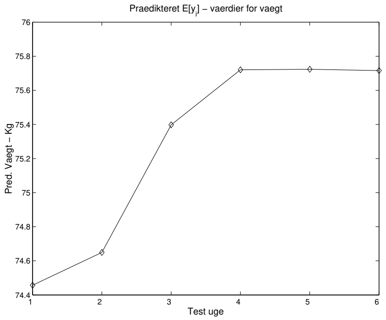
- De generelt høje p-værdier viser at der ikke synes at være nogen særlige signifikante forskelle på målingerne på de enkelte patienter mht. testdage.

- P-værdien for Temperatur-variablen er den mest signifikante af de alle fire, men som det ses af nedenstående graf over de prædikterede middelværdier, skyldes forskellen mellem de forskellige testdage snarere en variation i den målte kropstemperatur igennem hele testforløbet (projektfasen), end en egentlig forskel mellem målinger foretaget i begyndelsen og afslutningen af projektfasen. Dermed synes der ikke at være baggrund for at antage at selve projektfasen betyder en varig forandring i patientens kropstemperatur.



Figur 4.12: Prædikteret temperatur dag 1 til 40.

- For vægtmålingerne synes der at være særdeles signifikant forskel på de forskellige testuger. Af understående graf ses linien for de prædikterede middelværdier, og i understående tabel er ligeledes givet estimater for spredning, konfidensinterval samt ændringer mellem testuger.



Figur 4.13: Prædikteret vægt uge 1 til 6.

Model - 1	vægt = testnr		
	$E(y_i)$	$\hat{S}_d$	95% konfidensinterval
$uge1$	74.46	2.089	74.46 ± 4.15
$uge2$	74.65	2.087	74.65 ± 4.15
$uge3$	75.40	2.088	75.40 ± 4.15
$uge4$	75.72	2.089	75.72 ± 4.15
$uge5$	75.72	2.088	75.72 ± 4.15
$uge6$	75.72	2.089	75.72 ± 4.15
Δ_{scores}	$\Delta E(y_i)$		
uge_{1-2}	0.19		
uge_{2-3}	0.75		
uge_{3-4}	0.32		
uge_{4-5}	0.00		
uge_{5-6}	0.00		
uge_{1-6}	1.26		

Tabel 4.5: Parameterestimater og proc Mixed resulater for vægt-målinger.

Ses på ændringer i de prædikterede middelværdier i ovenstående tabel, synes patienterne at tage cirka 1.26 kg på, i uge 1 til uge 4. Dernæst holder patienternes vægt sig konstant.

Yderligere bemærkes at estimatet for spredningen samt konfidensintervallet holder sig meget stabilt for samtlige testuger.

Kapitel 5

Afsluttende behandling

I de forrige kapitler er resultater af spørgeskemadataene og de objektive data, blevet behandlet hver for sig. I dette kapitel vil resultater af analyser for begge slags data blive sammenlignet, med henblik på at give en overordnet fortolkning af resultater, samt fremlægge de problemstillinger der igennem rapporten er fundet relevante i forbindelse med behandling af data.

5.1 Resultater og sammenligninger

For spørgeskemadataenes vedkommende er der tre tests der har været anvendt:

- Parret t-test
- Korrelationstest
- Proc Mixed analyser

For de objektive data, er følgende tests blevet anvendt:

- Parret t-test
- Proc Mixed analyser

I nedenstående skema ses de helbredskategorier og patientmål, for hvilke der i de anvendte test og analyser, blev fundet signifikante forskelle mellem

tests (test_nr, testda og test_dag)[1]:

Test:	Helbredskategorier	Objektive mål
Parret t-test:	RP*, RE*	Styrke$(\Sigma)_{\alpha=0.001}$ $VO_2max_{\alpha=0.005}$
Proc Mixed:**	RP, QL2	Vægt

** signifikansniveau: α = 5%*
*** signifikansniveau: α < 1%*

Tabel 5.1: Oversigt over resultater.

Parret t-test: Ovenstående skema viser at der i forbindelse med projektfasen er tegn på forøgelse af patienternes fysiske styrke og VO_2max. I samme periode er der for helbredskategoriernes vedkommende, signifikante forbedringer af RP (fysisk betinget begrænsninger) og RE (psykisk betinget begrænsninger). Perioden efter projektfasen (uge 7-10) er ikke medtaget, da antallet af patienter med spørgeskemadata her falder fra 22-23 stk. til 12-13 stk. samt at de objektive data ophører efter uge 6.

Proc Mixed: Proc Mixed analyserne af udvalgte helbredskategorier gav signifikante forskelle mellem tests for helbredskategorierne RP og QL2. Det synes umiddelbart overraskende at analysen af helbredskategorien QL2 viste signifikante forskelle mellem tests, da analysen af denne ikke havde signifikante forskelle mellem test i den parrede t-test. I proc Mixed analysen inkluderes imidlertid test nr. 3, hvilket umiddelbart er årsagen, da de prædikterede ændringer mellem test for QL2, viste en relativ beskeden score forøgelse på 7.15 point mellem test 1 til 2, mod en score forøgelse på 14.34 point mellem test 2 og 3 (hhv. 21.13 og 16.25 point for helbredskategorien RP i samme perioder[2]). En generel øget score for helbredskategorien QL2, synes derfor i større grad at gøre sig gældende i perioden fra uge 7 til 10, end for perioden hvor patienterne er i den egentlige behandling (projektfasen (uge 1 til 6)).

For de objektive data viste resultater af proc Mixed analyserne, at patienternes vægt øges i projektfasens 3 til 4 uger med 1,26 kg[3], mens vægten for de resterende 2 uger holder sig stabilt.

[1] Helbredskategorier fra tabel: (3.18), afsnit: (3.3.2) Parret t-test, samt proc Mixed analyser.
[2] tabel: (3.19) for RP og tabel: (3.20) for QL2.
[3] Se tabel: (4.5).

Korrelationstest: Analyserne af korrelationerne mellem de 23 helbredskategorier, gav ikke entydige resultater. Dette viste sig ved at der var signifikante korrelationer mellem forskellige helbredskategorier, for hver af de to test (1 og 2) der blev analyseret. For SF-36 og EORTC-skemaerne er der på trods af de forskellige resultater, interessante korrelationer at påpege. Dette drejer sig om signifikante korrelationer mellem helbredskategorien QL2 (Quality of life 2) fra EORTC skemaet, og helbredskategorierne: "Generelt helbred" (GH), "energi" (VT) og "social funktion" (SF) fra SF-36 skemaet.

Ovenstående korrelationer betyder antageligt at helbredskategorier fra SF-36'skemaet kan anvendes til at sammenligne med den overordnede helbredskategori QL2 fra EORTC'skemaet. Denne problematik er relevant fordi de overordnede helbredsmål hhv. MCS og PCS fra SF-36'skemaet, i flere tilfælde giver anledning til meget abstrakt fortolkninger af patientens overordnede helbredsstatus, hvis der eksempelvis optræder forøgelse eller formindskelse af bestemte helbredskategorier[4].

5.2 Problemstillinger for behandling af data og resultater

Den statistiske behandling af data samt fortolkning af resultater fra test og analyser, har i rapporten givet anledning til en række af forskellige problemstillinger. Disse problemstillinger er i det følgende gennemgået nærmere.

Manglende kontrolgruppe: For at kunne fortolke om eventuelle forandringer af patienternes helbred, i form af spørgeskemaundersøgelser samt objektive målinger, kan skyldes effekter af et gennemført projektforløb, er det vigtigt at kunne udelukke effekter der kan skyldes naturlige variationer. For de foreliggende datas vedkommende er det ikke muligt at vurdere om de forandringer, der er fundet signifikante igennem denne rapport, i virkeligheden skyldes forhold der ligeledes optræder for patienter der ikke har gennemgået et lignende projektforløb. Dette problem ville antageligt kunne imødekommes hvis der indføres en kontrolgruppe at sammenligne målinger med.

[4]Se tabel: (2.7).

Missing data: Et problem med de anvendte data, er i stor grad de mange manglende data. Dette betyder at det er særdeles svært at analysere på specielt test nr. 3 for helbredskategoriernes vedkommende. For de objektive data betyder de manglende data at patienternes forløb er svære at sammenligne, da der er få data for de enkelte tidspunkter, hvor patienterne befinder sig på samme tidsmæssige stadie i projektforløbet og derved svækker det patientforløb, der er interessant specielt i forbindelse med anvendelsen af en SAS proc Mixed analyse.

Information i manglen af data: En interessant problemstilling for en fremtidig behandling af lignende data, kunne være at vurdere om der ligger information i manglen af data. En sådan information kunne eventuelt betyde at manglen af data kunne fortolkes, og derved indgå i de anvendte test og analyser med eksempelvis estimerede værdier.

Målingernes usikkerhed: Denne problemstilling er specielt relevant for de objektive datas vedkommende, hvilket skyldes at der ikke for nogle af disse målinger, er opgivet tal for eventuelle usikkerheder. Hvis der eksempelvis havde været opgivet et estimat for usikkerheden på en blodtryksmåling, ville dette betyder at en del af variationen i blodtryksmålingerne kunne tilskrives denne usikkerhed, og medtages i den statistiske behandling.

For spørgeskemaernes vedkommende er målingernes usikkerheder, dels forbundet med en gentagelses usikkerhed/ variation for patientens besvarelse af enslydende spørgesmål, og dels i hvilket omfang de enkelte helbredskategorier "måler" hvad de er tiltænkt. Dette betyder at der kan være tvivl om fx helbredskategorien QL2, måler en "livskvalitet" der relevant i forhold til projektet. Dette spørgsmål er ikke op til denne rapport at vurdere, men er et spørgsmål om validering af de anvendte spørgeskemaer[5].

5.3 Metodeovervejelser

Opdeling af patienter: De forskellige opdelinger/ grupperinger af patienterne (civilstatus, alder o.s.v.) var mht. proc Mixed analyserne ikke signifikant, med undtagelse af "test"[6] og "team"[7]. Dette betød, at for de øvrige

[5]Se evt. afsnit: (2.1.1) SF-36 -Spørgeskema om helbredsstatus. Underafsnit: Reliabilitet, Floor, Ceiling og Validitets begreber.
[6]Givet ved: Testnr og test_dag/ test_uge.
[7]Se tabel: (3.16).

opdelinger/ grupperinger vedkommende, var disse ikke relevante i forhold til at forklare variationen i det aktuelle datamateriale. Det kan imidlertid ikke udelukkes, at den pågældende opdeling stadig kan være relevant for et større datasæt.

Selektionsmetoder: Den anvendte "backward elimination", kan udvides til mere omfattende metoder. En sådan udvidelse er i forhold en analyse af fase 1 data, ikke vurderet relevant, da kun team- og testopdelingerne viste sig signifikante i de udførte proc mixed analyser. Det kan imidlertid ikke udelukkes at en mere omfattende metode kan være nyttig for et større datasæt.

Selektions-p-værdier: De anvendte selektions-p-værdier, fra de tre selektionstrin i proc Mixed analysen af helbredskategorierne, kan øges så der er mulighed for at opnå en bredere beskrivelse af variationen i scores for helbredskategorierne. Dette kan være relevant for større datasæt, med flere patienter for hver gruppering.

Statistisk styrkeberegning: For større datasæt kan det være relevant at beregne estimater for den statistiske styrke af de anvendte tests, for statistisk at underbygge påviste tendenser. Dette er i denne rapport ikke vurderet at være relevant, da der i forvejen er et relativt lille patientgrundlag at påvise tendenser med, og dermed udelukkende ses efter mulige tendenser.

Korrelationer: For større datasæt, kunne det være interessant at gå mere i dybden mht. korrelation imellem helbredskategorier, og eventuelt udvide til at kigge på korrelation mellem helbredskategorier og de objektive data. Dette vil antageligt kunne give et mere sammenhængende billede af dels spørgeskemadataene/ helbredskategorierne, samt af helbredskategorierne og de objektive patientmål.

Kapitel 6

Konklusion

Rapporten har benyttet en række test og analysemetoder, som har haft til formål at påpege tendenser i det anvendte datamateriale. Resultater af disse test og analyser, antyder at der er en tendens til at patienterne igennem projektfasen oplever en generel forøgelse af deres styrke og VO_2max, samt at deres vægt ligeledes øges en smule i projektetfasens første 3-4 uger.

Lader man patienterne selv evaluere deres helbred igennem projektforløbet vha. spørgeskemaerne SF-36 og EORTC QLQ-C30, viser disse at patienterne mener deres fysisk betinget begrænsninger mindskes, samt at deres overordnede helbred i form af helbredskategorien QL2 ligeledes forøges. For helbredskategorien QL2 er gældende, at denne hovedsageligt øges over et længere forløb, der strækker sig til tre måneder efter at den egentlige behandling har fundet sted.

For de øvrige helbredskategoriers vedkommende, her i blandt angst og depressions variable, viste analyse af disse data ikke tegn på signifikante forandringer igennem projektet, ligesom der heller ikke var signifikante forandringer at spore for målinger af patienternes blodtryk (højt/lavt) og hvilepuls. For målinger af patienternes kropstemperatur, syntes der at være signifikante forandringer under projektfasen men ikke nogen umiddelbart effekt i forhold til start og slutmålinger.

Den statistiske behandling af data i denne rapport, er i høj grad præget af en begrænset datamængde samt af en manglende kontrolgruppe. Det er

i det hele taget forbundet med en vis usikkerhed, ud fra denne rapport, at udtale sig om projektet “Krop & Kræft’s” indvirkning på de involveret patienters subjektive som objektive helbredsmål, og det er i høj grad diskutabelt om antallet af patienter er tilstrækkeligt for en statistisk vurdering af en effekt af den pågældende intervention.

De opnåede resultater skal på baggrund af ovenstående, udelukkende ses som tendenser i et datamateriale, mens den anvendte statistiske fremgangsmåde kan ses som en guide til hvordan efterfølgende datasæt kan behandles. Det anbefales at man for en lignende statistisk behandling, inddrager en kontrolgruppe at holde sine resultater op mod, samt benytter et et øget antal patienter. Sammensætningen af patienter for fase 1, synes ud fra en statistisk vurdering, repræsentativt velovervejet.

Bilag A

Tabel A.1: HAD score tabel.

spm	Skala		Svar nr.	Score
1	A			
			1	3
			2	2
			3	1
			4	0
2		D		
			1	0
			2	1
			3	2
			4	3
3	A			
			1	3
			2	2
			3	1
			4	0
4		D		
			1	0
			2	1
			3	2
			4	3
5	A			
			1	3
			2	2
			3	1
			4	0
6		D		
			1	3
			2	2
			3	1
			4	0
7	A			
			1	0
			2	1
			3	2
			4	3

spm	Skala		Svar nr.	Score
8		D		
			1	3
			2	2
			3	1
			4	0
9	A			
			1	0
			2	1
			3	2
			4	3
10		D		
			1	3
			2	2
			3	1
			4	0
11	A			
			1	3
			2	2
			3	1
			4	0
12		D		
			1	0
			2	1
			3	2
			4	3
13	A			
			1	3
			2	2
			3	1
			4	0
14		D		
			1	0
			2	1
			3	2
			4	3

Tabel A.2: SF-36 score tabel.

Guide til tabel:	
H	= Helbredskategori.
O-kode	= Oprindelig kode.
E-kode	= Endelig kode.

* = koden for spørgsmål 8 og 7, som begge er besvaret

H	Item	Svarmuligheder		O-kode	E-kode
PF	3a-j	Ja, meget begrænset		1	1
		Ja, lidt begrænset		2	2
		Nej, slet ikke begrænset		3	3
RP	4a-d	Ja		1	1
		Nej		2	2
BP	7	Ingen smerte		1	6
		Meget lette smerter		2	5.4
		Lette smerter		3	4.2
		Middelstærke smerter		4	3.1
		stærke smerter		5	2.2
		Meget stærke smerter		6	1
	8 (+7)		(8*)	(7*)	
		Slet ikke	1	1	6
		Slet ikke	1	2-6	5
		Lidt	2	1-6	4
		Noget	3	1-6	3
		En hel del	4	1-6	2
		Virkelig meget	5	1-6	1
	8 (−7)	Slet ikke	1		6
		Lidt	2		4.75
		Noget	3		3.5
		En hel del	4		2.25
		Virkelig meget	5		1

H	Item	Svarmuligheder		O-kode	E-kode
GH	1	Fremragende		1	5
		Vældigt godt		2	4.4
		Godt		3	3.4
		Mindre godt		4	2
		Dårligt		5	1
	11a, 11c	Helt rigtigt		1	1
		Overvejende rigtigt		2	2
		Ved ikke		3	3
		Overvejende forkert		4	4
		Helt forkert		5	5
	11b, 11d	Helt rigtigt		1	5
		Overvejende rigtigt		2	4
		Ved ikke		3	3
		Overvejende forkert		4	2
		Helt forkert		5	1
VT	9a, 9e	Hele tiden		1	6
		Det meste af tiden		2	5
		En hel del af tiden		3	4
		Noget af tiden		4	3
		Lidt af tiden		5	2
		På intet tidspunkt		6	1
	9g, 9i	Hele tiden		1	1
		Det meste af tiden		2	2
		En hel del af tiden		3	3
		Noget af tiden		4	4
		Lidt af tiden		5	5
		På intet tidspunkt		6	6
SF	6	Slet ikke		1	5
		Lidt		2	4
		Noget		3	3
		EN hel del		4	2
		Virkeligt meget		5	1
	10	Hele tiden		1	1
		Det meste af tiden		2	2
		Noget af tiden		3	3
		Lidt af tiden		4	4
		På intet tidpunkt		5	5

H	Item	Svarmuligheder		O-kode	E-kode
RE	5a-c	Ja		1	1
		Nej		2	2
MH	9b, 9c, 9f	Hele tiden		1	1
		Det meste af tiden		2	2
		En hel del af tiden		3	3
		Noget af tiden		4	4
		Lidt af tiden		5	5
		På intet tidspunkt		6	6
	9d, 9h	Hele tiden		1	6
		Det meste af tiden		2	5
		En hel del af tiden		3	4
		Noget af tiden		4	3
		Lidt af tiden		5	2
		På intet tidspunkt		6	1

H	Item	Svarmuligheder	O-kode	E-kode
HT	2	Meget bedre nu end for ét år siden	1	-
		Noget bedre nu end for ét år side	2	-
		Nogenlunde det samme	3	-
		Noget dårligere nu end for ét år siden	4	-
		Meget dårligere nu end for ét år siden	5	-

Bilag B

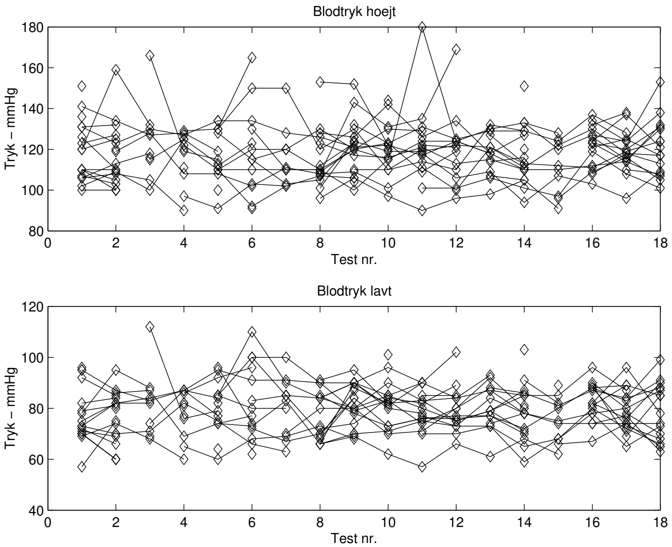
Tabel B.1: Backward elimination.

Y_i	Selektion nr.	Variabel	P-value
RP	1.	Sex	0.9394
	2.	Syg_st	0.8534
	3.	Age_40	0.3846
	4.	Team	0.2168
	5.	Aktiv_ba	0.2081
	6.	Civ_st	0.0182
	-	testnr	0.0034
BP	1.	Sex	0.8501
	2.	Age_40	0.7590
	3.	Aktiv_ba	0.6845
	4.	Team	0.1238
	5.	testnr	0.1514
	6.	Civ_st	0.0496
	7.	Syg_st	0.0278
RE	1.	Age_40	0.8988
	2.	Syg_st	0.8496
	3.	Aktiv_ba	0.6852
	4.	Sex	0.4448
	5.	Civ_st	0.3080
	6.	testnr	0.0316
	-	Team	0.0006
QL2	1.	Team	0.9589
	2.	Aktiv_ba	0.9002
	3.	Civ_st	0.7412
	4.	Sex	0.4233
	5.	Age_40	0.2042
	6.	Syg_st	0.1725
	-	testnr	0.0029
RF2	1.	Aktiv_ba	0.9901
	2.	Team	0.8987
	3.	Sex	0.7626
	4.	Age_40	0.5406
	5.	Civ_st	0.2700
	6.	Syg_st	0.2324
	7.	testnr	0.0319

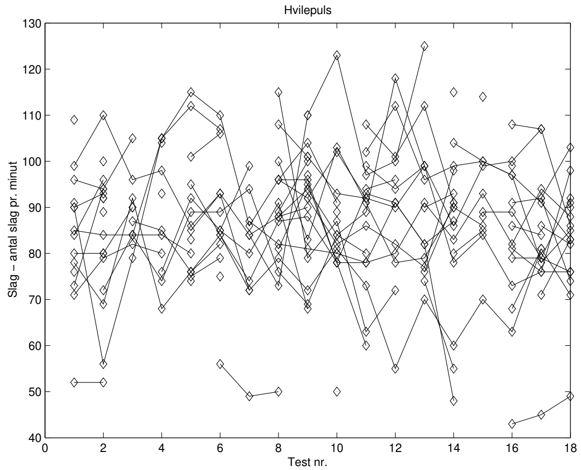
Variabel forkortelser	
Sygdomsstatus_ved_inklusion:	Syg_st
Aktivitet_baseline:	Aktiv_ba
Civilstatus:	Civ_st

Tabel B.2: Forklaring af variabelforkortelser anvendt i "backward elimination".

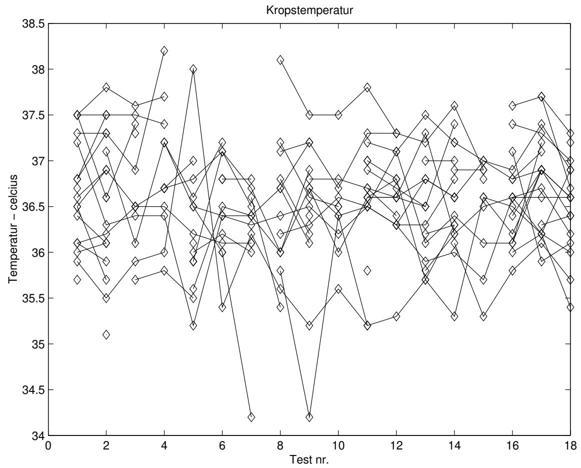
Bilag C



Figur C.1: Spaghettiplot af højt og lavt blodtryk.



Figur C.2: Spaghettiplot af hvilepuls.



Figur C.3: Spaghettiplot af kropstemperatur.

Litteratur

[Brendan, 1993] Brendan T. Carroll, M.D., Roger G. Kathol, M.D., Russell Noyes, Jr., M.D., Tina G. Wald, M.D., and Gerald H. Clamon, M.D.: *Screening for Depression and Anxiety in Cancer Patients Using the Hospital Anxiety and Depression Scale*. General Hospital Psychiatry, 15, 69-74, 1993.

[Cureton & Plowman, 2002] Kirk J. Cureton & Sharon A. Plowmanm: *FITNESSGRAM Reference Guide: Aerobic Capacity Assessment*. The Cooper Institute; Dallas TX, 2002. (www.Cooperinst.org)

[Diggle, 1994] Peter J. Diggle, Kung-yee Liang and Scott L. Zeger: *Analysis of Longitudinal Data*. Oxford science Publications, Oxford Statistical Science series 13, 1994.

[Douglas C. Montgomery, 2001] Douglas C. Montgomery: *Introduction to statistical quality control,4_th edition*. Arizona State University, John Wiley & sons, inc. 2001.

[EORTC scoring manual, 2001] Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group: *The EORTC QLQ-C30 Scoring Manual (3rd Edition)*. Published by: European Organisation for Research and Treatment of Cancer, Brussels 2001.

[EORTC QLQ-C30, 1993] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JCJM, Kaasa S, Klee MC, Osoba D, Razavi D, Rofe PB, Schraub S, Sneeuw KCA, Sullivan M, Takeda F.: *The European Organisation for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology*. Journal of the National Cancer Institute, 85, 365-376, 1993.

[H. Rootzén, 1999] Helle Rootzen (Holst): *Anvendt statistik og introduktion til SAS III*. IMM, DTU, 1999.

[Johnston, 2000] Marie Johnston, Beth Pollard, Peter Hennessey: *Contruct validation of the hospital anxiety and depression scale with clinical populations*. Journal of Psychosomatic Research, 48, 579-584, 2000.

[K. Conradsen, 1995] Knut Conradsen: *En introduktion til statistik, bind 1B, 6. udgave*. IMM, DTU, 1995.

[N.K. Aaronson, 1993] N.K. Aaronson: *Development of core questionnaire and lung module in lung cancer patients Publication: The European Organization for Research and Treatment of Cancer QLQ-C30: A Quality-of-Life Instrument for use in international clinical trials in oncology.*. National Cancer Institute, Vol 85, No. 5, March 1993.

[P. Thyregod, 1998] Poul Thyregod: *En Introduktion til statistik, Bin 3A-D*. IMM, DTU, 1998.

[SF-36 manual, 1997] Jakob Bue Bjørner, Mogens Trab Damsgaard, Torquil Watt, Per Bech, Niels Kristian Rasmussen, Tage Søndergaard, Jens Modvig, Kate Thunedborg: *Dansk manual til SF-36*. København, 1997.

[Skarstein, 2000] Jon Skarstein, Nina Aass, Sophie D. Fosså, Eva Skovlund, Alv A. Dahl: *Anxiety and depression in cancer patients: relation between the Hospital Anxiety and Depression Scale and the European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire*. Journal of Psychosomatic Research, 49, 27-34, 2000.

[Snaith & Zigmond, 1994] R. P. Snaith, A. S. Zigmond: *The hospital Anxiety and Depression Scale, with the irritability-Depression-Anxiety Scale, and The Leeds Situational Anxiety Scale, Manual*. NFER-NELSON, 1994.

[Thiébaut, 2002] Rodolphe Thiébaut, Héléne Jacqmin-Gadda, Geneviéve Chêne, Catherine Leport, Daniel Commenges: *Bivariate linear mixed models using SAS proc MIXED*. Computer Methods and Programs in Biomedicine, 69, 249-256, 2002.

[Wolfinger, 1997] Russell D. Wolfinger, Ph.D. : *An example of using mixed models and proc mixed for longitudinal data*. Journal of Biopharmaceutical Statistics, 7(4), 481-500. 1997.

[Zigmond & Snaith, 1983] A. S. Zigmond, R. P. Snaith: *The hospital Anxiety and Depression Scale*. Acta psychiatr. scand., 67, 361-370, 1983.