

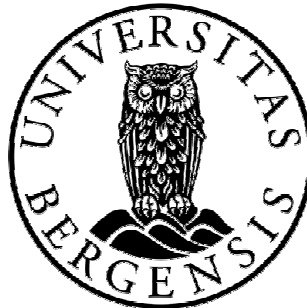
*Magnetisk resonanstomografi i
multisenterstudier*

Objektiv og subjektiv evaluering av bildekvalitet

Masteroppgave i helsefag – Radiograf

Ingeborg Elise Nakken

Høsten 2011



Institutt for samfunnsvitenskapelige fag

Universitetet i Bergen

St. Olavs Hospital, Klinikk for bildediagnostikk

Trondheim

Forord

Denne oppgaven ble til i forbindelse med et engasjement i en tverrfaglig nasjonal multisenterstudie. Den har således bidrag fra bildediagnostiske avdelinger i helseforetakene Helse Sør-Øst, -Vest og -Midt-Norge.

Muligheten til å ta en mastergrad ved siden av min radiografstilling ved St. Olavs Hospital er et privilegium. Det har vært en intens prosess, men en erfaring jeg ikke ville vært foruten. Takk til mine overordnede Karin I. Steen (klinikksjef, Klinikk for bildediagnostikk) og Bjarte D. Snekvik (seksjonsleder, Seksjon for MR-radiografer) for deres positive innstilling og organisering av det praktiske.

Hjertelig takk til mine to veiledere Jarle Rørvik (professor i radiologi, Institutt for kirurgiske fag, Universitetet i Bergen) og Asta Kristine Håberg (professor i neuroimaging, dr. med., Institutt for nevromedisin, NTNU) for deres tid og samarbeid. Takk til Una Ørvm Sølvik (førstemanuensis, koordinator for Masterprogram i Helsefag, studieretning RAB-fag, Institutt for samfunnsmedisinske fag, Universitetet i Bergen) for god oppfølging underveis.

Prosjektet hadde ikke latt seg gjennomføre uten økonomisk støtte. Klinikk for bildediagnostikk ved St. Olavs Hospital har dekket reisene til samlinger ved Universitetet i Bergen, samt perioder med permisjon. Nasjonalt kompetansesenter for aldring og helse ved Psykiatrien i Vestfold har dekket reisen til deltakersentrene. Nasjonalt kompetansesenter for funksjonell MRI i Trondheim har dekket utstyr og fraktkostnader. Kontaktutvalget mellom St. Olavs Hospital og Det medisinske fakultet ved Norges teknisk-naturvitenskapelige universitet (NTNU) i Trondheim har gitt et fullføringsstipend på kr 50.000,- Takk til Toril Åsnes (økonomirådgiver, Klinikk for bildediagnostikk, St. Olavs Hospital) som har administrert økonomien og ført regnskap for meg.

Takk til alle som arbeider ved sentrene som deltok i prosjektet. Uten deres hjelp og velvillighet i forbindelse med besøk og avbildning av frivillig og fantom ville ikke dette vært mulig. Takk til Heidi Henrikke Engebretsen (radiograf, Radiologisk avdeling,

Stavanger Universitetssykehus) for lån av fantom og for samarbeidet som innledet til masterprosjektet.

Takk til radiologene Kjell Arne Kvistad (seksjonsoverlege, dr. med., St. Olavs Hospital) og Jana Rydland (overlege, St. Olavs Hospital) for innspill og oppmuntring. Takk til Turi Olene Dalaker (lege i spesialisering, Radiologisk avdeling, Stavanger Universitetssykehus) for interesse og velvilje. Alle tre besvarte spørreskjemaet som ble laget i dette prosjektet med engasjement og grundighet.

Jeg vil også benytte anledningen til å rette en stor takk til alle ja-menneskene jeg har rundt meg ved St. Olavs Hospital. MR-teknisk gruppe ved seksjon for tekniske funksjoner er unike. Pål Erik Goa (MR-fysiker, dr. scient.), Anders Kristoffersen (fysiker, dr. ing.) og Roar Sunde (overingeniør) har en fagkunnskap, erfaring og imøtekommenhet uten sidestykke.

Takk til alle mine gode kolleger ved MR-seksjonen for gjennomlesning og oppmuntring. Takk til Børge min. Din støtte er uvurderlig.

Trondheim, 15. november 2011.

Ingeborg Elise Nakken

Innhold

Forord	i
Sammendrag	v
Abstract	vi
1 Introduksjon	1
1.1 Bakgrunn.....	2
1.2 Teori	3
1.2.1 Optimalisering, standardisering, og kvalitetskontroll.....	3
1.2.2 Bildekvalitet	5
1.2.3 Ulike aspekter ved bildekvalitet på MR	7
1.2.3.1 Signal/støy forhold og signalhomogenitet	7
1.2.3.2 Lavkontrastdetaljer, kontrast/støy-forhold og romlig oppløsning	10
1.2.3.3 Gradientlinearitet, snittykkelse og snittposisjon	11
1.2.3.4 Ghosting og kjemisk skift	11
1.2.3.5 T2-måling.....	13
1.3 Tidligere forskning omkring testing av MR-data i multisenterstudier.....	14
1.3.1 Bruk av fantomdata i multisenterstudier.....	14
1.3.2 Sammenligning av yteevne mellom ulike leverandører	15
1.3.3 Kvalitetskontroll og reproduksjon av funn	16
1.3.4 Testede aspekter ved bildekvalitet i ulike studier.....	17
2 Hensikt og problemstilling	19
3 Metode	20
3.1 Valg av forskningsdesign	20
3.2 Utvalg og utvalgsriterier	20
3.3 Testobjekter og variabler	22
3.4 Optimalisering og standardisering	23
3.5 Datainnsamling	25
3.5.1 Bildeopptak	25
3.6 Analyse	26
3.6.1 Uthenting av informasjon fra opptaks materialet	27
3.6.1.1 Fantomopptak.....	27
3.6.1.2 Hjerneopptak.....	28
3.6.1.3 Beregning av SNR, CNR og T2-verdier.....	30
3.6.2 Hypotesetesting	32
3.6.2.1 Analysemetoder.....	32
3.6.2.2 Beregning av måleusikkerhet.....	34
3.7 Etske betraktninger.....	35
4 Resultat	37
4.1 Variasjon mellom senter og sekvenser.....	37
4.1.1 Signal/støy forhold og signalhomogenitet	37
4.1.2 Lavkontrastdetaljer og romlig oppløsning	44
4.1.3 Gradientlinearitet, snittykkelse og snittposisjon	46
4.1.4 Ghosting.....	49
4.1.5 T2-måling.....	51
4.1.6 Volumetri	54
4.1.7 Visuell vurdering	55
4.2 Tomograf-resultater i forhold til utstyr og andre aspekter	58

4.2.1	Ulike kvalitetsaspekter i forhold til hverandre	58
4.2.2	Ulike kvalitetsaspekter i forhold til utstyr	59
4.3	Objektive målinger i fantomopptak versus hjerneopptak	60
4.3.1	SNR og CNR	60
4.3.2	Ghosting	61
4.3.3	T2-måling	61
4.4	Objektiv versus subjektiv bildekvalitet	61
4.4.1	Signal/støy forhold	62
4.4.2	Vevskontrast og differensierbarhet	62
4.4.3	Ghosting versus generelle forstyrrelser og artefakter	63
4.5	Beregning av måleusikkerhet og testing av bildekvalitet over tid	64
4.5.1	Måleusikkerhet	64
4.5.2	Bildekvalitet over tid (4–6 måneder)	65
5	Diskusjon	66
5.1	Diskusjon omkring funn i de ulike kvalitetsaspektene	66
5.1.1	Signalhomogenitet (<i>U</i>)	66
5.1.2	Signal/støy forhold	67
5.1.3	Kontrast/støy forhold	68
5.1.4	Lavkontrastdetaljer	69
5.1.5	Ghosting	69
5.1.6	T2-måling	70
5.1.7	Usikkerhetsfaktorer i volumetri	71
5.1.8	Sentre som utmerket seg	72
5.2	Generelle begrensninger	74
5.2.1	Størrelse på utvalg og datamateriale	74
5.2.2	Valg av hypotesetester og signifikansnivå	75
5.2.3	Standardisering av protokoller	76
5.2.4	Måleusikkerhet	77
5.2.5	ACR-kriterier og bruken av ACR-fantomet	77
5.2.6	Kvantifisering av kvalitative data	78
5.2.7	Bruken av spørreskjema	80
5.2.8	Omgivelser – temperatur og opptakstidspunkt	82
5.2.9	Sammenligning av objektiv og subjektiv bildekvalitet	83
5.2.10	Generelle utfordringer og veien videre	83
6	Konklusjon	85
	Referanser	86
	Vedlegg	90

Sammendrag

Bakgrunn: Magnetisk resonanstomografi (MR) er en avansert og teknisk sårbar bildeopptaksmetode. Optimalisering av sekvenser på tvers av utstyr øker bildekvaliteten og gir mer reproduserbare multisenterdata. Få har imidlertid testet og sammenlignet tomografrelatert yteevne, og det finnes få retningslinjer for hva som er god bildekvalitet og hvordan kvalitetskontroll bør gjennomføres.

Mål: Teste og sammenligne bildekvalitet og yteevne hos ulike tomografer, uavhengig av leverandør og modell.

Materiale og metode: Opptak av fantom og frisk frivillig i hodespole i åtte tomografer med magnetfeltstyrke 1,5 T fra tre ulike leverandører ble standardisert med tanke på sammenlignbarhet. Hjerneopptakene besto av et T1-vektet 3D-volum, en T2-vektet fluid attenuated inversion recovery (FLAIR)-sekvens, en dobbelekk (protontetthetsvektet og T2-vektet) turbo spinn-ekko (TSE)-sekvens, og en T2*-vektet gradientekko (GRE)-sekvens. Fantomopptakene ble gjennomført to ganger hvert sted med 4–6 måneders mellomrom, og besto av en dobbelekk TSE- og spinn-ekko (SE)-sekvens, samt en T1-vektet SE-sekvens. Signal/støy-forhold (SNR), kontrast/støy-forhold (CNR), T2-verdier, ghosting og geometriske størrelser ble målt i fantomopptak og hjerneopptak. Romlig oppløsning, lavkontrastdetaljer, signalhomogenitet, snittykkelse og snittposisjon ble målt i fantomopptak. Hjerneopptakene ble vurdert visuelt av tre radiologer, hvor differensierbarhet av vev og strukturer, artefakter, samt bildekvalitet som helhet ble vurdert på en skala 1–5.

Resultater: Et flertall av de testede kvalitetsaspektene viste signifikante forskjeller ($\alpha = 0,05$) mellom tomografer. Ulike tomografer skilte seg ut i positiv og negativ retning på ulike tester. Resultatene var stabile over tid.

Konklusjon: Tomografens spesifikasjoner (utstyr, gradientstyrke, og lignende) sier lite om bildekvaliteten, den må testes.

Nøkkelord: magnetisk resonanstomografi, multisenterstudie, kvalitetskontroll, bildekvalitet, fantom, frisk frivillig.

Abstract

Background: Magnetic resonance imaging (MRI) is a technically vulnerable acquisition method. Sequence optimization across equipment increase image quality and reproducibility of multicenter data. Comparisons of scanner-related performance are scarce, mainly due to few guidelines on assessing image quality.

Aim: Test and compare image quality and performance in different MRI scanners, independent of vendor and model.

Material and methods: A phantom and a healthy volunteer's brain were scanned in the head coil of eight MRI machines (field strength 1.5 Tesla) from three different vendors, with sequences standardized for comparability. The brain protocol included the following: T1-weighted 3D-volume, T2-weighted fluid attenuated inversion recovery (FLAIR)-sequence, dual echo (proton density (PD)- and T2-weighted) turbo spin-echo (TSE)-sequence, and T2*-weighted gradient-echo (GRE)-sequence. Phantom scans were repeated in each scanner after 4–6 months. The phantom protocol included a dual echo TSE- and spin echo (SE)-sequence, and a T1-weighted SE-sequence. Signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), T2-values, ghosting and geometric accuracy were measured in phantom- and brain images. Spatial resolution, low contrast object detectability, image intensity uniformity, slice thickness accuracy and slice position accuracy were measured in phantom images. Three radiologists visually evaluated the brain images on a five-point scale, focused on tissue conspicuity, artifacts and overall image quality.

Results: A majority of tested quality aspects were significantly different ($\alpha = 0,05$) between scanners. Different tests accentuated different scanners positively and negatively. Results were stable over time.

Conclusion: Scanner-specific features reveal little about image quality without testing.

Keywords: magnetic resonance imaging, multicenter study, image quality control, phantom, healthy volunteer.

1 Introduksjon

Magnetisk resonanstomografi (MR) av hjernen er blant annet kjent for sin detaljerte gjengivelse av anatomi. Teknologien har gjennomgått en enorm utvikling siden den første MR-tomografen kom til Norge i 1986, både når det gjelder utstyr, opptaksteknikker og bruken av MR-undersøkelser (Smith & Gjesdal, 2000). Ulike leverandører driver teknologien fram i konkurransen om å være først ute på markedet med nytt utstyr og nye metoder for bildeopptak. Dette kappløpet har imidlertid også ført til variasjon i blant annet utstyr og programvare mellom leverandører og modeller, noe som kompliserer muligheten til å sammenligne MR-opptak på tvers av sentre og over tid (Ashton, 2010).

Prognosis of depression in the elderly (PRODE) er en nasjonal multisenterstudie med blant annet MR-avbildning av eldre med førstegangs depresjon. Dette er så vidt meg bekjent første forsøk i Norge på å standardisere MR-opptak for en stor gruppe pasienter ved flere sykehus og på tvers av helseforetak. Det finnes riktignok studier med en felles grunnplattform eller minimumsstandard for bildeopptak med MR, blant annet innen multippel sklerose (MS). Det er imidlertid ikke innført en gjennomgående standardisering av hele MR-protokollen for MS-utredning (Simon, Li, Traboulsee, Coyle, Arnold, Barkhof et al., 2006).

Ett av målene i PRODE er å klargjøre sammenhenger mellom strukturelle hjerneforandringer og psykiatriske lidelser, samt eventuelle sammenhenger mellom dette og utvikling av Alzheimers sykdom (Håberg, Borza, Selbæk & Engedal, 2010). Nasjonalt kompetansesenter for funksjonell MR (fMRI) ved Klinikk for bildediagnostikk ved St. Olavs Hospital i Trondheim, fikk i 2010 ansvaret for MR-delen av utredningen av pasientene i PRODE-studien. Dette innebærer å finne fram til hvilke typer bildeopptak som skal gjøres, sette opp de aktuelle sekvensene, samt få i gang rutiner for kontinuerlig evaluering av MR-opptakene på tvers av tomografer ved sentre i henholdsvis Helse Sør-Øst, Helse Vest og Helse Midt-Norge.

1.1 Bakgrunn

En utvikling av mer standardiserte MR-protokoller kan ha både kliniske og forskningsmessige fordeler. For eksempel kan mer utredning og oppfølging av pasienter skje lokalt, selv om selve behandlingen foregår ved regionsykehusene. Ved å bygge kliniske samarbeidsrelasjoner på tvers av geografi og helseregioner, bygger vi samtidig en infrastruktur som støtter multisenterdata. Å trekke forskningen inn i klinikken på denne måten kan ha positive ringvirkninger både i forhold til kompetanseøkning, og for bedre utnyttelse av kapasitet og ressurser.

Helsedirektoratet har satt som et mål å få gjennomført standardisering av MR-opptak i flere pasientgrupper med hjernesykdom. For eksempel ble den nasjonale forskningssatsingen ”Alvorlige psykiske lidelser” i regi av Nasjonal samarbeidsgruppe for medisinsk og helsefaglig forskning (NSG) vedtatt i desember 2008. Her er ett av målene samarbeid omkring felles forskningsprotokoller, deriblant metoder for MR-avbildning (NSG, 2008).

Felles forskningsprotokoller for hjerneavbildning på tvers av senter og utstyr har klare statistiske fordeler. Flere pasienter, med ulik geografisk tilknytning, kan inkluderes og utredes uten at datainnsamlingen tar lenger tid, og større datamaterialer er som regel bedre for å finne sammenhenger (Polit & Beck, 2010). Målet med en standardisert opptaksprotokoll i PRODE-studien er å få kvantitativt sammenlignbare data på tvers av MR-klinikker, leverandører og modeller. Erfaringen fra PRODE MR kan danne grunnlag for videre arbeid med kvalitetssikring og sekvensoppsett i arbeidet mot nasjonale normer for MR av hjerne.

På grunn av variasjonen mellom leverandører og modeller, er det ikke mulig å få til en helhetlig standardisering av alle MR-sekvenser og -protokoller. Ulik oppbygning av MR-sekvenser kan gi tilsynelatende tilnærmet lik vektning (kontrast), og mange sekvenser er utviklet for å redusere opptakstiden. Det er imidlertid ikke alltid de nyeste, raskeste og mest sofistikerte sekvenser som gir mest sammenlignbare bilder (Ashton, 2010). Flere multisenterstudier har lagt mye arbeid i utvikling av standardiserte MR-sekvenser ved hjernesykdom (Van Horn & Toga, 2009). Et eksempel er Alzheimer’s Disease Neuroimaging Initiative (ADNI) som blant annet gjennomfører MR-avbildning

av hjernen hos eldre med kognitiv svikt (n = 400), Alzheimers sykdom (n = 200) og friske kontroller (n = 200) ved deltakersentre i USA og Canada. Ett av målene i ADNI-studien har vært å utvikle standardiserte MR-sekvenser for å følge endringer over tid i den aldrende hjernen i et stort datamateriale (Mueller, Weiner, Thal, Petersen, Jack, Jagust et al., 2005). I denne forbindelse har ADNI optimalisert to sekvenser med leverandørtilpassede parameter for å gi lik bildekontrast og oppløsning på tvers av leverandører og sentre. Disse er en T1-vektet 3D-sekvens og en dobbelekkosekvens¹ (ADNI, 2011). Dersom en undersøkelsesprotokoll også skal ha klinisk nytteverdi for pasienten, trengs det imidlertid flere sekvenser enn ADNI-sekvensene. PRODE MR er et forsøk på å optimalisere en undersøkelsesprotokoll som både har klinisk nytteverdi for pasienten, og samtidig har en bilde kvalitet som gjør det mulig å samle og sammenligne bildene fra alle sentrene kvantitativt i forskningsøyemed.

1.2 Teori

1.2.1 Optimalisering, standardisering, og kvalitetskontroll

Optimalisering

Hvilke sekvenser en MR-protokoll skal inneholde er begrunnet i hva man ønsker å undersøke, og avgjørelser tas av MR-radiologer. MR-radiografens jobb er å avgjøre hvilke parameter som bør brukes for å oppnå ønsket vektning i bildene. Sekvens- og parametervalg er godt dokumentert både i teori og praksis. Litteraturen er i store trekk samstemt når det gjelder hvilke sekvenser som egner seg best til å vise ulike former for sykdom som for eksempel hjernesvinn, blødning, hvitsubstanslesjoner, osv. (Reimer, Parizel & Stichnoth, 2006). Spillerommet i opptaksparameter for å oppnå ønsket vektning og vevskontrast ved gitte magnetfeltstyrker og sykdomsindikasjoner er også veldokumentert (Moeller & Reif, 2010; Westbrook, 2008). Generelle vurderinger omkring sekvensoppsett og parametervalg ved utredning av eldre deprimerte tas derfor ikke opp her. Spesifikke avveininger i forbindelse med parametervalg i PRODE MR tas opp under punkt 3.4 *Optimalisering og standardisering*.

¹ En dobbelekkosekvens gir to bildesett, et med protontetthetsvektning (PD-vektning, fra engelsk proton density), og et med T2-vektning, i samme opptak.

Standardisering

Når bildekvaliteten skal være stabil på tvers av utstyr og leverandør, bør leverandørspesifikke sekvenser unngås. Bruk av nøyaktig like parameter på alle tomografer gir imidlertid ikke samme bildekvalitet, blant annet fordi sekvensene har noe ulik oppbygning hos ulike leverandører. I tillegg tilbyr ikke alle leverandører samme trinnvise endring i parameter som for eksempel matrisestørrelse og båndbredde. Ulike antenntyper gir ulik grad av støy, og tilpasninger må gjøres for å sikre godt nok bildesignal på alle tomografer, uavhengig av antenntype. Ulike egenskaper ved gradientene tillater ulik snittykkelse og opptakshastighet. Kort sagt må spesifikke parameter tilpasses ulike leverandører og modeller for å få bilder med sammenlignbar bildekvalitet, en slik tilpasning kalles en standardisering (Ashton, 2010).

Kvalitetskontroll

For å undersøke i hvilken grad en standardisering har vært vellykket, bør bildene fra de ulike sentrene kvalitetskontrolleres. Kontrollen bør gjennomføres systematisk og med opptaksparametre og et materiale som er sammenlignbart fra gang til gang og over tid (Van Horn & Toga, 2009). En kvalitetskontroll har flere aspekter. I evalueringen om opptak er sammenlignbare er det hovedsakelig den generelle yteevnen til tomografens ulike komponenter, samt gjengivelsen av anatomi og vevskontrast, som er aktuelt å undersøke (Tofts, 1998). Det finnes geometriske fantomer utviklet for spesifikk testing av de ulike sidene tomografkomponentenes yteevne. American College of Radiology (ACR) startet i 1996 et systematisk opplegg for kvalitetsgodkjenning og sertifisering av MR-tomografer i USA. I den forbindelse har de utviklet et fantom til testing av tomografen, samt en veiledning for hvordan fantomopptakene skal gjøres, hvordan bildene skal vurderes, og kriterier for akseptabel bildekvalitet. Resultatet av disse testene kan sammenlignes med andre tomografer, og med tomografen selv over tid (ACR, 2002; ACR, 2005).

For sammenligning av gjengivelse av anatomi og vevskontrast ved ulike tomografer, bør man i tillegg til geometrisk fantom, gjøre opptak av en frisk frivillig i alle tomografene. Alternativt kan man gjøre opptak av en kadaverhjerne. En kadaverhjerne kan gi realistisk gjengivelse av anatomi, og også fremstilling av patologi, men på grunn av blant annet blodgjennomstrømning, molekylbevegelse og temperatur, er vektningen ulik den man finner i levende vev. I tillegg kan det være vanskelig å sikre at materialet

holder seg stabilt over tid (Tofts, 1998). En evaluering av bildekvalitet fra frivilligopptak på tvers av tomografer, bør omfatte både objektive målinger og subjektive vurderinger (jf. punkt 1.2.2 *Bildekvalitet*). Det finnes imidlertid ingen etablerte standardkriterier eller ”gullstandard” for bildekvalitet på MR, og heller ingen publiserte artikler som har benyttet visuell vurdering av hjerneavbildning (Ö. Smedby (professor i medisinsk radiologi og medisinsk bildevitenskap, Institutionen för medicin och hälsa, Linköpings universitet.), personlig e-post korrespondanse, 25. april 2011). Derfor er det i forbindelse med dette prosjektet utarbeidet et eget spørreskjema med en flertrinnskala for subjektiv visuell vurdering av bildekvalitet (Vedlegg 1). Sammen med en automatisk vurdering av bildene på samme måte som PRODE-studien planlegger å gjøre med pasientdata, vil en slik subjektiv vurdering være et supplement i evalueringen av bildekvalitet på tvers av tomografer. Mer om hvordan dette skal gjøres under punkt 3.6.1 *Uthenting av informasjon fra opptaks materialet*.

1.2.2 Bildekvalitet

Artefakter

Artefakter er forstyrrelser i form av bildesignaler som ikke representerer en sann gjengivelse av objektet (MR-TIP, 2011). Det er mange kilder til artefakter i et MR-bilde, men i store trekk kan artefakter relateres til fysiologiske faktorer som for eksempel pustebevegelse hos det avbildede individet, opptaks- og rekonstruksjonsmetoder, samt yteevnen til tomografens ulike komponenter. Endret yteevne kan skyldes skadde eller feilkalibrerte elektriske og elektroniske komponenter. Risikoen for teknisk svikt er alltid til stede, og kan oppstå plutselig eller over tid (Zhuo & Gullapalli, 2006). Nærvær av artefakter kan kvantifiseres både ved hjelp av objektive beregninger og subjektive vurderinger (ACR, 2005; Båth & Månsson, 2007; Båth, 2010).

Objektiv beregning av bildekvalitet

Ulike former for teknisk svikt forstyrrer bildekvaliteten i ulik grad, og kan derfor være vanskelig å oppdage uten fantomtesting (Zhuo & Gullapalli, 2006). For eksempel kan tomografens gjengivelse av geometriske størrelser, snittykkelse og snittposisjon kun beregnes ved hjelp av et fantom med kjente geometriske størrelser.

Det finnes ulike dataprogrammer for automatisk gjenkjennelse og måling av strukturer i hjernen. Et eksempel er NeuroQuant® (CorTechs Labs, San Diego, California, USA), et kommersielt segmenteringsprogram basert på FreeSurfer (Athina A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts, USA). Programmet er optimalisert for helautomatisk prosessering av ADNIs 3D T1-opptak på 1,5 Tesla (T). Ut fra 3D-serien beregner NeuroQuant® arealer og volumer av en rekke ulike hjernestrukturer, og sammenlikner disse volumene med normalverdier for pasientens alder (CorTechs Labs 2007; CorTechs Labs 2009).

Resultatet av segmenteringen er en såkalt generell morfometri-rapport med absolutte og relative størrelser oppgitt i henholdsvis cm^3 og prosentvis andel av totalt hjernevolum. Rapporten oppgir også prosentvis størrelsesforskjell mellom strukturer i henholdsvis venstre og høyre hjernehalvdel i forhold til deres gjennomsnitt. Opptakene rekonstrueres i tynne snitt hvor ulike vev og strukturer tildeles individuelle farger. På denne måten kan man bla gjennom hjerneopptaket og kontrollere segmenteringen visuelt (CorTechs Labs 2007).

Subjektiv vurdering av bildekvalitet

Når det gjelder vurdering av i hvilken grad artefakter eller komponentfeil har innvirkning på bildenes utseende, kan subjektive vurderinger være verdifulle. En slik form for kvalitativ vurdering må kvantifiseres for å kunne sammenlignes kvantitativt på tvers av tomografer (Båth & Månsson, 2007). Mer om hvordan dette skal gjøres under punkt 3.6.1.2 *Hjerneopptak* med underpunkter.

Bildekvalitet, i form av for eksempel gjengivelse av anatomi og nærvær av eventuelle artefakter, kan altså evalueres objektivt og subjektivt. En framgangsmåte for evaluering av gjengivelse av anatomi, kan for eksempel være testing med NeuroQuant®, eller systematisk visuell vurdering gjort av radiologer. Nærvær eller fravær av artefakter kan vurderes visuelt, eller testes med fantomopptak. I tillegg kan faste rutiner for fantomtesting avdekke andre underliggende problemer eller endringer over tid.

1.2.3 Ulike aspekter ved bildekvalitet på MR

De følgende underpunkter tar for seg ulike begreper innenfor bildekvalitet på MR. I den forbindelse nevnes noe om hvordan valg og innstillinger av parameter og sekvenser kan virke inn på bildekvalitet, samt årsaken til, og effekten av, ulike artefakter og generelle komponentfeil i de enkelte tilfellene. Denne oppgaven har imidlertid ikke plass til en detaljert forklaring av bildedannelsen i MR, og tar derfor utgangspunkt i at leseren innehar noe grunnleggende kunnskap om MR.

1.2.3.1 Signal/støy forhold og signalhomogenitet

Signal/støy forhold

Signal/støy forholdet (SNR, fra engelsk signal-to-noise ratio) beskriver forholdet mellom gjennomsnittlig signalintensitet og støynivå i MR-signalet. Støy er uønskede forstyrrelser i form av tilfeldig spredte frekvenser i tid og rom, og er definert som standardavviket til bakgrunnen ($SD_{bakgrunn}$, fra engelsk standard deviation). For å oppnå god kontrast- og detaljoppløsning i bildet bør bidraget fra støy være lavest mulig.

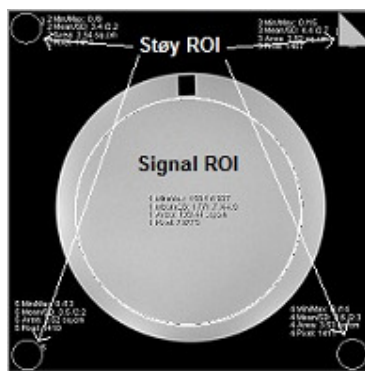
Generelt er SNR gitt av:

$$SNR = \frac{\text{mean}(ROI_{signal})}{SD(ROI_{bakgrunn})} \quad (1)$$

hvor ROI (interesseområde, fra engelsk region of interest) er et spesifikt måleområde for bestemmelse av SD og gjennomsnittlig (mean) pikselverdi, ROI_{signal} er måleområde for bildesignalet, og $ROI_{bakgrunn}$ er måleområde for støynivå (MR-TIP, 2011) (Figur 1). Underveis i bildedannelsen Fouriertransformeres MR-signalet fra tidsdomenet til frekvensdomenet. Man må imidlertid være klar over at Fouriertransformerte verdier ikke kan bli negative. Derfor er signalet utenfor objektet kunstig lavt, det har en såkalt Rayleigh-fordeling, i motsetning til signalet fra objektet som er Gaussisk fordelt. Dette må korrigeres for med faktor 0,665:

$$SNR = \frac{\text{mean}(ROI_{signal})}{SD(ROI_{bakgrunn})} \cdot 0,665 \quad (2)$$

(Haacke, Brown, Thompson & Venkatesan, 1999).



Figur 1: Plassering av ROI_{signal} og ROI_{bakgrunn} i fantombilde.

En del av støyen er pasientavhengig (for eksempel bevegelse), og noe er maskinavhengig (for eksempel magnetfelthomogenitet og antenntype). MR-systemet har alltid en viss grad av elektrisk bakgrunnsstøy, og denne varierer fra system til system (Westbrook, Kaut Roth & Talbot, 2005). SNR er en grunnleggende parameter med stor innvirkning på bildekvalitet, og bør kontrolleres regelmessig (McRobbie, Moore, Graves & Prince, 2007). SNR-målinger er sensitive, men uspesifikke. Lav SNR sier lite om hva som er galt med tomografen (Firbank, Harrison, Williams & Coulthard, 2000), fordi mange ulike parameter har innvirkning på signalet (Tabell I). Ved oppsett av sekvenser gjøres det alltid avveininger i forhold til ulike parametervalg og SNR. For mer om SNR, henvises til kapittel 5 og 11 i McRobbie et al. (2007).

Tabell I: Ulike parameters innvirkning på SNR og bildekvalitet (Sprawls, 2000).

Parametre som øker SNR	Virkning/konsekvens
↑ vokselstørrelse ⁽¹⁾ (↑ FOV ⁽²⁾ , ↑ snittykkelse, ↓ matrisestørrelse)	↓ romlig oppløsning
↑ antall målinger (NSA, fra engelsk number of signal averages)	↑ opptakstid (dobling av NSA ⇒ økt SNR med faktor $\sqrt{2}$)
↓ mottakerbåndbredde (rBW, fra engelsk receive bandwidth)	↑ kjemisk skift ↑ grad av bevegelsesartefakter
↑ repetisjonstid ⁽³⁾ (TR, fra engelsk repetition time)	↑ scantid ↓ grad av T1-veking
↓ ekkotid ⁽³⁾ (TE, fra engelsk echo time)	↓ grad av T2-veking
Flippvinkel 90° ⁽³⁾ (FA, fra engelsk flip angle)	FA bestemmer veking i en del sekvenser (generelt: ↓ FA ⇒ T2-veking, ↑ FA ⇒ T1-veking, finnes imidlertid en del unntak)

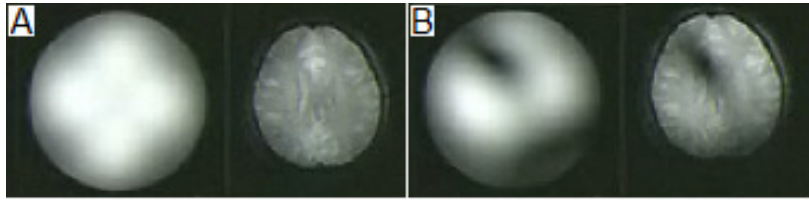
⁽¹⁾Voksel: tredimensjonal piksel (fra engelsk volume pixel). Vokselstørrelsen er bestemt av pikselstørrelsen samt snittykkelsen. ⁽²⁾FOV: field-of-view: størrelsen på bildeopptaksfeltet. ⁽³⁾TR, TE og FA har direkte innvirkning på vektingen i bildet (T1-, T2- og PD-veking), og bør derfor ikke endres, selv om det kan gi positiv effekt for SNR.

Signalhomogenitet

Bildesignalets ensartethet, signalhomogeniteten, har betydning for vurdering og gjengivelse av anatomi, vev og strukturer, samt eventuell patologi. Signalhomogenitet (U , fra engelsk uniformity) kan kvantifiseres ved å beregne prosentvis signalvariasjon i området hodespolen skal dekke (Ligning 3). U reflekterer tomografens evne til å generere en konstant signalrespons over et gitt volum fra et homogent fantom:

$$U = 1 - \frac{S_{\max} - S_{\min}}{S_{\max} + S_{\min}} \cdot 100 \quad (3)$$

hvor S_{\max} og S_{\min} er gjennomsnittlig pikselverdi i områdene med henholdsvis høyeste og laveste signalintensitet. Stor variasjon i signalintensitet innenfor området antennen skal dekke er uønsket fordi det kan etterligne eller skjule patologi (Figur 2). En slik variasjon kan for eksempel skyldes feil på antennen, gradientspolene, eller problemer i RF-kjeden (ACR, 2005).



Figur 2: A viser jevn signalintensitet i bildet, fantom til venstre og hjerne til høyre. B viser ujevn signalintensitet pga manglende antenneelement (Ibrahim, 2009). Ingen av antennetyperne i dette prosjektet tilsvarer den på figuren, dette er bare ment som en illustrasjon.

1.2.3.2 Lavkontrastdetaljer, kontrast/støy-forhold og romlig oppløsning

Lavkontrastdetaljer og kontrast/støy-forhold

Lavkontrastoppløsning er systemets evne til å skille områder med små kontrastforskjeller (McRobbie et al., 2007). Det er hovedsakelig kontrast/støy-forholdet (CNR, fra engelsk contrast-to-noise ratio) i bildet som bestemmer muligheten til å detektere lavkontrastdetaljer. CNR er forskjellen i SNR mellom to områder (Ligning 4), og styres derfor av mange av de samme parametrene som SNR (Westbrook et al., 2005). Selv med høy CNR og SNR, kan imidlertid andre forstyrrelser i bildet gjøre det vanskelig å detektere lavkontrastområder.

$$CNR_{AB} = \frac{mean(ROI_{signal_A}) - mean(ROI_{signal_B})}{SD(ROI_{bakgrunn})} = SNR_A - SNR_B \quad (4)$$

hvor $signal_A$ og $signal_B$ er signalintensiteten til vev A og B (Haacke et al., 1999).

Romlig oppløsning (detaljoppløsning)

Romlig oppløsning beskriver hvor små detaljer tomografen kan gjengi i et bilde, og er definert som den minste synlige avstanden mellom to punkter i et objekt (McRobbie et al., 2007). Dårlig romlig oppløsning medfører at strukturer “flyter sammen” (blurring) i stedet for å gjengis adskilt slik de faktisk ligger (Sprawls, 2000). Romlig oppløsning er gitt av vokselstørrelsen (Reimer et al., 2006), og står i direkte forhold til SNR (jf. Tabell I). Fordi signalgjengivelse og støybidrag varierer mellom tomografer, vil også romlig oppløsning være tomografavhengig.

1.2.3.3 Gradientlinearitet, snittykkelse og snittposisjon

Gradientlinearitet

MR-signalet oppstår som følge av at eksiterte hydrogenprotoner responderer på små endringer i et kraftig magnetfelt. En endring i magnetfeltstyrke mellom to punkter kalles en gradient, og er indusert av gradientspoler. Gradientlinearitet beskriver gradientspolenes evne til en jevn (lineær) magnetfeltendring (Westbrook et al., 2005), og er en grunnleggende spesifikasjon for en gradientspole (McRobbie et al., 2007).

Ikke-lineariteter gir geometriske forvrengninger (distorsjon) i magnetfeltet, og dermed feil i avlesningen og bildegjengivelsen av objektet eller strukturer i hjernen. Dette kan for eksempel medføre at størrelsen på strukturer eller områder med patologi ikke gjengis korrekt. Gradientspolene får strøm fra gradientforsterkere, og både spoler og forsterkere krever regelmessig ettersyn og justering fordi de blir mindre nøyaktige over tid. Kontroll av gradientsystemet inngår i det generelle vedlikeholdet som utføres av leverandøren hver 6. måned. Feilkalibrerte gradientspoler eller feil på gradientforsterkere kan imidlertid forekomme, og vil medføre feil i snittposisjon.

Snittykkelse og snittposisjon

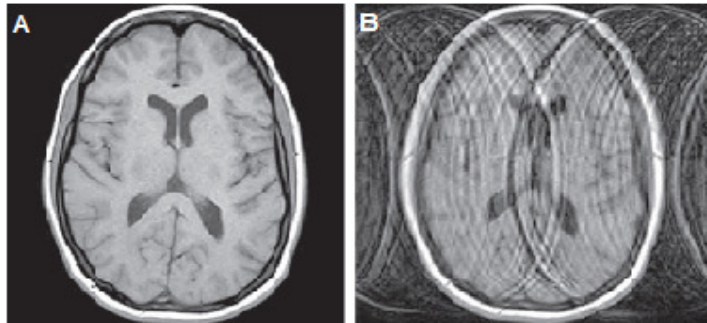
I et 2D-opptak er snittykkelse og snittposisjon bestemt av styrken på snittseleksjonsgradienten (ΔG_{ss}), båndbredden til radiofrekvenssignalet (RF-pulsen), og senterfrekvensen til RF-pulsen. ΔG_{ss} bestemmer hvilken båndbredde RF-pulsen må ha for å eksitere en viss snittykkelse. På samme måte er det ΔG_{ss} som avgjør hvor mye RF-pulsen må endres fra Larmor-frekvensen for å tilsvare angitt snittposisjon (Elster & Burdette, 2001). Dermed kan unøyaktig snittykkelse eller snittposisjon skyldes feil eller dårlig kalibrering av gradientsystemet eller RF-systemet. Hvilken av disse som forårsaker problemer kan man finne frem til ved feilsøking, men først må man bli klar over at det er et problem.

1.2.3.4 Ghosting og kjemisk skift

Ghosting

Ghosting er feilregistrering av signal som følge av fasevariasjoner mellom de ulike fasekodingstrinn. Det ser ut som en “skygge” eller dobbeltkontur av den opprinnelige

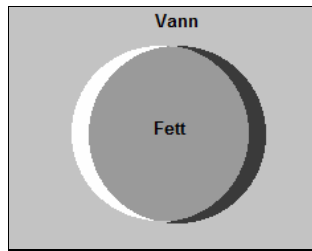
strukturen som gjentas gjennom hele bildet, hovedsakelig i faseretningen (Figur 3B). Det oppstår altså et misforhold mellom vokselposisjon under eksitasjon og signalavlesning, for eksempel som følge av bevegelse (pasientbevegelse eller vibrasjoner) (Haacke et al., 1999). Ghosting kan også skyldes en kabel eller kobling som ikke står riktig i kontakten, feil ved fasekodingsgradienten, mottakerspolen, eller problemer i RF-kjeden.



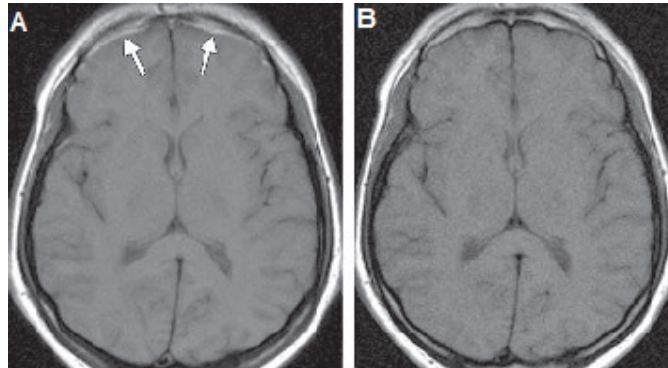
Figur 3: Eksempel på ghosting artefakt på grunn av pasientbevegelse. **A:** normalt bilde, **B:** bilde med ghosting (Morelli, Runge, Ai, Attenberger, Vu, Schmeets et al., 2011).

Kjemisk skift

Betegnelsen kjemisk skift kommer av pikselforskyvningen som oppstår som følge av hydrogenprotonenes resonansfrekvens i ulike vev. I kroppen er det hovedsakelig områder i overgangen mellom fett (lipider) og vann (H_2O) som skaper kjemisk skift artefakt (Figur 4 og 5). Larmorfrekvensen til lipider ligger omtrent 3,35 milliondeler (ppm, fra engelsk parts per million) lavere enn Larmor-frekvensen til vann, ved 1,5 T utgjør dette omtrent 214 Hz (Haacke et al., 1999). Årsaken til dette er forskjellen i molekylbindingen i fett og vann. Fett er store molekyler, og hydrogenprotonene som er bundet i fett er derfor omringet av en stor elektronsky som skjermer mot magnetfeltendringen som påføres under bildedannelsen. Dermed vil hydrogenprotonene som er bundet i fett reagere langsomt på endringer i magnetfeltet. I vann er situasjonen motsatt, vannmolekylene er små, elektronskyen er liten, og hydrogenprotonene i vann følger derfor magnetfeltendringer raskt. Disse ulike "reaksjonsmønstrene" gjør at signalet fra fett og vann feilplasseres i forhold til hverandre (Liang & Lauterbur, 2000).



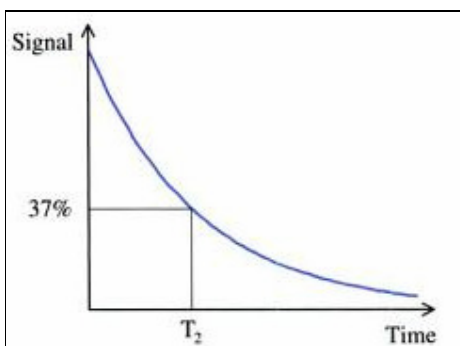
Figur 4: Skjematisk fremstilling av kjemisk skift artefakt. Fettsignalet er forskjøvet i forhold til vannsignalet på grunn av sin lavere resonansfrekvens i forhold til vann. Pikslene forskyves i bildet, og dermed vil signalet fra fett og vann nulles ut (høyre brem) eller forsterkes (venstre brem).



Figur 5: Kjemisk skift artefakt fra fettlaget i skalpen (A). Dette kan reduseres ved å øke båndbredden, men dette går på bekostning av SNR (B) (Morelli et al., 2011).

1.2.3.5 T₂-måling

T₂-relaksasjonstiden beskriver hvor hurtig MR-signalet henfaller til 37 % av sitt utgangspunkt (Figur 6). Henfallet i signal skyldes defasing av spinnene som følge av deres utveksling av magnetisering, samt inhomogeniteter i magnetfeltet. Ved T₂-måling er det førstnevnte effekt man er ute etter.



Figur 6: T₂ er punktet på tidskurven hvor signalet er 37 % av utgangspunktet (figur hentet fra McRobbie et al., 2007)

En turbo spinn-ekko (TSE)-sekvens egner seg til beregning av T₂-tid fordi ekkotoget fjerner effekten av inhomogeniteter i magnetfeltet. Dette skjer fordi en 180° RF-puls refokuserer spinnene en rekke ganger (7–16 i dette prosjektet) etter hverandre uten ny

eksitasjon. En slik refokusering gjør at den tilfeldige defasingen fra feltinhomogeniteter jevnes ut, mens den systematiske defasingen som skyldes biologiske faktorer gjentas og forsterkes. Dermed står man igjen med signalet fra defasingen som skyldes biologiske faktorer, altså vevets reelle T2-signal.

Opptaket må ha minimum to ulike ekkotider (TE) for å beregne T2-tid.

”Gullstandarden” for T2-måling med MR er en sekvens som inneholder flere enn to TE. Hver TE gir et punkt langs T2-kurven, og jo flere ekko, desto mer nøyaktig måling. Det er likevel ønskelig å gjøre pålitelige T2-målinger ved hjelp av dobbelekk-sekvenser, fordi slike sekvenser tar kort tid, dekker et stort anatomisk område og gir strukturell informasjon i tillegg til T2-informasjon (Bauer, Jara & Killiany, 2010). Derfor var det aktuelt å gjøre T2-målinger i dette prosjektet.

Ulike molekyler har ulike magnetiske egenskaper, noe som gir spinnene ulike utgangspunkt for utveksling av magnetisering. Derfor har ulike materialer og vevstyper ulik T2-tid, og et vevs T2-tider kan endres ved sykdom. For eksempel kan hjernesvinn og sykkelig lagring av jern være sykdomstegn og endrer T2-tidene i ulike retninger (Vymazal, Righini, Brooks, Canesi, Mariani, Leonardi et al., 1999). Estimert T2-verdi i henholdsvis hvit og grå normal hjernesubstans er 80 millisekunder (ms) og 100 ms (Haacke et al., 1999).

1.3 Tidligere forskning omkring testing av MR-data i multisenterstudier

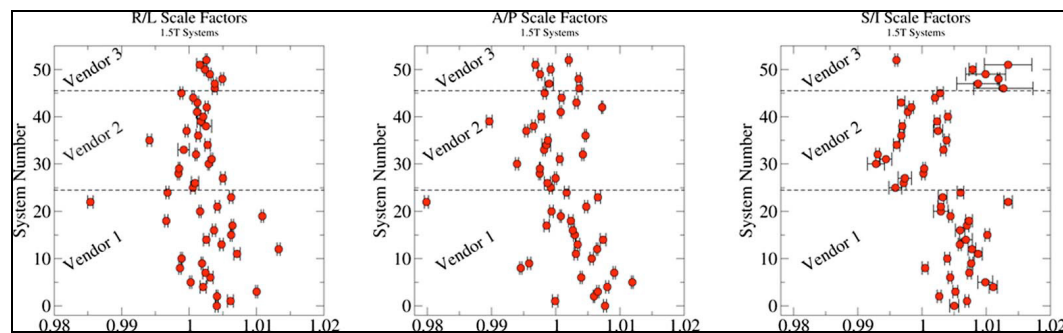
1.3.1 Bruk av fantomdata i multisenterstudier

ADNI-studien gjør opptak av et egenutviklet fantom etter hver pasient for å kontrollere SNR, CNR og gradientlinearitet. Alle sentrene har sitt eget fantom, men fantomdata analyseres sentralt ved hjelp av helautomatiske analysemetoder, og fantomene kontrolleres for ulikheter seg imellom (Jack, Bernstein, Fox, Thompson, Alexander, Harvey et al. 2008). ADNI inkluderer kun sentre med tomografer som innehar visse kriterier for yteevne, og derfor benyttes fantomopptakene hovedsakelig til feilsøking og kontroll av tomografenes yteevne og stabilitet over tid. Gunter et al. (2009) har analysert fantomdata fra ADNI-studien, samt tatt for seg praktiske utfordringer med kvantitative multisenterstudier med tomografer fra ulike leverandører. Sentral

overvåkning og kontroll av fantomopptakene medførte at studien mellom 2006 og 2008 påviste og rettet opp i komponentfeil som til sammen berørte over 25 % av ADNI-sentrene. I tillegg ble fantomopptak benyttet som en hjelp for å skille problemer med tomografen fra biologiske variasjoner og faktiske patologiske forandringer. Fant man noe unormalt i pasientbildene, gikk man tilbake i fantomdata og sjekket målingene. Noen av utfordringene ved multisenterstudier som ble påpekt i denne studien er, at til tross for at man har kontinuerlige fantomdata, kan det være vanskelig å vite nøyaktig hva som har blitt gjort med tomografen i en periode hvor man observerer endringer. Ulike sentre har ulik praksis angående prosedyrer for vedlikehold og registrering av avvik, og små endringer ved tomografens yteevne kan være vanskelig å oppdage visuelt.

1.3.2 Sammenligning av yteevne mellom ulike leverandører

ADNI-fantomet detekterer variasjon i gradientlinearitet i alle gradientretningene over tid. Gunter et al. (2009) har samlet data fra 2200 fantomopptak fordelt over 58 tomografer fra tre leverandører (Figur 7), og fant signifikante forskjeller i gradientenes yteevne hos ulike leverandører.



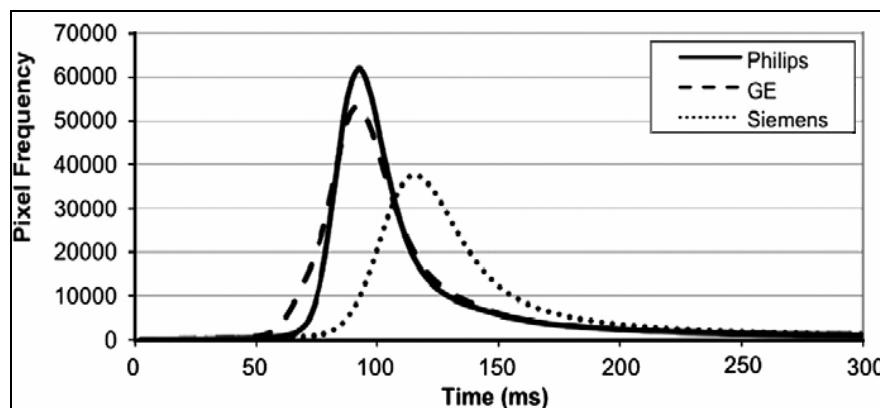
Figur 7: Pooled varians av gradientlinearitet med standardavvik blant 2200 fantomopptak fordelt over 58 tomografer fra tre leverandører. Gradientlinearitet langs x-aksen og tomografnummer langs y-aksen. Hvert diagram representerer en gradientretning. Innad for hver tomograf presterer gradientene relativt jevnt over tid, med unntak av S/I-retning (z-retning) blant leverandør 3-tomografene. Gjennomsnittlig variasjon er systematisk forskjellig hos ulike leverandører. Leverandørene som er testet er GE, Philips og Siemens, men det er ikke oppgitt hvem som Vendor 1, 2 og 3 (Gunter et al., 2009).

De nasjonale helseinstitutter i USA (National Institutes of Health, NIH) har samlet inn MR-bilder av hjernen til over 500 friske barn fra syv ulike senter (NIH Normal Pediatric Brain Development Database, NIHPD). De samme barna avbildes med jevne mellomrom over en 6-års periode. Kvalitetskontrollen foregår ved hjelp av fantomopptak med ACR-fantomet minst én gang pr. måned. En studie av Fu, Fonov,

Pike, Evans og Collins (2006) har undersøkt variasjonen mellom tomografer når det gjelder SNR, signalhomogenitet og geometrisk nøyaktighet ved hjelp av 1048 fantomopptak med T1-vektet 3D-volum. De fant statistisk signifikante forskjeller på tvers av senter, men ett senter skilte seg ut i alle testene og økte graden av variasjon. Endringer over tid ble fulgt ved hvert senter. Endringene var små, men statistisk signifikante. Studien sier imidlertid ikke noe om muligheten for at fantomene er ulike, hvordan de valgte ut deltakersentre, eller hvordan bildene påvirkes av resultatet fra fantomopptakene.

1.3.3 Kvalitetskontroll og reproduksjon av funn

Bauer, Jara og Killiany (2010) har undersøkt mulige leverandøravhengige systematiske ulikheter i kvantitativ T2-måling ved hjelp av ACR-fantom (fire tomografer) og normale kontroller fra ADNI-studien (18 tomografer). Det ble vist at T2-målinger på bakgrunn av TSE-sekvenser varierer signifikant mellom leverandører (Figur 8), både ved måling på fantom og individer. Forfatterne sier imidlertid at resultatet bør etterprøves med et større datamateriale.



Figur 8: Gjennomsnittlig T2-tid beregnet ved hjelp av en dobbelekk TSE-sekvens i 32 normale hjerner fordelt på tre leverandører. Kurvene er laget i samsvar med antall modeller og individer hos hver leverandør (Philips: 6 kvinner, 3 menn. GE: 5 kvinner, 10 menn. Siemens: 4 kvinner, 4 menn) (Bauer et al., 2010).

Ewers, Teipel, Dietrich, Schonberg, Jessen, Heun et al. (2006) har i en multisenterstudie undersøkt om tomografers yteevne korrelerer med muligheten til å gjøre korrekte volum-målinger i hjernen. Yteevnen til 11 1,5 T tomografer ble testet ved hjelp av ACR-fantomet, hvorav to tomografer ikke møtte ACR-kriteriene for snittposisjon, snittykkelse, lavkontrastdetaljer og romlig oppløsning. Ett opptak ved hvert senter ble

ansett som representativt for tomografenes yteevne, til tross for at andre studier har vist at målinger kan variere over tid. T1-vektede volumopptak ble gjort av en frisk frivillig og minimum 20 pasienter med lett kognitiv svikt eller Alzheimers sykdom ved seks av sentrene, inkludert dem som ikke oppfylte alle ACR-kriteriene. De ulike sentrene brukte noe ulike opptaksparameter. I frivilligopptakene varierte både manuelle målinger av hippocampusvolum og automatiske målinger av gråsubstansvolum med en variasjonskoeffisient på mindre enn 5 % mellom sentrene. Variasjonen var moderat, men større enn man kunne forvente ved gjentatte målinger ved ett senter, og ble derfor kalt multisentereffekten. I pasientopptakene var imidlertid ikke multisentereffekten systematisk til stede. Selv om målinger varierte mellom sentre, så det i denne studien ut til at faktiske biologiske variasjoner mellom pasientgrupper var større. Forfatterne anbefaler bruk av ACR-kriteriene for å vurdere yteevne mellom sentre. De kriteriene noen senter feilet på i denne studien, hadde imidlertid tilsynelatende ingen betydning for muligheten til å gjøre korrekte mål av hippocampus- og gråsubstansvolum i pasienter.

1.3.4 Testede aspekter ved bildekvalitet i ulike studier

Til tross for at flere studier har testet bildekvalitet, både ved hjelp av fantomopptak, frisk frivillig og pasienter, er det vanskelig å sammenligne funn i ulike studier på grunn av ulike målemetoder og ulikt forskningsdesign. Ewers et al. (2006) fikk signifikante avvik i målt snittposisjon, snittykkelse, lavkontrastdetaljer og romlig oppløsning på to av 11 tomografer ved opptak av ACR-fantomet på 1,5 T, uten at det hadde innvirkning på bildekvaliteten på pasienter. Jack, Slomkowski, Gracon, Hoover, Felmlee, Stewart et al. (2003) har gjort opptak av pasienter med Alzheimers sykdom over tid, og rapporterte at geometriske forvrengninger og lav SNR var de vanligste problemene med kvalitet.

Flere studier har brukt fantomopptak for å teste yteevnen til ulike tomografer. Det er imidlertid få som sammenligner yteevne med tanke på å få sammenlignbare bilder på tvers av utstyr. Chen, Wan, Wai og Liu (2004) har testet fire tomografer med ulik feltstyrke fra ulike leverandører ved hjelp av kriteriene fra ACR. To av fire tomografer besto alle testene. Etter å ha gjennomført feilsøk og foretatt justeringer, besto tre av fire tomografer alle testene. Geometrisk nøyaktighet, signalhomogenitet og snittykkelse var det som feilet oftest. En studie av Firbank et al. (2000), testet SNR, signalhomogenitet, snittykkelse, snittposisjon, ghosting, geometrisk nøyaktighet og romlig oppløsning ved

hjelp av ulike fantomer på en 1,0 T tomograf i løpet av ett år, og fant at alt unntatt SNR var stabilt mens målingene pågikk. En studie av Ihalainen, Sipila og Savolainen (2004) testet signalhomogenitet, ghosting, SNR, geometrisk nøyaktighet, snittykkelse, snittposisjon, romlig oppløsning og T1- og T2-nøyaktighet på fire 1,5 T tomografer. Fantomopptakene ble analysert automatisk, og forfatterne konkluderte med at fantomopptak med spesialiserte geometriske fantom avdekket flere problemer enn leverandørens egne rutiner for vedlikehold hadde gjort.

McRobbie og Quest (2002) testet 17 nyinstallerte MR-systemer med ulike feltstyrker i perioden 1994–2001. Studien brukte egenutviklede fantomer, egne protokoller for testing og egne kriterier for godkjenning. De testet geometrisk nøyaktighet, romlig oppløsning, snittposisjon, snittykkelse, SNR, signalhomogenitet og ghosting. Det ble avdekket og utbedret feil i 49 % av systemene. Geometrisk nøyaktighet og SNR var de kriteriene som feilet oftest, signalhomogenitet og ghosting feilet nest oftest. Resten av parametrene feilet sjelden.

Forfatterne påpekte også at deres kvalitetskontroll av nyinstallerte MR-systemer viste at færre systemer innfridde kriteriene de siste årene testingen pågikk sammenlignet med de første årene. Studien nevner imidlertid ikke noe om fantomstabilitet over årene testene foregikk, eller om det ble laget flere testfantom.

2 Hensikt og problemstilling

Optimaliseringen av PRODE MR-protokollen og standardiseringen av parametre ble klart i løpet av høsten 2010, og parametrene er lagt inn ved de ulike deltakersentrene. Når ulike MR-tomografer skal benyttes i innsamlingen av strukturelle hjernedata, er det vesentlig å sikre at tomografene avbilder og gjengir det samme. Hver enkelt tomograf må betraktes på linje med andre måleinstrumenter. Yteevnen, som gjenspeiles i bildekvaliteten, må testes og sammenlignes for å sikre at data samlet inn ulike steder er sammenlignbar.

Man bør være oppmerksom på eventuelle sammenhenger mellom tomograf og bildekvalitet, og deres betydning ved planlegging, gjennomføring og oppfølging av multisenterstudier. I og med at det ikke er utarbeidet noen generelle kriterier som brukes som grunnlag i vurdering av bildekvalitet på MR, ønsket man gjennom dette prosjektet å gå bredt ut og samle data som både favner objektive og subjektive aspekter omkring bildekvalitet. Det betyr ikke at alt som er gjort i forbindelse med dette prosjektet er hensiktsmessig eller praktisk gjennomførbart i andre sammenhenger. Håpet er at prosjektets systematiske tilnærming gjør at det innsamlede materialet kan være til hjelp for andre som skal i gang med multisenterstudier med MR-avbildning. I tillegg kan en slik tilnærming kanskje avdekke områder oppmerksomheten bør rettes mot i innføringen av generelle rutiner for kontinuerlig evaluering av bildekvalitet ved multisenterstudier.

Overordnet hensikt i dette prosjektet er å besvare om standardiserte opptaksprotokoller gir sammenlignbar bildekvalitet fra ulike tomografer, uavhengig av leverandør og modell. Ved hjelp av objektive beregninger og subjektive vurderinger vil man:

- identifisere individuelle ulikheter i bildekvalitet på grunnlag av en rekke kvantitative og kvalitative kvalitetsaspekter i opptak av fantom og frisk frivillig
- teste samvariasjoner mellom
 - bildekvalitet og tomografenes tekniske og utstyrmessige spesifikasjoner
 - objektive målinger i fantomopptak og hjerneopptak
 - objektive beregninger og subjektive vurderinger av ulike kvalitetsaspekter

3 Metode

3.1 Valg av forskningsdesign

Datainnsamling og analyse er i dette prosjektet foretatt gjennom en kvantitativ, ikke-eksperimentell tilnærming. Det ble også benyttet en viss grad av blinding i deler av prosjektet (Polit & Beck, 2010). Prosjektet er først og fremst en tverrsnittstudie av bildekvaliteten i MR-opptak på et bestemt tidspunkt. Tomografene ble fulgt opp med fantomopptak etter 4–6 måneder for å kontrollere stabiliteten i fantommålingene.

Evalueringen av opptaks materialet foregikk både ved hjelp av objektive og subjektive metoder. I objektive metoder inngikk både automatiske og manuelle metoder. Alle evalueringsmetodene ga kvantitative verdier, men på ulike målenivåer. Noen aspekter ble målt på en kontinuerlig tallskala, andre vurdert på en ordinalskala. Bruken av ulike evalueringsmetoder er en form for metodetriangulering med formål å kartlegge i hvilken grad metodene evnet å fange opp ulike aspekter omkring bildekvalitet.

3.2 Utvalg og utvalgskriterier

Utvalget var tomografene ved de ulike MR-klinikkene som bidrar til datainnsamlingen i PRODE MR. En oversikt over leverandører, modeller og utstyr er presentert i Tabell II. Det var til sammen åtte tomografer, fra tre ulike leverandører. Tomografene ble valgt på bakgrunn av at sin nærhet til en av de alderspsykiatriske avdelingene som henviser pasienter til PRODE studien. Det eneste inklusjonskriteriet utenom dette, var en magnetfeltstyrke på 1,5 T.

Tabell II: Tomograftype, årsmodell, antenntype og gradientsystem ved hvert senter.

Senter-nummer	Leverandør og modell	Årsmodell og versjon	Antenne-type	Gradientsystem, amplitude ⁽²⁾ , slew rate ⁽³⁾ , rise time ⁽⁴⁾
Senter 1	GE Signa	2000	8-kanalers hodespole	HDx
	General Electric Healthcare, Chalfont St. Giles, Storbritannia	Oppgradert til HDx 2007		33 mT/m 120 T/m/s 275 μ s
Senter 2	Philips Intera	2001	6-kanalers hodespole	Pulsar HP
	Philips Medical Systems, Best, Nederland	Oppgradert til R2.6.3 2010		33 mT/m 180 T/m/s 180 μ s
Senter 3	Siemens Avanto	2006	12-kanalers hodespole	Q-engine
	Siemens, Erlangen, Tyskland	Versjon B17		33 mT/m 125 T/m/s 264 μ s
Senter 4	Siemens Avanto	2008	12-kanalers hodespole	Q-engine
	Siemens, Erlangen, Tyskland	Versjon B17		33 mT/m 125 T/m/s 264 μ s
Senter 5	Siemens Symphony Vision	2003	8-kanalers hodespole	Quantum
	Siemens, Erlangen, Tyskland	Versjon A25		30 mT/m 125 T/m/s 240 μ s
Senter 6	Siemens Avanto	2006	12-kanalers hodespole	SQ-engine
	Siemens, Erlangen, Tyskland	Versjon B17		45 mT/m 200 T/m/s 200 μ s
Senter 7	Siemens Avanto	2007	12-kanalers hodespole	SQ-engine
	Siemens, Erlangen, Tyskland	Versjon B17		45 mT/m 200 T/m/s 200 μ s
Senter 8	Siemens Symphony	2002	CP ⁽¹⁾ hodespole	Sprint
	Siemens, Erlangen, Tyskland	Versjon A30A		30 mT/m 75 T/m/s 400 μ s

⁽¹⁾CP: sirkulært polarisert (fra engelsk circular polarized). ⁽²⁾Gradientamplitude er utslaget til gradientfeltet i positiv eller negativ retning. Den angir spillerommet i magnetfeltendring til tomografen. ⁽³⁾Slew rate (stighastighet) og ⁽⁴⁾rise time (stigetid) er stigningen til gradienten fra null til maksimum angitt i henholdsvis magnetfeltstyrke og tid.²

² En CP spole har to elementer som mottar signal 90° faseforskjøvet i forhold til hverandre (senter 8). Senter 1–7 har såkalte phased-array spoler, dette er en antenne hvor flere antenne-elementer (kanaler) arbeider sammen i signalmottak og billedannelse. Hovedtanken bak ulike antenntyper er at flere elementer gir bedre signal. Hovedtanken bak gradientstyrke er at kraftige og raske gradienter (høy amplitude og slew rate) øker romlig oppløsning og senker optakstiden. Rise time = amplitude/slew rate.

3.3 Testobjekter og variabler

Testobjekter

Standardiserte tester ved hjelp av opptak av spesialiserte fantom er et godt utgangspunkt for sammenligning av objektiv bildekvalitet på tvers av tomografer. ACR-fantomet ble valgt som geometrisk fantom i dette prosjektet. Som tidligere nevnt benyttes ACR-fantomet som en del av sertifiseringssystemet for MR-tomografer i USA. Det er derfor velprøvd, og mye brukt også i andre sammenhenger enn sertifisering (jf. punkt 1.3 *Tidligere forskning omkring kvalitetssikring av MR-data i multisenterstudier* med underpunkter). ACR har gitt ut en veiledning for fantomopptak. Denne inneholder informasjon om fantomet og dets innhold, i tillegg til en grundig rettleiding og grenseverdier for akseptabel bildekvalitet (ACR, 2005). I dette prosjektet ble to ACR-fantom av typen "Large MRI accreditation phantom" benyttet i datainnsamlingen. Fantomene har serienummer J8915 og J9834, og er levert av JM Specialty Parts Inc. (San Diego, California, USA). Det ble gjort opptak med begge fantomene samme dag ved ett senter for å avdekke eventuelle forskjeller mellom disse to.

Som tidligere nevnt, er det visse begrensninger i hva man kan evaluere hvis man kun undersøker tomografens evne til å avbilde fantomer. For å få et mer realistisk inntrykk av tomografens evne til å gjengi anatomi og vevskontrast, må man gjøre hjerneopptak. Materiale som forandres over tid (sykdom), eller ulikt materiale (flere individer), ville imidlertid ikke gitt et tilfredsstillende sammenligningsgrunnlag for dette prosjektet. Det ble derfor gjennomført systematiske hjerneavbildninger ved alle sentrene ved hjelp av en frisk frivillig. Mer om dette under punkt 3.7 *Etiske betraktninger*.

Variabler

ACR sin veiledning for fantomopptak foreslår måling av følgende syv avhengige variabler: geometrisk nøyaktighet, romlig oppløsning, snittykkelse, snittposisjon, signalhomogenitet, grad av ghosting og lavkontrastdetaljer (ACR, 2005). I tillegg ble fantomopptakene brukt til beregning av SNR, CNR og T2-tid.

Hjerneopptakene ble brukt til beregning av SNR, CNR, T2-tider, ghosting, samt størrelse på ulike strukturer i hjernen. Alle de nevnte variablene ble testet mot de

uavhengige variablene gradientamplitude, slew rate og type mottakerspole. I tillegg ble de avhengige variablene testet mot hverandre for å se etter sammenhenger mellom ulike aspekter.

Bildekvaliteten på hjerneopptakene ble også vurdert visuelt, mer om dette under punkt 3.6.1.2 Hjerneopptak.

3.4 Optimalisering og standardisering

Oppsettet av grunnsekvensene ble gjort i samarbeid med radiograf Heidi Engebretsen ved Stavanger Universitetssykehus (SUS) i perioden 1. september–1. desember 2010. Årsaken til dette er at vi har erfaring med ulike leverandører, noe som forenklet standardiseringsprosessen. Protokollen består av standard kliniske sekvenser som blant annet brukes i forbindelse med utredning av demens (Reimer et al., 2006), i tillegg til ADNI-sekvensene. Grunnprotokollen i PRODE MR er som følger:

1. Sagittal T1-vektet 3D-volum ("ADNI-volum")
2. Transversal dobbelekk TSE-sekvens (basert på ADNIs dobbelekk-sekvens)
3. Transversal T2 FLAIR-sekvens (fluid attenuated inversion recovery)
4. Transversal T2*-vektet sekvens ("hemosekvens")
5. Transversal diffusjonsavbildning (diffusion weighted imaging, DWI)

Serie 1 er den samme som benyttes i ADNI-studien, og fullstendig parameteroversikt til serie 1 for ulike tomografer og modeller er tilgjengelig fra ADNIs hjemmeside (ADNI, 2011). En oversikt over de viktigste³ parametrene i serie 2–4 er vedlagt (Vedlegg 4).

Som tidligere nevnt, ser PRODE MR hovedsakelig etter morfologiske (strukturelle) endringer i hjernen hos eldre med depresjon, og dette er utgangspunktet for grunnsekvensene. Protokollen skal fungere som en klinisk undersøkelse, men skal i tillegg gi god og stabil nok bildekvalitet til å kunne analyseres kvantitativt i forbindelse med PRODE-studien. Kravene til bildekvalitet må imidlertid alltid veies opp mot opptakstiden. De parametrene som styrer bildekvalitet, styrer også opptakstiden, og en

³Alle parametrene tilpasset hver leverandør og modell tar for stor plass til å legges ved i sin helhet. Vedlegg 4 er ment å gi en generell oversikt over opptaksparametre, noe som er godt nok siden det er generell yteevne og ikke parametervalg som har fokus i dette prosjektet.

optimalisert MR-protokoll har et balansert forhold mellom bildekvalitet og total opptakstid (Sprawls, 2000). Dersom opptaket tar for lang tid, kan pasientene for eksempel få problemer med å ligge rolig, noe som igjen medfører redusert bildekvalitet. Opptakstiden avhenger imidlertid ikke bare av hva pasienten klarer å gjennomføre. Også generell kapasitet på MR, samt hvor mange pasienter man ønsker å inkludere i forhold til hvor lenge studien skal vare, kan ha innvirkning på lengden på opptaksprotokollen. I multisenterstudier kan undersøkelsens varighet også ha betydning for hvor mange senter som velger å delta i datainnsamlingen.

Man planlegger at dataanalysen av datamaterialet fra PRODE MR skal skje ved automatisk sammenligning av bildesignal voksel for voksel. Siden NeuroQuant® allerede er optimalisert mot prosessering av ADNIs 3D T1-opptak på 1,5 T, var hovedfokus på serie 2, 3 og 4. Målet var at disse sekvensene skulle ha lik vokselstørrelse ved alle sentrene, med godt nok SNR innenfor en rimelig opptakstid (~ 30 minutter sammenlagt). I tillegg må opptakene være uten manipulasjon i form av for eksempel interpolering eller lignende for å sikre sammenlignbarhet. Parametrene som styrer vokselstørrelse er snittykkelse, FOV og matrisestørrelse (jf. Tabell I). Disse størrelsene ble bestemt med utgangspunkt i ADNIs dobbelekkosekvens. Et annet viktig fokus var vektingen i bildet. Den er i hovedsak styrt av TR, TE og FA. Disse parametrene krever en viss grad av tilpasning til hver enkelt leverandør og modell for å få likt resultat. Deltakersentrene ble bedt om å fylle ut et skjema med generell informasjon om tomografen (Vedlegg 2 og Tabell II). Ut fra denne informasjonen ble det satt opp forslag til parameter man antok ville være innenfor hva utstyret ved de enkelte tomografene ville akseptere. Kort oppsummert ble alle sekvensene satt opp med mest mulig sammenlignbare parametre, med noen leverandørspesifikke variasjoner. To pasienter gjennomgikk opptak ved SUS og St. Olavs Hospital før den ferdige protokollen ble distribuert i januar 2011. Lokale nevroradiologer ble konferert underveis for bekreftelse på at visuell bildekvalitet var som ønsket.

3.5 *Datainnsamling*

3.5.1 **Bildeopptak**

Første opptak med ACR-fantom og frisk frivillig ble gjennomført ved alle sentrene i perioden 12. januar–14. juni 2011. I denne runden ble fantom- og snittplassering gjort av samme personer (H. Engebretsen og I. Nakken) i alle opptakene. I perioden 30. juni–15. oktober 2011 ble fantomet sendt pr. post rundt til sentrene for nye opptak. De samme sekvensene ble brukt, men fantom- og snittplassering ble nå utført av radiografer ved hvert senter som hadde vært til stede under første opptak. Det ble derfor utarbeidet en oppskrift som fulgte fantomet (Vedlegg 3). De to fantomopptakene ble foretatt med 4–6 måneders mellomrom ved hver tomograf. På denne måten kunne man følge opp tomografenes generelle yteevne over tid, samt få kontrollert resultatene fra første opptak. Ved ett av sentrene (senter 6) ble det gjort flere opptak på samme dag. Dette var for å kontrollere stabiliteten og reproduserbarheten til målinger i forhold til selve opptaksprosessen, samt reposisjonering i hodespolen. Mer om dette under punkt 3.6.2.2 *Beregning av måleusikkerhet*.

Følgende opptak ble gjort:

ACR-fantom (angitt vekt 100 kg):

1. Sagittal T1-vektet spinn-ekko (SE) oversikt
2. Transversal T1-vektet SE-sekvens
3. Transversal dobbelekk SE-sekvens
4. Transversal dobbelekk TSE-sekvens

Frisk frivillig (angitt vekt 70 kg):

Opptak 1–4 i grunnprotokollen i PRODE MR (se side 23).

Opptaksparameter for fantomopptak 1–3 ble hentet fra rettleddningen fra ACR (ACR, 2002) (Vedlegg 4). Serie 4 ble kjørt som tillegg fordi det gir muligheten til å kontrollere TSE i forhold til SE. SE gir sikrere beregninger i fantomet, men TSE er mer

sammenlignbar med hjerneopptakene. Opptaksparameter for frivilligopptakene var de samme som parametrene i PRODE MR (Vedlegg 4).

Digital måler for temperatur og luftfuktighet fulgte med fantomet og ble plassert i maskinrommet mens opptakene foregikk. Alle opptakene ble brent på CD og tatt med til Trondheim for videre uthenting av data og analyse. I Trondheim ble også tomograf-spesifikk informasjon fjernet fra frivilligopptakene før disse ble distribuert til observatørene som skulle utføre visuell vurdering. Tomografinformasjonen ble erstattet med senternummer, for sikkerhets skyld et annet senternummer enn det som er brukt i resten av prosjektet, før bildene ble distribuert til observatørene.

3.6 Analyse

Dataanalysen ble gjennomført i to omganger. I første omgang ble informasjon hentet ut fra opptaksmaterialet, blant annet ved hjelp av manuelle målinger av strukturer og plassering av ROI i bildene. Denne informasjonen ble lagret i JPEG-format (Joint Photographic Experts Group) på en minnepinne (SanDisc Cruzer, 16 gigabyte kapasitet). Deretter ble målingene overført manuelt til Microsoft Excel, hvor utregninger og sortering ble gjort. Excel-filene ble skrevet ut på papir og sammenlignet med JPEG-filene for å kontrollere for eventuelle feil i overføringen av målingene. Også svarverdiene fra visuell vurdering ble overført manuelt til Excel, og overføringen ble kontrollert på samme måte.

Andre omgang av dataanalysen omfattet videre sammenligning av data og hypotesetesting. Dette ble gjort i Statistical Package for the Social Sciences (SPSS). En mer detaljert oversikt over programvaren som ble benyttet i analysens ulike trinn er presentert i Tabell III.

Tabell III: Oversikt over programvare brukt i prosjektet.

Programvarenavn	Versjon og utgivelsesår	Utgiver	Brukt til
Siemens Avanto Syngo MR B17	Versjon 9.0 Utgitt juni 2009	Siemens, Erlangen, Tyskland	Uthenting av informasjon fra opptaksmaterialet (lengdemål, tegning av ROI, osv.)
nordicICE ⁽¹⁾	Versjon 2.3.12 Utgitt mai 2011	NordicImagingLab AS, Bergen, Norge	T2-måling
NeuroQuant®	Versjon 1.0.1. Utgitt oktober 2007	CorTechs Labs, San Diego, California, USA	Volumberegninger fra 3D T1-sekvens
Microsoft Excel	2002, Service Pack 3 Utgitt 2002	Microsoft Corp., Redmond, Washington, USA	Beregninger og lagring av informasjon fra opptaksmaterialet
IBM® ⁽²⁾ SPSS®	Statistics 19.0 Utgitt august 2010	IBM, Armonk, New York, USA	Dataanalyse

⁽¹⁾ICE: Image control and evaluation. ⁽²⁾IBM: International Business Machines.

3.6.1 Uthenting av informasjon fra opptaksmaterialet

3.6.1.1 Fantomopptak

De to aspektene lavkontrastdetaljer og romlig oppløsning er lagt opp som en visuell vurdering med opptelling av antall synlige strukturer i bildene. Dette kan ikke sammenlignes med visuell vurdering av hjerneopptak, fordi ACR gir helt klare instruksjoner på hvordan fantombildene skal vurderes og hva som aksepteres. Resten av ACR-testene ble gjort ved hjelp av mål og beregninger som beskrevet i teoridelen og i veiledningen fra ACR. Tabell IV gir en oversikt over aspektene som kan testes med ACR-fantomet, samt grenseverdier for akseptabel bildekvalitet. I tillegg til de syv testene foreslått av ACR ble SNR, CNR og T2-verdier målt, se beskrivelse under punkt 3.6.1.3 *Beregning av SNR, CNR og T2-verdier.*

De manuelle målingene ble gjort to ganger av samme radiograf (I. Nakken) for hvert opptak ved hvert senter. 95 % konfidensintervall (KI_{95}) omkring de to målingene ble beregnet for å få bedre oversikt over manuell måleusikkerhet. Tosidig t-fordelingstabell ble brukt til beregning av KI_{95} , da denne tar hensyn til antall målinger i beregningen.

Tabell IV: Kvalitetsaspekter testet med ACR-fantomet. ACR sine grenseverdier for akseptabel bildekvalitet, og hvilke sekvenser grenseverdiene gjelder for er brukt i prosjektet.

Kvalitetsaspekt ACR-fantom	Minimumskrav til tomografen	Sekvenser som testes
Signalhomogenitet	$\geq 87,5 \%$	Transversal T1 og T2 SE
Lavkontrastdetaljer	> 9 punkter (maksimum 40)	Transversal T1, T2 SE og T2 TSE
Romlig oppløsning	$\leq 1,0$ mm	Transversal T1 og T2 SE
Geometrisk nøyaktighet	$\leq \pm 2$ mm	Sagittal og transversal T1
Snittykkelse	$\leq \pm 0,7$ mm	Transversal T1 og T2 SE
Snittposisjon	< 4 mm avvik	Transversal T1 og T2 SE
Grad av ghosting	$< 2,5 \%$	Transversal T1
SNR	ikke spesifisert	ikke spesifisert
CNR	ikke spesifisert	ikke spesifisert
T2-verdi	ikke spesifisert	ikke spesifisert

3.6.1.2 Hjerneopptak

SNR, ghosting og T2-tid ble beregnet med tilnærmet samme metode som i ACR-opptakene. I tillegg gjorde tre radiologer en kvalitativ vurdering av bildekvalitet uavhengig av hverandre. Ved bruk av frisk frivillig i stedet for pasienter, mister man muligheten til å vurdere synligheten av ulike former for patologi. Muligheten til å detektere patologi korrelerer imidlertid med gjengivelsen av anatomi. Aktuelle aspekter var derfor generell synlighet og gjenkjennelse av ulike vev og strukturer, nærvær/fravær av støy og ulike artefakter, samt bildekvalitet som helhet. For å kvantifisere slike subjektive vurderinger, kan man, i følge Båth (2010), for eksempel be observatøren identifisere noen utvalgte og relevante anatomiske strukturer, og gradere synligheten av disse på en flertrinnskala.

Utarbeidelse av spørreskjema

Som tidligere nevnt, var bakgrunnen for å lage et eget spørreskjema til dette prosjektet at det ikke fantes noen velprøvde skjema fra før, verken norske eller utenlandske. Andre studier har imidlertid tatt i bruk spørreskjema som tar opp lignende tema andre steder i kroppen, deriblant prostata (Wagner, Rief, Busch, Scheurig, Taupitz, Hamm et al., 2010).

Formuleringen av noen generelle bildekvalitetskriterier på MR av hjerne, ble utarbeidet i samarbeid med en nevroradiolog ved St. Olavs Hospital. Kriteriene kunne deles inn i tre hovedgrupper av spørsmål. Første gruppe omhandlet differensiering av ulike vev og

strukturer, andre gruppe omhandlet artefakter. Tredje gruppe omhandlet bildekvaliteten som helhet, altså helhetsinntrykket av den aktuelle sekvensen. I denne gruppen ble spørsmål som har betydning for bildekvalitet, men som ikke kunne kategoriseres under differensierbarhet eller artefakter, tatt opp. Her kom blant annet SNR inn.

Det ble lagt opp til korte og konsise spørsmål, samt en nøytral spørsmålsstilling. I følge Halvorsen (1996) bør setningsbyggingen være enkel, og ikke mer enn 20 ord pr. spørsmål. Hvert spørsmål var på 10–15 ord, og for hvert spørsmål skulle ett aspekt ved bildekvalitet vurderes på en skala 1–5, hvor 1 var dårligst og 5 var best. Det var ingen alternativ for ”vet ikke”, men det var et åpent spørsmål til slutt med plass for kommentarer. Fempunkts svars skala er mye brukt i brukerundersøkelser (Nasjonalt kunnskapssenter for helsetjenesten, 2009). En større skala kan gjøre det vanskeligere å definere hvert trinn på skalaen. I tillegg kan flere valgmuligheter gi respondenten større problemer med å bestemme seg, noe som kan redusere reproduserbarheten ved gjentatte vurderinger (Fayers & Machin, 2007).

Spørreskjemaet skulle gi et mest mulig nyansert bilde av observatørens inntrykk, med minst mulig rom for misforståelser. Det var nødvendig å benytte fagtermer i spørsmålene, og til tross for at ordene vil være kjent for de fleste radiologer, ble det laget en ordliste for å unngå eventuelle feiltolkninger. Også betydningen i hvert trinn på svars skalaen var nøye definert. Spørreskjemaet med forklaringer og definisjoner er vedlagt i sin helhet som Vedlegg 1.

Det ble lagt opp til blinding i forbindelse med besvarelsen av spørreskjemaet. Radiologene fikk ingen opplysninger om tomografene hvor opptakene var gjort. Dette var for å unngå systematiske feil, for eksempel som følge av at besvarelsene ble farget av observatørens kjennskap til tomografene fra andre sammenhenger eller lignende.

Volumetri

3D T1-opptakene ble overført til NeuroQuant®, analysen tok ~10 minutter pr. opptak. NeuroQuant® beregnet volumer (volumetri) av hjernestrukturer, samt hjernen som helhet. Følgende strukturer ble målt: lillehjernen (cerebellum), grå hjernesubstans i hjernebarken (cortical gray matter), regioner av grå hjernesubstans dypt i hjernen (nucleus caudatus, putamen, globus pallidus og thalamus), samt deler av det limbiske

system (hippocampus og amygdala). Både målte størrelser (cm^3) og relative størrelser (% av totalt hjernevolum) ble sammenlignet. Alle de nevnte strukturene var aktuelle å sammenligne i dette prosjektet.

3.6.1.3 Beregning av SNR, CNR og T2-verdier

SNR

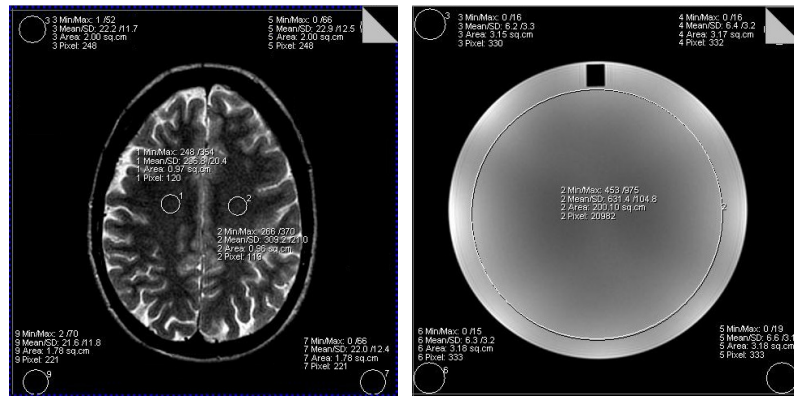
Årsaken til at et såpass viktig aspekt ved bildekvalitet som SNR ikke er tatt med i veiledningen fra ACR, er trolig at det er vanskelig å sette et fast kriterium for hva som er god nok SNR. SNR må imidlertid være relativt høy for å kunne gjennomføre flere av de andre ACR-testene, så en lav score på ACR-testene kan skyldes lav SNR. En sammenligning av SNR mellom ulike tomografer, forutsetter bruk av samme type MR-sekvens i opptakene, samt samme testobjekt fra måling til måling. På hjerneopptak må det forventes noe lavere SNR-verdier, da blant annet bevegelse fra for eksempel svelging, blinking, blodgjennomstrømning og pulsasjon bidrar til økt støy i opptaket (Kaufman, Kramer, Crooks & Ortendahl 1989).

ACR-fantomet inneholder nikkellorid og natriumklorid løst i vann (konsentrasjon: 10 mM NiCl_2 og 75 mM NaCl)⁴ (ACR, 2005). Dette er en flytende ionisk løsning med prototetthet, relaksasjonstider, samt magnetiske og ledende egenskaper som simulerer egenskapene og ledeevnen i biologisk vev. Dermed gir fantomet samme induktive belastning (load) i antennen som pasienten gir. Da oppnås omtrent samme reduksjon i Q-faktor (coil quality factor) og grad av støy i bildet, som man kan forvente ved biologiske opptak (Lerski & de Certaines, 1993; McRobbie et al., 2007). SNR kan derfor beregnes ved hjelp av fantomopptak.

SNR ble beregnet med målinger av pikselintensitet i bildet. $\text{ROI}_{\text{signal}}$ ble lagt i fantomopptakets snitt #7, da dette er et homogent område, sentralt i antennen. ROI ble plassert på samme sted i bildene fra de ulike tomografene. $\text{ROI}_{\text{signal}}$ ble lagt slik at den dekket ~70 % av fantomet i det aktuelle snittet (~200 cm^2). På denne måten unngikk man usikkerhet i målingen som følge av ujevn signalfordeling i antennen. I hjernen var $\text{ROI}_{\text{signal}} \sim 1 \text{ cm}^2$ og ble lagt midt i hvitsubstansområdet i storehjernen (sentrum

⁴ M er mol/liter, mM er millimol/liter.

semiovale) (Figur 9). Ved riktig posisjonering av kroppen og hodet befinner dette området seg omtrent midt i antennen. Både i fantombildet og hjernebildet ble støyen målt som et gjennomsnitt av ROI_{bakgrunn} plassert i bildets fire hjørner, utenfor områder med ghosting av noe slag. Verdiene fra målingene ble satt inn i Ligning 2.



Figur 9: Illustrasjon på plassering av ROI_{signal} og ROI_{bakgrunn} i hjerne- og fantomopptak. ROI_{signal} dekket et lite, men homogent område i hjernen, mens den dekket ~70 % av fantomets tverrsnitt. ROI_{bakgrunn} var plassert på samme måte i begge opptakene.

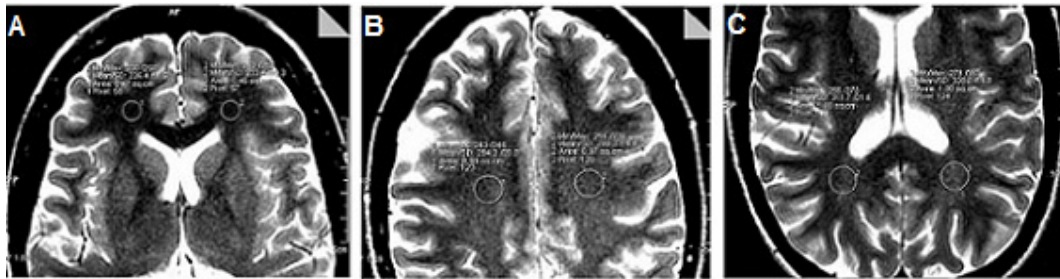
CNR

Som tidligere nevnt, er ACR-fantomet fylt med 10 mM NiCl_2 og 75 mM NaCl . Fantomet er også utstyrt med en liten beholder fylt med 20 mM NiCl_2 og 15 mM NaCl . Disse to løsningene har ulik konsentrasjon, og dermed ulik relaksasjonstid og signalintensitet. CNR ble beregnet i fantomopptakene ved å beregne SNR i de to løsningene og sette verdiene inn i Ligning 4. I hjerneopptakene ble CNR beregnet for grå og hvit substans, målt i thalamus og sentrum semiovale (Figur 9 – venstre og Figur 11).

T2-målinger

De to løsningene nevnt ovenfor har T2-tid henholdsvis ~135 ms og ~70 ms (Leppert, Almlı, McKinstry, Mulkern, Pierpaoli, Rivkin et al., 2009). T2-tidene i ulike vev er også kjent, men varierer noe fra person til person. T2-målingene ble gjort på T2 TSE-sekvensen i fantom- og frivilligopptakene. Et T2-kart ble laget med programvaren nordicICE (Tabell III), og ROI ble lagt manuelt i aktuelle områder i fantomet og i hjernen. I hjernen ble T2-tider målt i hvit- og grå substans. For måling i hvitsubstans ble ROI lagt frontalt, sentralt og occipittalt (trigonum cerebrale) i sentrum semiovale (Figur

10). Måling i gråsubstans ble gjort i thalamus, da dette er en struktur som er stor nok til å legge ROI uten å komme borti andre strukturer (Figur 11).



Figur 10: Illustrasjon på hvordan ROI ble lagt i hvitsubstans i ulike deler av sentrum semiovale. A: frontalt. B: mediant. C: occipitalt (trigonum). ROI ble lagt både i høyre og venstre hjernehalvdel.



Figur 11: Illustrasjon på hvordan ROI ble lagt i thalamus.

Det ble laget histogrammer av signalfordelingen i hver ROI for å sjekke at denne var normalfordelt. Avvik i normalfordelingskurven kan blant annet skyldes at ROI inneholder områder med andre signaler enn det aktuelle vevet man er ute etter, som for eksempel cerebrospinalvæske fra deler av ventriklene. En slik blanding av vev med ulike T2-tider i samme ROI var ikke ønskelig.

3.6.2 Hypotesetesting

3.6.2.1 Analysemetoder

Dataanalyse

De åtte tomografene ble testet mot hverandre ved hjelp av ett opptak fra hver tomograf. Beregningene i bildene ble gjort to ganger i hvert opptak og disse to målingene var grunnlag for analysen. Linear mixed model (LMM) ble valgt som analysemetode for å

teste variasjonen mellom hvert senter. LMM fungerer omtrent som en variansanalyse, men uten samme strenge forutsetning om uavhengige normalfordelte verdier med like varianser (Norusis, 2008; Wenstøp, 2006). LMM er en mer sensitiv metode fordi den tar høyde for at disse kriteriene ikke er oppfylt (Nahhas, 1999). Som post-hoc test ble Fishers least significant difference (LSD) valgt. Denne testen gjør en parvis sammenligning av alle mulige gruppepar (senterpar) som var inkludert i analysen, med signifikansnivå 0,05 (Polit, 2010).

Fantomopptakene ble gjentatt ved hvert senter etter 4–6 måneder. Disse opptakene ble benyttet for å verifisere opptak fra første omgang, samt teste stabiliteten over tid. Valgt signifikansnivå var 0,05.

Variansanalyse – visuell vurdering

Variasjon i verdier vurdert på en skala 1–5 kunne ikke analyseres ved hjelp av LMM. Dette er ordinale tall, såkalte ikke-parametriske verdier (rangeringer). Motstykket til enveis variansanalyse for ikke-parametriske verdier er Kruskal-Wallis test, og Bjørndal og Hofoss (2004) foreslår denne testen til ordinale data. Testen sammenligner medianene i flere ulike grupper. Denne testen bør kun brukes dersom man har fem eller flere verdier i hver gruppe (Polit, 2010). Med tre respondenter fikk hvert spørsmål i spørreskjemaet tre svarverdier. Derfor ble spørsmålene delt inn i følgende tre hovedkategorier: differensiering av ulike vev og strukturer (4 spørsmål, 12 verdier for hvert senter), artefakter i bildet (7 spørsmål, 21 verdier for hvert senter), og bildekvalitet som helhet (4 spørsmål, 12 verdier for hvert senter) (Vedlegg1).

Korrelasjon og regresjon

Samvariasjoner ble testet ved hjelp av enkel lineær regresjonsanalyse, signifikansnivå 0,05. To og to variabler ble testet, hovedsakelig på grunn av prosjektets begrensede utvalg med få måleverdier fra hvert senter.

Sammenheng mellom sentervariasjon og utstyr

Siden opptakene var gjort med standardiserte parametre for å bli så like som mulig, hadde det liten hensikt å lete etter sammenhenger mellom opptaksparameter og bildekvalitet. Det var isteden ønskelig å se etter sammenhenger mellom grunnleggende forskjeller ved tomografene og bildekvalitet. Regresjonsanalysene ble konsentrert mot

gradientstyrke og type mottakerspole (antenne). Sammenhenger mellom ytre forhold og bildekvalitet ble testet i form av regresjonsanalyse med romtemperatur som variabel.

Sammenheng mellom avhengige variabler

Flere bildekvalitetsaspekter (avhengige variabler) ble testet mot hverandre. På denne måten kunne man for eksempel teste om det var sammenheng mellom beregnet snittykkelse og snittposisjon, eller mellom snittposisjon og geometri. Dette ble gjort for å teste om det var noen sammenheng mellom variablene, altså om resultatet av noen kvalitetsaspekter kan forutsis ut fra resultatet av noen andre.

Sammenheng mellom objektiv og subjektiv bildekvalitet

Testing for eventuelle sammenhenger mellom objektive beregninger og subjektive vurderinger ble undersøkt med beregning av gjennomsnitt, median og typetall for hver kategori (differensiering av vev og strukturer, artefakter og bildekvalitet som helhet). Disse beregningene ble gjort både for hver radiolog for seg og for radiologenes samlede vurdering.

3.6.2.2 Beregning av måleusikkerhet

Det er først ved å estimere innslag av måleusikkerhet at man kan vurdere om dette representerer en alternativ tolkning av mellomgruppevariasjoner (Martinussen 2010). I dette prosjektet var total måleusikkerhet sammensatt av flere forhold. I tillegg til varierende stabilitet i selve opptaksprosessen, er posisjonering i spolen, snittplassering og manuell analyse eksempler på forhold som kan gi varierende resultater fra gang til gang og dermed er med på å øke måleusikkerheten. I forsøk på å finne et estimat på hvor stor denne måleusikkerheten var, ble det forsøkt å separere ut ulike bidrag til usikkerhet. Testingen ble gjort med ett av ACR-fantomene, ved ett senter (senter 6), og ved hjelp av følgende forsøksserie:

1. gjentatte beregninger i samme opptak
2. gjentatte opptak og beregninger uten reposisjonering i spolen, med samme snittplassering
3. gjentatte opptak og beregninger etter reposisjonering i spolen med ulike personer som gjorde reposisjoneringen

Opptakene ble gjort på samme dag, og opptak og analyser ble gjentatt åtte ganger for hvert punkt 1–3. Gjennomsnitt og SD ble beregnet for hvert aspekt innenfor alle punktene, og resultatet ble sammenlignet med gjennomsnitt og SD mellom senter.

3.7 Ethiske betraktninger

Godkjenning

Regional komité for medisinsk og helsefaglig forskningsetikk i Midt-Norge (REK midt) har vurdert dette prosjektet som utviklingsarbeid, og det er derfor ikke søkepliktig (Vedlegg 5). Prosjektet faller altså inn under definisjonen på kvalitetssikring i Helse- og omsorgsdepartementets veileder til helseforskningsloven (2010), kapittel 2, punkt 2.4 *Nærmere om forholdet mellom forskning og kvalitetssikring:*

Kvalitetssikring som er en del av helsetjenesten faller i utgangspunktet utenfor helseforskningslovens virkeområde. Slik kvalitetssikring trenger derfor, i motsetning til medisinsk og helsefaglig forskning, ikke forhåndsgodkjenning fra REK (s. 8). (...) Kvalitetssikring kan defineres som prosjekter, undersøkelser, evalueringer o.l. som har som formål å kontrollere at diagnostikk og behandling faktisk gir de intenderte resultater. Nasjonale tiltak for å sikre og forbedre kvaliteten i tjenestene inkluderer utvikling av nasjonale kvalitetsindikatorer, samordning og styrking av medisinske kvalitetsregistre og å utarbeide gode faglige retningslinjer. Kvalitetsarbeidet må baseres på systematisk dokumentasjon (s. 9).

Den studien som kvalitetskontrollen inngår i (PRODE), er godkjent av REK sør-øst.

Prosjektet er meldt til personvernombudet for forskning/Norsk samfunnsvitenskapelig datatjeneste (NSD) (Vedlegg 6). Siden prosjektet bare har benyttet en frisk frivillig, var det ikke behov for noen form for identifisering av personen i bildene. De eneste opplysningene som ble registrert i forbindelse med avbildning var senterinformasjon. Siden bildene altså ikke på noe tidspunkt inneholdt personopplysninger eller andre sensitive opplysninger, kunne de sendes pr. post mellom sentre og observatører til visuell vurdering.

Bruk av frisk frivillig

Det er ingen biologisk risiko ved å gjennomføre en MR-undersøkelse så lenge man ikke har noen kontraindikasjoner mot MR (Kanal, Barkovich, Bell, Borgstede, Bradley, Froelich et al., 2007). Det kan imidlertid forekomme at friske frivillige har sykdommer som ikke gir symptomer, eller at man finner tegn på mulighet for sykdom i framtiden. Bildene ble ikke gransket spesielt for å avdekke eventuelle sykdommer eller anomalier. Siden bildekvaliteten skulle vurderes av tre radiologer, er det usannsynlig at eventuelle sykdomstegn ville blitt oversett. I tilfeller hvor sykdomstegn oppdages, har man plikt til å informere den frivillige slik at vedkommende kan få forsvarlig oppfølging av helsevesenet. Det er derfor alltid viktig å informere frivillige om muligheten for å oppdage sykdom, opplysningsplikten og eventuelle følger, slik at deltakelsen er grundig gjennomtenkt. Den frivillige i dette prosjektet har gjennomgått MR-undersøkelse tidligere i forbindelse med annen utprøving og studier, og vet derfor at man ikke vil gjøre uventede funn av sykdom, anomalier eller andre ting som gjør vedkommende uegnet som testperson. Personen bruker ingen medisiner som kunne vært en bias. Personen fikk ingen økonomisk kompensasjon for å delta.

Anonymisering

Ingen av bildeopptakene inneholder noen form for personinformasjon om den frivillige. Verken oppgitt fødselsdato eller kjønn var korrekt i opptakene. Registrert fødselsdato var den datoen opptaket ble gjort, og registrert pasientnavn var senternummer. Dermed er det umulig å spore opptakene tilbake til den frivillige. Bildene ble ikke lagret på tomografen, men eksportert til en CD for oppbevaring til masterprosjektet er gjennomført.

4 Resultat

Måleverdiene som er oppgitt i forbindelse med fantomopptak i dette kapittelet er gjennomsnittsverdier fra de to målingene fra første runde med fantomopptak. KI_{95} for målingene er tatt med i figurene, unntatt ved målinger med så smale intervaller at de skjules av datapunktene.

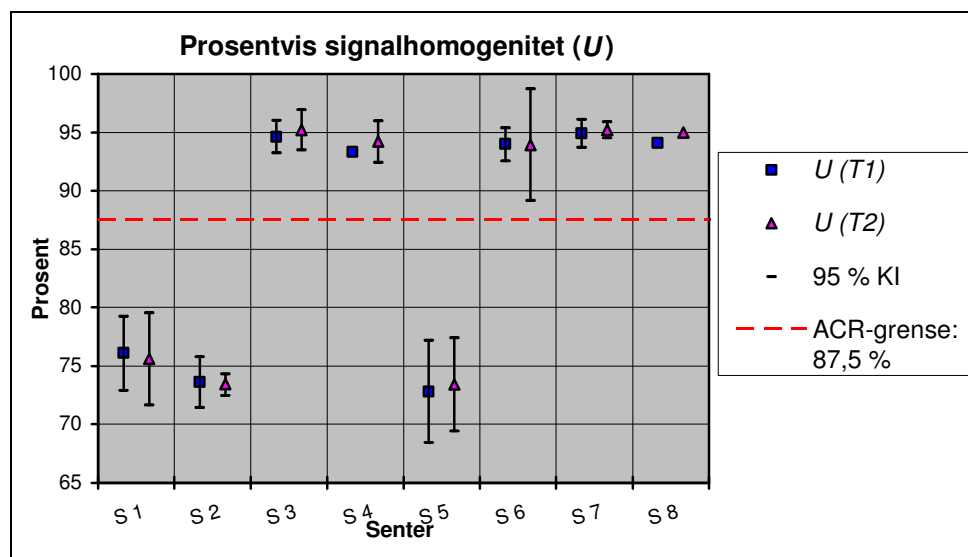
4.1 Variasjon mellom senter og sekvenser

Verdiene som var hentet ut fra opptaksmaterialet ble analysert ved hjelp av LMM-analyse for å avdekke signifikante individuelle ulikheter i bildekvalitet mellom sentrene. Valgt signifikansnivå var 0,05.

4.1.1 Signal/støy forhold og signalhomogenitet

Signalhomogenitet (U)

Senter 1, 2 og 5 skilte seg ut med en score på mer enn 10 prosentpoeng lavere enn 87,5 % som er ACR-kravet for tilfredsstillende U for en hodespole ved magnetfeltstyrke 1,5 T. Ved de resterende sentrene var $U > 93$ %. En oversikt over signalhomogenitet ved de ulike sentrene er gitt i Figur 12.



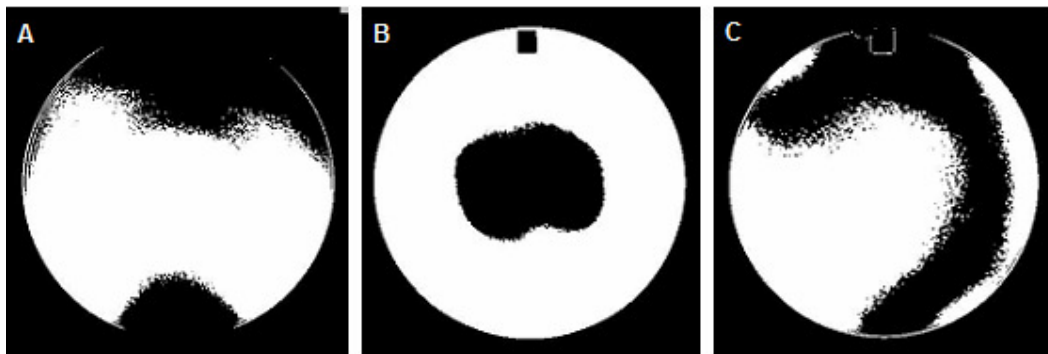
Figur 12: Beregnet prosentvis signalhomogenitet (U) for T1 og T2 SE med tilhørende KI_{95} .

Konfidensintervall for andelen tomografer som ikke nådde ACR-grensen for U

Resultatet ble gjort om til en binomial fordeling, med verdiene 1–Bestått og 2–Ikke-bestått. KI_{95} for andelen ikke-bestått ble beregnet i SPSS. $KI_{95}(2) = [4,0-71,1]$.

Intervallbredden avhenger både av utvalgsstørrelse og andel ikke-bestått, bredden øker med minkende utvalg og med andeler opp mot 50 %. Til tross for at beregningen ga et upresist svar på hvor stor andel av et utvalg man kan forvente faller utenfor ACR-kravet for U , ga den imidlertid noe: KI_{95} inkluderte ikke 0, sannsynligheten for å få et utvalg hvor alle tomografene klarer kravet for U er altså $< 5 \%$.

Ulike spoletyper hadde ulik signalfordeling. Dette ble visualisert i fantomopptaket ved å sette skjermens gråtoneskala til null, og justere skjermkontrasten til signalfordelingen trådte frem. Lyse og mørke områder var henholdsvis områder med høy og lav signalintensitet (Figur 13).



Figur 13: Signalfordeling i tre ulike spoletyper. A: CP-spole (senter 8), B: 6-kanalers hodespole (senter 2), C: 12-kanalers hodespole (senter 3). Signalfordelingen ved senter 1 og 5 var tilsvarende til bilde B, og signalfordelingen ved senter 4, 6 og 7 var tilsvarende til bilde C.

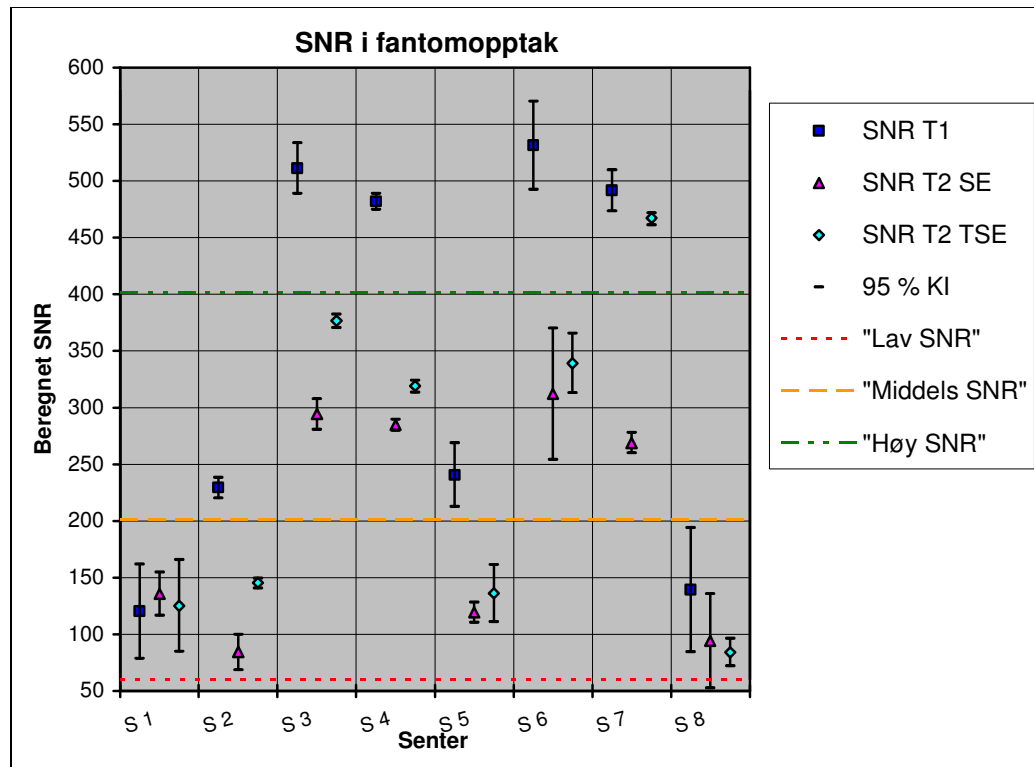
Signal/støy forhold

Det var signifikante forskjeller i SNR ved ulike senter ($p < 0,0001$). Resultatene fra analyse med LMM i henholdsvis fantom- og hjerneopptak er gitt i Tabell V og VI.

SNR i fantomopptak

For lettere å kunne beskrive hvordan sekvensene fordelte seg med hensyn til SNR-verdier beregnet i fantomopptak innad ved hvert senter, ble verdiene delt inn i tre kategorier: lav SNR: 60–200, middels SNR: 200–400, høy SNR: 400–580 (Figur 14). Kategoriene favnet sentrenes KI_{95} , men var en grov inndeling med variasjoner opp mot 15 % innad i enkelte av kategoriene.

Senter 1 og 8 skilte seg ut med lav SNR både for T1 og T2. Senter 1 hadde liten forskjell i SNR mellom de ulike sekvensene, beregnede SNR verdier for T1, T2 SE og T2 TSE var henholdsvis 120,3, 126,8 og 125,4. Senter 8 hadde utvalgets laveste verdi på T2 TSE (84,2), og utvalgets nest laveste verdi på T2 SE (94,1), men verdien på T1 lå nesten 50 % høyere enn dette (139,3). Senter 2 og 5 hadde middels verdier for T1 (229,3 og 240,7), og lave verdier for T2 SE (84,3 og 119,4) samt T2 TSE (145,6 og 136,4). Senter 3, 4 og 6 hadde høye verdier for T1 (511,3, 482,2 og 531,4), og middels verdier for T2 SE (294,4, 284,9 og 312,2), samt T2 TSE (376,6, 319,4 og 339,3). Senter 7 skilte seg ut med høye verdier både for T1 og T2 TSE (491,7 og 467,4), og middels verdi for T2 SE (269,1).



Figur 14: Beregnet SNR for T1 og T2 med tilhørende KI_{95} . Nedre kategorigrenser for kategoriene lav (60–200), middels (200–400) og høy SNR (> 400) er markert.

Tabell V: LMM analyse for beregnede SNR-verdier i fantomopptak.

Sekvens	p-verdi	F-verdi	LSD ⁽¹⁾
T1	< 0,0001	2231,2	S4 og S7 (p = 0,110)
T2 SE	< 0,0001	1789,3	S1 og S5 (p = 0,059)
T2 TSE	< 0,0001	7409,6	Ingen

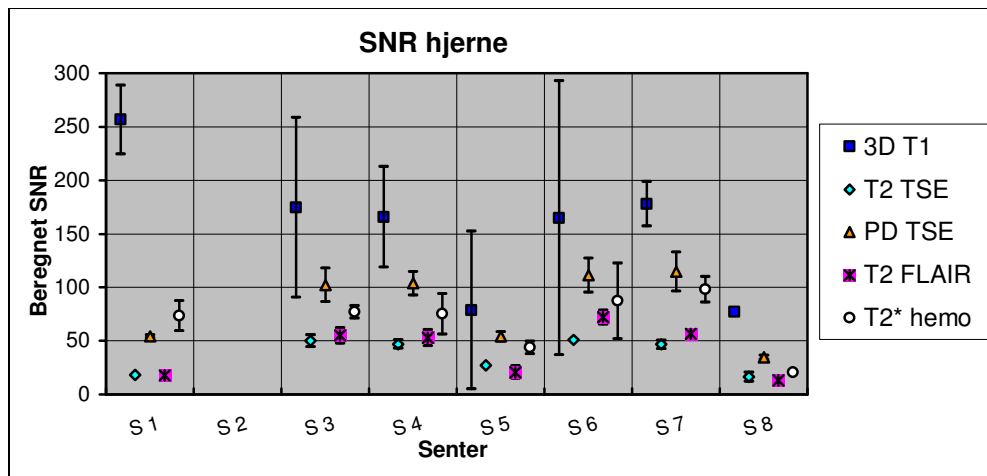
⁽¹⁾LSD = least significant difference. Senterpar med sammenligbare verdier er tatt med i denne kolonnen.

SNR i hjerneopptak

I hjerneopptakene befant alle SNR-verdiene seg i området 12,9–256,8, altså et annet intervall enn i fantomopptakene (84,2 – 531,4). Rekkefølgen på sekvenser i forhold til SNR-verdi var ulik ved ulike senter. 3D T1-sekvensen hadde høyest SNR ved alle sentrene, men noen steder var T2* den sekvensen med nest høyest SNR, mens andre steder hadde PD-sekvensen nest høyest SNR (Figur 15).

Det lot seg ikke gjøre å beregne SNR for hjerneopptak fra senter 2. Signalintensiteten og SD i ROI_{bakgrunn} var null, og dermed ble SNR-verdien uendelig stor⁵. Derfor er senter 2 utelatt fra denne vurderingen. Blant de resterende sentre skilte senter 8 seg ut med utvalgets laveste SNR-verdier. Senter 8 lå jevnt med senter 5 på 3D T1 (77,2 og 78,7), men hadde SNR < 35 (spredningsbredde 12,9–34,6) på alle de andre sekvensene. Senter 1 og 5 hadde utvalgets nest laveste verdier på T2 FLAIR (17,6 og 20,7), T2 (18,4 og 27,2) og PD (54,3 og 54,1). Samtidig hadde senter 1 utvalgets høyeste verdi på 3D T1 (256,8). Verdiene ved senter 3 og 4 lå jevnt (T2 50,1 og 47,0, T2 FLAIR 55,0 og 52,7, T2* 77,0 og 75,3, PD 102,1 og 103,8, 3D T1 174,8 og 165,9). Også noen av verdiene ved senter 6 og 7 lå jevnt (3D T1 164,9 og 178,0, T2 51,2 og 47,0, PD 111,4 og 114,8). Senter 7 hadde utvalgets høyeste verdi for T2* (98,1), etterfulgt av senter 6 (87,3). Senter 6 hadde utvalgets høyeste verdi for T2 FLAIR (72,1), etterfulgt av senter 7 (56,6). Ved senter 3, 4, 6 og 7 var sekvensene fordelt likt i forhold til SNR: T2 lå lavest, etterfulgt av T2 FLAIR, T2*, PD og 3D T1. Også senter 1 og 5 hadde tilnærmet lik fordeling av sekvenser i forhold til SNR, med T2 FLAIR lavest, etterfulgt av T2, PD og 3D T1. Unntaket var T2*, som hadde høyest verdi blant sekvensene ved senter 1 (73,5), og lå mellom T2 og T2 FLAIR ved senter 5 (43,9).

⁵ Årsaken til dette er at Philips (senter 2) benytter SENSE (sensitivity encoding) rekonstruksjon. Alle signaler lavere enn en viss verdi tildeles verdien 0 for å forenkle rekonstruksjonen (MR-TIP, 2011).



Figur 15: Beregnet SNR i ulike sekvenser i hjerneopptak med tilhørende KI_{95} .

Tabell VI: LMM analyse for beregnede SNR-verdier i hjerneopptak.

Sekvens	p-verdi	F-verdi	LSD
3D T1	< 0,0001	468,5	S3 og S4 (p = 0,064), S3 og S7 (p = 0,465), S4 og S6 (p = 0,822), S5 og S8 (0,715)
T2 TSE	< 0,0001	3483,0	S4 og S7 (p = 1,00)
PD TSE	< 0,0001	1284,1	S1 og S5 (p = 0,88), S3 og S4 (p = 0,26)
T2 FLAIR	< 0,0001	3371,8	Ingen
T2*	< 0,0001	394,1	S1 og S3 (p = 0,064), S1 og S4 (p = 0,281), S3 og S4 (p = 0,320)

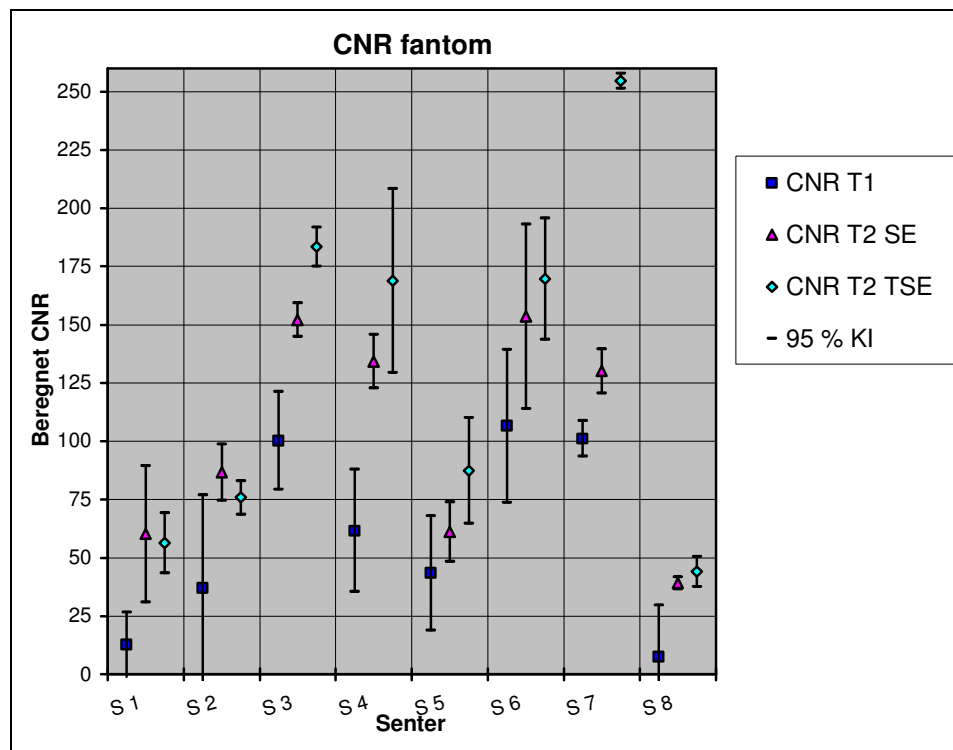
Kontrast/støy forhold

CNR i fantomopptak

Innad ved hvert senter var det T1-sekvensen som hadde lavest CNR. Ved senter 1 og 2 var denne etterfulgt av T2 TSE-sekvensen, mens den ved senter 3–8 var etterfulgt av T2 SE-sekvensen (Figur 16).

Senter 8 utmerket seg med lavest CNR-verdi i alle sekvensene (T1 7,6, T2 SE 39,2 og T2 TSE 44,1). Senter 1 hadde også lav verdi på T1 (12,7), men noe høyere verdier på T2 SE og TSE (60,2 og 56,3). Senter 2 og 5 lå jevnt på T1 (36,9 og 43,5), men lenger fra hverandre på T2 SE (86,7 og 61,1) og på T2 TSE (75,9 og 87,3). Senter 6 hadde utvalgets høyeste verdi på T1 og T2 SE (106,6 og 153,6). Senter 7 hadde utvalgets høyeste verdi på T2 TSE (254,6), etterfulgt av senter 3 (183,5). Senter 3 og 7 lå jevnt med senter 6 på T1 med verdiene 101,1 og 100,3. Senter 3 lå jevnt med senter 6 på T2 SE med verdien 152,1. Senter 4 og 6 lå jevnt på T2 TSE (168,9 og 169,7), og senter 4

og 7 lå jevnt på T2 SE (134,3 og 130,1). Verdiene ved senter 3 og 4 var signifikant ulike ($p = 0,001$), men forholdet mellom dem var jevnt (justert $r^2 = 0,99$, $p = 0,038$, $F = 278,2$). Signifikansverdier mellom senter er gitt i Tabell VII.



Figur 16: Beregnet CNR for T1 og T2 i fantomopptak med tilhørende KI_{95} .

Tabell VII: LMM analyse for beregnede CNR-verdier i fantomopptak.

Sekvens	p-verdi	F-verdi	LSD
T1	< 0,0001	74,6	S1 og S8 ($p = 0,547$), S2 og S5 ($p = 0,283$), S3 og S6 ($p = 0,358$), S3 og S7 ($p = 0,899$), S6 og S7 ($p = 0,423$)
T2 SE	< 0,0001	488,5	S1 og S5 ($p = 0,746$), S3 og S6 ($p = 0,611$), S4 og S7 ($p = 0,075$)
T2 TSE	< 0,0001	1717,0	S4 og S6 ($p = 0,719$)

CNR i hjerneopptak

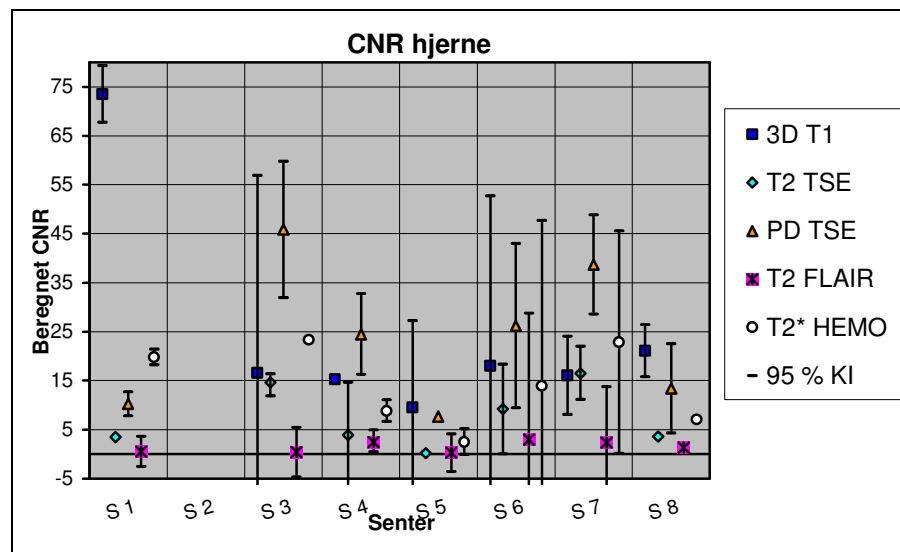
Rekkefølgen på sekvenser i forhold til CNR-verdi var ulik mellom ulike senter. T2 FLAIR-sekvensen hadde lav CNR-verdi ved alle senter (variasjonsbredde 0,3–3,0), og det var ikke signifikant forskjell mellom senter for denne sekvensen (Tabell VIII).

Senter 1 utmerket seg med utvalgets høyeste verdi i 3D T1-sekvensen. Resten av sentrene lå jevnt i CNR-verdi for denne sekvensen (variasjonsbredde 9,5–21,1), og med senter 1 utelatt fra analysen var det ikke signifikant forskjell mellom senter i denne

sekvensen. Siden det ikke lot seg gjøre å beregne SNR for senter 2, kunne man heller ikke beregne CNR ved dette senteret (Figur 17).

Senter 5 utmerket seg med lavest verdi i alle sekvensene. Her var CNR-verdien i T2-sekvensen helt nede på 0,2, mens T2*, PD og 3D T1 hadde verdiene 2,5, 7,7 og 9,5. Senter 1 og 8 lå jevnt med senter 5 på PD med verdiene 10,2 og 13,4. I T2-sekvensen lå senter 1, 4 og 8 jevnt med utvalgets nest laveste verdier (3,5, 3,9 og 3,6). I T2*-sekvensen lå senter 4 og 8 jevnt med utvalgets nest laveste verdier (8,8 og 7,1).

Senter 7 utmerket seg med utvalgets høyeste verdi i T2-sekvensen (16,5), etterfulgt av senter 3 og 6 (14,6 og 9,2). På T2* lå senter 3 og 7 jevnt med utvalgets høyeste verdier (23,3 og 22,8), etterfulgt av senter 1 og 6 (19,8 og 13,9). Senter 3 hadde utvalgets høyeste verdi i PD-sekvensen (45,8). Nest høyeste verdi i PD hadde senter 7 (38,7), deretter lå senter 4 og 6 jevnt med verdiene 24,5 og 26,2.



Figur 17: Beregnet CNR i ulike sekvenser i hjerneopptak med tilhørende KI₉₅.

Tabell VIII: LMM analyse for beregnede CNR-verdier i hjerneopptak.

Sekvens	p-verdi	F-verdi	LSD
3D T1	< 0,001	22,3	Alle unntatt S1 (p = 0,254, F = 1,8)
T2 TSE	< 0,0001	186,4	S1 og S4 (p = 0,508), S1 og S8 (p = 0,940), S4 og S8 (p = 0,555)
PD TSE	< 0,0001	69,6	S1 og S5 (p = 0,335), S1 og S8 (p = 0,235), S5 og S8 (p = 0,052), S4 og S6 (p = 0,512)
T2 FLAIR	0,338	1,4	
T2*	< 0,001	32,2	S1 og S3 (p = 0,139), S1 og S7 (p = 0,201), S3 og S7 (p = 0,001)

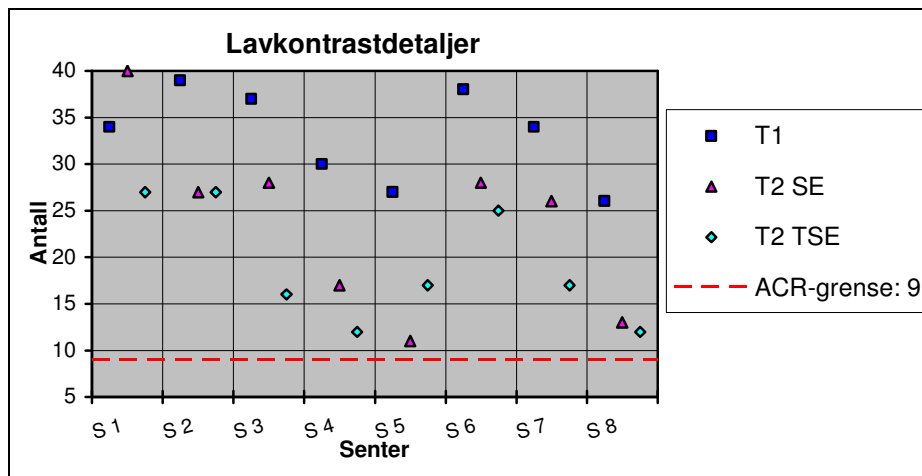
4.1.2 Lavkontrastdetaljer og romlig oppløsning

Lavkontrastdetaljer

Alle sentrene var innenfor ACR-kriteriet (> 9 synlige punkter) for 1,5 T tomografer i denne testen (Tabell IX og Figur 18). Med unntak av senter 1 hadde alle sentrene flere synlige detaljer i T1- enn i T2-sekvensen. Synligheten av detaljer i T2 SE- og TSE-sekvensen varierte innad ved hvert senter. Ved senter 2 hadde disse samme score, mens ved senter 5 hadde T2 SE-sekvensen lavest score. Ved de resterende sentre hadde T2 SE-sekvensen høyere score enn TSE-sekvensen. Det var signifikante forskjeller mellom senter (Tabell X).

Tabell IX: Totalt antall lavkontrastdetaljer ved de ulike sentrene, rangert etter score (lav-høy).

T1	T2 SE	T2 TSE
S8 (26)	S5 (11)	S4 (12)
S5 (27)	S8 (13)	S8 (12)
S4 (30)	S4 (17)	S3 (16)
S1 (34)	S7 (26)	S5 (17)
S7 (34)	S2 (27)	S7 (17)
S3 (37)	S3 (28)	S6 (25)
S6 (38)	S6 (28)	S1 (27)
S2 (39)	S1 (40)	S2 (27)



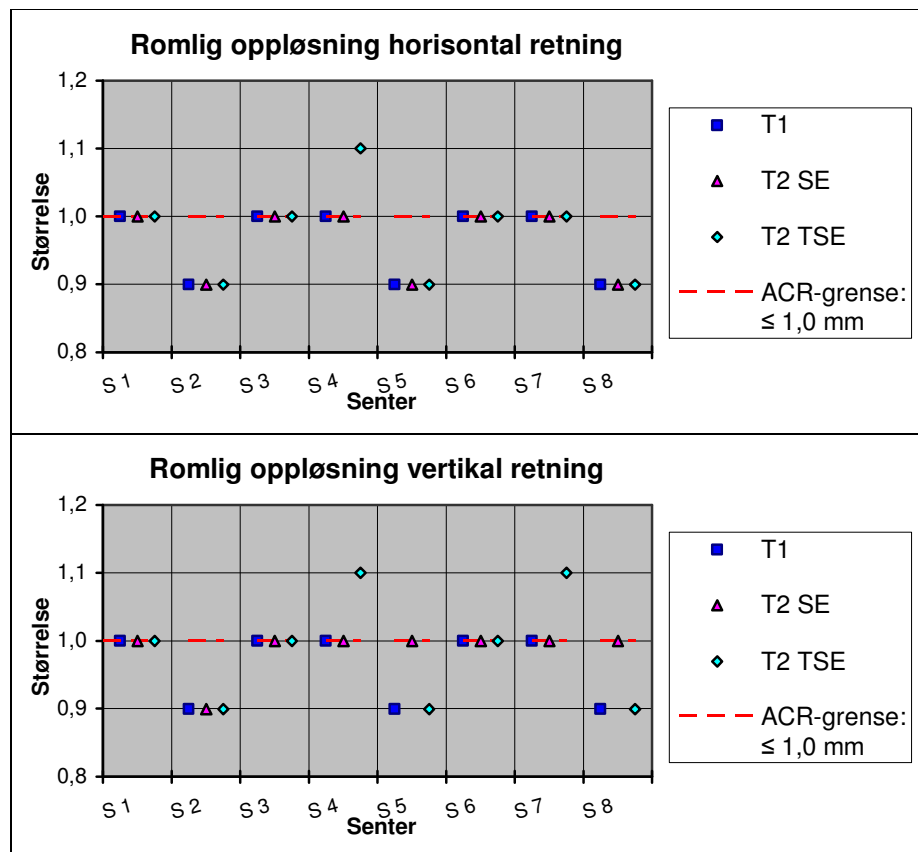
Figur 18: Totalt antall synlige lavkontrastpunkter ved hvert senter for T1, T2 SE og T2 TSE.

Tabell X: LMM analyse for lavkontrastdetaljer i fantomopptak.

Sekvens	p-verdi	F-verdi	LSD
T1	< 0,0001	141,7	S1 og S7 (p = 0,425), S2 og S6 (p = 0,425), S3 og S6 (p = 0,134), S5 og S8 (p=0,425)
T2 SE	< 0,0001	85,5	S2 og S7 (p = 0,306), S3 og S6 (p = 1,00), S3 og S7 (p = 0,185), S6 og S7 (p = 0,185), S4 og S5 (p = 0,109), S5 og S8 (p = 0,486)
T2 TSE	< 0,0001	73,3	S1 og S2 (p = 1,00), S1 og S6 (p = 0,096), S2 og S6 (p = 0,096), S3 og S5 (p = 0,650), S3 og S7 (p = 0,373), S5 og S7 (p = 0,650), S4 og S8 (p = 0,650)

Romlig oppløsning

Alle sentrene var innenfor ACR-kravet for tilfredsstillende bildekvalitet på dette kriteriet ($\leq 1,0$ mm), og det var ikke signifikante forskjeller mellom sentre ($p = 0,86-0,99$). Selv om forskjellene ikke var signifikante, skilte senter 2, 5 og 8 seg positivt ut i begge retninger (Figur 19).



Figur 19: Romlig oppløsning ved hvert senter for T1 og T2.

4.1.3 Gradientlinearitet, snittykkelse og snittposisjon

Gradientlinearitet – geometri

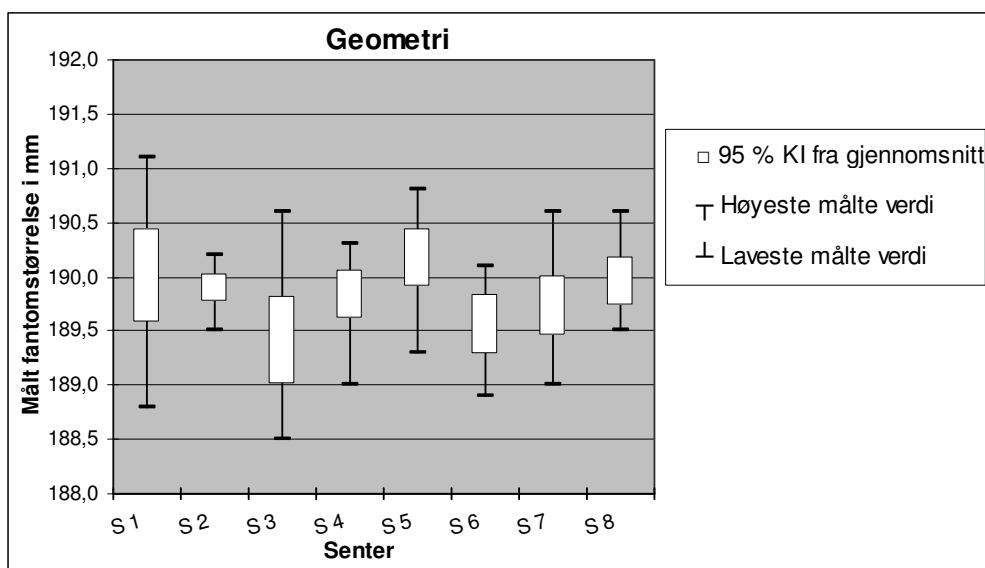
Fantomets gjengitte størrelse i T1-bildet ble målt i ulike retninger. Avvik mellom målt størrelse og reell størrelse i de aktuelle retningene er gitt i Tabell XI.

Tabell XI: Avvik i mm mellom målte størrelser og reell fantomstørrelse. Høydemål ble gjort i sagittalt (grå kolonne) og diametermål i transversalt (hvite kolonner) T1-opptak.

Senter nr.	avvik høydemål***	sn.#1 v.d.	sn.#1 h.d.	sn.#5 v.d.	sn.#5 h.d.*	sn.#5 1.diag.	sn.#5 2.diag.*	gj.sn. diam.*
S 1	- 1,0	- 0,9	0,8	- 0,5	0,8	- 0,1	0,1	0,0
S 2	- 1,5	0,1	- 0,1	- 0,4	- 0,2	0,1	- 0,2	- 0,1
S 3	0,4	0,5	- 0,6	- 0,5	- 1,2	- 0,3	- 1,2	- 0,6
S 4	- 0,3	0,1	0,0	- 0,4	0,1	- 0,6	- 0,1	- 0,2
S 5	- 1,0	0,3	0,7	0,1	0,3	- 0,6	0,3	0,2
S 6	- 0,4	- 0,1	- 0,3	- 0,9	- 0,3	- 0,5	- 0,4	- 0,4
S 7	- 0,2	- 0,3	0,3	- 0,5	- 0,3	- 0,3	- 0,4	- 0,3
S 8	- 0,5	0,2	0,2	0,0	- 0,1	- 0,4	0,0	0,0

sn.# = snittnummer, v.d. = vertikal diameter, h.d. = horisontal diameter, 1. og 2. diag. = 1. og 2. diagonale diameter (henholdsvis nedre venstre til øvre høyre og øvre venstre til nedre høyre), gj.sn. diam. = gjennomsnittsavvik for diametermål. * = $p < 0,05$, ** = $p < 0,01$, *** = $p < 0,001$.

Majoriteten av de målte avvikene (36 av 56) var negative, hvilket betyr at de fleste sentrene hadde en tendens til å gjengi objekter litt mindre enn deres reelle størrelse. Dette gjaldt imidlertid ikke senter 5. Her var de fleste målingene positive. Ved senter 1 og 8 var målingene fordelt symmetrisk rundt 0. Senter 1 og 3 hadde størst spredning på målingene. Ved senter 3 og 6 omfavnet ikke KI₉₅ fantomets virkelige størrelse (Figur 20). Alle verdiene var imidlertid godt innenfor kriteriene fra ACR, som angir avvik $\leq \pm 2$ mm som tilfredsstillende i denne testen.



Figur 20: Diamettermål i ulike retninger og på ulike snitt i det transversale T1-opptaket. Reell fantomdiameter var 190 mm. ACR-grensene på ± 2 mm utgjør øvre og nedre grense på y-aksen.

Både høydemål og gjennomsnittlig diamettermål var signifikant ulikt mellom senter ($p < 0,0001$ og $p = 0,021$). Fullstendig analyse er gitt i Tabell XII.

Tabell XII: LMM analyse for målt fantomstørrelse. Avvik fra reell verdi ble benyttet i analysen.

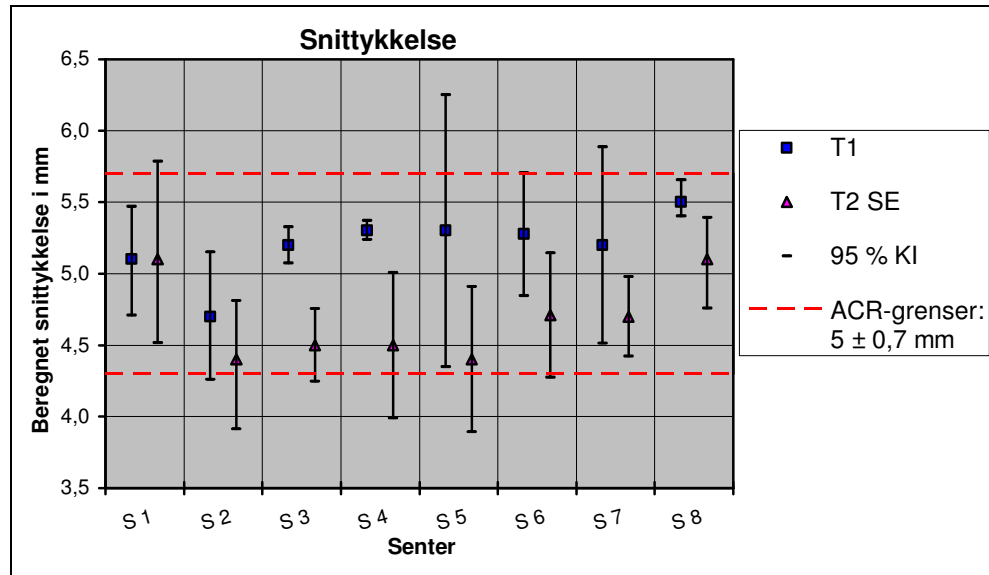
Måleretning	p-verdi	F-verdi	LSD
høydemål	< 0,0001	39,2	Alle unntatt S1, S2 og S5 ($p = 0,483$, $F = 1,1$)
sn.#1 v.d.	0,038	4,3	Alle unntatt S1 ($p = 0,124$, $F = 2,5$)
sn.#1 h.d.	0,215	1,8	
sn.#5 v.d.	0,052	3,7	Alle unntatt S6 ($p = 0,325$, $F = 1,5$)
sn.#5 h.d.	0,056	3,4	Alle unntatt S3 ($p = 0,583$, $F = 0,8$)
sn.#5 1.diag.	0,534	0,9	
sn.#5 2.diag.	0,017	5,2	Alle unntatt S3 ($p = 0,509$, $F = 1,0$)
gj.sn. diam.	0,021	4,8	Alle unntatt S3 og S6 ($p = 0,729$, $F = 0,6$)

Forkortelser har samme betydning som i Tabell XI. Signifikante p-verdier er uthevet.

Snittykkelse

Beregnet snittykkelse var innenfor ACR-grensen ($5 \pm 0,7$ mm) både for T1 og T2 SE ved alle sentrene. KI_{95} var imidlertid ikke innenfor grensene ved alle sentrene (Figur 21). I T1-sekvensen var beregnet snittykkelse > 5 mm ved alle sentrene unntatt senter 2 (4,7 mm). Senter 8 hadde mest avvikende verdi (5,5 mm), og senter 1 mest presis verdi (5,1 mm) i denne sekvensen. Sentrene 3–7 hadde verdier i området 5,2–5,3 mm. I T2 SE-sekvensen var beregnet snittykkelse < 5 mm ved alle sentrene unntatt senter 1 og 8.

Disse to sentrene hadde mest presise verdier (5,1 mm), etterfulgt av senter 6 og 7 med verdien 4,7 mm. Sentrene 2–5 hadde verdier i området 4,5–4,6 mm. Ulikhetene mellom senter var statistisk signifikante (Tabell XIII).



Figur 21: Beregnet gjengitt snittykkelse i mm ved de ulike sentrene med KI₉₅.

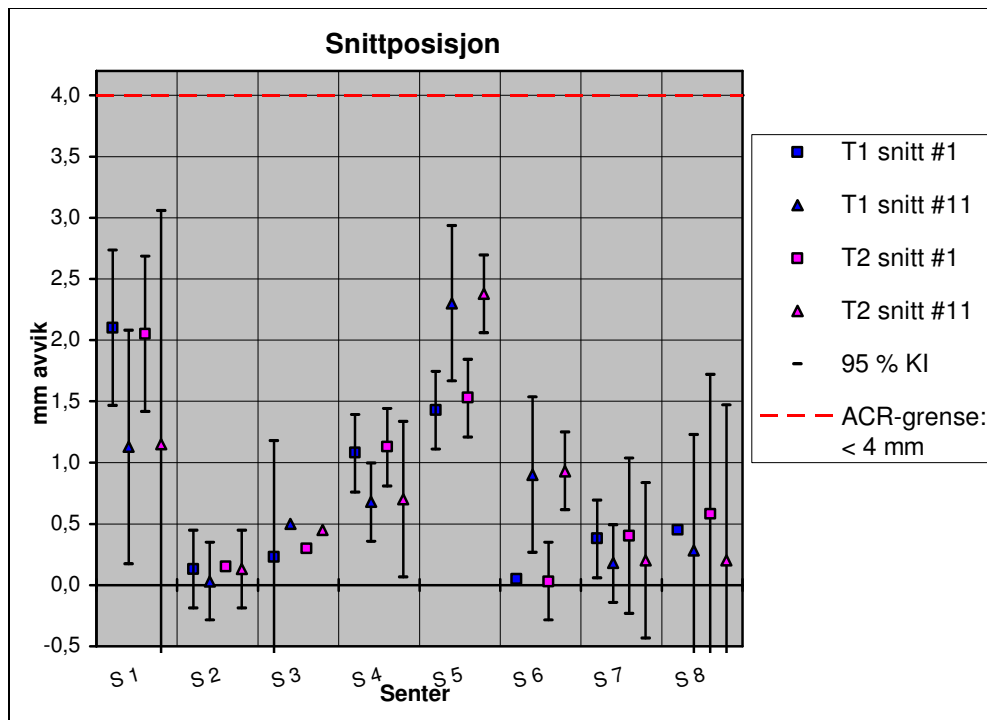
Tabell XIII: LMM analyse for snittykkelse. Beregnet med avviksverdier fra reell snittykkelse.

Sekvens	p-verdi	F-verdi	LSD
T1	0,003	10,4	Alle unntatt S1 og S8 (p = 0,842, F = 0,4)
T2 SE	< 0,0001	54,9	S1 og S8 (p = 0,569), S2 og S5 (p = 0,647), S3 og S4 (p = 0,647), S6 og S7 (p = 0,318)

Snittposisjon

Beregnet snittposisjon med KI₉₅ var innenfor ACR-grensen (< 4 mm) ved alle sentrene. Ved hvert senter var det større avvik i beregnet snittposisjon mellom opptakets første og siste snitt, enn mellom T1- og T2-sekvensen (Figur 22). Det var signifikant forskjell mellom alle senter (p < 0,0001) (Tabell XIV).

Senter 1 og 5 utmerket seg med avvik > 2 mm på henholdsvis første og siste snitt. Ellers var avvikene i snittposisjon < 1,5 mm. Senter 2, 3, 7 og 8 hadde minst avvik i snittposisjon (0,03–0,58).



Figur 22: Beregnet mm avvik i gjengitt snittposisjon i første og siste snitt i T1- og T2-sekvensen med tilhørende KI_{95} . Absoluttverdien ble registrert, og avvikene kan dermed ikke bli negative.

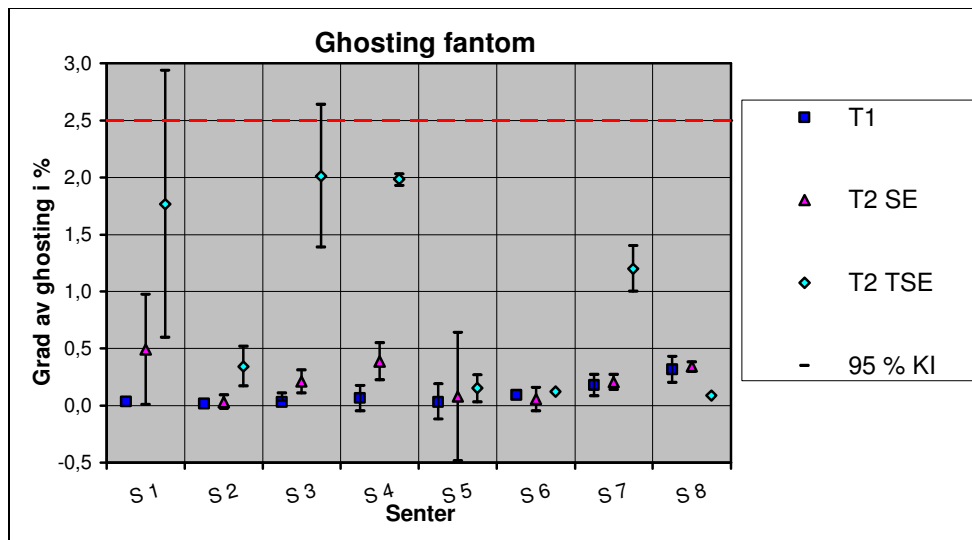
Tabell XIV: LMM analyse for snittposisjon.

Sekvens	p-verdi	F-verdi	LSD
T1 sn #1	< 0,0001	876,6	S2 og S6 (p = 0,070), S7 og S8 (p = 0,070)
T1 sn #11	< 0,0001	261,9	S7 og S8 (p = 0,155)
T2 sn #1	< 0,0001	113,5	S2 og S3 (p = 0,158), S2 og S6 (p = 0,199), S3 og S7 (p = 0,329), S7 og S8 (p = 0,107)
T2 sn #11	< 0,0001	115,4	S1 og S6 (p = 0,066), S2 og S7 (p = 0,468), S2 og S8 (p = 0,468), S7 og S8 (p = 1,00)

4.1.4 Ghosting

Ghosting i fantomopptak

Alle sentrenes grad av ghosting lå langt innenfor ACR-kriteriet (< 2,5 % i T1-sekvensen), med følgende verdier (i %) for senter 1–8: 0,04, 0,02, 0,03, 0,07, 0,03, 0,09, 0,18 og 0,32 (Figur 23). Ghosting ble også beregnet i T2 TSE-opptakene til sammenligning med hjerneopptakene. Her skilte senter 1, 3, 4 og 7 seg ut med størst grad av ghosting (1,77, 2,01, 1,98 og 1,2). Forskjellen mellom sentrene var signifikant (Tabell XV).



Figur 23: Beregnet grad av ghosting med tilhørende KI₉₅ i T1, T2 SE- og T2 TSE-sekvensen i fantomopptak.

Tabell XV: LMM analyse for ghosting i fantomopptak.

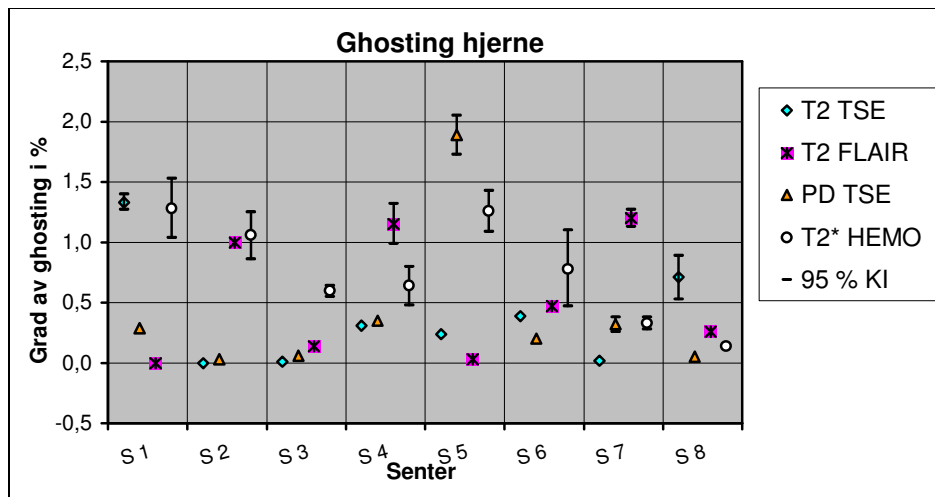
Sekvens	p-verdi	F-verdi	LSD
T1	< 0,0001	206,3	S1, S2, S3 og S5 (p = 0,463, F = 1,0)
T2 SE	< 0,0001	84,7	S2 og S5 (p = 0,093), S2 og S6 (p = 0,378), S3 og S7 (p = 0,850), S4 og S8 (p = 0,137), S5 og S6 (p = 0,384)
T2 TSE	< 0,0001	595,1	S3 og S4 (p = 0,586), S8 og S5 (p = 0,257), S8 og S6 (p = 0,127)

Ghosting i hjerneopptak

Resultatet fra beregning av ghosting i de transversale hjerne-opptakene er gitt i Tabell XVI og Figur 24. Forskjellene mellom sentrene var signifikante ($p < 0,0001$) (Tabell XVII). Senter 5 skilte seg ut med noe høyere grad av ghosting i det PD-vektede opptaket. Alle andre sekvenser og senter hadde grad av ghosting < 1,5 %.

Tabell XVI: Beregnet grad av ghosting på de ulike transversale sekvensene i hjerneopptakene.

Senternr.	Sekvens			
	T2 TSE	PD TSE	T2 FLAIR	T2* HEMO
S 1	1,33	0,29	0,00	1,28
S 2	0,00	0,03	1,00	1,06
S 3	0,01	0,06	0,14	0,60
S 4	0,31	0,35	1,15	0,64
S 5	0,24	1,89	0,03	1,26
S 6	0,39	0,20	0,47	0,78
S 7	0,02	0,32	1,20	0,33
S 8	0,71	0,05	0,26	0,14



Figur 24: Beregnet grad av ghosting med KI₉₅ i de ulike transversale hjerneopptakene.

Tabell XVII: LMM analyse for ghosting i hjerneopptak.

Sekvens	p-verdi	F-verdi	LSD
T2 TSE	< 0,0001	6451,1	S2 og S3 (p = 0,110), S2 og S7 (p = 0,064), S3 og S7 (p = 0,734)
PD TSE	< 0,0001	20446,0	S2 og S3 (p = 0,066), S2 og S8 (p = 0,236), S3 og S8 (p = 0,424), S4 og S7 (p = 0,207)
T2 FLAIR	< 0,0001	11635,2	Ingen
T2*	< 0,0001	1245,5	S1 og S5 (p = 0,182)

4.1.5 T2-måling

T2-verdier i fantom

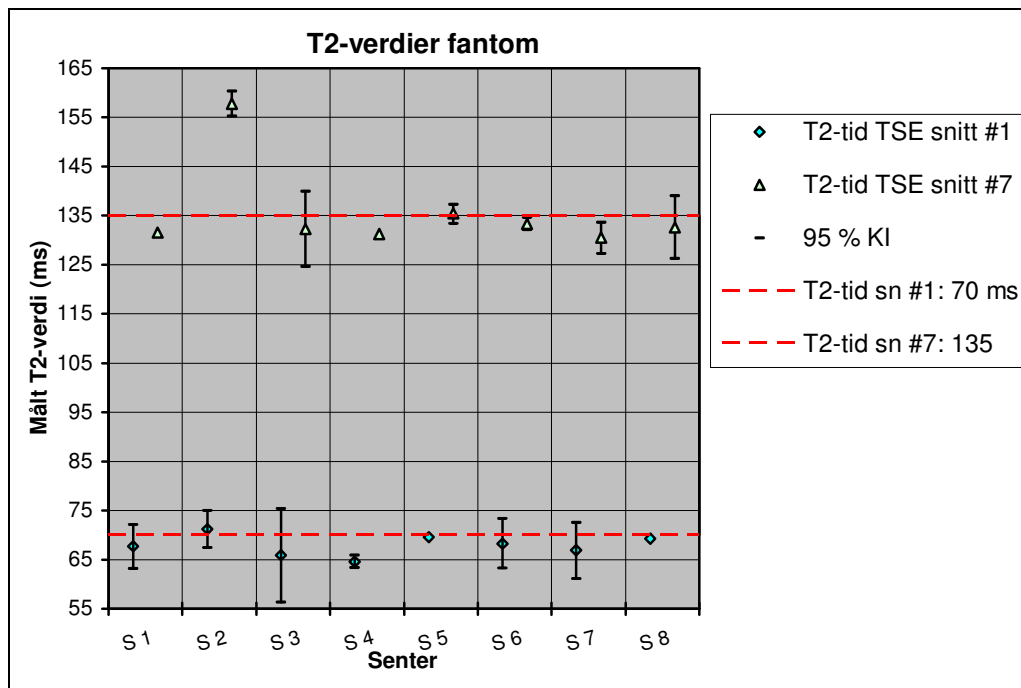
T2-verdier var signifikant ulike mellom senter (p < 0,0001) (Tabell XVIII).

Snitt #1 – reell T2-tid 70 ms:

Med verdiene 65,9 og 64,6 utmerket senter 3 og 4 seg med T2-målinger lengst unna fantomets reelle T2-tid, etterfulgt av senter 7 og 1 (66,9 og 67,7). Senter 5 og 8 var nærmest med verdiene 69,6 og 69,3, etterfulgt av senter 2 og 6 (71,2 og 68,3).

Snitt #7 – reell T2-tid 135 ms:

Senter 2 utmerket seg med verdien 157,8. Blant de resterende sentre, var senter 4 og 7 lengst unna fantomets reelle T2-tid med verdiene 131,3 og 130,5, etterfulgt av senter 1 (131,6). Senter 5 var nærmest med verdien 135,5, etterfulgt av senter 6 (133,3), 8 (132,6) og 3 (132,3) (Figur 25).



Figur 25: Målte T2-verdier i fantomopptakets T2 TSE-sekvens, snitt nummer 1 og 7 med KI₉₅.

Tabell XVIII: LMM analyse for T2-verdier i fantomopptak.⁽¹⁾

Tid	p-verdi	F-verdi	LSD
70 ms	< 0,0001	57,7	S1 og S2 (p = 0,065), S1 og S6 (p = 0,261), S1 og S7 (p = 0,175), S2 og S5 (p = 0,175), S2 og S6 (p = 0,380), S2 og S8 (p = 0,380), S3 og S7 (p = 0,100), S5 og S8 (p = 0,592), S6 og S8 (p = 0,100)
135 ms	< 0,0001	776,8	S1 og S3 (p = 0,122), S1 og S4 (p = 0,508), S3 og S8 (p = 0,508), S4 og S7 (p = 0,102), S6 og S8 (p = 0,145)

⁽¹⁾Beregnet med avviksverdier fra reell T2-verdi.

T2-verdier i hjernen

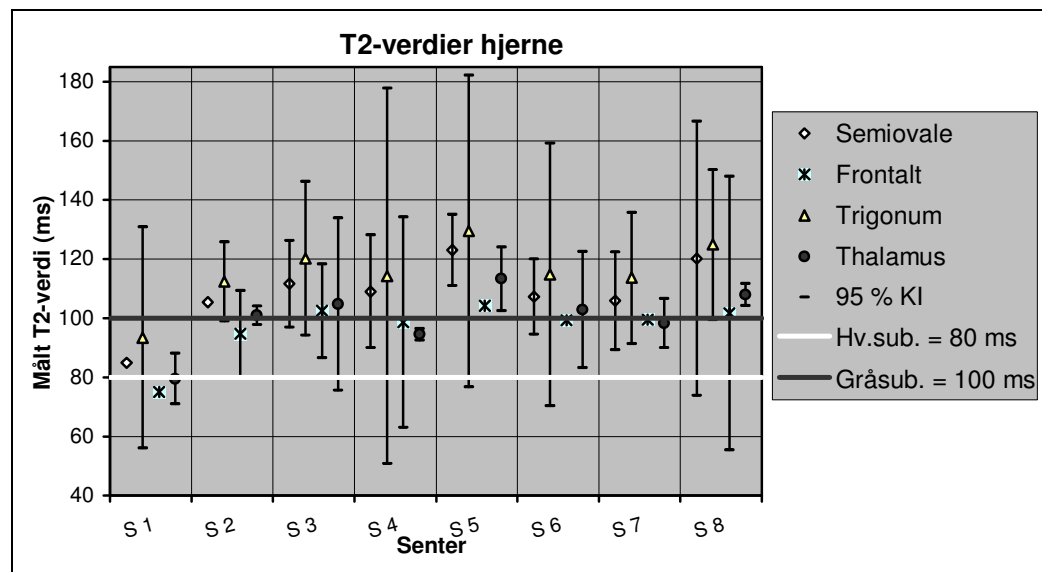
Flere av T2-verdiene målt i hvitsubstans var høyere enn de som ble målt i gråsubstans (thalamus). De omtalte anatomiske områdene er vist i Figur 10 og 11, resultatene er gitt i Tabell XIX og Figur 26. Sentervariasjonen er gitt i Tabell XX.

Med utgangspunkt i at den frivilliges hjerne hadde T2-verdi tilnærmet estimert verdi for normale hjerner (80 ms i hvit substans og 100 ms i grå substans), var det senter 1 og 2 som hadde mest presise verdier for henholdsvis hvit og grå substans. Senter 5 hadde minst presise gjengivelser av alle verdier. Dette var motsatt av fantom hvor senter 5 var mest presis, og senter 2 minst presis på den lengste T2-verdien. Samtidig som senter 1

utmerket seg som eneste senter med T2-verdier nær normalverdier i hvitsubstans, var det dette senteret som var lengst unna normalverdien i grå substans.

Tabell XIX: Målte T2-verdier i hjerneopptak, oppgitt i ms. Thalamus er grå substans, resten er hvitsubstansområder.

Senternr.	Anatomisk område			
	Semiovale	Trigonom	Frontalt	Thalamus
S 1	85,0	93,5	75,1	79,5
S 2	105,5	112,5	94,8	101,0
S 3	111,7	120,3	102,6	104,8
S 4	109,1	114,3	98,7	94,6
S 5	123,1	129,6	104,3	113,4
S 6	107,3	114,8	99,4	103,0
S 7	105,9	113,7	99,5	98,4
S 8	120,3	124,9	101,8	108,0



Figur 26: Målte T2-verdier i ulike områder i hjernen med tilhørende KI₉₅.

Tabell XX: LMM analyse for T2-verdier i hjerneopptak.

Område	p-verdi	F-verdi	LSD
Semiovale	< 0,0001	52,0	Alle unntatt S1, S5 og S8 (p = 0,051, F = 5,1)
Frontalt	0,001	12,3	Alle unntatt S1 (p = 0,103, F = 2,8)
Trigonom	0,001	12,2	Ingen
Thalamus	< 0,0001	37,6	Ingen

4.1.6 Volumetri

Resultatet av volummålingene er gitt i Tabell XXI. De volumene som skilte seg ut med signifikante forskjeller mellom sentrene var totalt hjernevolum (ICV, fra engelsk intracerebral volume), lillehjernens størrelse, andelen gråsubstans i hjernebark og amygdalavolum (Tabell XXII). I totalt hjernevolum inngår alle deler av hjernen, deriblant hele storehjernen og hjernestammen. Det var altså flere strukturer som kunne være kilde til ulikheter enn de som er med i tabellene under, men det var utenfor dette prosjektets målsetting å teste alle. Særlig senter 1, 2 og 5 hadde verdier som skilte seg ut fra resten av sentrene. Forskjellene i ICV var ikke lenger signifikante når disse sentrene ble tatt ut av analysen ($p = 0,462$, $F = 1,1$). Senter 1 utmerket seg med utvalgets minste volummål på lillehjernen, men samtidig var andel gråsubstansvolum beregnet til utvalgets største. Amygdalavolumer var ikke signifikant ulike på tvers av senter når senter 1 ble tatt ut av analysen (absolutte størrelser (cm^3): $p = 0,163$, $F = 2,3$, relative størrelser (% av ICV): $p = 0,051$, $F = 4,2$).

Tabell XXI: Volum-mål på totalt hjernevolum samt ulike strukturer i hjernen, oppgitt i cm^3 .

Anatomisk område	Senternummer							
	S 1	S 2	S 3	S 4	S 5	S 6	S 7	S 8
ICV	1401,78	1429,25	1451,78	1455,43	1478,29	1453,54	1461,16	1451,64
Lillehjernen	62,52	63,97	64,30	65,82	65,84	64,74	65,54	65,83
Cort. GM	247,72	237,59	239,47	243,91	237,44	241,53	238,92	241,76
Caudatus	3,94	3,85	3,88	3,94	3,96	3,99	3,98	3,99
Putamen	5,11	5,22	4,92	5,14	4,84	5,02	5,13	5,08
Pallidus	1,09	1,15	1,01	1,11	1,07	1,06	1,00	1,09
Thalamus	8,02	7,45	7,82	7,81	7,58	7,90	7,83	7,81
Hippo-campus	3,82	3,65	3,70	3,76	3,88	3,78	3,71	3,71
Amygdala	1,45	1,71	1,69	1,66	1,63	1,64	1,61	1,62

Tabell XXII: Variasjonsbredde med gjennomsnitt og standardavvik (SD) for beregnede størrelser ved alle sentrene i ulike områder i hjernen. Verdiene er oppgitt i absolutte og relative volumer (cm³ og % av ICV), med tilhørende p-verdier.

Anatomisk område	Variasjonsbredde (cm ³)	Gj.snitt og SD	p-verdi	Variasjonsbredde (% av ICV)	Gj.snitt og SD	p-verdi
ICV	1401,8–1478,3	1447,86 ± 22,97	0,000			
Lillehjernen	61.68–66.93	64,82 ± 1,36	0,000	4,40–4,61	4,48 ± 0,07	0,029
Cort. GM ⁽¹⁾	236,96–249,68	241,04 ± 3,59	0,002	16,03–17,81	16,68 ± 0,53	0,000
Caudatus	3,66–4,14	3,94 ± 0,14	0,871	0,25–0,28	0,27 ± 0,01	0,756
Putamen	4,51–5,38	5,06 ± 0,24	0,145	0,31–0,38	0,35 ± 0,02	0,097
Pallidus	0,97–1,18	1,07 ± 0,06	0,253	0,07–0,08	0,07 ± 0,01	0,202
Thalamus	7,07–8,23	7,78 ± 0,30	0,267	0,49–0,59	0,54 ± 0,03	0,074
Hippocampus	3,46–4,03	3,75 ± 0,21	0,320	0,24–0,28	0,26 ± 0,02	0,500
Amygdala	1,44–1,73	1,62 ± 0,09	0,003	0,10–0,12	0,11 ± 0,01	0,001

⁽¹⁾Cort. GM = cortical gray matter – gråsubstans i hjernebark. p-verdier < 0,05 er uthevet.

4.1.7 Visuell vurdering

En samlet oversikt over aspektene som ble vurdert, median og typetall for vurderingene, samt signifikansnivå er gitt i Tabell XXIII. For fullstendig svars skjema, se ”Svarskjema” i Vedlegg 1.

Det var flere signifikante forskjeller mellom de tre radiologenes vurderinger, enn mellom deres samlede vurderinger av hvert senter. Hvilke senter som ble vurdert mest forskjellig varierte med spørsmålene. Det var egentlig utenfor oppgavens målsetting å teste enigheten mellom radiologer. Det ble imidlertid vanskelig å se bort fra at diskrepansen i vurderingene kunne ha betydning for sammenligning av senter. Tabell XXIII inneholder derfor en kolonne som viser senter med signifikante diskrepanser for hver sekvens i hver spørsmålskategori.

Differensiering av ulike vev og strukturer

I vurderingen av differensiering av ulike vev og strukturer var det ingen signifikante forskjeller mellom senter i T2*- T2- og PD-sekvensen (p = 0,807, 0,452 og 0,177). Det var heller ikke signifikant forskjellige vurderinger mellom radiologer i PD-sekvensen. I T2*- og T2-sekvensen var det signifikante forskjeller mellom radiologenes vurderinger ved alle senter. I T2 FLAIR- og 3D T1-sekvensen var det både signifikante forskjeller mellom senter (p = 0,015 og < 0,0001), og mellom radiologenes vurderinger. Det var oftest signifikant ulike vurderinger for senter 2, 3 og 6, og sjeldnere for senter 5 i denne spørsmålskategorien.

Artefakter

I vurderingen av artefakter hadde både median og typetall verdien 5, og det var ingen signifikante forskjeller mellom senter ($p = 0,646-0,998$). Det var imidlertid signifikante forskjeller mellom radiologenes vurderinger, unntatt for 3D T1-sekvensen. I denne spørsmålskategorien var det oftest signifikant ulike vurderinger for senter 3, 6 og 7, og sjeldnere for senter 4, 5 og 8.

Bildekvalitet som helhet

I vurderingen av bildekvalitet som helhet var T2 eneste sekvens uten signifikante forskjeller mellom senter ($p = 0,318$). Det var imidlertid signifikante forskjeller mellom radiologenes vurderinger ved flere senter for denne sekvensen. Alle de andre sekvensene hadde signifikante forskjeller mellom senter ($p = 0,000-0,008$), og blant disse var det PD-sekvensen som hadde færrest senter med signifikant diskrepans mellom radiologvurderinger. Det var oftest signifikant ulike vurderinger for senter 6, og sjeldnere for senter 3 i denne spørsmålskategorien.

Tabell XXIII: Samlet vurdering av sentre med variasjonsbredde av vurderinger, samt median og typetall for score innad i hver spørsmålsgruppe. Signifikansverdier for senter og radiolog er beregnet ved hjelp av Kruskal-Wallis test.

Sekvens og kvalitetsaspekt	Median / Typetall	Score-bredde ⁽¹⁾	Senter med radiolog-diskrepans	p-verdi senter
Sag 3D T1				
1. Differensiering*** vev og strukturer	4 / 4	2–5	S2*, S3*, S6*	0,000
2. Artefakter	5 / 5	2–5	Ingen senter ($p = 0,174-0,329$)	0,993
3. Helhetlig bildekvalitet***	4 / 4	2–5	S5*, S6*, S7*, S8*	0,000
Tra PD TSE				
1. Differensiering vev og strukturer	4 / 4	2–5	Ingen senter ($p = 0,052-0,937$)	0,177
2. Artefakter	5 / 5	3–5	S2*, S3*, S6*, S7*	0,949
3. Helhetlig bildekvalitet**	4 / 4	3–5	S4*, S6*	0,003
Tra T2 TSE				
1. Differensiering vev og strukturer	4 / 4	2–5	Alle unntatt S5 ($p = 0,007-0,028$)	0,452
2. Artefakter	5 / 5	2–5	Alle senter ($p = 0,001-0,038$)	0,998
3. Helhetlig bildekvalitet	4 / 4	3–5	Alle unntatt 5 og 8 ($p = 0,011-0,040$)	0,318
Tra T2 FLAIR				
1. Differensiering* vev og strukturer	4 / 4	1–5	Alle unntatt S5 ($p = 0,005 - 0,049$)	0,015
2. Artefakter	5 / 5	3–5	Alle unntatt S5	0,976

			(p = 0,002–0,038)	
3. Helhetlig bildekvalitet**	4 / 4	2–5	S2*, S4*, S5*, S6**	0,008
Tra T2* hemo				
1. Differensiering vev og strukturer	4 / 4	1–5	Alle senter (p = 0,005–0,36)	0,807
2. Artefakter	5 / 5	3–5	S1**, S3*, S5**, S6*, S7*	0,646
3. Helhetlig bildekvalitet**	4 / 4	3–5	Alle unntatt S3 og S7 (p = 0,004–0,049)	0,003

⁽¹⁾Spørsmålene ble besvart på en skala 1–5 hvor 1 var dårligst og 5 var best.

* = p < 0,05, ** = p < 0,01, *** = p < 0,001.

Samlet inntrykk

På tvers av senter var det mest samsvar mellom radiologenes vurderinger for 3D T1- og PD-sekvensen. På tvers av sekvenser var det opptakene fra senter 5 og 8 som hadde mest samsvar mellom radiologenes vurderinger.

Alle sentrene ble rangert etter samlet score, hvor både samlet poengsum, gjennomsnitt, median, typetall og variasjonsbredde ble tatt med i rangeringen. Ulike senter skilte seg ut i positiv og negativ retning avhengig av kategori og sekvens. En oversikt over de to sentrene som hadde henholdsvis høyeste og laveste samlede score i hver kategori og sekvens er gitt i Tabell XXIV. Senter 1 og 5 skilte seg ut med like mange høye som lave utmerkelse. Senter 2 og 6 hadde flest høye score, og senter 6 hadde ingen laveste score. Senter 8 hadde flest lave score, og ingen høyeste score. Delkapittel 4.4 *Objektiv versus subjektiv bildekvalitet* med underpunkter tar opp flere enkeltaspekter omkring de subjektive vurderingene.

Tabell XXIV: Oversikt over hvilke senter som fikk henholdsvis høyest og lavest samlet score i hver spørsmålskategori.

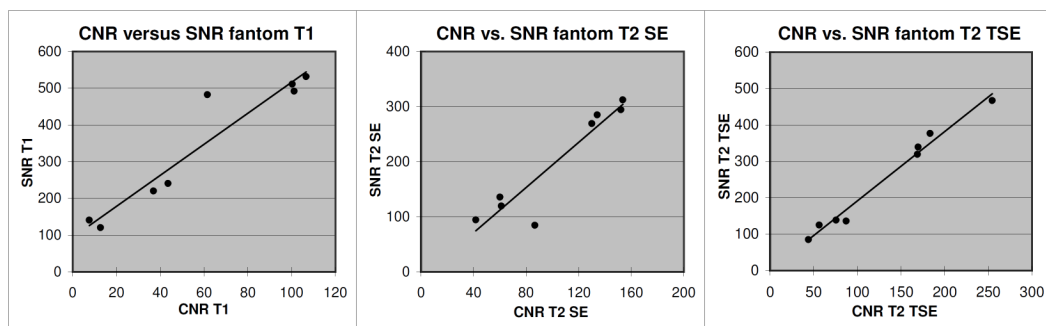
Sekvens	Differensiering vev og strukturer		Artefakter		Helhetlig bildekvalitet	
	Høyest score	Lavest score	Høyest score	Lavest score	Høyest score	Lavest score
Sag 3D T1	S1, S2, S3	S7, S8	S1, S2, S6	S5, S8	S1, S2	S5, S8
Tra PD TSE	S3, S6	S1, S8	S2, S5	S3, S4	S3, S6	S1, S8
Tra T2 TSE	S2, S6	S1, S8	S2, S6	S5, S8	S2, S6	S1, S5
Tra T2 FLAIR	S3, S6	S1, S8	S1, S2	S5, S8	S2, S5	S4, S8
Tra T2* hemo	S4, S5	S2, S3	S1, S5	S2, S8	S1, S5	S2, S4

4.2 Tomograf-resultater i forhold til utstyr og andre aspekter

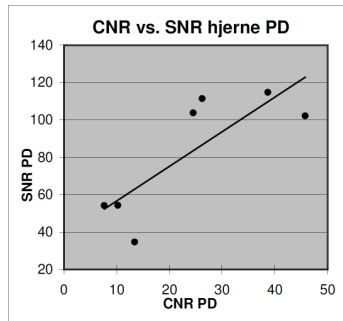
4.2.1 Ulike kvalitetsaspekter i forhold til hverandre

Som nevnt kan ulike problemer med tomografen ha mange årsaker. Derfor ble alle de testede kvalitetsaspektene (Tabell IV) testet mot hverandre med tanke på om resultatene fra ett aspekt kunne ha betydning for utfallet av et annet. Siden dette prosjektet hadde et lite utvalg med få måleverdier fra hvert senter, ble ett og ett aspekt testet mot hverandre ved hjelp av enkel lineær regresjonsanalyse, signifikansnivå 0,05. De eneste to kvalitetsaspektene med signifikant korrelasjon, var SNR og CNR (Figur 27). Alle sekvensene i fantomopptakene hadde positiv korrelasjon mellom CNR og SNR ($r = 0,95, 0,94$ og $0,99$, justert $r^2 = 0,89, 0,86$ og $0,98$ for henholdsvis T1-, T2 SE- og T2 TSE-sekvensen). Korrelasjonene var signifikante med $p < 0,001$ og $F = 58,0, 42,7$ og $330,3$ for hver av de tre sekvensene.

Sammenhengen mellom CNR og SNR i hjerneopptakene var ikke like entydig. Det var positiv korrelasjon for alle sekvensene, men kun PD-sekvensen var signifikant ($p = 0,028, F = 9,4$) (Figur 28). For de andre sekvensene i hjerneopptakene var $p = 0,07-0,09$.



Figur 27: Spredningsplott med tegnede regresjonslinjer mellom CNR og SNR i henholdsvis T1-, T2 SE- og T2 TSE-sekvensen i fantomopptakene. Merk at skalaen langs x- og y-aksen er tilpasset sekvensene.



Figur 28: Spredningsplott med tegnet regresjonslinje mellom CNR og SNR i PD-sekvensen.

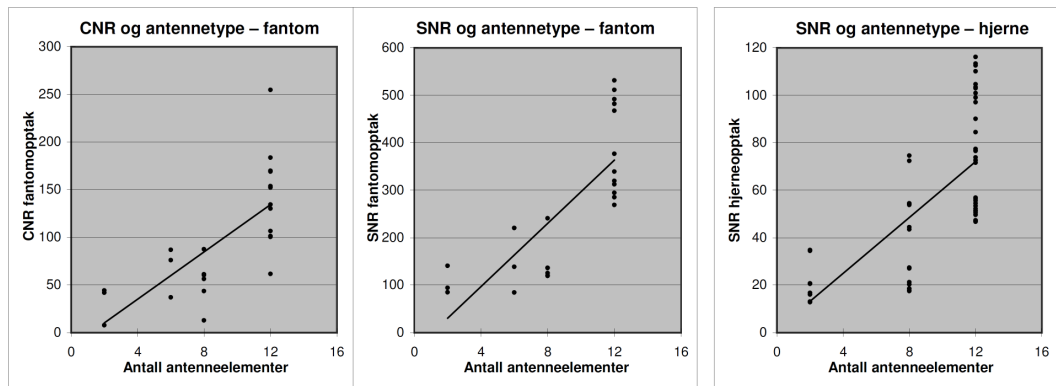
4.2.2 Ulike kvalitetsaspekter i forhold til utstyr

Utstyr – gradientsystem og antenntype

Betydning av forskjeller i gradientstyrke og antenntype ble testet mot de ulike kvalitetsaspektene (Tabell IV). Med gradientstyrke menes gradientamplitude og slew rate. Rise time ble ikke testet separat da denne avhenger av de to førstnevnte. Testing av antenntype ble avgrenset til antall kanaler. Blant sentrene i dette utvalget var det ett senter med 6-kanalers hodespole, ett senter med CP hodespole (2 kanaler), to senter med 8-kanalers hodespole og fire senter med 12-kanalers hodespole. (jf. Tabell II). Få sammenhenger ble funnet. Antenntype så ut til å betyding for SNR og CNR, og gradientsystem hadde ingen signifikant betydning for noen av de testede kvalitetsaspektene.

Antenntype

Mulige sammenhenger mellom antenntyper og bildekvalitetsaspekter ble kun funnet for SNR og CNR i fantomopptak, og kun for SNR i hjerneopptak. Figur 29 viser spredningsplott og tegnet regresjonslinje for disse tre. I fantomopptak er alle sekvensene (T1, T2 SE og T2 TSE) samlet i samme figur. I hjerneopptak er alle de transversale sekvensene (T2, PD, T2 FLAIR og T2*) samlet i samme figur. Korrelasjonen var signifikant ($p < 0,0001$, $F = 23,9-47,0$) for alle plottene.



Figur 29: Spredningsplott med tegnede regresjonslinjer mellom beregnet CNR og SNR og antall antennekanaler.

Til tross for at regresjonsanalysen viste signifikant positiv korrelasjon, var det ikke signifikant forskjell mellom SNR-verdi og antenntype for alle typene. I fantomopptakene var det ikke signifikant forskjell i CNR-verdi mellom CP-, 6-kanalers og 8-kanalers hodespole. CNR-verdier beregnet i 12-kanalers hodespoler skilte seg signifikant fra de andre typene ($p < 0,0001$, $F = 32,2$). SNR-verdier var signifikant ulike i ulike antenntyper i fantomopptakene ($p < 0,0001$, $F = 43,4$). I hjerneopptakene var det derimot kun forskjell i SNR-verdi mellom 8- og 12-kanalers hodespole. ($p < 0,0001$, $F = 26,6$). SNR-verdier i CP- og 8-kanalers hodespole var ikke signifikant ulike ($p = 0,067$).

4.3 Objektive målinger i fantomopptak versus hjerneopptak

Som tidligere nevnt ble en dobbelekk TSE-sekvens gjort både i fantom og frisk frivillig. Dermed var det T2 TSE-sekvensen som var mest aktuell for sammenligning av de kvalitetsaspektene som ble målt både i fantom- og hjerneopptak.

4.3.1 SNR og CNR

Som sagt ble ikke SNR-verdier i hjerneopptak beregnet ved senter 2, så regresjonsanalyse ble utført ved de resterende sentre. Regresjonsanalyse viste at det var signifikant positiv korrelasjon mellom beregnet SNR i fantomopptak og i hjerneopptak ($p = 0,003$, $F = 29,0$, $r = 0,92$, justert $r^2 = 0,82$). Sammenhengen var signifikant, men ikke like tydelig for CNR ($p = 0,02$, $F = 11,0$, $r = 0,83$, justert $r^2 = 0,63$).

4.3.2 Ghosting

Det var liten grad av ghosting i alle sekvensene, både i fantom- og hjerneopptak. Dette var altså en parameter med liten betydning for bildekvaliteten. I tillegg viste regresjonsanalyse ingen sammenheng mellom ghosting i fantom- og hjerneopptak ($r = 0,01$, $p = 0,99$, $F = 0,0$).

4.3.3 T2-måling

Målingene i fantomopptak hadde generelt mindre avvik fra reell verdi enn målingene i hjerneopptak. I fantom var gjennomsnittlig avvik $5,2 \pm 7,2$ ms for 135 ms, og $2,4 \pm 1,8$ ms for 70 ms. I hjerneopptakene var det verdier målt i hvitsubstans frontalt som var nærmest normalverdi, med gjennomsnittlig avvik $18,3 \pm 6,1$. Thalamus var eneste gråsubstansområde hvor T2-verdier ble målt, og gjennomsnittlig avvik var her $6,2 \pm 6,7$. Det var imidlertid ingen sammenheng mellom målte verdier i fantom og hjerne ($p = 0,06$, $F = 3,8$, $r = 0,34$, justert $r^2 = 0,08$).

4.4 *Objektiv versus subjektiv bildekvalitet*

De subjektive vurderingene ga data på ordinalnivå. Dermed kunne ikke objektive og subjektive mål sammenlignes direkte med regresjonsanalyse. Testene ble derfor sammenholdt i forhold til om de pekte mot de samme tomografene i positiv og negativ retning. Følgende objektive og subjektive kvalitetsaspekter ble vurdert opp mot hverandre:

- SNR-måling versus SNR-vurdering
- CNR, lavkontrastdetaljer og romlig oppløsning versus kontrast/skarphet og differensiering av ulike vev og strukturer
- Ghosting versus evaluering av forstyrrelser fra bevegelse, flow, blurring og stripeartefakt

4.4.1 Signal/støy forhold

Objektiv måling av SNR

De ulike sekvensene sett under ett, var det i fantomopptakene senter 1 og 8 som hadde lavest SNR, mens senter 3, 4, 6 og 7 hadde høyest SNR. I hjerneopptakene var det senter 8 som hadde lavest SNR, og senter 3, 4 og 6 som hadde høyest SNR.

Subjektiv vurdering av SNR

Når det gjaldt subjektiv vurdering av SNR i hjerneopptakene, kom senter 8 dårligst ut, mens senter 2 kom best ut på tvers av sekvenser. Enigheten blant radiologene var imidlertid signifikant ulik ved senter 2 ($p = 0,043$).

4.4.2 Vevskontrast og differensierbarhet

Testing av lavkontrastdetaljer i fantomopptak

På T1 fikk senter 5 og 8 lavest score på dette kriteriet, mens senter 2, 3 og 6 fikk høyest score. På T2 SE fikk senter 5 lavest score, og senter 3, 6 og 7 fikk høyest score. På T2 TSE fikk senter 4 og 8 lavest score, men senter 1, 2 og 6 fikk høyest score.

Testing av romlig oppløsning i fantomopptak

Senter 4 og 7 skilte seg ut med lavest romlig oppløsning på T2 TSE. Senter 2, 5 og 8 skilte seg ut med høy oppløsning i alle sekvensene.

Objektiv måling av CNR

De ulike sekvensene sett under ett, var det i fantomopptakene senter 1, 2, 5 og 8 som hadde lavest CNR, mens senter 3, og 7 hadde høyest CNR. I hjerneopptakene var det senter 5 og 8 som hadde lavest CNR, og senter 3 og 7 som hadde høyest CNR. Senter 1 utmerket seg i hjerneopptakene med noen av utvalgets laveste verdier for PD- og T2-sekvensen, men noen av utvalgets høyeste verdier for T2*- og 3D T1-sekvensene.

Subjektiv vurdering av kontrast og skarphet i hjerneopptak

Alle sekvensene sett under ett, kom senter 8 dårligst ut i vurderingen av kontrast og skarphet i hjerneopptakene, mens senter 6 kom best ut. Det var enighet blant radiologene i denne vurderingen ($p = 0,38$ for senter 8 og $0,09$ for senter 6).

Subjektiv vurdering av ulike vev og strukturers differensierbarhet

Differensierbarheten av ulike vev og strukturer ble ulikt vurdert i forhold til struktur og sekvens ved de ulike sentrene. Det var imidlertid noen senter som skilte seg ut ved at de hadde fått flere lave vurderinger enn andre. Ved å finne snittverdien til alle sekvensene sammenlagt ved hvert senter i form av typetall og median, kom det frem at senter 8, 1 og 5 hadde blitt tildelt flest vurderinger lavt på skalaen. De sentrene som hadde fått tildelt best score i vurderingen av differensierbarhet, var senter 2, 3 og 6. Unntaket var T2*-sekvensen, hvor senter 2 og 3 kom dårligst ut på differensierbarhet.

4.4.3 Ghosting versus generelle forstyrrelser og artefakter

Objektiv måling av ghosting

Senter 7 og 8 hadde mest ghosting i fantomopptakene. I hjerneopptakene hadde senter 5 mest ghosting på grunn av den høye målingen i PD-sekvensen. Ellers var det ingen senter som skilte seg ut.

Subjektiv vurdering av ghosting

Det som var mest sammenlignbart med ghosting blant de subjektive målene, var bevegelse, flow, blurring og stripeartefakt. Det er kun de to første som kan kalles ghosting i ordets rette betydning, men alle disse aspektene ble sett på i forbindelse med vurdering av ghosting fordi det kan se likt ut i bildet.

Vurdering av bevegelse: senter 1 og 2 kom best ut, men alle sentrene ble vurdert ganske likt. Enigheten blant radiologene var signifikant ulik ved senter 1 ($p = 0,006$).

Vurdering av flow: senter 1 og 2 kom best ut, men alle sentrene ble vurdert ganske likt. Enigheten blant radiologene var signifikant ulik ved senter 1 ($p = 0,026$).

Vurdering av blurring: senter 4 og 8 kom dårligst ut, men alle sentrene ble vurdert ganske likt. Det var enighet blant radiologene i denne vurderingen ($p = 0,429-0,926$).

Vurdering av stripeartefakt: senter 8 kom dårligst ut, og senter 2 ble vurdert som best. Enigheten blant radiologene var imidlertid signifikant ulik ved senter 2 ($p = 0,026$).

4.5 Beregning av måleusikkerhet og testing av bildekvalitet over tid

4.5.1 Måleusikkerhet

Som tidligere nevnt ble de følgende tre punkter testet ved ett senter (senter 6). Resultatene er presentert som Test 1–3 i Tabell XXV.

1. åtte gjentatte beregninger i samme opptak (Test 1)
2. åtte gjentatte opptak og beregninger uten reposisjonering i spolen, med samme snittplassering (Test 2)
3. åtte gjentatte opptak og beregninger etter reposisjonering i spolen med ulike personer som gjorde reposisjoneringen (Test 3)

Tabell XXV: Gjennomsnitt og standardavvik for måleusikkerhet. Test 1–3 kun gjort ved senter 6.

Kvalitetsaspekt	Test 1 ⁽¹⁾	Test 2 ⁽²⁾	Test 3 ⁽³⁾	Sentervariasjon
Geometri	189,4 mm ± 0,34	189,7 mm ± 0,02	189,2 mm ± 0,15	189,8 mm ± 0,24
Avviksstørrelse	0,9 mm ± 0,50	0,3 mm ± 0,02	0,8 mm ± 0,09	0,2 mm ± 0,19
Snittykkelse T1	5,2 mm ± 0,05	5,3 mm ± 0,07	5,3 mm ± 0,08	5,2 mm ± 0,24
Avviksstørrelse	0,2 mm ± 0,05	0,3 mm ± 0,07	0,3 mm ± 0,08	0,3 mm ± 0,13
Snittposisjon				
T1 snitt #1	0,0 mm ± 0,04	0,4 mm ± 0,05	0,7 mm ± 0,46	0,7 mm ± 0,73
T1 snitt #11	0,9 mm ± 0,06	0,4 mm ± 0,05	0,5 mm ± 0,38	0,7 mm ± 0,73
Signalhomogenitet				
T1	93,8 % ± 0,14	94,4 % ± 0,28	93,4 % ± 0,40	86,7 % ± 10,41
Ghosting				
T1	0,08 ± 0,02	0,06 ± 0,03	0,05 ± 0,03	0,10 ± 0,10
T2 TSE	0,10 ± 0,02	0,14 ± 0,03	0,15 ± 0,04	0,96 ± 0,88
Lavkontrastdetaljer				
T1	37,4 ± 1,34	37,0 ± 0,00	35,0 ± 1,47	32,9 ± 4,97
T2 TSE	24,2 ± 1,79	16,0 ± 1,70	11,8 ± 0,29	19,0 ± 6,42
SNR T1	536,3 ± 6,02	514,5 ± 3,61	537,0 ± 14,29	342,3 ± 177,83
SNR T2 TSE	341,1 ± 3,62	323,8 ± 6,28	327,8 ± 22,71	248,5 ± 143,48
CNR T1	113,5 ± 4,74	98,0 ± 9,26	119,4 ± 35,55	58,8 ± 40,11
CNR T2 TSE	172,7 ± 3,96	160,6 ± 7,95	170,9 ± 20,19	130,0 ± 74,68
T2-måling				
70 ms	68,5 ms ± 0,12	67,4 ms ± 0,13	68,1 ms ± 0,09	67,9 ms ± 2,14
135 ms	133,3 ms ± 0,00	131,3 ms ± 0,36	132,1 ms ± 0,07	135,6 ms ± 9,10

⁽¹⁾Test 1: flere beregninger i samme opptak. ⁽²⁾Test 2: flere opptak uten reposisjonering. ⁽³⁾Test 3: flere opptak med reposisjonering.

4.5.2 Bildekvalitet over tid (4–6 måneder)

Beregninger fra andre omgang med fantomopptak viste at det var de samme sentrene som skilte seg ut i positiv og negativ retning. Sentrene hadde også samme resultat i forhold til ACR-kriterier som man så i første omgang. Det som varierte mest var SNR-verdi. Dette var også den parameteren som varierte mest i forhold til måleusikkerhet. Regresjonsanalyse viste imidlertid godt samsvar i SNR-målinger over tid, med $r = 0,98$, justert $r^2 = 0,95$, $p < 0,0001$ og $F = 127,4$.

5 Diskusjon

5.1 *Diskusjon omkring funn i de ulike kvalitetsaspektene*

Variasjonen mellom sentrene var ulik for de ulike kvalitetstestene. Ved de geometriske testene gradientlinearitet, snittykkelse og snittposisjon (punkt 4.1.3) var det signifikante forskjeller mellom sentrene, men variasjonene var innenfor ACR-grensene. I testen for romlig oppløsning (punkt 4.1.2), var det ikke signifikante forskjeller mellom sentrene. Disse aspektene blir dermed ikke så mye omtalt i diskusjonsdelen, da oppgaven tar sikte på å peke på mer utpregede ulikheter som kan være av klinisk betydning.

5.1.1 Signalhomogenitet (U)

Som tidligere vist, var signalfordelingen forskjellig i ulike spoletyper (Figur 13). Beregningen av U er imidlertid uavhengig av signalfordelingen i spolen. Siden man måler i området med lavest signalintensitet samt området med høyest signalintensitet og beregner prosentforholdet mellom disse, vil antenner med ulike mønster av signalrespons stille likt, og lav U avsløres uavhengig av signalfordelingen i spolen. Det kunne se ut til at høy grad av inhomogeniteter kan ha betydning for kvantitativ bildeanalyse. Senter 1, 2 og 5 skilte seg negativt ut i denne testen, og disse tre sentrene skilte seg også ut fra de andre sentrene i NeuroQuant®-beregningene.

Signalhomogenitet så ikke ut til å ha betydning i visuell vurdering. Senter 1 og 5 hadde mange både høye og lave score, mens senter 2 skilte seg ut som ett av sentrene med høyest score (Tabell XXIV). Dette kan tyde på at dette ikke er et aspekt som har betydning for radiologenes evne til å differensiere strukturer, i hvert fall ikke med $U > 70\%$ som var tilfelle i dette prosjektet.

Etter min mening var det alvorlig at tre av åtte senter havnet så langt under ACR-grensen i denne testen. Utvalget var imidlertid for lite til å kunne si noe om generaliserbarheten av dette resultatet. Dersom en utvalgstilfeldighet har inkludert flere senter med lav U enn ”normalt”, utgjør det en større andel i et lite utvalg. Beregningen av KI_{95} for andelen av tomografer som ikke nådde ACR-grensen for U ga også lite i et så lite utvalg.

Alle sentrene som kom dårlig ut i denne testen lå langt under ACR-grensen, og det var samsvar mellom første og andre opptak. Dette tyder på at resultatet ikke skyldes målefeil eller tilfeldigheter, men en reell svikt i antennen eller et annet sted i signaloverføringen. Tekniske tester må gjøres av leverandøren for å finne feilen.

5.1.2 Signal/støy forhold

SNR-verdiene var varierende mellom tomografer. Både i fantom- og hjerneopptak var variasjonsbredden større enn det som er rapportert i andre studier. Blant annet testet Chen et al. (2004) tre 1,5 T tomografer, og variasjonsbredden i SNR-verdier målt i ACR-fantom var 205–425, mot 85–530 i dette prosjektet.

De signifikante ulikhetene i SNR beregnet i fantom ved hvert senter kan ikke forklares ved hjelp av ulik signalfordeling i ulike spoletyper. Målingene ble gjort i snitt #7, og ROI dekket 70 % av fantomets tverrsnitt (Figur 9 – høyre). Dermed ble både områder med lav og høy signalintensitet tatt med i signalberegningen. Årsaker til at variasjonsbredden er større i dette prosjektet enn i andre studier, kan være at enkelte studier har sammenlignet få tomografer, og at de testede tomografene ikke var så ulike i type og alder som tomografene i dette prosjektet.

I hjerneopptakene ble ROI_{signal} lagt i hvit substans i sentrum semiovale (Figur 9 – venstre). Flere andre har brukt dette området til å måle signalintensitet i hjernen i forbindelse med beregning av SNR, deriblant Gedamu, Collins og Arnold (2008). Når pasienten er plassert korrekt i hodespolen ligger dette området omtrent midt i spolen. I tillegg er det såpass høyt i hjernen at det ligger over øyne, ventrikler og store blodkar. Dermed er det et område lite berørt av bevegelse og flow som kan forstyrre målingene.

Det var positiv korrelasjon mellom SNR beregnet i fantom og i hjerne på T2 TSE-sekvensen. Dermed så det ut til at fantomberegninger kan benyttes til å forutsi SNR-verdier i hjerneopptak. Det var ingen direkte lineær sammenheng, ved noen senter var forskjellen mellom SNR i hjerne- og fantomopptak stor, og mindre ved andre. En av årsakene til dette kan være at i fantomopptaket ble hele fantomet dekket av en stor ROI ($\sim 200 \text{ cm}^2$), mens i hjernen ble en liten ROI ($\sim 1 \text{ cm}^2$) lagt i et homogent område. Som

tidligere vist var signalfordelingen ulik i ulike antenntyper, og dermed kan en liten ROI bli offer for signalinhomogeniteter i antennen.

Senter 8 kom dårligst ut både i målt og subjektivt vurdert SNR. I subjektive vurderinger kom senter 2 best ut, og dette senteret manglet objektiv måling av SNR i hjerneopptak. Nest best i subjektive vurderinger var senter 3 og 6, og disse sentrene var blant de med høyest målte SNR-verdi. Dette viser at variasjonen i SNR-verdier var så stor at den kunne avdekkes ved visuell vurdering av bildene.

Det kunne se ut til å være en sammenheng mellom antall spoleelementer og SNR. Utvalget var imidlertid for lite til å si noe sikkert. Med data fra ett senter med en CP hodespole, og ett med en 6-kanalers hodespole, kunne man ikke trekke noen sikre slutninger fra disse to. Når det gjelder 8 og 12 kanaler var det litt flere av hver type, henholdsvis to tomografer med 8-kanalers hodespole og fire tomografer med 12-kanalers hodespole. Her så det ut til å være en sammenheng. I dette prosjektet var alle de fire spolene med 12 kanaler av samme type. Dette er av betydning, fordi selv om to spoler har like mange kanaler, trenger de ikke være like. For eksempel kan elementenes størrelse være ulik, og dette har betydning for antennens rekkevidde. I tillegg kan elementene være koblet sammen på forskjellige måter i opptaket. For med sikkerhet å fastslå betydningen av antall kanaler, bør man derfor ha data fra flere ulike spoletyper med like mange kanaler.

5.1.3 Kontrast/støy forhold

Det ligger mer enn CNR i begrepet differensierbarhet, men det kunne likevel se ut til å være en sammenheng mellom målt CNR og visuell vurdering av ulike vev og strukturer. Senter 5 og 8 kom dårlig ut både i målt CNR og i subjektiv vurdering av differensierbarhet. I subjektiv vurdering av differensierbarhet kom senter 2, 3 og 6 best ut. Senter 2 manglet objektiv måling av CNR i hjerneopptak, og her kom senter 3 og 7 best ut.

Det var positiv korrelasjon mellom CNR beregnet i fantom og i hjerne på T2 TSE-sekvensen. Dermed så det ut til at fantomberegninger kan benyttes til å forutsi CNR-verdier i hjerneopptak. Det er imidlertid ikke alle sekvenser som kan forventes å ha like

høy CNR i samme områder. For eksempel er T2 FLAIR-sekvensen laget for å visualisere hvit substanslesjoner og vaskulære lesjoner. Selv om CNR mellom grå og hvit substans er lav i en FLAIR-sekvens, vil CNR mellom de nevnte lesjoner og omkringliggende vev være høy. Dette kan imidlertid ikke vurderes med bildeopptak av en frisk hjerne, og derfor ble det kun gjort målinger av CNR mellom grå og hvit hjernesubstans.

5.1.4 Lavkontrastdetaljer

I følge Ihalainen, Lönnroth, Peltonen, Uusi-Simola, Timonen, Kuusela et al. (2011), har SNR direkte innvirkning på lavkontrastdetaljer. Dette var imidlertid ikke tilfelle i dette prosjektet. Studien av Chen et al. (2004) fant heller ingen sammenheng mellom SNR og lavkontrastdetaljer.

Det var stort sprang mellom høyeste og laveste score i denne testen. På tvers av sekvenser var variasjonsbredden i dette prosjektet 11–40. Til sammenligning var variasjonsbredden i Chen et al. (2004) 11–39. Alle sentrene var innenfor ACR-kriteriet (> 9). ACR har imidlertid ikke opplyst om denne grensen er satt for at flest mulig tomografer skal klare testen, eller om det er en grense på når bildekvaliteten blir for dårlig til å se ønskede detaljer i hjerneopptak. På tvers av sekvenser kom senter 4, 5 og 8 dårligst ut i denne testen, mens senter 1, 2 og 6 kom best ut. Dette resultatet samsvarer ikke med resultater fra visuell vurdering eller volumetri. Som sagt var hjerneopptakene uten patologi, og et annet resultat kan ikke utelukkes dersom det hadde vært patologi til stede i hjernebildene.

5.1.5 Ghosting

Ulike sekvenser har ulik grad av ghosting. Dette kommer blant annet av ulik følsomhet for bevegelse, ulik grad av vibrasjon, samt ulike måter å samle inn signal på. På grunn av at det er mer bevegelse i en hjerne, som for eksempel øyebevegelser, pustebevegelser og flow fra blodgjennomstrømning og ventrikler, blir det ikke helt riktig å sammenligne ghosting fra fantomet med ghosting i hjernen. Det ble også gjort opptak med andre sekvenser i hjernen enn i fantom, bortsett fra dobbelekk TSE. Man kan gå ut fra at siden det er samme hjerne som er avbildet ved alle sentrene, vil flow være omtrent den

samme. Bevegelse er derimot noe mer tilfeldig. I dette prosjektet var det liten grad av ghosting både i fantom- og hjerneopptak. Variasjonsbredde på 0,04–0,32 % i fantomopptak var lavere enn hos Chen et al. (2004) (0,04–0,80 %).

5.1.6 T2-måling

Det er forskjell på grå substans ulike steder i hjernen. For eksempel har grå substans i thalamus ikke samme sammensetning som grå substans i hjernebarken. Thalamus er imidlertid den største samlingen av grå substans på ett sted i hjernen, og egner seg derfor til plassering av ROI. Hensikten med denne testen var ikke å beregne T2-tiden i den avbildede hjernen, men å teste om T2-tider var reproduserbare i ulike tomografer. Selv om verdiene var signifikant ulike på tvers av tomografer, var KI₉₅ overlappende ved mange senter (Figur 26).

Med noen få unntak var beregnet T2-verdi høyere i hvit enn i grå substans. I teorien skulle hvit substans være ~80 ms, og grå substans ~100 ms. Bauer et al. (2010) fant i sin studie at tomografer fra Siemens hadde en tendens til å beregne høyere T2-verdier enn GE og Philips (Figur 8). I dette prosjektet var det bare én tomograf fra henholdsvis GE og Philips, men disse to hadde gjennomgående lavere T2-verdier i hvitsubstans enn det som ble målt på Siemens. I thalamus var GE lavest, deretter kom to målinger fra Siemens (senter 4 og 7) foran verdien beregnet på Philips. Bauer et al. (2010) beregnet histogrammer for hjernen som helhet, mens i dette prosjektet ble det målt T2-verdi i ulike steder i hjernen. Til tross for et lite utvalg i dette prosjektet, og beregninger i kun én hjerne, kan det se ut til å være samsvarende funn. I følge Leppert et al. (2009), gjengir tomografer fra GE T2-tider systematisk lavere enn Siemens tomografer. Dette var i samsvar med funnene fra hjerneopptak i dette prosjektet, men stemte ikke like godt for fantomopptak.

Henfall av signal følger en eksponentiell kurve. Programvaren som benyttes til å lage T2-kart lager en så god tilnærming til denne kurven som mulig ut fra tilgjengelige opplysninger. Når beregningene gjøres på en dobbelekkosekvens har programmet bare to TE å beregne T2-kurven fra. Som tidligere nevnt, vil en sekvens som benytter flere TE i opptaket gi en mer nøyaktig kurve, men det ble ikke gjort i dette prosjektet. En slik sekvens kan ikke benyttes til annet enn T2-måling. Den sekvensen som ble benyttet i

dette prosjektet skal i utgangspunktet være god nok til T2-målinger, og i tillegg har den strukturell nytteverdi. Flere artikkelforfattere, deriblant Bauer et al. (2010), har tatt til orde for at T2-målinger bør kunne gjøres på en alminnelig dobbelekkosekvens. Registrering av T2-tider er interessant i mange sykdomsforløp, deriblant hos deltakerne i PRODE MR. Det var derfor av interesse å teste reproduserbarheten av T2-målinger på tvers av tomografer med dobbelekkosekvensen fra PRODE studien.

Det kunne vært nyttig å teste reproduserbarheten av T2-målingene i hjerne ved hver enkelt tomograf. Dette lot seg ikke gjøre siden man bare hadde ett hjerneopptak fra hvert senter. Det kunne også vært nyttig å ha gjort opptak med en sekvens med flere enn to TE og sammenlignet resultatene fra de to. Dette har imidlertid vært gjort i tidligere studier, deriblant Leppert et al. (2009), så det ville ikke vært avgjørende i så måte.

I følge McRobbie et al. (2007) er det flere feilkilder ved T2-måling. For eksempel kan ujevne snittprofiler og partiell volum effekter minske målenøyaktigheten. Beregnet T2-tid kan også være avhengig av hvor i antennen beregningene gjøres, ulik grad av signal og støy kan gjøre T2-beregningen mer unøyaktig. Det var imidlertid ingen sammenheng mellom beregnet snittykkelse eller snittposisjon og T2-målinger. Variasjonen i beregnede T2-verdier hadde heller ingen sammenheng med variasjoner i SNR-verdi eller U .

5.1.7 Usikkerhetsfaktorer i volumetri

Alle målemetoder har fordeler og ulemper, og automatiske metoder er ikke nødvendigvis bedre enn manuelle. Det er raskt og effektivt, men det er for eksempel ikke mulig å gjøre manuelle endringer i de segmenterte strukturene fra NeuroQuant®-analysen, og det kan være vanskelig å verifisere resultatet av segmenteringen. I følge Vangberg (2011), som blant annet har benyttet NeuroQuant® og FreeSurfer i behandlingen av MR-opptak i Tromsøundersøkelsen, kan NeuroQuant® ha en tendens til å overestimere små volum og underestimere større volum. Dette skulle imidlertid ikke ha betydning i dette prosjektet, da volumene er fra samme hjerne og burde bli like store, uavhengig av tendenser til over- eller underestimering. Når NeuroQuant® benyttes i pasientstudier er det vanlig å benytte relative volumer i dataanalysen, siden

ulike hjerner er ulik i størrelse. Dersom opptak fra ulike tomografer gir ulik segmentering, vil ikke dette kunne fjernes ved bruk av relative størrelser.

Bortsett fra sykdommer som medfører endring av selve hjernestrukturen (for eksempel hjernetumor), er bevegelse og artefakter de største årsakene til feil i den automatiske segmenteringsprosessen. Opptaket må stemme overens med templatet NeuroQuant® er utstyrt med, og derfor må hele hjernen være avbildet, og bildet må være uten aliasing. I diagnostisering av Alzheimers sykdom er NeuroQuant® kjent for å være robust. Det skal mye til før vev og strukturer ikke blir identifisert, og økt generell hjerneatrofi i forhold til alder og kjønn kan detekteres tidlig (McEvoy & Brewer 2010). Dette er imidlertid ingen garanti for at beregnede størrelser er korrekte. Hjernesegmenteringen fremstilles som fargebilder i tynne snitt gjennom hele hjernen. Denne visualiseringen gjør at man kan bla gjennom serien og kontrollere segmenteringen visuelt. I tillegg til morfologisk informasjon, sammenligner programmet den aktuelle hjernen med normale volumstørrelser i hjerner i samme aldersgruppe. Dersom normalvariasjon er større enn eventuelle strukturelle endringer, er man helt avhengig av oppfølging av pasienten over tid for å kunne si noe om eventuelle endringer som følge av sykdom.

5.1.8 Sentre som utmerket seg

Fishers LSD-test ble valgt som post-hoc test for å teste om noen av de samme sentrene skilte seg ut i de ulike kvalitetstestene. Håpet var at noen av de samme senterparene skulle forekomme i flere av testene, eller at ett senter stadig skulle skille seg signifikant fra resten. Det var imidlertid få tilfeller hvor kun ett senter eller senterpar skilte seg fra resten. Det var ingen klar sammenheng i hvilke senter som var sammenlignbare, sammenlignbare senter varierte både mellom tester og mellom sekvenser. I følge Peat og Barton (2005) krever LSD-test relativt få grupper, med noenlunde lik varians, noe som ikke var tilfelle for dataene i dette prosjektet. Det er mulig at en mer konservativ test som for eksempel Bonferronis post-hoc test ville vært mer passende til dette prosjektet.

Senter 1, 2 og 5 utmerket seg ved at de havnet betydelig under ACR-grensen for signalhomogenitet, og også langt under resultatet ved de resterende sentre i denne

testen. Sammen med senter 8 var det også Senter 1, 2 og 5 som skilte seg negativt ut når det gjaldt SNR. Sammen med senter 4 var det senter 5 og 8 som skilte seg ut negativt når det gjaldt lavkontrastdetaljer og snittykkelse. Senter 4 skilte seg ikke ut i noen retning i de andre testene. På snittposisjon var det igjen senter 1 og 5 som kom dårligst ut.

Det var altså senter 1, 5 og 8 som sammenlagt kom dårligst ut i ACR-testene. Senter 2 og 8 utmerket seg imidlertid blant annet som best på romlig oppløsning og geometrisk nøyaktighet. Senter 2 scoret også høyest på snittposisjon og lavkontrastdetaljer.

Det at noen senter scorer lavest på enkelte tester og høyest på andre tester, samsvarer blant annet med funn fra Bourel, Gibon, Coste, Daanen og Rousseau (1999). Dette kan være noe av årsaken til at man fant få sammenhenger mellom flere kvalitetsaspekter. Tanken med å teste ulike kvalitetsaspekter mot hverandre, var at dersom man fant sammenheng mellom to eller flere kvalitetsaspekter, kunne dette være en hjelp til å avgrense og konkretisere årsakene bak eventuelle problemer med bildekvaliteten. I tillegg kan det være nyttig å vite hvilke andre aspekter det kan gå utover dersom tomografen har et kjent problem med ett av kvalitetsaspektene.

På visuell vurdering fikk senter 2 og 3 flest høye score sammenlagt, mens senter 1, 5 og 8 fikk lavest score sammenlagt.

Senter 6 utmerket seg som generelt bra på alt. Det var dette senteret som ble valgt ut for tettere oppfølging for å se på stabiliteten til tomografen samt validiteten på målingene. Senter 6 ble valgt fordi det er den tomografen jeg har best tilgang til. Det var også et passende valg fordi senter 3, 4 og 7 hadde tilsvarende tomografer. Personlig hadde jeg forventet flere sammenlignbare verdier blant tomografene av samme type.

Generelt så det ut til at det var de eldste tomografene som klarte seg dårligst i dette prosjektet. Senter 5 og 8 var henholdsvis fra 2003 og 2002. Senter 1 og 2 var henholdsvis fra 2000 og 2001, men disse ble oppgradert i 2007 og 2010.

5.2 Generelle begrensninger

5.2.1 Størrelse på utvalg og datamateriale

Utvalget i dette prosjektet var relativt lite. Dette gjorde at det var flere tomografer som det kun var én type av i utvalget. I følge personlig e-post korrespondanse pr. 21. februar 2011 med medarbeidere i henholdsvis GE, Philips og Siemens, er fordelingen av 1,5 T tomografer i Norge omtrent slik: GE 16 %, Philips 30 % og Siemens 54 % av markedet (henholdsvis 19, 36 og 65 tomografer installert). I dette prosjektet var den prosentvise fordelingen slik: GE 12,5 %, Philips 12,5 % og Siemens 75 %. Skjevfordelingen i dette prosjektets materiale representerer altså fordelingen av tomografer i Norge ganske godt. For å få like mange tomografer fra hver leverandør må man altså ikke bare utvide utvalget, men også gjøre en mer strategisk utvelgelse, noe som ikke reflekterer virkeligheten i samme grad.

På grunn av det smale utvalget kunne det være vanskelig å avgjøre om resultater fra den enkelte tomograf var reelle eller tilfeldige, samt om tilsynelatende reelle funn skyldtes problemer med akkurat den tomografen, eller for modellen generelt. Dermed er en del av funnene i dette prosjektet ikke nødvendigvis gjeldende for en større gruppe eller en annen gruppe enn den utvalget representerer. Det er imidlertid svært vanskelig å finne identisk utstyr blant klinisk tilgjengelige 1,5 T tomografer. Foruten leverandør og modell, vil alder, oppgraderinger, utstyr, vedlikeholdsrutiner osv. være avhengig av lokale faktorer utenfor forskerens kontroll.

I tillegg til at det var et lite utvalg, var det også samlet inn relativt få data fra hvert senter. Dette skyldtes i hovedsak tiden man hadde til rådighet, spesielt tiden de ulike sentrene kunne avse til prosjektet. I en travel klinisk hverdag har sentrene begrenset kapasitet til å gjøre slike ekstra opptak. Også kostnadsbegrensninger omkring reiser og postforsendelser har spilt inn på datamengden. Det var minst reising og opptakstid involvert i forbindelse med fantomopptak. Derfor ble fantomavbildning gjennomført på to ulike tidspunkt ved hvert senter. Dette ga muligheten til å kontrollere funn gjort i første opptak og ble spesielt viktig for verifisering av senter som skilte seg ut, både i positiv og negativ retning.

For å kontrollere for eventuelle feil eller uregelmessigheter i alle tomografene kunne det vært gjort to fantomopptak rett etter hverandre ved hvert senter. Dette hadde også generert mer data. For å unngå å doble opptakstiden, kunne det vært et alternativ å gjennomføre noen utvalgte sekvenser to ganger i fantomet. Resultatene fra testing av måleusikkerhet ved ett senter ga imidlertid liten grunn til å tro at flere opptak etter hverandre kunne ha betydning for resultatet, og man valgte derfor å begrense seg til ett opptak ved hvert senter. Stabile resultater fra første og andre runde med fantomopptak bygger opp under denne antakelsen.

I den visuelle vurderingen ble det samlet inn data fra tre observatører. Det var flott at tre erfarne radiologer stilte gratis opp i en travel hverdag. Det er ingen garantier for at flere observatører ville gitt mer enhetlige data. Tanken med visuell vurdering var ikke en evaluering av subjektive meninger omkring bildekvalitet. Hvis det hadde vært det sentrale måtte man hatt flere respondenter. Wagner et al. (2010) brukte to radiologer sin vurdering av bildekvalitet i prostata.

Datamaterialet var stort nok til å avdekke forskjeller mellom tomografer. Faren for type I feil er imidlertid større ved et lite utvalg og få data. Man er mer tilbøyelig til å gjøre signifikante funn som egentlig ikke er der, enn å overse signifikanser. Derfor kan det hende at noen av ulikhetene som er signifikante i dette prosjektet ikke ville vært det dersom datamaterialet hadde vært større.

5.2.2 Valg av hypotesetester og signifikansnivå

5 % ble valgt som signifikansnivå. Ved enkelte av aspektene var det liten tvil om at det var signifikante forskjeller mellom tomografene. Dette gjaldt for eksempel for signalhomogenitet (Figur 12). Det er imidlertid alltid en mulighet for "falske signifikanser". Dette gjaldt spesielt i tilfeller hvor signifikansverdien var lav ($0,05 < p < 0,06$), men ikke svært lav (for eksempel $0,05 > p > 0,01$). Faren for å gjøre både type I og type II feil er størst i disse områdene.

Variablene ble testet hver for seg for å unngå multippel sammenligningsproblematikk. Dette gjorde at analysen tok lenger tid, men ga mer nøyaktige p-verdier for hvert enkelt

kvalitetsaspekt. Selv om ikke alle sammenligningene ga like entydige signifikansverdier, ble det avdekket at bildekvalitet som helhet var tomografavhengig.

5.2.3 Standardisering av protokoller

Det som i denne oppgaven har blitt omtalt som en standardisering av sekvenser som inngikk i PRODE MR, var egentlig mer en optimalisering og harmonisering av parametre på tvers av det utstyret vi hadde å jobbe med. Det er ikke noe som omtales som "PRODE-standard" i Norge, kun ved de åtte sentrene hvor denne studien er implementert, og da kun gjeldende for de pasientene som inngår i PRODE studien. Dersom det hadde vært andre senter med annet utstyr involvert i prosjektet, kan det ikke utelukkes at det ville vært nødvendig med endringer i forhold til de parametrene som ble benyttet i dette prosjektet. Alle hovedleverandørene av MR-utstyr i Norge var med i utvalget, men på langt nær alle modellene. Dette var derfor ikke ment å være en standard på tvers av alt utstyr som finnes i Norge, og resultatene er derfor heller ikke direkte overførbare til annet utstyr.

Selve optimaliseringsprosessen har ikke blitt tatt opp i særlig stor grad i denne oppgaven. Det er fordi hovedfokus for mastergradsoppgaven var selve kvalitetstesting og resultatet av denne. Optimaliseringsprosessen var naturligvis et viktig forstadium til kvalitetstesting, og ideen til mastergradsoppgaven hadde utspring i ønsket om å teste i hvilken grad optimaliseringen hadde lyktes. Det ble imidlertid vanskelig å si noe om dette før man visste mer om individuelle ulikheter blant tomografene de optimaliserte sekvensene skulle implementeres på. Derfor startet dette prosjektet som en kartlegging av tomografene i utvalget ved hjelp av avbildning av ACR-fantomet og hjernen til en frisk frivillig som et grunnlag for sammenligning. Da det viste seg at tomografene hadde større grad av individuelle ulikheter enn ventet, ble det behov for en mer systematisk kartlegging, og dette ble hovedfokus for masterprosjektet. Kvalitetstesting har også vært et selvstendig arbeid i mye større grad enn optimaliseringsprosessen var, og fyller derfor kriteriet om at mastergradsoppgaven skal være et individuelt arbeid bedre.

Siden ACR-fantomet kun inneholder strukturer til bruk i beregningen av ulike kvalitetsaspekter, trenger ikke opptaksprotokollen for fantom å optimaliseres på samme måte som for hjerne for å få like bilder. Siden grunnlaget for fantomsekvensene er SE-

sekvenser, er det heller ikke så mange valgmuligheter for ulike parameter som man har ved andre opptaksmetoder. TSE-sekvensen som ble gjort i fantom var den samme som ble gjort i hjerne i dette prosjektet for å ha et bedre sammenligningsgrunnlag mellom fantom- og hjerneopptak.

5.2.4 Måleusikkerhet

Estimert grad av måleusikkerhet burde vært beregnet ved alle sentrene. Som nevnt er total måleusikkerhet sammensatt av flere faktorer, deriblant posisjonering i spolen, snittplassering og manuell analyse. Dette kommer i tillegg til systemets egen måleusikkerhet, altså små variasjoner i hvert opptak. De tre førstnevnte faktorene vil trolig være relativt like, uavhengig av hvilket system de testes på. Men stabiliteten i selve opptaksprosessen vil variere mellom tomografer. Estimert måleusikkerhet ble kun beregnet ved ett senter, hovedsakelig fordi det var en tidkrevende prosess. I ettertid viste det seg at det senteret hvor opptakene ble gjennomført (senter 6), var ett av dem med jevnest bildekvalitet. Beregningen ga imidlertid en pekepinn på graden av måleusikkerhet, selv om man ikke vet hvor representativ den var for opptaksstabiliteten ved de andre sentrene.

5.2.5 ACR-kriterier og bruken av ACR-fantomet

I testene av fantomopptak ble ACR-grensene benyttet som grenseverdier for tilfredsstillende bildekvalitet. Disse kriteriene er imidlertid satt som et minstekrav for hva man kan forvente av en velfungerende MR-tomograf. At en tomograf havner like innenfor ACR-grensen, er altså ikke ensbetydende med normal ytelse (ACR, 2005). Det er uklart hvordan ACR har kommet fram til disse grenseverdiene. Dersom de er satt slik at flest mulig tomografer skal bestå kriteriene, er det desto mer alvorlig når en tomograf ikke klarer kriteriene. I tillegg til de kvantitative testene beskrevet i veiledningen fra ACR, inngår også visuell vurdering av fantomopptakene i godkjenningen. I USA har ACR egne observatører som evaluerer bildene. Rene kvantitative tester vil ikke nødvendigvis fange opp alle aspekter omkring bildekvalitet (ACR, 2005).

5.2.6 Kvantifisering av kvalitative data

Objektiv testing og subjektiv vurdering tilsvarer hver sin ende i bredden av muligheter for hvordan bildekvalitet kan undersøkes. De ulike aspektene omkring bildekvalitet er i utgangspunktet en rekke kvalitative egenskaper. Utviklingen av fantomer og analyseprogrammer er forsøk på å redusere disse til kvantitative verdier for dermed å gjøre dem lettere tilgjengelig for vurdering, sammenligning og analyse. Kvalitative data har imidlertid flere ting ved seg enn det som egentlig kan kvantifiseres. Kvantifisering er en forenkling. Dette prosjektet har tatt for seg tomografrelaterte årsaker til forskjeller mellom sentre. Men det er mye annet som kan spille inn i en multisenterstudie. For eksempel er pasientspesifikke årsaker til forskjeller som plassering, stabilisering og sentrering er ikke tatt med her.

ACR-fantomet inneholder geometriske strukturer som gir muligheten til å måle aspekter ved bildekvalitet som ikke er målbart i morfologiske opptak. Størrelser og andre verdier (for eksempel T2-verdier) er kjent, og tomografens nøyaktighet i gjengivelsen av objekter kan dermed testes kvantitativt.

Verdiene beregnet ved hjelp av NeuroQuant® er basert på automatisk gjenkjenning og måling av vokselvis signalintensitet. Vokslenes signalintensitet endres gradvis i overgangen mellom ulike strukturer, og avgrensningen av ulike strukturer er derfor en form for ”automatisert vurdering” utført ved hjelp av et dataprogram. Resultatet avhenger av programmets nøyaktighet, men også av tomografens nøyaktighet i gjengivelsen av objekter.

Subjektive, visuelle vurderinger er vanskelig å sammenligne uten noen form for kvantifisering. Det var likevel ønskelig å inkludere subjektive vurderinger i prosjektet, da dette potensielt kunne forklare flere sider omkring fenomenet bildekvalitet enn objektive tester. Det var også ønskelig å sammenholde de subjektive vurderingene med resultatene fra objektive tester for å se om registrerte variasjoner var store nok til å ha betydning i visuell vurdering. Spørreskjema ble valgt som måleinstrument for subjektive vurderinger. Mer om dette valget under punkt 5.2.7 *Bruken av spørreskjema*.

Etter min mening var den objektive tilnærmingen til evalueringen av kvalitetsaspektene den mest praktiske og effektive. Det faktum at det finnes mange ulike kvantitative metoder for måling av bildekvalitet, viser at det er flere som mener det samme. Mangelen på standardiserte metoder viser imidlertid også at det er delte meninger i fagmiljøet om hva som er mest hensiktsmessig ved evaluering av bildekvalitet. Det finnes fantomer som er svært avanserte i bruk, men dermed er de også høyere i pris og mindre utbredt. Det finnes enklere fantomer enn ACR-fantomer, men bruken av disse øker sjansen for å overse viktige aspekter ved bildekvaliteten.

Man mister noen sider av vurderingene ved bruk av en ren objektiv fremgangsmåte, men man tjener noe ved at fremgangsmåten er universell og kan gjentas av andre eller gjøres flere ganger. For eksempel kan yteevne følges over tid. Tidsaspektet var ikke hovedfokus i dette prosjektet, men i en pasientstudie bør det gjennomføres jevnlig kvalitetskontroll som er sammenlignbar fra gang til gang. Dette er forskerens eneste mulighet til å overvåke tomografenes yteevne og rette opp i eventuelle feil mens studien pågår. Som tidligere nevnt kan det ta tid før feil på utstyr blir så alvorlig at det blir synlig i bildet, men kvantitative målinger i fantomopptak kan avdekke en gradvis nedgang i kvalitet. Dermed har man ved slik oppfølging mulighet til å finne ut når eventuelle feil oppsto. Dette avhenger av at kvalitetskontroller er gjennomført gjentatte ganger, og at det ikke har gått for lang tid mellom. Forstyrrelser i bildekvalitet kan være kilder til feil når pasientdata slås sammen, selv om de ikke er synlige i bildet. Derfor kan også kvalitetskontroll være nyttig i vurderingen av om all innsamlet pasientdata skal tas med i dataanalysen, eller om noe bør holdes utenfor.

I PRODE studien følges sentrene fortsatt opp med kvalitetskontroll, uavhengig av dette masterprosjektet. Sentrene som falt utenfor ACR-grensen i beregningen av signalhomogenitet har fått beskjed om dette. Nye kontroller gjennomføres dersom nytt utstyr tas i bruk, etter vedlikeholdsarbeid og større oppdateringer, og minst hver sjettemåned. Datakvaliteten i PRODE MR får altså løpende oppmerksomhet, selv om masterprosjektet er avsluttet.

5.2.7 Bruken av spørreskjema

Hovedårsaken til at spørreskjema ble valgt til kartlegging av subjektive vurderinger, var at svarene skulle kunne analyseres kvantitativt, og at analysen skulle kunne gjøres direkte, uten noen form for tolkning eller koding av svarene. I tillegg var tanken at faste spørsmål krever mindre av tid og motivasjon fra respondentene, og dermed kunne flere radiologer være villige til å svare.

Det er imidlertid noen ulemper med spørreskjema. Blant annet kan ikke misforståelser eller feiltolkninger av spørsmål avklares umiddelbart. Radiologene kunne kontakte meg underveis ved spørsmål, men det er ikke alltid man vet at man har misforstått noe. En del av spørsmålene hadde motstridende svar blant radiologene, noe som kan tyde på at spørsmålene ble oppfattet ulikt. Ulike svar kan imidlertid også skyldes motstridende meninger og oppfatninger av aspektene omkring bildekvalitet, samt bildekvalitet generelt. Som tidligere nevnt finnes det ingen gullstandard for god bildekvalitet på MR av hjerne. Det er altså liten enighet om hva som er god kvalitet, og da må man også forvente uoverensstemmelser i svarene når man spør etter ulike radiologers meninger om bildekvalitet.

Selve spørsmålsstillingen i spørreskjemaet ble lagt opp slik at radiologen skulle vurdere grad av ulike aspekter omkring kvalitet på en skala 1–5. Hvert trinn på denne skalaen var nøyaktig definert. På denne måten fikk hvert tall samme betydning for alle respondentene, og det lå mer informasjon i svaret enn et tall. Målet med dette var å få et bedre bilde av hva radiologene mente om det aktuelle kvalitetsaspektet. Noen av fallgruvene med dette er at upresise formuleringer kan gi rom for feiltolkning, og at definisjonene ikke leses nøye nok. Alternativet var kun å definere ytterpunktene på skalaen, og la radiologen selv vurdere verdien av de resterende trinnene. Etter min mening ville dette gitt enda en faktor blant årsakene til ulike svar radiologene imellom, og derfor var dette uaktuelt.

Vurderingene fra alle radiologene ble analysert samlet for å få nok verdier i analysen. Denne framgangsmåten var planlagt på forhånd, og spørsmålene var gruppert slik at det skulle la seg gjøre. Det ble imidlertid ikke testet om radiologene oppfattet spørsmålene likt, eller om de sjekket og vurderte de ulike kvalitetsaspektene på samme måte. Man

vet derfor egentlig ikke om svarene deres er sammenlignbare, eller i hvilken grad. Hensikten med spørreskjemaet var ikke å teste pålitelighet på tvers av respondentene. For å kunne gjøre det, måtte man for eksempel testet deres evne til å finne patologi, ulike artefakt eller lignende. I dette prosjektet stolte man på at radiologene kjente igjen de ulike strukturene det ble spurt etter, samt at de kjente til og kunne identifisere de ulike artefaktene. Kombinasjonen av kunnskap omkring patologi, anatomi og bildekvalitet er radiologenes fagfelt. Det var derfor denne faggruppen ble valgt til subjektiv vurdering.

Respondentene

Man var ute etter radiologenes profesjonelle meninger om bildekvaliteten. Aktuelle respondenter var derfor radiologer som vurderer hjerneopptak. I tillegg var det ønskelig med radiologer som jobber med ulikt utstyr til daglig. Av de tre som svarte, jobber en med MR-utstyr fra Philips og to med MR-utstyr fra Siemens til daglig. De tre hadde ulik erfaring og fartstid (to overleger og en lege i spesialisering).

Radiologene besvarte skjemaene hver for seg, og de visste ikke hva andre hadde svart. Vurderingene ble gjort på ulike skjermer og ved hjelp av ulike visningsprogrammer. Også vurderingen ble gjort på ulike måter. Noen vurderte hvert senter for seg, mens andre hadde bildene fra de ulike sentrene ved siden av hverandre på samme skjerm.

Oppfølging og redigering i ettertid

På en del av spørsmålene var svarene signifikant ulike mellom respondentene. De ble derfor kontaktet for å undersøke om spørsmålene var blitt forstått forskjellig. Det kom da fram at ulike svar ikke skyldtes ulik tolkning av spørsmålene, men ulike meninger om opptakene.

Som tidligere nevnt finnes det ikke noe validert spørreskjema for evaluering av bildekvalitet i hjerneopptak på MR, og det var derfor det ble laget et skjema til dette prosjektet. Det var i hovedsak to ting som ikke ble fanget opp før spørreskjemaene ble distribuert.

For T2*-sekvensen ble ett av spørsmålene utelatt fra analysen. Dette var spørsmålet om susceptibilitetsartefakt. T2*-sekvensen er laget for å øke susceptibilitetsartefakter fordi

dette er med på å fremheve områder med små blødninger. Vurderingen av susceptibilitetseffektene i denne sekvensen er viktig fordi det sier noe om i hvilken grad sekvensen viser det den er ment å vise. Derfor ble det feil å ta med dette spørsmålet under artefakter, da definisjonen på artefakter i denne forbindelse var forstyrrelser som degraderer bildekvaliteten. Susceptibilitetseffekter er en stor grad av helheten i denne sekvensen, og kom dermed inn i vurderingen av vektingen, noe som var et spørsmål under kategorien ”Bildekvalitet som helhet”. Det var derfor ikke aktuelt å lage et eget spørsmål omkring susceptibilitet som var bedre tilpasset denne sekvensen. At spørsmålet kom med i kategorien ”Artefakter” også for denne sekvensen, var en ren redigeringsfeil. Siden spørsmålet ikke ble fjernet på forhånd, ble svarene fjernet i ettertid.

5.2.8 Omgivelser – temperatur og opptakstidspunkt

Romtemperatur og luftfuktighet kan ha betydning for bildekvalitet. Årsaken til dette er at luftfuktigheten endres med temperatur, og tørr luft kan være opphav til statisk elektrisitet, noe som har innvirkning på MR-signalet. Noen leverandører utstyrer MR-rommet med luftfukter for å minske problemer med statisk elektrisitet. Luftfuktigheten skulle derfor måles for å teste sammenhenger mellom yteevne og luftfuktighet. Luftmåleren ble imidlertid ødelagt underveis i prosjektet. Den ble erstattet, men også denne ble ødelagt. Målingene ble derfor for usikre og mangelfulle til å kunne vurderes eller analyseres i forhold til yteevne. Dette er en variasjon som kunne hatt betydning for forskjeller mellom sentre.

Opptak ble gjort på ulike tidspunkt av døgnet på de ulike sentrene. Årsaken til dette var at opptakene ble lagt inn på den måten som passet best i forhold til resten av dagsprogrammet. Det er spesielt gradientforsterkerne som kan ha varierende yteevne dersom de ikke får tid til å varmes opp og stabiliseres. For å sikre at gradientlinearitet måles under optimale forhold, bør tomografen være ferdig oppvarmet før fantomopptakene starter. Selv om man hadde kontroll på tidspunktet opptakene ble gjort, ble det ikke grundig sjekket om tomografen var oppvarmet. Med ett unntak ble alle opptakene gjort mellom pasienter, og det er derfor grunn til å tro at opptakene ble gjort på et tidspunkt hvor tomografen var oppvarmet. Unntaket var senter 4, hvor opptaket ble gjort før første pasient. Dette ser imidlertid ikke ut til å ha hatt betydning

for resultatet, da det ikke var forskjell mellom første og andre opptak, og dette ikke var et senter som skilte seg ut i nevneverdig grad.

5.2.9 Sammenligning av objektiv og subjektiv bildekvalitet

Det er ikke alle aspektene som er direkte sammenlignbare. Kvalitetsaspektene kan ikke skilles like enkelt ved å vurdere bildenes utseende som ved objektiv testing. For eksempel kan vanskeligheter med å differensiere ulike vev og strukturer ha mange årsaker. Det kan blant annet skyldes problemer med SNR, CNR, lavkontrastdetaljer og romlig oppløsning. Det kan skyldes en av disse tingene, eller en kombinasjon. Svarene gjenspeiler både radiologenes profesjonelle mening og personlige preferanser. Fant jo frem til noen av de samme sentrene, noe som kan tyde på at objektiv og subjektive evalueringer kan supplere hverandre.

5.2.10 Generelle utfordringer og veien videre

PRODE studien er et forsøk på å standardisere kliniske undersøkelser slik at opptak fra ulike senter kan samles til et større materiale og analyseres i forskningsøyemed. Det er altså et skritt på veien mot klinisk nyttige forskningsdata. Det bør være mulig å standardisere enkelte sekvenser på nasjonalt plan, slik ADNI har gjort med sitt T1-vektede 3D-volum. Flere standardiserte sekvenser kan etter hvert settes sammen til standard protokoller. Dersom man skal klare dette, må man også ha en nasjonal standard for kvalitetskontroll og vedlikehold. Alt utstyr må testes og feilsøking må foretas jevnlig for å fange opp små endringer over tid. Det er mange aspekter omkring bildekvalitet som ikke fanges opp ved å studere pasient- eller frivilligdata. Det er imidlertid også noen aspekter som ikke fanges opp av fantomdata, og man må være klar over disse. Det var en praktisk utfordring å frakte fantomet rundt til alle deltakersentrene, og å få data tilbake fra deltakersentrene.

Dette prosjektet har hatt hovedfokus på de største ulikhetene mellom tomografer. Det er imidlertid også flere kvalitetsaspekter med små ulikheter, noe som styrker troen på at multisenterstudier vil gi pålitelige data. Slik jeg ser det var det største problemet i dette prosjektet ulikheter i signalhomogenitet. Dette skyldes trolig feil på antennen, og er noe

som kan testes av leverandøren selv ved periodisk vedlikehold. Antennen vil da repareres eller erstattes med en ny.

Når man setter i gang en multisenterstudie, bør målet med studien og metoder for dataanalyse være klart før deltakersentre testes og inkluderes. Man kan blant annet ikke stole på at MR-data egner seg til automatisk volumetri basert på visuelle vurderinger gjort av erfarne nevreradiologer. Ulike kvalitetsaspekter må testes avhengig av endepunktet for studien. Jack et al. (2008) påpeker at ulik bruk og analyse av datamaterialet stiller ulike krav til sekvensene. Mens en erfaren nevreradiolog klarer å tolke bildene til tross for mindre artefakter, har automatiske segmenterings- og analyseprogrammer et helt annet utgangspunkt. Dette må tas hensyn til både i optimalisering av sekvenser og protokoller på tvers av utstyr, men også i inkluderingen av sentre til multisenterstudier.

6 Konklusjon

Tomografene i dette utvalget hadde grunnleggende forskjeller som ble gjenspeilet i bildekvaliteten. Objektive tester og subjektive vurderinger fremhevet i store trekk de samme tomografene i positiv og negativ retning. De største forskjellene ble avdekket for signalhomogenitet, SNR, CNR, lavkontrastdetaljer og differensiering. Testresultatene hadde få sammenhenger med tomografenes spesifikasjoner, men samlet sett kom de eldste tomografene dårligst ut. I hvilken grad resultatene har betydning for evaluering og beskrivelse av sykdom kan ikke besvares på grunnlag av materialet i dette prosjektet.

Betydningen av tomografers ulikheter i forhold til bildekvalitet bør ikke ignoreres, verken i multi- eller monosenterstudier. Resultatene fra dette prosjektet viser at jevnlig og systematisk kontroll av yteevne er et viktig ledd i datainnsamlingen. En av utfordringene er manglende retningslinjer, noe som kan gjøre kvalitetskontrollen uforholdsmessig kostbar og tidkrevende.

Referanser

- Alzheimer's Disease Neuroimaging Initiative. (2011). *MRI protocols*. I: Laboratory of Neuro Imaging. (red.). Los Angeles, California: Laboratory of Neuro Imaging, UCLA School of Medicine. Tilgjengelig fra: http://adni.loni.ucla.edu/research/protocols/mri-protocols/#MRI_Scanner_Protocols (lest 5.mars 2011).
- American College of Radiology. (2002). *Site scanning instructions for use of the MR phantom for the ACR MRI accreditation program*. Reston, Virginia: American College of Radiology.
- American College of Radiology. (2005). *Phantom test guidance*. Reston, Virginia: American College of Radiology.
- Ashton, E. (2010). Quantitative MR in multi-center clinical trials. *J. Magn. Reson. Imaging*, 31 (2): 279–288.
- Bauer, C. M., Jara, H. & Killiany, R. (2010). Whole brain quantitative T2 MRI across multiple scanners with dual echo FSE: applications to AD, MCI, and normal aging. *NeuroImage*, 52 (2): 508–514.
- Bjørndal, A. & Hofoss, D. (2004). *Statistikk for helse- og sosialfagene*. 2. utg. Oslo: Gyldendal akademisk. 269 s. s.
- Bourel, P., Gibon, D., Coste, E., Daanen, V. & Rousseau, J. (1999). Automatic quality assessment protocol for MRI equipment. *Medical physics*, 26 (12): 2693–2700.
- Båth, M. & Månsson, L. G. (2007). Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation. *The British journal of radiology*, 80 (951): 169–176.
- Båth, M. (2010). Evaluating imaging systems: practical applications. *Radiation protection dosimetry*, 139 (1–3): 26–36.
- Chen, C. C., Wan, Y. L., Wai, Y. Y. & Liu, H. L. (2004). Quality assurance of clinical MRI scanners using ACR MRI phantom: preliminary results. *Journal of digital imaging: the official journal of the Society for Computer Applications in Radiology*, 17 (4): 279–284.
- CorTechs Labs, Inc. (2007). *NeuroQuant™ user manual. Version 1.0.1*. San Diego, California: CorTechs Labs, Inc.
- CorTechs Labs, Inc. (2007–2009). *NeuroQuant®. I: CorTechs Labs, Inc. (red.)*. La Jolla, San Diego, California: CorTechs Labs, Inc. Tilgjengelig fra: <http://www.cortechs.net/products/neuroquant.php> (lest 9. mai 2011).
- Elster, A. D. & Burdette, J. H. (2001). *Questions & answers in magnetic resonance imaging*. 2. utg. St. Louis, Missouri: Mosby. XIV, 333 s. s.
- Ewers, M., Teipel, S. J., Dietrich, O., Schonberg, S. O., Jessen, F., Heun, R., Scheltens, P., van de Pol, L., Freymann, N. R., Moeller, H. J. & Hampel, H. (2006). Multicenter assessment of reliability of cranial MRI. *Neurobiology of aging*, 27 (8): 1051–1059.
- Fayers, P. M. & Machin, D. (red.). (2007). *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. 2. utg. Chichester: John Wiley. XXI, 544 s. s.
- Firbank, M. J., Harrison, R. M., Williams, E. D. & Coulthard, A. (2000). Quality assurance for MRI: practical experience. *The British journal of radiology*, 73 (868): 376–383.
- Fu, L., Fonov, V., Pike, B., Evans, A. C. & Collins, D. L. (2006). Automated analysis of multi site MRI phantom data for the NIHPD project. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on*

- Medical Image Computing and Computer-Assisted Intervention*, 9 (Pt 2): 144–151.
- Gedamu, E. L., Collins, D. L. & Arnold, D. L. (2008). Automated quality control of brain MR images. *Journal of magnetic resonance imaging : JMRI*, 28 (2): 308–19.
- Gunter, J. L., Bernstein, M. A., Borowski, B. J., Ward, C. P., Britson, P. J., Felmlee, J. P., Schuff, N., Weiner, M. & Jack, C. R. (2009). Measurement of MRI scanner performance with the ADNI phantom. *Medical physics*, 36 (6): 2193–2205.
- Haacke, E. M., Brown, R. W., Thompson, M. R. & Venkatesan, R. (1999). *Magnetic resonance imaging: physical principles and sequence design*. New York: Wiley-Liss. XXVII, 914 s. s.
- Halvorsen, K. (1996). *Forskningsmetode for helse- og sosialfag. En innføring i samfunnsvitenskapelig metode*. Oslo: Cappelen akademisk forlag. VIII, 186 s. s.
- Hashemi, R. H., Lisanti, C. J. & Bradley, W. G. (2004). *MRI: the basics*. 2. utg. Philadelphia: Lippincott Williams & Wilkins. XIII, 353 s. s.
- Helse- og omsorgsdepartementet (2010). *Veileder til lov 20. juni 2008 nr. 44 om medisinsk og helsefaglig forskning (helseforskningsloven)*. Veiledninger og brosjyrer. Oslo: Helse- og omsorgsdepartementet. 40 s.
- Håberg, A., Borza, T., Selbæk, G. & Engedal, K. (2010). *Informasjonsskriv om PRODE, delstudie med MR*. Upublisert rundskriv.
- Ibrahim, T. (2009, 12.–14. september). *Subject insensitive RF arrays for multi-transmit applications*. Ultra High Field User Meeting, Leipzig: Siemens Healthcare.
- Ihalainen, T. M., Lönnroth, N. T., Peltonen, J. I., Uusi-Simola, J. K., Timonen, M. H., Kuusela, L. J., Savolainen, S. E. & Sipilä, O. E. (2011). MRI quality assurance using the ACR phantom in a multi-unit imaging center. *Acta Oncologica*, 50 (6): 966-972.
- Ihalainen, T., Sipilä, O. & Savolainen, S. (2004). MRI quality control: six imagers studied using eleven unified image quality parameters. *European radiology*, 14 (10): 1859–1865.
- Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L. Whitwell, J., Ward, C., Dale, A. M., Felmlee, J. P., Gunter, J. L., Hill, D. L. G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C. S., Gunnar, K., Ward, H. A., Metzger, G. J., Scott, K. T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J. P., Fleisher, A. S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J. & Weiner, M. W. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27 (4): 685-691.
- Jack, C. R., Jr., Slomkowski, M., Gracon, S., Hoover, T. M., Felmlee, J. P., Stewart, K., Xu, Y., Shiung, M., O'Brien, P. C., Cha, R., Knopman, D. & Petersen, R. C. (2003). MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. *Neurology*, 60 (2): 253–260.
- Kanal, E., Barkovich, A. J., Bell, C., Borgstede, J. P., Bradley, W. G., Jr., Froelich, J. W., Gilk, T., Gimbel, J. R., Gosbee, J., Kuhni-Kaminski, E., Lester, J. W., Jr., Nyenhuis, J., Parag, Y., Schaefer, D. J., Sebek-Scoumis, E. A., Weinreb, J., Zaremba, L. A., Wilcox, P., Lucey, L. & Sass, N. (2007). ACR guidance document for safe MR practices: 2007. *AJR. American journal of roentgenology*, 188 (6): 1447–1474.
- Kaufman, L., Kramer, D. M., Crooks, L. E. & Ortendahl, D. A. (1989). Measuring signal-to-noise ratios in MR imaging. *Radiology*, 173 (1): 265-267.

- Leppert, I. R., Almli, C. R., McKinstry, R. C., Mulkern, R. V., Pierpaoli, C., Rivkin, M. J. & Pike, G. B. (2009). T2 relaxometry of normal pediatric brain development. *Journal of Magnetic Resonance Imaging*, 29 (2): 258–267.
- Lerski, R. A. & de Certaines, J. D. (1993). II. Performance assessment and quality control in MRI by Eurospin test objects and protocols. *Magnetic Resonance Imaging*, 11 (6): 817–833.
- Liang, Z.-P. & Lauterbur, P. C. (2000). *Principles of magnetic resonance imaging. A signal processing perspective*. New York: SPIE Optical Engineering Press; IEEE Press. XV, 416 s. s.
- Magnetic Resonance – Technology Information Portal. (2003–2011). *Artifact*. I: SoftWays' Medical Imaging Group. (red.): SoftWays' Medical Imaging Group. Tilgjengelig fra: <http://www.mr-tip.com/serv1.php?type=db1&gid=2> (lest 18. april 2011).
- Magnetic Resonance – Technology Information Portal. (2003–2011). *Signal to noise ratio*. I: SoftWays' Medical Imaging Group. (red.): SoftWays' Medical Imaging Group. Tilgjengelig fra: [http://www.mr-tip.com/serv1.php?type=db1&dbs=Signal to Noise Ratio](http://www.mr-tip.com/serv1.php?type=db1&dbs=Signal%20to%20Noise%20Ratio) (lest 22. februar 2011).
- Martinussen, M. (red.). (2010). *Kvantitativ forskningsmetodologi i samfunns- og helsefag*. Bergen: Fagbokforlaget. 343 s. s.
- McEvoy, L. K. & Brewer, J. B. (2010). Quantitative structural MRI for early detection of Alzheimer's disease. *Expert Review of Neurotherapeutics*, 10 (11): 1675–1688.
- McRobbie, D. W. & Quest, R. A. (2002). Effectiveness and relevance of MR acceptance testing: results of an 8 year audit. *The British journal of radiology*, 75 (894): 523–531.
- McRobbie, D. W., Moore, E. A., Graves, M. J. & Prince, M. R. (2007). *MRI: from picture to proton*. 2. utg. Cambridge: Cambridge University Press. XII, 393 s. s.
- Moeller, T. B. & Reif, E. (2010). *MRI parameters and positioning*. 2. utg. Stuttgart: Thieme. X, 341 s. s.
- Morelli, J. N., Runge, V. M., Ai, F., Attenberger, U., Vu, L., Schmeets, S. H., Nitz, W. R. & Kirsch, J. E. (2011). An image-based approach to understanding the physics of MR artifacts. *Radiographics : a review publication of the Radiological Society of North America, Inc*, 31 (3): 849–866.
- Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W. & Beckett, L. (2005). The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin N Am.*, 15 (4): 869–877.
- Nahhas, R. (1999, 19. juli). *Mixed linear models*. Ohio State University, Biostatistics Program, Columbus, Ohio. Tilgjengelig fra http://www.stat.osu.edu/~biostat/newsletters/volume3_1/article_vol3_1.html (lest 11. november 2011).
- Nasjonalt samarbeidsgruppe for medisinsk og helsefaglig forskning (2008). *Forslag til nasjonalt satsingsområde: Alvorlige psykiske lidelser*. [Grunnlagsdokument]. Nasjonalt samarbeidsgruppe for medisinsk og helsefaglig forskning. 15 s.
- Nasjonalt kunnskapssenter for helsetjenesten (2009). *Sjekkliste: Planlegging av lokale, kvantitative brukerundersøkelser*. Oslo: Nasjonalt kunnskapssenter for helsetjenesten, Seksjon for brukererfaringsundersøkelser. 10 s.
- Norušis, M. J. (2008). *SPSS statistics 17.0. Advanced statistical procedures companion*. Upper Saddle River, New Jersey: Prentice Hall. XIV, 418 s. s.
- Peat, J. & Barton, B. (2005). *Medical statistics: a guide to data analysis and critical appraisal*. Malden: Blackwell. XII, 324 s. s.

- Polit, D. F. (2010). *Statistics and data analysis for nursing research*. 2. utg. Upper Saddle River, New Jersey: Pearson. XII, 440 s., ill. s.
- Polit, D. F. & Beck, C. T. (2010). *Essentials of nursing research: appraising evidence for nursing practice*. 7. utg. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins. XIV, 610 s. s.
- Reimer, P., Parizel, P. M. & Stichnoth, F.-A. (red.). (2006). *Clinical MR imaging: a practical approach*. 2. utg. Berlin, Heidelberg: Schering. v.: digital s.
- Simon, J. H., Li, D., Traboulsee, A., Coyle, P. K., Arnold, D. L., Barkhof, F., Frank, J. A., Grossman, R., Paty, D. W., Radue, E. W. & Wolinsky, J. S. (2006). Standardized MR imaging protocol for multiple sclerosis: consortium of MS centers consensus guidelines. *American Journal of Neuroradiology*, 27 (2): 455–461.
- Smith, H. J. & Gjesdal, K. (2000). Magnetisk resonans – historikk og teoretisk grunnlag. *Tidsskrift for den norske legeforening*, 120 (8): 931–935.
- Sprawls, P. (2000). *Magnetic resonance imaging: principles, methods, and techniques*. Madison, Wisconsin: Medical Physics Pub. XVII, 173 s. s.
- Tofts, P. S. (1998). Standardisation and optimisation of magnetic resonance techniques for multicentre studies. *Journal of neurology, neurosurgery, and psychiatry*, 64 Suppl 1: 37–43.
- Vangberg, T. R. (2011). *MRI i Tromsøundersøkelsen, hva kan befolkningsundersøkelser bidra med?* Gardermoen (Foredrag NORNI-seminar – Norwegian NeuroImaging Initiative 8. september).
- Van Horn, J. D. & Toga, A. W. (2009). Multisite neuroimaging trials. *Current opinion in neurology*, 22 (4): 370–378.
- Vymazal, J., Righini, A., Brooks, R. A., Canesi, M., Mariani, C., Leonardi, M. & Pezzoli, G. (1999). T1 and T2 in the brain of healthy subjects, patients with Parkinson disease, and patients with multiple system atrophy: relation to iron content. *Radiology*, 211 (2): 489–495.
- Wagner, M., Rief, M., Busch, J., Scheurig, C., Taupitz, M., Hamm, B. & Franiel, T. (2010). Effect of butylscopolamine on image quality in MRI of the prostate. *Clinical Radiology*, 65 (6): 460–464.
- Wenstøp, F. (2006). *Statistikk og dataanalyse*. 9. utg. Oslo: Universitetsforlaget. 424 s. s.
- Westbrook, C., Kaut Roth, C. & Talbot, J. (2005). *MRI in practice*. 3. utg. Oxford: Blackwell Publishing. XII, 410 s. s.
- Westbrook, C. (2008). *Handbook of MRI technique*. 3. utg. Oxford: Wiley-Blackwell. X, 414 s. s.
- Zhuo, J. & Gullapalli, R. P. (2006). AAPM/RSNA physics tutorial for residents: MR artifacts, safety, and quality control. *Radiographics: a review publication of the Radiological Society of North America, Inc*, 26 (1): 275–297.

Vedlegg

Visuell vurdering av bildekvalitet

Ulike aspekter ved bildene fra senter nummer I-VIII skal vurderes på en skala fra 1 til 5 hvor 1 = dårlig kvalitet og 5 = utmerket kvalitet. Under følger en nærmere klarlegging av hva som ligger i hvert av trinnene på skalaen.

Differensiering av ulike vev og strukturer

- 1 = ikke differensierbar: strukturen er ikke synlig og kan ikke evalueres
- 2 = vanskelig differensierbar: strukturen er vanskelig å visualisere
- 3 = moderat differensierbar: strukturen visualiseres i noe grad
- 4 = godt differensierbar: strukturen ses avgrenset fra omkringliggende vev og strukturer
- 5 = utmerket differensierbarhet: strukturen er tydelig avgrenset fra omkringliggende vev og strukturer

Artefakter i bildet

- 1 = artefaktet er gjennomgående: fullstendig ødeleggende for kvaliteten
- 2 = artefaktet er til stede: delvis til hinder for korrekt visualisering av anatomi, vev og strukturer
- 3 = artefaktet er moderat til stede: forstyrrer visualiseringen av anatomi, vev og strukturer i noe grad
- 4 = artefaktet er nærmest ikke-eksisterende: forstyrrer ikke korrekt visualisering av anatomi
- 5 = artefaktet er ikke til stede i bildet

Bildekvalitet som helhet

- 1 = fullstendig utilfredsstillende: ikke klinisk brukbart
 - 2 = ikke tilfredsstillende: vanskelig å visualisere anatomi, vev og strukturer
 - 3 = moderat tilfredsstillende: anatomi, vev og strukturer visualiseres i noe grad tross forstyrrelsene
 - 4 = tilfredsstillende: godt nok for korrekt visualisering
 - 5 = utmerket: svært tilfredsstillende
-

Spørsmål

Det er 4 sett med bilder fra de 8 sentrene. Spørsmål 1-7 graderes i følge skalaen, spørsmål 8 er et åpent spørsmål. Neste side inneholder avklaring av ord og begreper. Svarskjema finnes på side 4. På forhånd takk for hjelpen!

<p>Sagittal T1-vektet 3D-volum <i>Differensiering av ulike vev og strukturer</i></p> <ol style="list-style-type: none"> 1. I hvilken grad kan grå- og hvit hjernesubstans differensieres 2. Hvor synlig/iøynefallende/fremtredende er <ol style="list-style-type: none"> a. Hippocampus b. Thalamus c. Basalganglier <p><i>Artefakter i bildet</i></p> <ol style="list-style-type: none"> 3. I hvilken grad er følgende artefakter synlige <ol style="list-style-type: none"> a. Bevegelse b. Flow c. Susseptibilitetseffekter d. Kjemisk skift e. Aliasing f. Blurring g. Stripeartefakt <p><i>Bildekvalitet som helhet</i></p> <ol style="list-style-type: none"> 4. Kontrastvurdering/skarphet som helhet 5. Vekting i bildet 6. Vurdering av signal/støy forhold som helhet 7. Bildekvalitet som helhet 8. Andre forstyrrelser som ikke er nevnt her 	<p>Transversal Pd/T2 Dobbelekk <i>Differensiering av ulike vev og strukturer</i></p> <ol style="list-style-type: none"> 1. I hvilken grad kan grå- og hvit hjernesubstans differensieres 2. Hvor synlig/iøynefallende/fremtredende er <ol style="list-style-type: none"> a. Hippocampus b. Thalamus c. Basalganglier <p><i>Artefakter i bildet</i></p> <ol style="list-style-type: none"> 3. I hvilken grad er følgende artefakter synlige <ol style="list-style-type: none"> a. Bevegelse b. Flow c. Susseptibilitetseffekter d. Kjemisk skift e. Aliasing f. Blurring g. Stripeartefakt <p><i>Bildekvalitet som helhet</i></p> <ol style="list-style-type: none"> 4. Kontrastvurdering/skarphet som helhet 5. Vekting i bildet 6. Vurdering av signal/støy forhold som helhet 7. Bildekvalitet som helhet 8. Andre forstyrrelser som ikke er nevnt her
<p>Transversal FLAIR <i>Differensiering av ulike vev og strukturer</i></p> <ol style="list-style-type: none"> 1. I hvilken grad kan grå- og hvit hjernesubstans differensieres 2. Hvor synlig/iøynefallende/fremtredende er <ol style="list-style-type: none"> a. Hippocampus b. Thalamus c. Basalganglier <p><i>Artefakter i bildet</i></p> <ol style="list-style-type: none"> 3. I hvilken grad er følgende artefakter synlige <ol style="list-style-type: none"> a. Bevegelse b. Flow c. Susseptibilitetseffekter d. Kjemisk skift e. Aliasing f. Blurring g. Stripeartefakt <p><i>Bildekvalitet som helhet</i></p> <ol style="list-style-type: none"> 4. Kontrastvurdering/skarphet som helhet 5. Vekting i bildet 6. Vurdering av signal/støy forhold som helhet 7. Bildekvalitet som helhet 8. Andre forstyrrelser som ikke er nevnt her 	<p>Transversal T2* (Hemosekvens) <i>Differensiering av ulike vev og strukturer</i></p> <ol style="list-style-type: none"> 1. I hvilken grad kan grå- og hvit hjernesubstans differensieres 2. Hvor synlig/iøynefallende/fremtredende er <ol style="list-style-type: none"> a. Hippocampus b. Thalamus c. Basalganglier <p><i>Artefakter i bildet</i></p> <ol style="list-style-type: none"> 3. I hvilken grad er følgende artefakter synlige <ol style="list-style-type: none"> a. Bevegelse b. Flow c. Susseptibilitetseffekter d. Kjemisk skift e. Aliasing f. Blurring g. Stripeartefakt <p><i>Bildekvalitet som helhet</i></p> <ol style="list-style-type: none"> 4. Kontrastvurdering/skarphet som helhet 5. Vekting i bildet 6. Vurdering av signal/støy forhold som helhet 7. Bildekvalitet som helhet 8. Andre forstyrrelser som ikke er nevnt her

Ordforklaringer

Generelle avklaringer

Artefakt: Forstyrrelser i form av bildesignaler som ikke representerer en sann gjengivelse av objektet.

Differensierbarhet: I hvilken grad man kan skjelve strukturer fra bakgrunnen.

Generelle forstyrrelser: Teoretisk kan alle former for artefakter og tekniske problemer oppstå i MR-bildene. Dette skjemaet har imidlertid bare tatt med begreper omkring bildekvalitet som man på forhånd kan forvente å finne i ulik grad. Derfor har svarskjemaene en rubrikk for eventuelle kommentarer eller tilføyelser.

Nærmere klarlegging av hva som ligger i ulike begreper

Aliasing: Også kalt "fold-over", "folding" eller "wrap-around". Oppstår som følge av at vev utenfor field-of-view (FOV) eksiteres og avgir signal. Signalet feiltolkes som tilhørende i bildet, og feilplasseres.

Bevegelse: Artefakter som følge av uro i objektet eller vibrasjoner.

Blurring: Små strukturer flyter sammen i stedet for å vises adskilt.

Flow: Artefakter som følge av blodgjennomstrømming og pulsasjon.

Kjemisk skift: Piksselforskyvning som følge av frekvensfeil på grunn av ulik resonansfrekvens i ulike vev. Ses som signalbortfall langs én side av strukturer, mens motsatt side tildeles økt signalintensitet.

Signal/Støy-forhold (SNR): MR-bildet inneholder alltid en viss grad av støy. I hvilken grad bildesignalet trenger gjennom og "overdøver" varierer mellom tomografer og sekvenser.

Stripeartefakt: Stripper i bildene kan oppstå av flere årsaker. Her nevnes noen av de vanligste typene: *Spikes* som følge av feil under fyllingen av k-space; *zipper* som følge av lekkasje av radiofrekvent stråling, enten fra systemet selv eller utenfra; *trunkasjon* (Gibb's ringing) som følge av vevsoverganger med brå endringer i signalintensitet; *ghosting* i form av "fartsstriper" som følge av høy reduksjonsfaktor ved bruk av parallell bildedannelse. Disse kan være vanskelig å skille fra hverandre, og alle former for striper vurderes derfor under ett.

Susceptibilitetseffekter: Artefakter som følge av frekvensfeil på grunn av ujevn shimming i overganger mellom vev med ulik protontetthet. Ses som distorsjon (forvrengninger) og/eller signalbortfall i bildet.

Vekting: Vevets iboende kontrastegenskaper (T1-, T2- og Protonvekting), og i hvilken grad denne er fremstilt.

Svarskjema

Sagittal T1-vektet 3D-volum

Spørsmål	Senter 1 (VI)	Senter 2 (I)	Senter 3 (VII)	Senter 4 (VIII)	Senter 5 (V)	Senter 6 (II)	Senter 7 (III)	Senter 8 (IV)
<i>Differensiering av ulike vev og strukturer</i>								
1. Gråhvit substans	5	5	5	5	4	4	4	4
2a. Hippocampus	5	5	5	5	4	5	5	4
2b. Thalamus	4	5	5	4	5	4	4	5
2c. Basalganglier	5	5	5	4	5	4	4	5
<i>Artefakter i bilder</i>								
3a. Bevegelse	5	5	5	5	5	5	5	5
3b. Flow	5	5	5	5	5	5	5	5
3c. Susseptibilitet	5	5	4	5	4	5	4	4
3d. Kjemisk skift	5	4	5	5	4	5	5	4
3e. Aliasing	5	5	5	5	5	5	5	5
3f. Blurring	5	5	5	4	5	4	5	5
3g. Striper	5	5	5	5	5	4	4	5
<i>Bildekvalitet som helhet</i>								
4a. Kontrast/skarphet	5	5	5	5	4	4	5	4
4b. Vekting	5	5	5	5	5	4	5	5
4c. SNR	5	5	5	5	4	4	4	4
4d. Helhetsinntrykk	5	5	5	4	4	4	4	4

Transversal PD

Spørsmål	Senter 1 (VI)	Senter 2 (I)	Senter 3 (VII)	Senter 4 (VIII)	Senter 5 (V)	Senter 6 (II)	Senter 7 (III)	Senter 8 (IV)
<i>Differensiering av ulike vev og strukturer</i>								
1. Gråhvit substans	4	4	3	4	4	4	5	4
2a. Hippocampus	3	4	3	3	4	3	5	4
2b. Thalamus	3	4	3	3	5	4	3	4
2c. Basalganglier	5	4	3	5	5	4	5	4
<i>Artefakter i bilder</i>								
3a. Bevegelse	5	5	4	5	5	5	3	5
3b. Flow	5	5	4	5	5	3	5	5
3c. Susseptibilitet	5	5	4	5	5	4	5	5
3d. Kjemisk skift	5	4	4	5	5	5	4	4
3e. Aliasing	5	5	5	5	5	5	5	5
3f. Blurring	3	4	5	5	5	5	4	4
3g. Striper	5	5	4	5	5	4	3	5
<i>Bildekvalitet som helhet</i>								
4a. Kontrast/skarphet	3	5	4	4	4	4	5	5
4b. Vekting	4	5	4	5	4	4	5	5
4c. SNR	4	4	4	5	5	4	5	3
4d. Helhetsinntrykk	3	4	4	4	5	4	4	3

Transversal T2

Spørsmål	Senter 1 (VI)	Senter 2 (I)	Senter 3 (VII)	Senter 4 (VIII)	Senter 5 (V)	Senter 6 (II)	Senter 7 (III)	Senter 8 (IV)
<i>Differensiering av ulike vev og strukturer</i>								
1. Gråhvit substans	3	4	3	3	5	4	3	5
2a. Hippocampus	2	4	3	2	5	4	2	5
2b. Thalamus	3	4	3	3	5	4	3	4
2c. Basalganglier	3	5	3	4	5	4	4	5
<i>Artefakter i bilder</i>								
3a. Bevegelse	5	5	4	5	5	4	5	5
3b. Flow	5	5	4	5	5	4	5	5
3c. Susseptibilitet	5	5	4	5	5	4	5	5
3d. Kjemisk skift	5	4	4	5	5	4	5	4
3e. Aliasing	5	5	4	5	5	4	5	5
3f. Blurring	4	4	4	5	4	4	4	4
3g. Striper	3	5	4	5	5	4	2	5
<i>Bildekvalitet som helhet</i>								
4a. Kontrast/skarphet	3	4	3	3	5	4	3	5
4b. Vekting	4	5	4	5	4	4	5	4
4c. SNR	3	5	3	4	5	3	3	3
4d. Helhetsinntrykk	3	4	3	3	5	4	3	4

Transversal T2 FLAIR

Spørsmål	Senter 1 (VI)	Senter 2 (I)	Senter 3 (VII)	Senter 4 (VIII)	Senter 5 (V)	Senter 6 (II)	Senter 7 (III)	Senter 8 (IV)
<i>Differensiering av ulike vev og strukturer</i>								
1. Gråhvit substans	3	3	3	3	5	4	4	5
2a. Hippocampus	2	4	3	3	5	4	3	4
2b. Thalamus	1	5	3	2	5	4	3	5
2c. Basalganglier	2	3	3	3	5	4	3	5
<i>Artefakter i bilder</i>								
3a. Bevegelse	5	5	4	5	5	4	4	5
3b. Flow	5	4	4	5	4	4	5	3
3c. Susseptibilitet	5	5	4	5	5	4	5	5
3d. Kjemisk skift	5	5	4	5	5	4	5	5
3e. Aliasing	5	5	5	5	5	5	5	5
3f. Blurring	5	5	3	4	5	3	4	5
3g. Striper	5	5	4	4	5	4	3	5
<i>Bildekvalitet som helhet</i>								
4a. Kontrast/skarphet	2	5	3	3	5	4	4	3
4b. Vekting	5	5	4	4	5	4	4	3
4c. SNR	4	5	4	4	5	4	3	5
4d. Helhetsinntrykk	3	4	3	3	5	3	4	4

Transversal T2* (hemosekvens)

Spørsmål	Senter 1 (VI)	Senter 2 (I)	Senter 3 (VII)	Senter 4 (VIII)	Senter 5 (V)	Senter 6 (II)	Senter 7 (III)	Senter 8 (IV)
<i>Differensiering av ulike vev og strukturer</i>								
1. Gråhvit substans	4	4	4	2	3	4	3	4
2a. Hippocampus	1	4	4	1	3	4	1	4
2b. Thalamus	2	4	4	1	5	4	2	5
2c. Basalganglier	3	5	4	2	4	4	2	5
<i>Artefakter i bilder</i>								
3a. Bevegelse	5	5	4	5	4	4	5	5
3b. Flow	5	5	4	5	4	4	5	5
3c. Susseptibilitet	3	5	3	3	4	4	3	5
3d. Kjemisk skift	5	5	5	5	5	5	5	5
3e. Aliasing	5	5	5	5	5	5	5	5
3f. Blurring	5	5	4	3	4	4	4	4
3g. Striper	5	5	4	5	5	4	5	5
<i>Bildekvalitet som helhet</i>								
4a. Kontrast/skarphet	4	5	4	3	4	3	3	4
4b. Vekting	5	5	4	5	4	4	5	4
4c. SNR	4	5	4	3	4	3	4	4
4d. Helhetsinntrykk	5	5	4	3	4	3	4	4

MR scanner info

Navn på avdeling og sykehus:

MR-scanner (fabrikant, modell og feltstyrke):

Gradient system (inkluder styrke, rise time etc):

Scanner software (versjon):

Spole (hode spole eller hode-nakke, antall kanaler):

Mulighet for DTI: Ja/ Nei/ Vet ikke

Mulighet for fMRI: Ja/ Nei/ Vet ikke

(Ad fMRI: trenger en T2*-vektet sekvens med EPI, trenger IKKE stimuleringsutstyr da personene kun skal ligge rolig i scanneren og tenke på "ingenting" i 7-10 min mens fMRI opptaket gjøres).

Har T2* hemo sekvens: Ja/ Nei/ Vet ikke

Har Susceptibility Weighted Imaging (SWI) sekvens: Ja/ Nei/ Vet ikke

Kontakt personer ved avdelingen med e-post og telefonnummer:

Radiolog:

Navn:

e-post

telefon:

Radiograf:

Navn:

e-post

telefon:

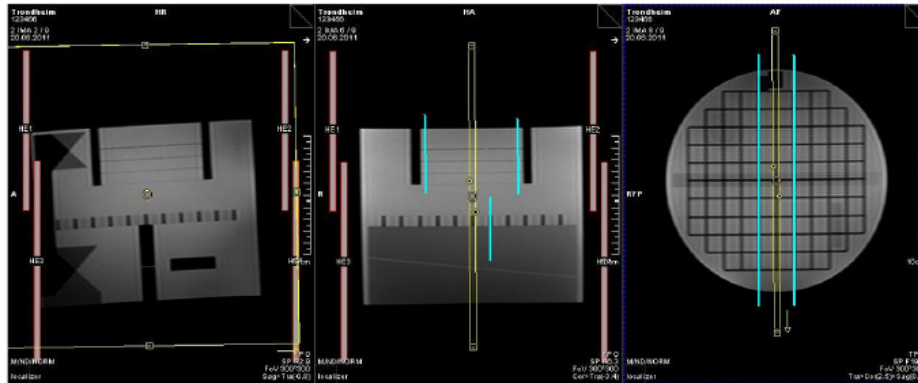
Annen informasjon som du/dere anser for viktig (f.eks. oppgradering av systemet, utfasing av MR systemet neste 4 år, spesielle utfordringer mht drift el.l.):

Kontaktperson ved Kompetansesenteret for fMRI, St. Olavs hospital:
Asta Håberg, professor neuroimaging, Institutt for nervemedisin/NTNU og
MR senteret, St. Olavs hospital, 7006 Trondheim.
Telefon: 73551352
e-post: asta.haberg@ntnu.no

Fantomopptak med ACR-fantom			
<p>Hvis spørsmål/problemer: Kontakt Ingeborg Nakken, mob:48023327, epost: inakken@hotmail.com</p>			
<p>SCANNING:</p> <p>Temperaturregistrering: Plasser termometeret inne i scannerrommet før scanningen starter, f.eks på en hylle ved veggen. Les av temperatur og luftfuktighet etter gjennomført scanning.</p> <p>Plassering: Legg fantomet i headcoilen som om det var et hode, med "NOSE" (nese) og "CHIN" (hake) plassert slik pasientene ligger i scanneren. Listen som er montert på fantomets "hals" skal ligge vannrett (bruk vater).</p> <p>Stabilisering: Støtt opp med puter og lignende slik at fantomet blir liggende fast og i vater.</p> <p>Sentrering: Laserlysene skal treffe det sorte krysset ved "NOSE", samt midten av coilen.</p>			
<p>REGISTRERING PÅ SCANNER:</p> <p><u>Navn:</u> Stedet opptakene gjøres</p> <p><u>Vekt:</u> 100 kg</p>			
<p>LAGRING AV BILDER OG UTFYLLING AV SKJEMA:</p> <p>Alle bildene brennes på CD som merkes godt, eller legges inn på vedlagte minnepinne. <i>Vedlagte registreringskjema fylles ut ved hvert senter.</i></p>			
<p>SCAN PROTOKOLL: Se neste side for snittplassering! Scanparameter på eget ark.</p> <ol style="list-style-type: none"> 1. Localizer/Survey 2. Sagittal SE T1 (1 snitt) 3. Axial T1 SE (11 snitt) 4. Axial SE Pd/T2 Dual echo (11 snitt) 5. Axial TSE/FSE Pd/T2 Dual echo (11 snitt) 			
<p>LOGISTIKK:</p> <p>Fantomet transporteres mellom sentrene ved hjelp av Bring Budservice. Adresselapper medfølger eller tilsendes underveis.</p>			
<p>PAKKEN INNEHOLDER:</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top;"> <ul style="list-style-type: none"> • Fantom • Vater • Termometer • USB-pinne </td> <td style="width: 50%; vertical-align: top;"> <ul style="list-style-type: none"> • Registreringskjema • Oppskrift • Oversikt over pakkens reiserute </td> </tr> </table>		<ul style="list-style-type: none"> • Fantom • Vater • Termometer • USB-pinne 	<ul style="list-style-type: none"> • Registreringskjema • Oppskrift • Oversikt over pakkens reiserute
<ul style="list-style-type: none"> • Fantom • Vater • Termometer • USB-pinne 	<ul style="list-style-type: none"> • Registreringskjema • Oppskrift • Oversikt over pakkens reiserute 		

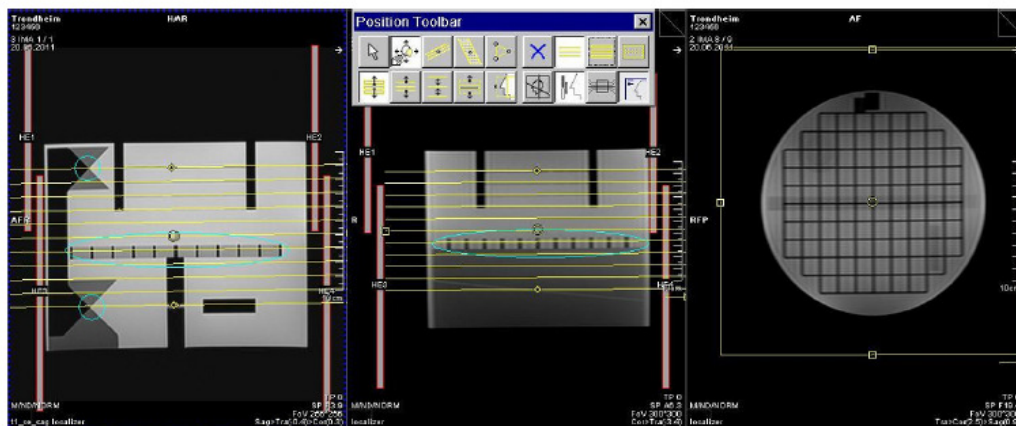
Sag T1

Snittpakken vinkles etter fantomets linjer (se turkise hjelpelinjer).



AXIALE SNITT

Snitt 1 og 11 treffer krysset i kilene til venstre (turkise sirkler). Vinkles etter gitter i midten (turkise ovaler).



Registreringsskjema for fantomopptak

Senter:

Radiograf:

Dato:

Klokkeslett:

Temperatur scannerrom:

Luftfuktighet scannerrom:

Opptaksparameter:

1) Båndbredde

Sag T1:

Aksial T1 SE:

Aksial pd+T2 SE:

Aksial pd+T2 TSE/FSE:

2) Turbofaktor

Aksial pd+T2 TSE/FSE:

Dato forrige hovedservice:

Generelle parameter fantomopptak				
Sekvens	Transversale sekvenser (T1, SE og TSE dobbelekk)		Sagittal T1	
Antall snitt	11		1	
Snittykkelse	5 mm		5 mm	
Snittmellomrom	5 mm		5 mm	
FOV	250 x 250 mm		250 x 250 mm	
Matrisestørrelse	256 x 256 px ⁽¹⁾		256 x 256 px	
Sekvensspesifikke parameter fantomopptak				
Sekvens	Sagittal T1	Tra T1	Tra dobbelekk	Tra dobbelekk
Sekvenstype	SE	SE	SE	TSE
Opptakstid	~ 1 min	~ 2:30 min	~ 8:30 min	~ 2:40 min
TR	200 ms	500 ms	2000 ms	2000 ms
TE 1	20 ms	20 ms	20 ms	20 ms
TE 2	-	-	80 ms	80 ms
Antall ekko	-	-	-	7-16
rBW	150-180 Hz/px	140-150 Hz/px	130-150 Hz/px	161-165 Hz/px

⁽¹⁾px = piksler

Generelle parameter hjerneopptak			
Sekvens	Transversale sekvenser (dobbelekk, T2 Flair og T2*)		
Antall snitt	44-48		
Snittykkelse	3 mm		
Snittmellomrom	0 mm		
FOV	230 x 230 mm		
Matrisestørrelse	256 x 256 px		
Vokselstørrelse	0,9 x 0,9 x 3,0 mm		
Sekvensspesifikke parameter hjerneopptak			
Sekvens	Dobbelekk	T2 Flair	T2*
Sekvenstype	TSE	IR TSE	GRE / EPI
Opptakstid	~5:30 min	~ 4-6 min	~ 5-6 min
TR	3000-3030 ms	9000-10000 ms	683-835 ms
TE 1	11-13 ms	107-121 ms	21-26 ms
TE 2	96-102 ms	-	-
TI	-	2500 ms	-
FA	150°-180°	150°-180°	20°-30°
Antall ekko	7-16	15-34	-
rBW	161-165 Hz/px	174-305 Hz/px	80-122 Hz/px (EPI: 770 Hz/px)

Fra: Regional komite for medisinsk og helsefaglig forskningsetikk REK midt

Til:

Ingeborg.Nakken@stolav.no

rek-4@medisin.ntnu.no

Dokumentreferanse: 2010/3140-2

Dokumentdato: 01.12.2010

**BILDEKVALITET PÅ MR VED MULTISENTERSTUDIER UTVIKLINGSARBEID;
IKKE SØKEPLIKTIG**

Hei, viser til e-post mottatt 24.11.10 om spørsmål om søkeplikt til REK. Komiteens leder har vurdert din forespørsel og kommet frem til at det skisserte prosjektet er utviklingsarbeid. I Norsk Lovkommentar til Helseforskningsloven står det følgende: "Utviklingsarbeid ligger i grenseland, men vil ifølge forarbeidene i utgangspunktet kun omfattes dersom målet er å frembringe kunnskap, og ikke dersom målet bare er å justere instrumenter e.l..." Ditt prosjekt er altså vurdert til å havne i sistnevnte kategori, og er dermed ikke søkepliktig. Vi minner imidlertid om at du må melde ditt prosjekt til personvernombudet/NSD ved St. Olavs Hospital.

Vennlig hilsen

Hilde Eikemo, rådgiver REK Midt

Norsk samfunnsvitenskapelig datatjeneste AS
NORWEGIAN SOCIAL SCIENCE DATA SERVICES



Harald Hårfagres gate 29
N-5007 Bergen
Norway
Tel: +47-55 58 21 17
Fax: +47-55 58 96 50
nsd@nsd.uib.no
www.nsd.uib.no
Org.nr. 985 321 884

Ingeborg Nakken
Klinikk for bildeagnostikk
St. Olavs Hospital HF
Olav Kyrresgt. 17
7006 TRONDHEIM

Vår dato: 17.10.2011

Vår ref: 28065 / 3 / MAB

Deres dato:

Deres ref:

KVITTERING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 18.09.2011. Meldingen gjelder prosjektet:

28065	<i>Kvalitetssikring av data innsamlet ved hjelp av magnetiske resonanstomografi ved multisenterstudier</i>
Behandlingsansvarlig	<i>St. Olavs Hospital HF, ved institusjonens overste leder</i>
Daglig ansvarlig	<i>Ingeborg Nakken</i>

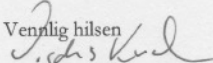
Personvernombudet har vurdert prosjektet og finner at behandlingen av personopplysninger er meldepliktig i henhold til personopplysningsloven § 31. Behandlingen tilfredsstiller kravene i personopplysningsloven.

Personvernombudets vurdering forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, eventuelle kommentarer samt personopplysningsloven/-helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, http://www.nsd.uib.no/personvern/forsk_stud/skjema.html. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, <http://www.nsd.uib.no/personvern/prosjektoversikt.jsp>.

Personvernombudet vil ved prosjektets avslutning, 15.11.2011, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen

Vigdis Namtvedt Kvalheim


Marte Bertelsen

Kontaktperson: Marte Bertelsen tlf: 55 58 33 48
Vedlegg: Prosjektvurdering

Avdelingskontorer / District Offices:

OSLO: NSD, Universitetet i Oslo, Postboks 1055 Blindern, 0316 Oslo. Tel: +47-22 85 52 11. nsd@uio.no

TRONDHEIM: NSD, Norges teknisk-naturvitenskapelige universitet, 7491 Trondheim. Tel: +47-73 59 19 07. kyrre.svarva@svt.ntnu.no

TROMSØ: NSD, HSL, Universitetet i Tromsø, 9037 Tromsø. Tel: +47-77 64 43 36. martin-arne.andersen@uit.no