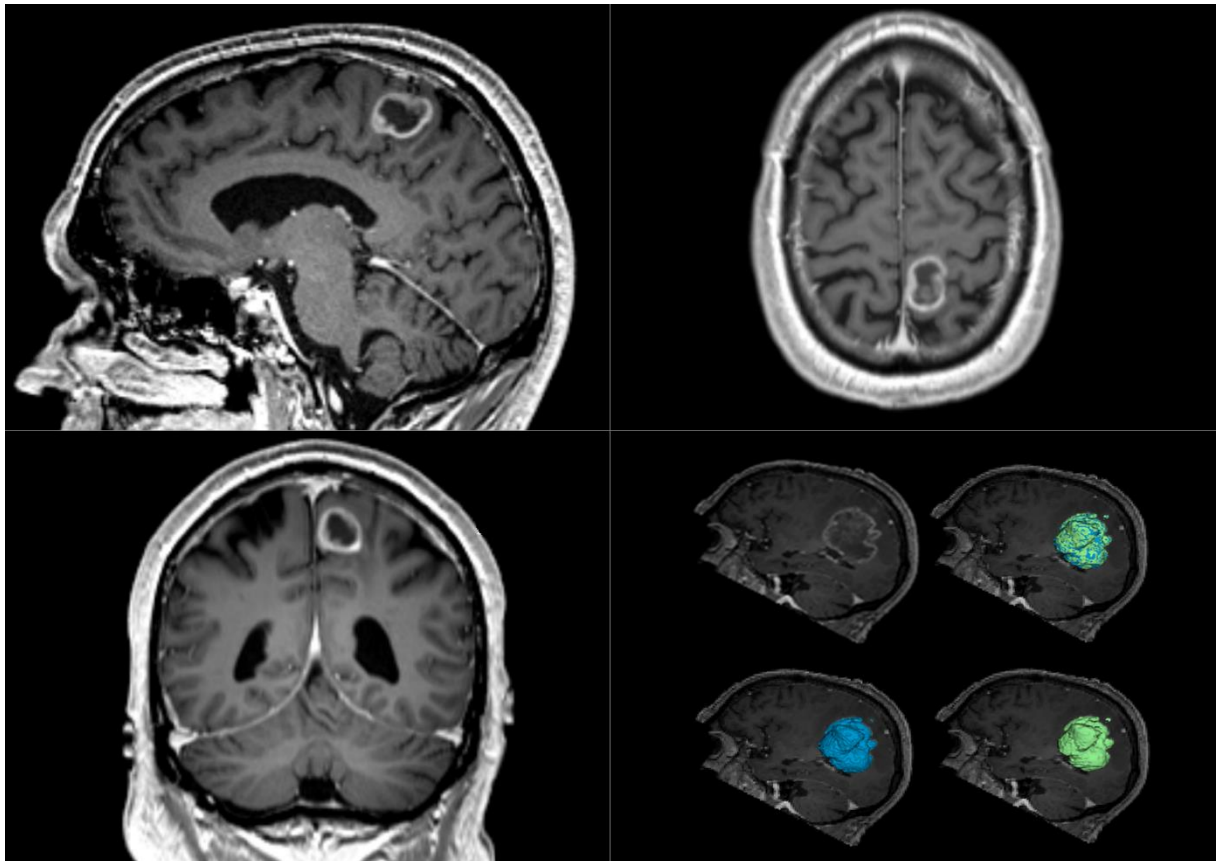


# Radiologisk vekstmåling av glioblastomer



# Innholdsfortegnelse

<b>FORORD</b> .....	<b>3</b>
<b>FORKORTELSER</b> .....	<b>5</b>
<b>SAMMENDRAG</b> .....	<b>7</b>
<b>1. BAKGRUNN</b> .....	<b>9</b>
1.1 GLIOBLASTOMER .....	9
1.2 BILDEDIAGNOSTIKK AV GLIOBLASTOMER .....	9
1.3 ESTIMERING AV TUMORSTØRRELSE .....	11
1.3.1 WHO kriteriene .....	11
1.3.2 Macdonald kriteriene.....	12
1.3.3 RANO-kriteriene.....	13
1.3.4 Volumsegmentering.....	16
1.4 VARIABILITET VED ESTIMERING AV TUMORSTØRRELSE.....	17
1.5 MÅLSETTING OG HYPOTESE .....	18
<b>2. MATERIALE OG METODE</b> .....	<b>19</b>
2.1 BEDØMMING AV TUMORVEKST .....	19
2.1.1 Radiologisk bedømming ved operasjonstidspunkt .....	20
2.1.2 Ny visuell radiologisk bedømming .....	21
2.1.3 RANO-måling .....	21
2.2 VOLUMSEGMENTERING .....	22
2.3 STATISTISKE ANALYSER .....	23
2.3.1 Frekvenstabeller.....	23
2.3.2 Krysstabeller .....	23
2.3.3 Kappa.....	23
2.3.4 Sensitivitet, spesifisitet, PPV og NPV .....	24
<b>3. RESULTATER</b> .....	<b>25</b>
3.1 RADIOLOGISK BEDØMMING AV VEKST VED OPERASJONSTIDSPUNKT .....	25
3.2 INTER-RATER VARIABILITET .....	26
3.2.1: Inter-rater variabilitet ved Visuell radiologisk bedømming.....	27
3.2.2 Inter-rater variabilitet ved RANO-måling .....	28
3.3 VISUELL RADIOLOGISK BEDØMMING VS. VOLUMSEGMENTERING .....	29
3.4 RANO-MÅLING VS. VOLUMSEGMENTERING .....	31
<b>4. DISKUSJON</b> .....	<b>33</b>
4.1 KLINISK PRAKSIS I DAG.....	33
4.2 INTER-RATER VARIABILITET .....	34
4.2.1 Visuell radiologisk bedømming.....	34
4.2.2 RANO-måling .....	37
4.2.3 Visuell radiologisk bedømming eller RANO-måling? .....	37
4.3 VISUELL RADIOLOGISK BEDØMMING VS. VOLUMSEGMENTERING .....	38
4.4 RANO-MÅLING VS. VOLUMSEGMENTERING .....	39
4.5 VISUELL RADIOLOGISK BEDØMMING, RANO-MÅLING ELLER VOLUMSEGMENTERING? .....	40
<b>5. KONKLUSJON</b> .....	<b>43</b>
<b>6. REFERANSER</b> .....	<b>45</b>
<b>APPENDIX</b> .....	<b>47</b>

## **Forord**

Vi vil rette en stor takk til vår hovedveileder, førsteamanuensis/overlege Erik Magnus Berntsen, for tett oppfølging og konstruktive tilbakemeldinger. Vi ønsker også å takke nevreradiologene Pål Christensen og Viggo Moholdt for nevreradiologisk ekspertise og den radiologiske gjennomgangen av materialet. Vi vil også takke medisinstudent/forskerlinjestudent Anne Line Stensjøen, nevreradiolog/seksjonsoverlege Kjell Arne Kvistad og biveileder professor/overlege Ole Solheim for tilgang til bakgrunnsmateriale og verdifulle innspill underveis.



# **Forkortelser**

<b>CT</b>	Computertomografi ( <i>engelsk: Computed Tomography</i> )
<b>GBM</b>	Glioblastoma Multiforme
<b>MR</b>	Magnetisk Resonanstomografi ( <i>engelsk: MRI – Magnetic Resonance Imaging</i> )
<b>NPV</b>	Negativ Prediktiv Verdi
<b>PACS</b>	Picture Archiving and Communication System
<b>PPV</b>	Positiv Prediktiv Verdi
<b>RANO</b>	Response Assessment in Neuro-Oncology



# Sammendrag

## **Bakgrunn:**

Anbefalt metode for bedømming av vekst av glioblastomer er i dag todimensjonale mål i henhold til RANO-kriteriene. En viktig svakhet ved denne metoden er at den bruker todimensjonale mål på en irregulær tredimensjonal struktur. Volumsegmentering blir regnet som gullstandard og brukes noe innen forskning, men er for ressurskrevende for klinisk bruk. Vi har i denne oppgaven sammenlignet to ulike metoder for radiologisk vekstvurdering av glioblastomer med volumsegmentering, samt undersøkt hvordan veksten av disse tumorene har blitt beskrevet i klinisk praksis

## **Metode:**

Vi har tatt utgangspunkt i 106 pasienter med histologisk verifisert glioblastom hvor volumsegmenteringen er blitt utført tidligere. Vi har først undersøkt hvordan veksten av disse har blitt beskrevet i klinisk praksis, ved å gå gjennom de radiologiske beskrivelsene før primæroperasjonen. Deretter har to nevreradiologer gjort en ny vekstvurdering på følgende to måter: 1) Ny visuell bedømming ved kun å bla i MR-bildene, og 2) Ny estimering av tumorstørrelse etter RANO-kriteriene (todimensjonale mål). Disse to metodene har så blitt sammenlignet med volumsegmentering av de samme tumorene.

## **Resultat:**

Når vi ser på metodenes evne til å angi progresjon, oppnår visuell radiologisk bedømming og RANO-måling tilnærmet like resultater, med en sensitivitet på ca. 0,60, en spesifisitet på ca. 0,90, negativ prediktiv verdi på ca. 0,50 og positiv prediktiv verdi på ca. 0,90. Inter-rater variabilitet er moderat ved begge metodene, med Kappa-verdier på ca. 0,70.

## **Konklusjon:**

Visuell radiologisk bedømming og RANO-måling gir upresise vekstestimer av glioblastomer, og omtrent 40% av svulstene med reell radiologisk vekst vil ikke bli fanget opp med disse metodene. Volumsegmentering er en betydelig mer nøyaktig metode, men er per i dag lite egnet for klinisk bruk grunnet høyt tidsforbruk.





# **1. BAKGRUNN**

## **1.1 Glioblastomer**

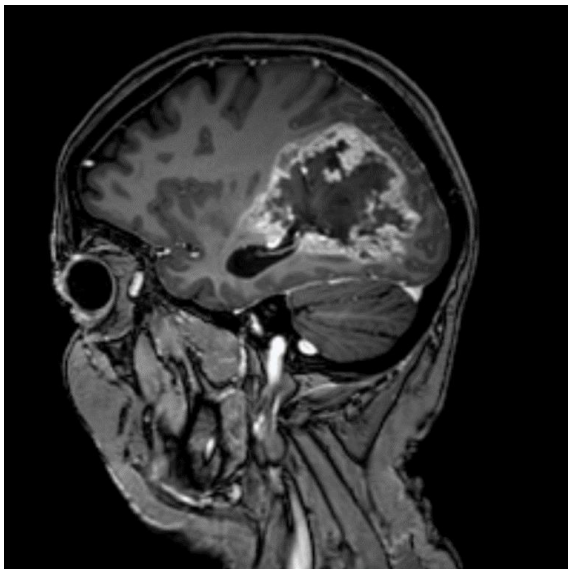
Glioblastomer er den hyppigste og mest aggressive primære hjernesvulsten hos mennesker. Den kalles også for astrocytom grad IV, da den er gradert til en WHO grad IV tumor, og ble tidligere kalt for glioblastoma multiforme (GBM). Glioblastomer kan deles inn i primære og sekundære, avhengig av hvilke celler de utvikler seg fra. De primære glioblastomene oppstår sannsynligvis direkte fra nevralt stamceller, mens de sekundære utvikler seg fra lavere grads astrocytomer (grad II og III) [1]. Insidensen av glioblastomer hos voksne er 3,19/100 000 [2]. Det finnes per i dag ingen kurativ behandling for glioblastomer. Standard behandling er kirurgisk reseksjon med påfølgende strålebehandling og kjemoterapi [3], men noen har for utbredt sykdom ved diagnosetidspunkt eller er for syke til å gjennomgå denne behandlingen. Gjennomsnittlig median overlevelse er 10 måneder, og øker til 16 måneder for de som gjennomgår kirurgisk reseksjon og påfølgende stråleterapi og Temozolomide kjemoterapi [4]. Glioblastomer vokser infiltrerende i omkringliggende vev og radikal kirurgi er derfor ikke mulig. Målet er likevel å fjerne så mye av svulsten som mulig, uten at det går utover viktige nevrologiske funksjoner som språk, motorikk og syn, da dette er vist å bedre overlevelsen [5]. Videre onkologisk behandling, stråling og kjemoterapi, igangsettes etter histologisk undersøkelse. Postoperativ strålebehandling er brukt i behandling av glioblastomer siden 1970-tallet, og siden 2005 har også kjemoterapi vært en sentral del av behandlingen. Under hele behandlingen følges klinisk og radiologisk respons nøye [6]. Ved progresjon under pågående behandling vil det ofte være aktuelt å avslutte adjuvant kreftbehandling. Etter avsluttet primærbehandling kan re-kirurgi, re-bestråling eller andrelinje kjemoterapi være aktuelt ved radiologisk progresjon. Nevroradiologiske undersøkelser er derfor en viktig del av det kliniske beslutningsgrunnlaget.

## **1.2 Bildediagnostikk av glioblastomer**

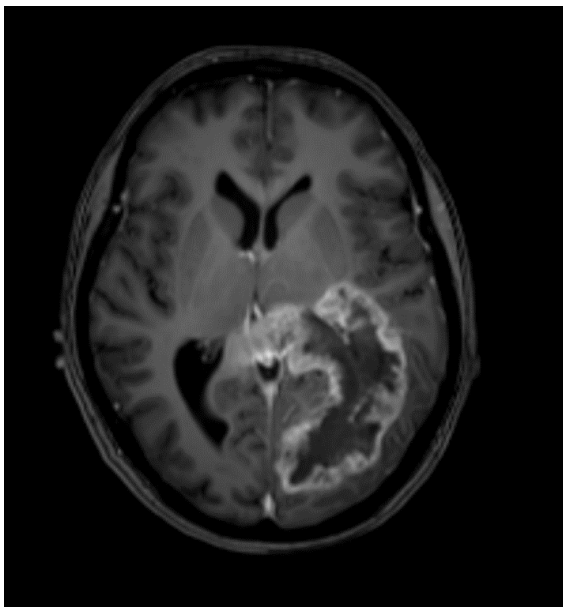
Den bildediagnostiske utredningen av alle hjernesvulster gjøres i dag med MR (Magnet Resonanstomografi (*engelsk*: MRI – Magnetic Resonance Imaging), både med og uten intravenøs MR-kontrast, med mindre det foreligger sterke kontraindikasjoner mot MR. Glioblastomer har nydannede patologiske blodårer som gjør at kontrastmidlet lekker ut og konsentrerer seg i svulsten. Typiske radiologiske trekk ved et glioblastom på

kontrastforsterkede  $T_1$ -vektede bilder er en sterkt kontrastladende irregulær ringformet tumor med en sentral ikke-kontrastladende nekrotisk kjerne med omliggende ødem [1] (Figur 1 og 2).

Svulstene vokser diffust og infiltrativt i normalt hjernevev og det er alltid tumorceller også utenfor de kontrastladende områdene. Dette gjør at den reelle tumorstørrelsen alltid er større enn det man klarer å fremstille med dagens bildeteknologi. Det er likevel som oftest den kontrastladende delen av tumor som brukes som utgangspunkt for estimering av tumorstørrelse, progresjon og behandlingsresponser [7, 8].



*Figur 1: Kontrastforsterket sagittalt  $T_1$  vektet MR-bilde av et histologisk verifisert glioblastom.*



*Figur 2: Kontrastforsterket transversalt  $T_1$  vektet MR-bilde av et histologisk verifisert glioblastom.*

## **1.3 Estimering av tumorstørrelse**

Siden de første kriteriene for estimering av tumorstørrelse ble lansert i 1979 (WHO-kriteriene), har det blitt utviklet flere ulike systemer og kriterier for vekstvurdering og responsevaluering av glioblastomer [9-11]. Metodene for estimering av tumorstørrelse har gjennomgått flere endringer og forbedringer i løpet av årene. Dette skyldes både endringer i bildeteknologi og kirurgisk teknologi, økende kunnskap om hvordan steroider, anti-angiogenetiske medikamenter, immunmodulerende medikamenter og annen behandling påvirker svulstene, men også en økende forståelse av kriterienes begrensninger [8, 12, 13]. De siste retningslinjene for responsevaluering av høygradige gliomer ble publisert av «*Response Assessment in Neuro-Oncology (RANO) Working Group*» i 2010. RANO-kriteriene bygger på de forutgående Macdonald-kriteriene, som igjen bygger på WHO-kriteriene [8].

RANO-kriteriene brukes lite i vanlig klinisk praksis, men er hyppigere brukt i studie- og forskningssammenheng, for eksempel ved multisenter-studier for da å nøyaktigere kunne sammenligne effekten av ulike behandlinger opp mot hverandre. I klinisk praksis er det nok heterogene lokale tradisjoner for hvordan vekst/progresjon bedømmes, for eksempel kun visuell bedømming (såkalt «eye-balling»), eventuelt i kombinasjon med ikke-systematiserte diametermålinger.

I dag er volumsegmentering av svulster tilgjengelig ved mange bildediagnostiske senter, og denne metoden for å estimere tumorstørrelse blir per i dag sett på som den mest nøyaktige (gullstandard) [14]. På grunn av flere ulike utfordringer ved volumsegmentering, som beskrives i avsnitt «1.3.4 Volumsegmentering», er den likevel ikke standard vurderingsmetode for tumorvekst i klinisk praksis, hovedsakelig på grunn av det er en tidkrevende prosedyre [8].

### **1.3.1 WHO kriteriene**

WHO-kriteriene fra 1979 var et tidlig forsøk på å implementere objektive mål på behandlingsrespons i onkologi. Kriterienes responsevaluering baserte seg blant annet på endringer i tumorstørrelse. Tumorstørrelsen ble definert ved hjelp av både en- og todimensjonale mål av tumor, der todimensjonale mål ble definert som produktet av den største diameteren multiplisert med diameteren perpendikulær på denne. Ut fra størrelsesendringer ble svulstene delt inn i ulike responskategorier. Følgende kategorier ble brukt: Komplet respons

(ingen gjenværende lesjon), Partiell respons ( $\geq 50\%$  reduksjon i tumorstørrelse), Stabil sykdom (mellom 50% reduksjon og 25% økning i tumorstørrelse) og Progressiv sykdom ( $\geq 25\%$  økning i tumorstørrelse) [9-11]. Dagens kriterier for responseevaluering av høygradige gliomer (RANO-kriteriene) bygger på disse WHO-kriteriene.

### **1.3.2 Macdonald kriteriene**

I 1990 publiserte Macdonald et al. Macdonald-kriteriene for responseevaluering av høygradige gliomer [15]. Kriteriene baserte seg primært på todimensjonale mål, og som i WHO-kriteriene ble tumorstørrelsen angitt som produktet av den største diameteren i et tverrsnitt av tumor og diameteren perpendikulær til denne. Målingene ble utført på den kontrastladende delen av tumor på computertomografi-bilder (CT-bilder, *engelsk*: Computed Tomography). Macdonald-kriteriene hadde flere forbedringer som gjorde det lettere å sammenligne pasienter i kliniske studier, blant annet at man nå også tok høyde for bruk av kortikosterioder og endringer i nevrologisk status hos pasienten. Kriteriene ble i utgangspunktet utviklet for CT-bilder, men har senere blitt tatt i bruk på MR-bilder, da dette nå er standard bildemodalitet for responseevaluering av høygradige gliomer [8].

Macdonald-kriteriene hadde flere begrensninger. Blant annet vanskeligheter med å måle irregulære tumorer, variasjon i målinger mellom radiologer, ingen vurdering av ikke-kontrastoppladende deler av tumor og manglende kriterier for vurdering av multifokale tumorer [8]. I tillegg så man store begrensninger ved vurdering av behandlingseffekt på grunn fenomenene pseudoprogresjon og pseudorespons. Pseudoprogresjon er en forbigående økning i kontrastopptak på MR-bilder som senere forsvinner uten endring i terapi. Dette fenomenet skyldes trolig økt permeabilitet i tumorvevets karforsyning i uker/måneder etter stråleterapi, og kan trolig potenseres av kjemoterapi med Temozolomide [16]. Pseudoprogresjon av tumor forekommer hos inntil 50% av pasienter med påvist progresjon som får postoperativ radio- og kjemoterapi [17]. Pseudorespons derimot simulerer behandlingseffekt ved at antiangiogenetisk behandling (Bevacizumab®) normaliserer hyperpermeable kar i tumor og gir en redusert kontrastoppladning i tumor uten en faktisk reduksjon i tumorvev [8, 18]. Bevacizumab er imidlertid lite brukt etter at to randomiserte studier ikke kunne påvise effekt på overlevelse [19].

I nesten 20 år var Macdonald-kriteriene mest brukt i responsevaluering og studier av høygradige gliomer. Begrensningene ved kriteriene gjorde at de ble revidert, og RANO-kriteriene ble presentert i 2010 [13].

### **1.3.3 RANO-kriteriene**

I 2010 ble “The Revised Assessment in Neuro-Oncology” (RANO-kriteriene) publisert. Denne oppdateringen, som bygger på Macdonald-kriteriene, inneholdt blant annet presise definisjoner av hva som er målbare og ikke-målbare kontrastladende lesjoner, samt egne kriterier for å vurdere de ikke-kontrastladende lesjonene. Det ble også innført kriterier for hvordan man skulle velge lesjoner ved multifokal sykdom, samt utviklet egne kriterier for å vurdere pasienter som hadde gjennomgått radio- og/eller kjemoterapi. I tillegg fikk man mer presise definisjoner for respons og progresjon, samt kriterier for å kunne inkludere pasienter i kliniske studier [8, 13].

RANO-kriteriene brukes til å avgjøre om det har foreligget endring i sykdomsomsfang. Dette gjøres ved å dele svulstene inn i ulike responskategorier etter ulike krav og parametere. Eksempel på slike parametere er størrelsesendringen på den kontrastladende- og ikke-kontrastladende delen av tumor, eventuelle nye tumorer/lesjoner, bruk av kortikosteroider og klinisk status. Dette vises i oversikten under, som inneholder kriteriene for responsevaluering utarbeidet av RANO Working Group [8].

## Criteria for Response Assessment Incorporating MRI and Clinical Factors

Response	Criteria
<i>Complete response</i>	Requires all of the following: complete disappearance of all enhancing measurable and nonmeasurable disease sustained for at least 4 weeks; no new lesions; stable or improved nonenhancing (T2/FLAIR) lesions; patients must be off corticosteroids (or on physiologic replacement doses only); and stable or improved clinically. Note: Patients with nonmeasurable disease only cannot have a complete response; the best response possible is stable disease.
<i>Partial response</i>	Requires all of the following: $\geq 50\%$ decrease compared with baseline in the sum of products of perpendicular diameters of all measurable enhancing lesions sustained for at least 4 weeks; no progression of nonmeasurable disease; no new lesions; stable or improved nonenhancing (T2/FLAIR) lesions on same or lower dose of corticosteroids compared with baseline scan; the corticosteroid dose at the time of the scan evaluation should be no greater than the dose at time of baseline scan; and stable or improved clinically. Note: Patients with nonmeasurable disease only cannot have a partial response; the best response possible is stable disease.
<i>Stable disease</i>	Requires all of the following: does not qualify for complete response, partial response, or progression; stable nonenhancing (T2/FLAIR) lesions on same or lower dose of corticosteroids compared with baseline scan. In the event that the corticosteroid dose was increased for new symptoms and signs without confirmation of disease progression on neuroimaging, and subsequent follow-up imaging shows that this increase in corticosteroids was required because of disease progression, the last scan considered to show stable disease will be the scan obtained when the corticosteroid dose was equivalent to the baseline dose.
<i>Progression</i>	Defined by any of the following: $\geq 25\%$ increase in sum of the products of perpendicular diameters of enhancing lesions compared with the smallest tumor measurement obtained either at baseline (if no decrease) or best response, on stable or increasing doses of corticosteroids*; significant increase in T2/FLAIR nonenhancing lesion on stable or increasing doses of corticosteroids compared with baseline scan or best response after initiation of therapy* not caused by comorbid events (eg, radiation therapy, demyelination, ischemic injury, infection, seizures, postoperative changes, or other treatment effects); any new lesion; clear clinical deterioration not attributable to other causes apart from the tumor (eg, seizures, medication adverse effects, complications of therapy, cerebrovascular events, infection, and soon) or changes in corticosteroid dose; failure to return for evaluation as a result of death or deteriorating condition; or clear progression of nonmeasurable disease.

NOTE. All measurable and nonmeasurable lesions must be assessed using the same techniques as at baseline. Abbreviations: MRI, magnetic resonance imaging; FLAIR, fluid-attenuated inversion recovery.  
\*Stable doses of corticosteroids include patients not on corticosteroids.

Ved bruk av RANO-kriteriene vurderer man altså eventuell progresjon blant annet ut fra endring i tumorstørrelse på målbare lesjoner. Under følger en sammenfatning av kriteriene på forrige side, som tydeliggjør størrelsesendringen som er med på å avgjøre om en tumor havner i den ene eller andre responsgruppen [8]:

<b>Responskategori</b>	<b>Estimert tumorstørrelse</b>
Komplett respons	Ingen gjenværende kontrastladende lesjon
Partiell respons	$\geq 50\%$ reduksjon i estimert tumorstørrelse
Stabil sykdom	Mellom 25% økning og 50% reduksjon i estimert tumorstørrelse
Progresjon	$\geq 25\%$ økning i estimert tumorstørrelse

Tumorstørrelsen blir, som ved Macdonald-kriteriene, angitt som produktet av to målte diametere, den største diameteren og den perpendikulære diameteren til denne. Målingene gjøres på kontrastladende del av tumor på aksiale snitt. Disse målene brukes til å estimere tumorstørrelsen, og anslå om den kontrastladende delen av tumor har redusert i størrelse, har progrediert eller forblitt uendret. ***Det er måling av denne størrelsesendringen på den kontrastladende delen av tumoren vi ser på i vår hovedoppgave, hvor vi evaluerer hvor nøyaktig ulike metoder bedømmer denne størrelsesendringen.***

RANO-kriteriene skiller mellom målbare og ikke-målbare kontrastladende lesjoner. For å kunne anse en lesjon som målbar, må følgende kriterier være oppfylt [8]:

1. Kontrastladende lesjon med tydelig avgrensning, med to perpendikulære diametere på minst 10 mm hver, målt på aksiale snitt
2. Lesjonen skal være synlig på minst to aksiale snitt med maksimum 5 mm tykkelse, uten avstand mellom snittene (ved tykkere snitt eller avstand mellom snittene må tumoren ved «baseline» ha diametere som er 2 ganger større enn summen av snittykkelsen og avstanden mellom snittene)
3. Cyster og operasjonskaviteter skal ikke inngå i tumorstørrelse, og er bare målbare dersom den kontrastladende delen har nodulære fortykkelser  $> 10$  mm.
4. Ved multiple kontrastladende lesjoner skal minst 2 og maks 5 lesjoner måles, og produktet av hver lesjon sin størrelse skal summeres, hvorpå summen skal representere estimert tumorstørrelse

Fylles ikke disse kriteriene blir lesjonen definert som “ikke-målbar kontrastladende lesjon”. Det kan for eksempel være lesjoner som vises på kun ett snitt, lesjoner uten klar avgrensing og lesjoner med minst en av de perpendikulære diameterene under 10mm [8].

Til tross for at RANO-kriteriene inneholdt flere forbedringer og strengere retningslinjer for responseevaluering, har de noen svakheter. Med RANOs todimensjonale størrelsesestimat blir høygradige gliomer betraktet som et todimensjonalt objekt, noe som ikke samsvarer med virkeligheten. Glioblastomer er diffust infiltrerende svulster med utbredelse i alle retninger og kan anta en rekke ulike geometriske former. RANO Working Group vektlegger imidlertid at det ikke er tilstrekkelig standardiserte og tilgjengelige metoder for volumetrisk estimering av tumorstørrelsen til at de ønsker å ta inn volumsegmentering som en del av vurderingen. De åpner imidlertid for at dette kan være en mulighet i framtiden når volumsegmentering blir mer tilgjengelig og standardisert [8].

#### **1.3.4 Volumsegmentering**

Både Macdonald- og RANO-kriteriene bygger på todimensjonale mål av svulsten ved estimering av tumorstørrelse. Volumetriske metoder for vekstmåling ses på som mer nøyaktige og reproducerbare enn todimensjonale metoder, da man ved volumetriske målinger bygger på tredimensjonale målinger av svulsten [13]. Mange segmenteringsprogramvarer er gratis tilgjengelig som for eksempel 3D Slicer ([www.slicer.org](http://www.slicer.org)) og ITK-Snap ([www.itksnap.org](http://www.itksnap.org)), men det finnes ingen standardiserte eller godt validerte full-automatiske segmenteringsmetoder. Manuell segmentering ansees derfor som gullstandard, men er for tidkrevende til at det gjøres rutinemessige i klinisk praksis [14].

De volumetriske målingene gjøres vanligvis ved at et dataprogram foreslår en grense mellom kontrastladende og ikke-kontrastladende deler av tumor. Denne grensen kan senere justeres manuelt av en nevrolog, såkalt semi-automatisk segmentering. Dataprogrammet kan videre regne ut totalt volum av tumor, samt kontrastladende volum og ikke-kontrastladende volum (for eksempel sentral nekrose eller cystiske deler av tumor) [20].

Den viktigste fordelene ved volumsegmentering er at svulstenes volum blir målt nøyaktig, særlig ved irregulære svulster og svulster med mye nekrose. Man kan også utelukke cystiske komponenter fra volummålet [21]. Flere mulige fordeler er at man enklere finner og måler små lesjoner, reduserer inter- og intra-rater variabilitet, samt at man tidligere ser endringer i



tumorstørrelse [13, 20]. Noen ulemper ved volumsegmentering er at det er tidkrevende å utføre og man trenger spesialkompetanse og egen dataprogramvare, og metoden er derfor ressurskrevende. Det kan i visse tilfeller også være utfordrende å segmentere gliomer der perifer kontrastladning er svært irregulær eller kun flekkvis [20]. Manglende standardisering og høyt tidsforbruk gjør at volumsegmentering ikke brukes rutinemessig i klinisk praksis, til tross for at dette er den mest nøyaktige metoden for responseevaluering av høygradige gliomer. Det kan være mulig å implementere volumsegmentering som standard metode i fremtiden, ikke minst dersom tilgjengelighet og standardisering forbedres [8].

## **1.4 Variabilitet ved estimering av tumorstørrelse**

Reproduserbarhet er en viktig faktor i alle typer målinger, og denne svekkes av stor variasjon i målingene. Teknikker som i stor grad baserer seg på manuelle målinger, som blant annet RANO-kriteriene, er utsatt for en slik variasjon. Dette gjelder både mellom de ulike personene som utfører målingene, samt innad hos en og samme person, henholdsvis inter- og intra-rater variabilitet. Tidligere studier har vist både dårlig inter- og intra-rater variabilitet ved manuelle målinger. I en studie der åtte radiologer gjorde manuelle bidimensjonale diometermålinger av glioblastomer, fant man enighet i vurdering av progresjon (definert som vekst  $\geq 25\%$ ) i kun 10 av 22 tilfeller (45%) [22]. I de andre 12 tilfellene var det minst en radiolog som vurderte at svulsten hadde vokst når de andre ikke fant tegn til progresjon. De fant dermed en 14% risiko for falsk-positiv progresjon hos pasienter med stabil sykdom. En annen studie som undersøkte intra-rater variabilitet hos syv radiologer, viste at i 76% av tilfellene var det  $\geq 25\%$  forskjell mellom ulike målinger i minst ett plan, gjort på samme MR-bilde [23]. Omtrent 25% av tilfellene hadde  $\geq 25\%$  forskjell i to plan. Studien viste at man dermed kan måle vekst  $\geq 25\%$  (progresjon) ved en tilfeldighet, og konkluderte med at intra-observer variabilitet kan være en viktig begrensning ved RANO-kriteriene.

I en studie fra vårt fagmiljø har man undersøkt grad av reproduserbarhet ved segmentering av glioblastomer ved bruk av semi-automatiske segmenteringsmetoder og manuelle segmenteringsmetoder. Det viste seg at reproduserbarhet ved manuell tumorsegmentering er noe bedre enn ved semi-automatiske metoder. De manuelle målingene er derimot svært tidkrevende, og er ikke særlig egnet for klinisk bruk i dag [24].

## **1.5 Målsetting og hypotese**

Målsettingen med denne oppgaven er å undersøke hvor god tradisjonell radiologisk vekstvurdering av glioblastomer er sammenlignet med volumsegmentering (gullstandard). Med tradisjonell radiologisk vekstvurdering mener vi her visuell radiologisk bedømming uten målinger og bedømming ved bruk av RANO-kriteriene. Vi ønsker også å se på grad av samsvar (inter-rater variabilitet) mellom nevreradiologene.

På denne måten ønsker vi å belyse styrker og svakheter ved dagens metoder for vekstvurdering sammenholdt med volumsegmentering.

### **Nullhypotese**

*Størrelsesendring av glioblastomer er svært vanskelig å estimere nøyaktig med tradisjonelle metoder grunnet irregulær tredimensjonal utbredelse av svulstene. Det vil derfor presumptivt ofte være diskrepans mellom hva som bedømmes som vekst ved visuell bedømming og ved bruk av RANO-kriteriene, sammenlignet med volumetriske beregninger.*

## **2. MATERIALE OG METODE**

Denne hovedoppgaven bygger på et materiale fra en forskerlinjeoppgave ved det medisinske fakultet ved NTNU. Forskerlinjeoppgaven studerte veksthastighet av glioblastomer hos 106 pasienter behandlet ved St. Olavs Hospital mellom januar 2004 og mai 2014 [7, 25]. Pasientene som ble inkludert i studien var pasienter over 18 år med ubehandlet kontrastladende histologisk verifisert glioblastom. De hadde ingen tidligere sykehistorie med hjernesvulst, og gjennomgikk førstegangs hjernekirurgi. Pasientene hadde gjennomført minst to MR-undersøkelser med kontrastforsterkede T<sub>1</sub>-serier før kirurgi, hvor tidsintervallet mellom undersøkelsene var minst to uker. Den første MR-undersøkelsen ble gjort i den initiale utredningen av pasienten (“diagnostisk MR”), og den andre vanligvis få dager før kirurgi (“preoperativ MR”) [7]. MR-bildene fra disse pasientene ble brukt i det videre arbeidet i denne hovedoppgaven. Totalt var det 262 pasienter med GBM som ble vurdert for inklusjon, der 106 (41%) fylte kriteriene og ble inkludert [7].

### **2.1 Bedømming av tumorvekst**

For å vurdere ulike metoder for radiologisk vekstmåling av glioblastomer, undersøkte vi tre forskjellige vekstmålinger for de 106 glioblastom-pasientene:

1. Om det var gjort noen vurdering av tumorvekst i den radiologiske beskrivelsen av undersøkelsene gjort ved operasjonstidspunktet.
2. Ny visuell radiologisk bedømming av vekst gjort av to nevreradiologier ved å kun se på bildene uten å måle (bare ved å «scrolle» i MR-bildene), såkalt «eye-balling»
3. Ny radiologisk estimering av tumorstørrelse etter RANO-kriteriene, gjort av de samme to nevreradiologene ved å måle i MR-bildene (RANO-måling)

Punkt 1 ble utført av studentene. Forut for den nye radiologiske bedømmingen (punkt 2 og 3), ble MR-undersøkelsene klargjort for nevreradiologene. MR-bildene ble hentet fra det digitale røntgenarkivet og “hengt” i PACS (Picture Archiving and Communication System), som er radiologenes dataprogram for bildehandtering. Det ble da hentet ut fire T<sub>1</sub>-vektede, kontrastforsterkede bilder for hver pasient. To diagnostiske og to preoperative bilder, der ett av hver var i sagittalplanet og ett av hver i transversalplanet. Hos noen pasienter, særlig fra de eldste

undersøkelsene, fantes ikke fire bilder som oppfylte disse kravene. I disse tilfellene ble bare de bildene som oppfylte kravene ”hengt” opp.

Nevroradiologene fikk så et skjema for hver pasient som de skulle fylle ut (appendix 1) hvor de dokumenterte antall kontrastladende lesjoner og størrelsen av dem. Dette beskrives nærmere i avsnitt «2.1.2 Ny visuell radiologisk bedømming» og «2.1.3 RANO-måling».

De lesjonene med en diameter under 10 mm ble ekskludert fra deler av analysene, da de ikke oppfylte størrelseskravene for målbare lesjoner ifølge RANO-kriteriene (Se avsnitt «1.3.3 RANO-kriteriene»). Alle analyser som involverer volumsegmentering, RANO-kriteriene og/eller visuell radiologisk bedømming, inkluderer derfor bare svulster med begge diametere over 10 millimeter (i henhold til RANO-kriteriene).

### **2.1.1 Radiologisk bedømming ved operasjonstidspunkt**

Studentene gikk gjennom de preoperative radiologiske beskrivelsene av de 106 pasientene i PACS. Her ble det registrert om det hadde vært vurdert en endring av tumorstørrelsen fra diagnosetidspunktet («diagnostisk MR») til operasjonstidspunktet («preoperative MR»). Dette ble gjort ved å lese beskrivelsen på de preoperative undersøkelsene og eventuelle demonstrasjonsbeskrivelser forut for operasjonen. Det ble også registrert hvordan endringen var bedømt. Dette ble ført inn i et eget skjema (appendix 2) med spørsmål “Er tumorvekst vurdert?”, med mulige svaralternativer “Nei”, “Ja, reduksjon”, “Ja, uendret” og “Ja, vekst”. Dersom vekst var vurdert, var neste spørsmål “Er vekstmål oppgitt?”, med mulige svaralternativer “I antall millimeter”, “I relativ vekst” og “Annet”.

Dersom nevreradiologene hadde skrevet på eksakte mål for hvor mye vekst som var tilkommet, ble det registrert under “Antall millimeter”. Hadde veksten blitt registrert som for eksempel dobling, ble det registrert under “Relativ vekst”. Under punktet “Annet” registrerte vi de tumorene hvor det ikke var oppgitt spesifikke vekstmål, men hvor det ble beskrevet i teksten at hadde vært en endring i størrelse.

Denne informasjonen ble så brukt til å se hvor godt det samsvarte med den volumetrisk beregnede størrelsesendringen.

### **2.1.2 Ny visuell radiologisk bedømming**

To nevreradiologer gikk uavhengig av hverandre gjennom de diagnostiske og preoperativ kontrastforsterkede T<sub>1</sub>-vektende MR-bildene for de 106 pasientene, og anslo vekst i prosent basert på visuell bedømming. Veksten ble anslått i følgende intervaller: Reduksjon i tumorvolum, 0-25% vekst, 25-100% vekst, 100-200% vekst og > 200% vekst. Dette ble registrert på egne skjema (appendix 1, punkt 1). Det viste seg etter datainnsamlingen at det hadde vært mer hensiktsmessig å klassifisere kategoriene for størrelsesendring mer nøyaktig etter RANO-kategoriene, da det er RANO-kriteriene som er dagens standard. De tumorene som var bedømt til å ha reduksjon i tumorvolumet ble derfor gjennomgått en andre gang, hvor nevreradiologene anslo reduksjonen til å være 0-50 %, 50-100 % eller komplett. Dermed kunne vi rekategorisere dataene til å samsvare med RANO-kriterienes kategorier (partiell respons/«reduksjon» (-100% til -50%), stabil sykdom/«uendret» (-50% til +25%) og progresjon/«vekst» (> 25%)) når vi skulle analysere både inter-rater variabilitet og samsvaret mellom visuell radiologisk bedømming og volumsegmentering. Da vi så på inter-rater variabilitet, valgte vi imidlertid å også beholde de prosentvise kategoriene, for å tydeliggjøre variasjonen mellom nevreradiologene og vise spredningen i resultatene bedre.

Resultatene fra dette ble analysert med fokus på følgende:

1. Hvor godt samsvarer visuell radiologisk bedømming med størrelsesestimering ved volumsegmentering (gullstandard)?
2. Hvor godt er samsvaret mellom nevreradiologene (inter-rater variabilitet)?

### **2.1.3 RANO-måling**

De to nevreradiologene har uavhengig av hverandre estimert tumorstørrelsen i henhold til RANO-kriteriene (produktet av de to største perpendikulære diameterne). Der det forelå multifokalitet, ble de enkelte lesjonene summert sammen. Dette ble registrert på egne skjema (appendix 1, punkt 2), der alle aktuelle diameterer ble notert. Som beskrevet i avsnitt «2.1.2 Ny visuell radiologisk bedømming» var det mer hensiktsmessig for deler av analysene å rekategorisere dataene i samsvar med RANO-kriteriene, så det ble også gjort for disse målingene.

Resultatene fra dette ble analysert med fokus på følgende:

1. Hvor godt samsvarer størrelsesestimatet på tumor i henhold til RANO-kriteriene (RANO-måling) med volumsegmentering (gullstandard), når progresjon er definert som beregnet volumøkning over en grense for «sikker vekst»?
2. Hvor godt er samsvaret mellom nevreradiologene (inter-rater variabilitet)?

Det ble i forbindelse med den nevnte forskerlinjeoppgaven beregnet en grense for hva som ble oppfattet som signifikant vekst («sikker vekst») ved volumsegmentering. Det ble brukt enten «Limits of Agreement» eller «Reproducibility Coefficient», avhengig av datagrunnlaget [7].

For å beregne disse grensene ble det segmentert 8 radiologisk typiske glioblastomer med 2 ukers mellomrom. Først ved bruk av tynne snitt på 1mm, deretter de samme glioblastomene reformatert til tykke snitt på 5mm. Ved segmentering med tynne snitt ble det regnet ut en reproduksjonskoeffisient på 1,72 milliliter, der endringer over denne verdien ble vurdert til å være signifikante. Det ble også sammenlignet segmentering av tynne og tykke snitt, der det ble regnet ut «Limits of Agreement», som var  $1,46\text{mL} \pm 2,42$ . Dermed ble en forskjell på  $-0,97$  milliliter eller  $+3,99$  milliliter sett på som signifikante. Begge disse målene, «Reproducibility Coefficient» og «Limits of agreement» ble da altså brukt for å vurdere om størrelsesendringen var tilstrekkelig for å kalle det «sikker vekst», der all progresjon over dette blir sett på som progresjon av tumor [7].

## **2.2 Volumsegmentering**

De volumetriske beregningene som brukes i denne hovedoppgaven ble utført i forbindelse med den tidligere nevnte forskerlinjeoppgaven. Tumorvolum ble da definert til å inkludere både kontrastoppladende del av tumor, samt ikke-kontrastoppladende del omringet av kontrast (sentral nekrose). I tilfeller hvor det var flere kontrastladende lesjoner, ble alle lesjoner inkludert i totalvolumet, men segmentert separat [7].

Det er dette som er vår gullstandard, som vi vil sammenligne metodene i denne oppgaven med.

## **2.3 Statistiske analyser**

Statistiske analyser ble utført med IBM SPSS 23.0, Statistical Package for the Social Science.

### **2.3.1 Frekvenstabeller**

Det ble brukt frekvenstabeller til å fremstille i hvor mange tilfeller det var blitt vurdert endret størrelse av de 106 glioblastomene, fra diagnostisk til preoperativ MR. De tumorene det var vurdert vekst hos, ble så kategorisert i en ny frekvenstabell ut fra hvordan denne veksten var vurdert.

### **2.3.2 Krysstabeller**

Det ble brukt krysstabeller for å kunne vurdere de ulike metodene for vekstmåling opp mot hverandre. Alle svarene fra de ulike vekstvurderingsskjemaene ble sortert i kategorier for å kunne sammenlignes. Dette inkluderte svarene fra den radiologiske beskrivelsen av vekst gjort ved operasjonstidspunktet, svarene fra ny visuell radiologisk bedømming og RANO-målingen, samt svarene fra volumsegmenteringen.

Krysstabellene viste grad av enighet mellom ulike metoder og mellom de to nevroradiologene (inter-rater variabilitet). De viste også grad av enighet fordelt i de ulike kategoriene innad i hver metode. Resultatene fra krysstabellene ble videre brukt til å regne ut samsvar i prosent.

### **2.3.3 Kappa**

Da vi så på inter-rater variabilitet, beregnet vi Kappa-verdier for å uttrykke graden av samsvar. Disse beregningene gjorde vi i SPSS. Kappaverdien vil variere fra 0-1, med følgende grenser for grader av enighet [26]:

<b>Kappaverdi</b>	<b>Grad av enighet</b>
<b>0–0.20</b>	Ingen
<b>0.21–0.39</b>	Minimal
<b>0.40–0.59</b>	Svak
<b>0.60–0.79</b>	Moderat
<b>0.80–0.90</b>	Sterk
<b>&gt;0.90</b>	Tilnærmet perfekt

### **2.3.4 Sensitivitet, spesifisitet, PPV og NPV**

Da vi sammenlignet vekstvurdering ved bruk av visuell radiologisk bedømming med volumsegmentering, og RANO-måling med volumsegmentering, regnet vi ut sensitivitet, spesifisitet, positiv prediktiv verdi (PPV) og negativ prediktiv verdi (NPV). Dette gjorde vi fordi vi ønsket å se på hvor god metodene var til å skille mellom “progresjon” og “ikke-progresjon” og brukte da volumsegmentering som gullstandard. For å undersøke dette slo vi sammen kategoriene “reduksjon” og “stabil sykdom” til en kategori, slik at vi kunne sette opp en 2x2-tabell. De svulstene nevreradiologene vurderte at det hadde vært en reduksjon eller stabil sykdom på, og volumsegmentering også vurderte det til å være reduksjon eller stabil sykdom, utgjør da de “sanne negative”. Der nevreradiologene anga reduksjon eller stabil sykdom, men volumsegmentering viste progresjon, utgjør de “falske negative”. De svulstene der volumsegmentering viste reduksjon eller stabil sykdom, og nevreradiologene vurderte det til å være progresjon, utgjør da de “falske positive”. De svulstene der både nevreradiologene og volumsegmenteringen vurderte det til å vært en progresjon mellom de to bildeundersøkelsene, utgjør da de “sanne positive”. Vi kunne ut fra dette regne ut sensitivitet, spesifisitet, PPV og NPV, og dermed se hvor gode metodene var på å skille mellom “progresjon” og “ikke-progresjon”.



## 3. RESULTATER

### 3.1 Radiologisk bedømming av vekst ved operasjonstidspunkt

I den radiologiske beskrivelsen var det for 92 av 106 glioblastomer (87 %) angitt en vurdering av endringen i tumorstørrelse mellom diagnostisk og preoperative MR-undersøkelse. Hos disse 92 pasientene var det angitt at svulsten hadde redusert i størrelse hos en pasient (1 %), mens svulsten var uendret hos 38 (41%) og hadde økt i størrelse hos 53 (58 %) (tabell 1). Hos de samme 92 pasientene var det, ved bruk av volumsegmentering, 6 tumorer som hadde reduksjon i størrelse (7 %), 19 var uendret (21 %) og 67 hadde økt (73 %).

<b>Tabell 1: Radiologisk bedømming av vekst ved operasjonstidspunkt sammenholdt med volumsegmentering (n = 106)</b>		
<b>Vekst vurdert ved operasjonstidspunktet? [Volumetrisk bedømt]</b>		
Nei	14 (13 %)	[R: 1, U: 9, V: 4]
Ja	92 (87 %)	
Reduksjon	1 (1 %)	[R: 0, U: 1, V: 0]
Uendret	38 (41 %)	[R: 3, U:14, V:21]
Vekst	53 (58 %)	[R: 3, U: 4, V:46]

Tabell 1: Kolonnen til venstre viser antall svulster der størrelsesendring er vurdert, mens kolonnen til høyre viser hva volumsegmentering anslo den til å være. R: Reduksjon. U: Uendret. V:Vekst.

I 58 av de 92 (63%) glioblastomene der det var vurdert vekst, var det ikke dokumentert kvantitative mål for vekstendring, men beskrivelser som «det har foreligget en vekst av tumor» eller at «tumorstørrelsen er uendret». Der vekstmål var blitt oppgitt, var det oppgitt en, to eller tre diametere i henholdsvis 21, 8 og 5 pasienter (tabell 2).

<b>Tabell 2: Ble vekstmål angitt? (n=92)</b>	
Nei	58 (63%)
Ja, en diameter	21 (23%)
Ja, to diametere	8 (9%)
Ja, tre diametere	5 (5%)
<b>Total</b>	<b>92 (100%)</b>

Tabell 2: Bruk av vekstmål i de radiologiske beskrivelsene ved operasjonstidspunktet

## **3.2 Inter-rater variabilitet**

Vi har i denne oppgaven valgt å bruke to ulike kategoriseringer i presentasjonene av resultatene for inter-rater variabilitet. Dette fordi vi ønsker å fremheve variasjonene i bedømming mellom nevreradiologene når vi bruker de opprinnelige kategoriene for størrelsesendring, samtidig som vi ønsker å sammenligne resultatene med RANO-kategoriseringen. Ved rekategorisering slik at de passer med RANO-kategoriene, har vi i tillegg slått sammen responskategoriene partiell- og komplett respons til en samlet kategori som vi kalte “reduksjon”, da ingen svulster havnet i kategorien komplett respons.

11 svulster oppfylte ikke størrelseskravene for målbare lesjoner i følge RANO-kriteriene og ble dermed fjernet. Videre analyser som involverer visuell radiologisk bedømming, RANO-målinger og/eller volumsegmentering inkluderer derfor bare 95 pasienter.

### 3.2.1: Inter-rater variabilitet ved Visuell radiologisk bedømming

Ved bruk av den originale kategoriseringen, som ble brukt under datainnsamlingen, var den totale enigheten mellom de to nevrordiologene på 59% (56 av 95 svulster), svarende til en Kappa-verdi på 0,38 (tabell 3). Det var uenighet i alle kategoriene.

Tabell 3: Inter-Rater Variabilitet - Visuell radiologisk bedømming							
		Nevroradiolog 2					Total
		-50-0%	0-25%	25-100%	100-200%	>200%	
Nevro-radiolog 1	-50-0%	0	0	0	0	0	0
	0-25%	10	35	7	0	0	52
	25-100%	0	6	16	4	1	27
	100-200%	0	0	2	1	2	5
	>200%	0	0	3	4	4	11
Total		10	41	28	9	7	95
Total enighet							56/95 (59%)
Kappa							0,38

Tabell 3: Inter-rater variabilitet ved bruk av visuell vurdering av tumorstørrelse. De estimatene de er enig om er markert.

Da vi rekategoriserte dataene etter RANO sine responskategorier fikk vi en prosentvis enighet mellom nevrordiologene på 86% (82 av 95 svulster), svarende til en Kappa-verdi på 0,72 (tabell 4).

Tabell 4: Inter-Rater Variabilitet - Visuell radiologisk bedømming					
		Nevroradiolog 2			Total
		Reduksjon	Uendret	Vekst	
Nevro-radiolog 1	Reduksjon	0	0	0	0
	Uendret	0	45	7	52
	Vekst	0	6	37	43
Total		0	51	44	95
Total enighet					82/95 (86%)
Kappa					0,72

Tabell 4: Inter-rater variabilitet ved bruk av visuell vurdering av tumorstørrelse. Resultatene er delt inn i responskategorier etter RANO-kriteriene. De estimatene de er enig om er markert.

### 3.2.2 Inter-rater variabilitet ved RANO-måling

Ved bruk av den originale kategoriseringen, som ble brukt ved datainnsamlingen, var den totale enigheten mellom nevrordiologene på 72% (68 av 95 svulster), svarende til en kappa-verdi på 0,60 (tabell 5).

Tabell 5: Inter-Rater Variabilitet – RANO-måling							
		Nevroradiolog 2					Total
		-50-0%	0-25%	25-100%	100-200%	>200%	
Neuro-radiolog 1	-50-0%	6	2	0	0	0	8
	0-25%	4	27	6	1	0	38
	25-100%	1	6	21	2	0	30
	100-200%	0	1	3	5	0	9
	>200%	0	0	1	0	9	10
Total		11	36	31	8	9	95
Total enighet							68/95 (72%)
Kappa							0,60

Tabell 5: Inter-rater variabilitet ved RANO-måling. De estimatene de er enig om er markert

Da vi rekategoriserte dataene etter RANO sine responskategorier fikk vi en prosentvis enighet mellom nevrordiologene på 84% (80 av 95 svulster), svarende til en Kappa-verdi på 0,68 (tabell 6).

Tabell 6: Inter-Rater Variabilitet – RANO-måling					
		Nevroradiolog 2			Total
		Reduksjon	Uendret	Vekst	
Neuro-radiolog 1	Reduksjon	0	0	0	0
	Uendret	0	39	7	46
	Vekst	0	8	41	49
Total		0	47	48	95
Total enighet					80/95 (84%)
Kappa					0,68

Tabell 6. Inter-rater variabilitet ved RANO-måling. Resultatene er delt inn i responskategorier etter RANO-kriteriene. De estimatene de er enig om er markert

### 3.3 Visuell radiologisk bedømming vs. volumsegmentering

#### **Nevroradiolog 1**

Nevroradiolog 1 samsvarer med volumsegmentering i 59 av 95 svulster (62%) (tabell 7). Dersom man betrakter visuell radiologisk bedømming som en diagnostisk test på tumorvekst eller ikke, gir dette en sensitivitet på 0,60, spesifisitet på 0,89, samt positiv prediktiv verdi på 0,93 og negativ prediktiv verdi på 0,48 (tabell 8).

#### **Nevroradiolog 2**

Nevroradiolog 2 samsvarer med volumsegmentering i 60 av 95 svulster (63%) (tabell 7). Dersom man betrakter visuell radiologisk bedømming som en diagnostisk test på tumorvekst eller ikke, gir dette en sensitivitet er 0,61, spesifisitet er 0,89, samt positiv prediktiv verdi på 0,93 og negativ prediktiv verdi på 0,49 (tabell 9).

<b>Tabell 7: Visuell radiologisk bedømming vs. volumsegmentering</b>					
		<b>Volumsegmentering</b>			Total
		Reduksjon	Uendret	Vekst	
<b>Nevro-radiolog 1</b>	Reduksjon	0	0	0	0
	Uendret	6	19	27	52
	Vekst	1	2	40	43
Total		7	21	67	95
Total enighet					59/95
Prosent enighet					62%
<b>Nevro-radiolog 2</b>	Reduksjon	0	0	0	0
	Uendret	6	19	26	51
	Vekst	1	2	41	44
Total		7	21	67	95
Total enighet					60/95
Prosent enighet					63%

Tabell 7: Grad av samsvar mellom den visuelle radiologiske bedømmingen til nevreradiolog 1 og 2 og volumsegmentering (gullstandard)

Tabell 8: Visuell radiologisk bedømming nevreradiolog 1 vs. volumsegmentering			
		Volumsegmentering	
		Vekst	Ikke vekst
Visuell radiologisk bedømming Nevreradiolog 1	Vekst	Sanne positive a            40	Falske positive b            1+2
	Ikke vekst	Falske negative c            27+0	Sanne negative d            19+6+0+0

Tabell 8: 2x2 tabell for visuell radiologisk bedømming gjort av nevreradiolog 1, mot volumsegmentering. Tallene i tabellen er så brukt for å regne ut sensitivitet, spesifisitet, PPV og NPV.

Sensitivitet:  $a/(a+c) = 40/(40+27+0) = 0,60$

Spesifisitet:  $d/(b+d) = (19+6+0+0)/(1+2+19+6+0+0) = 0,89$

PPV:  $a/(a+b) = 40/(40+1+2) = 0,93$

NPV:  $d/(c+d) = (19+6+0+0)/(27+0+19+6+0+0) = 0,48$

Tabell 9: Visuell radiologisk bedømming nevreradiolog 2 vs. volumsegmentering			
		Volumsegmentering	
		Vekst	Ikke vekst
Visuell radiologisk bedømming Nevreradiolog 2	Vekst	Sanne positive a            41	Falske positive b            1+2
	Ikke vekst	Falske negative c            26+0	Sanne negative d            19+6+0+0

Tabell 9: 2x2 tabell for visuell radiologisk bedømming gjort av nevreradiolog 2, mot volumsegmentering. Tallene i tabellen er så brukt for å regne ut sensitivitet, spesifisitet, PPV og NPV.

Sensitivitet:  $a/(a+c) = 41/(41+26+0) = 0,61$

Spesifisitet:  $d/(b+d) = (19+6+0+0)/(1+2+19+6+0+0) = 0,89$

PPV:  $a/(a+b) = 41/(41+1+2) = 0,93$

NPV:  $d/(c+d) = (19+6+0+0)/(26+0+19+6+0+0) = 0,49$

### 3.4 RANO-måling vs. volumsegmentering

#### **Nevroradiolog 1:**

Nevroradiolog 1 samsvarer med volumsegmentering i 60 av 95 svulster (63%) (tabell 10).

Dersom man betrakter visuell radiologisk bedømming som en diagnostisk test på tumorvekst eller ikke, gir dette en sensitivitet på 0,66, spesifisitet på 0,82, samt positiv prediktiv verdi på 0,90 og negativ prediktiv verdi på 0,50 (tabell 11).

#### **Nevroradiolog 2:**

Nevroradiolog 2 samsvarer med volumsegmentering i 58 av 95 svulster (61%) (tabell 10).

Dersom man betrakter visuell radiologisk bedømming som en diagnostisk test på tumorvekst eller ikke, gir dette en sensitivitet på 0,63, spesifisitet på 0,79, samt positiv prediktiv verdi på 0,88 og negativ prediktiv verdi på 0,47 (tabell 12).

<b>Tabell 10: RANO-måling vs- volumsegmentering</b>					
		<b>Volumsegmentering</b>			<b>Total</b>
		<b>Reduksjon</b>	<b>Uendret</b>	<b>Vekst</b>	
<b>Nevroradiolog 1</b>	Reduksjon	0	0	0	0
	Uendret	7	16	23	46
	Vekst	0	5	44	49
<b>Total</b>		7	21	67	95
Total enighet		60/95			
Prosent enighet		63%			
<b>Nevroradiolog 2</b>	Reduksjon	0	0	0	0
	Uendret	6	16	25	47
	Vekst	1	5	42	48
<b>Total</b>		7	21	67	95
Total enighet		58/95			
Prosent enighet		61%			

Tabell 10: Grad av samsvar mellom den radiologiske bedømmingen til nevreradiolog 1 og 2 ved RANO-måling og volumsegmentering

Tabell 11: RANO-måling nevreradiolog 1 vs- volumsegmentering			
		Volumsegmentering	
		Vekst	Ikke vekst
RANO Nevreradiolog 1	Vekst	Sanne positive a            44	Falske positive b            5+0
	Ikke vekst	Falske negative c            23+0	Sanne negative d            16+7+0+0

Tabell 11: 2x2 tabell for RANO-måling gjort av nevreradiolog 1, mot volumsegmentering. Tallene i tabellen er så brukt for å regne ut sensitivitet, spesifisitet, PPV og NPV for vekstvurdering gjort ved RANO-måling.

Sensitivitet:  $a/(a+c) = 44/(44+23+0) = 0,66$

Spesifisitet:  $d/(b+d) = (16+7+0+0)/(5+0+16+7+0+0) = 0,82$

PPV:  $a/(a+b) = 44/(44+5+0) = 0,90$

NPV:  $d/(c+d) = (16+7+0+0)/(23+0+16+7+0+0) = 0,50$

Tabell 12: RANO-måling nevreradiolog 2 vs- volumsegmentering			
		Volumsegmentering	
		Vekst	Ikke vekst
RANO Nevreradiolog 2	Vekst	Sanne positive a            42	Falske positive b            1+5
	Ikke vekst	Falske negative c            25+0	Sanne negative d            16+6+0+0

Tabell 12: 2x2 tabell for RANO-måling gjort av nevreradiolog 2, mot volumsegmentering. Tallene i tabellen er så brukt for å regne ut sensitivitet, spesifisitet, PPV og NPV for vekstvurdering gjort ved RANO-måling.

Sensitivitet:  $a/(a+c) = 42/(42+25+0) = 0,63$

Spesifisitet:  $d/(b+d) = (16+6+0+0)/(1+5+16+6+0+0) = 0,79$

PPV:  $a/(a+b) = 42/(42+1+5) = 0,88$

NPV:  $d/(c+d) = (16+6+0+0)/(25+0+16+6+0+0) = 0,47$



## **4. DISKUSJON**

Målsettingen med denne oppgaven var å sammenligne ulike radiologiske metoder for vekstvurdering av glioblastomer med volumsegmentering, som regnes som gullstandard. Den radiologiske vekstvurderingen bestod både av visuell radiologisk bedømming ved å bla gjennom bildeseriene, samt RANO-måling (bidimensjonale målinger i henhold til RANO-kriteriene). Vi ønsket også å se på graden av samsvar (inter-rater variabilitet) mellom de to nevreradiologene.

Vår hypotese var at veksten av glioblastomer er vanskelig å estimere nøyaktig grunnet irregulær tredimensjonal utbredelse av svulstene, og at det derfor ville være stor diskrepans mellom den radiologiske vekstvurderingen og volumsegmenteringen. Våre resultater bekrefter i stor grad dette, da nesten 40 % av pasientene ble klassifisert forskjellig av de to nevreradiologene ved både visuell radiologisk bedømming og ved RANO-måling, sammenlignet med volumsegmentering. Det er interessant at visuell bedømming ikke er særlig dårligere enn kriteriebasert metodikk med RANO-måling. Klinikere legger vekt på radiologisk vekst i kliniske beslutninger, noe som gir støtte til å uttrykke seg med forsiktighet ved antatt mindre vekstendringer eller å ta i bruk segmenteringsmetoder i klinisk praksis. Den dårlige reliabiliteten er også noe man bør ha i mente ved bruk av RANO i kliniske studier hos disse pasientene. I dette materialet så man bare på ubehandlede glioblastomer og det er grunn til å frykte at validitet og reliabilitet ved vekstestimerer er dårligere hos behandlede pasienter. Disse restsvulstene er ofte mer irregulære i fasong, der det er kontrastlekkasje etter kirurgisk betingede sirkulasjonsforandringer eller pseudoprogresjon kan forekomme.

### **4.1 Klinisk praksis i dag**

Resultatet fra gjennomgangen av de radiologiske beskrivelsene viser at i vår glioblastompopulasjon ble det gjort en vekstvurdering mellom diagnosetidspunktet og operasjonstidspunktet i 87% av pasientene (tabell 1). Der hvor det ble gjort en vekstvurdering, ble det i 63% av tilfellene ikke oppgitt vekstmål. Formuleringer som ofte ble brukt var ”betydelig vekst av tumor”, ”tumorstørrelse er uendret” og lignende. Der det ble oppgitt vekstmål, ble det oppgitt i en, to eller tre diametere (tabell 2). Bare i 5 tilfeller ble tre diametere oppgitt, noe som betyr at det i svært få svulster er benyttet tredimensjonale mål for å bedømme tumorveksten. Videre ser vi også at ved vurderingene utført i klinisk praksis ble 53 av 92

glioblastomer (58%) vurdert til å ha vekst, men hvis vi sammenligner med volumsegmentering ser vi at 67 av 92 tumorer (73%) har vokst. De er enige i 46 av disse. 21 glioblastomer som ifølge gullstanderen har vokst, er beskrevet som uendret i klinikken. Dette betyr altså at det som gjøres i klinisk praksis i dag er mer unøyaktig sammenlignet med gullstandard. Disse funnene tyder på at kan være behov for implementering av mer konkrete retningslinjer for vekstvurdering av glioblastomer i klinisk praksis, dersom veksten skal vurderes. Vekstvurdering har en særlig stor klinisk betydning i postoperativ fase for å tilrettelegge videre behandling.

## **4.2 Inter-rater variabilitet**

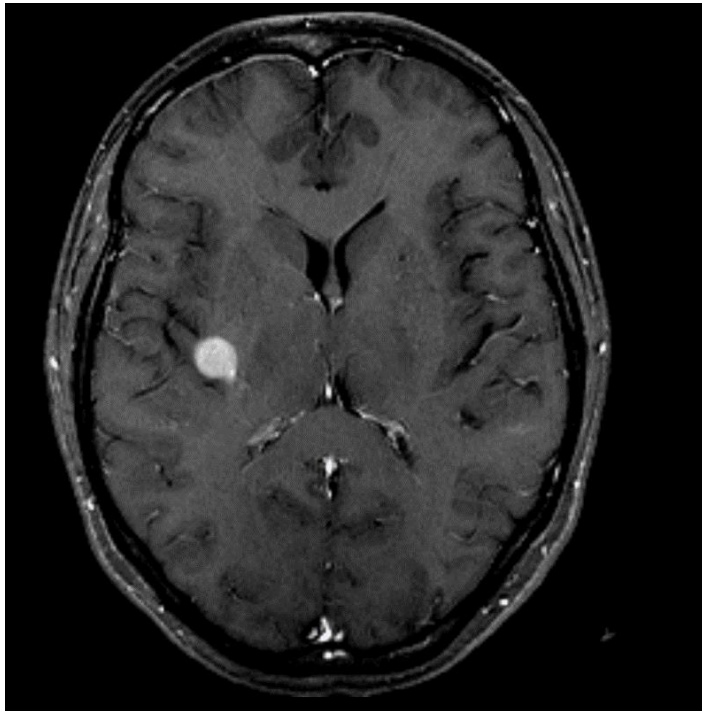
Vi har sammenlignet graden av samsvar mellom de to nevreradiologene (inter-rater variabilitet). Som beskrevet i metoddelen har vi brukt to ulike kategoriseringer for å fremstille dette. Først delte vi inn i prosentvise kategorier like de som er brukt da nevreradiologene foretok den visuelle radiologiske bedømmingen (tabell 3 og 5), deretter i responskategorier etter RANO-kriteriene (partiell respons/«reduksjon», stabil sykdom/«uendret» og progresjon/«vekst») (tabell 4 og 6).

### **4.2.1 Visuell radiologisk bedømming**

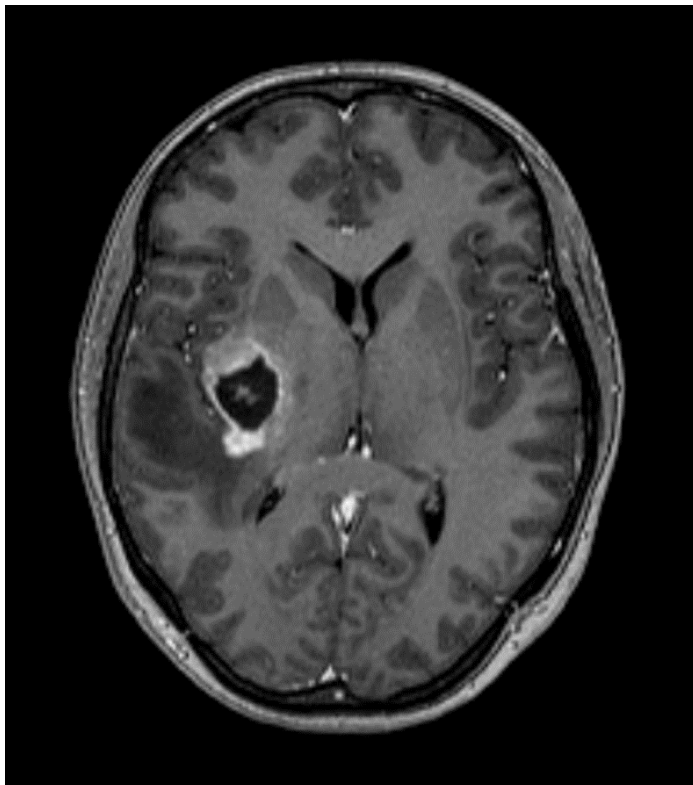
Vi forventet at vi ville finne betydelig variasjon mellom nevreradiologenes vekstmålinger, og at resultatene trolig ville vise tilsynelatende økt grad av enighet ved kategorisering etter RANO-kriterienes responskategorier.

Ved bruk av prosentvise kategorier er graden av enighet mellom nevreradiologene 59% (56 av 95 svulster), med en Kappa-verdi på 0,38 (tabell 3). Dette tilsvarer «minimal» grad av enighet. Vi ser stor grad av uenighet i alle vekstkategoriene. Vi ser blant annet at nevreradiolog 2 anslo reduksjon mellom -50% og 0% i 10 svulster, mens nevreradiolog 1 ikke anslo reduksjon i noen svulster. Nevreradiolog 1 anslo vekst mellom 25% og 100% i 27 tumorer, men nevreradiolog 2 gjorde det i 28 svulster, men bare 16 av tumorene var de samme. Figur 5 og 6 viser et eksempel der nevreradiologene bedømmer veksten ulikt, og den radiologiske MR-beskrivelsen ikke samsvarer med noen av nevreradiologene. Da vi strukturerte data etter RANO-kriterienes responskategorier (tabell 4), økte graden av enighet mellom nevreradiologene fra 59% til 86%

(82 av 95 svulster). Kappaverdien økte da fra 0,38 til 0,72, som tilsvarer at graden av enighet går fra “minimal” til “moderat”. Ved å bruke RANO-kriterienes grenseverdier for hva som anses som progresjon, ser det tilsynelatende ut som nevreradiologene er mer enig, men dette skyldes at RANO bruker kategorier med videre intervaller enn hva vi brukte da vi samlet inn dataene.



*Figur 5: T<sub>1</sub> vektet kontrastforsterket MR-bilde (transversalplanet). Diagnostisk bilde.*



*Figur 6: T<sub>1</sub> vektet kontrastforsterket MR-bilde (transversalplanet). Preoperativt bilde. Nevroradiolog 1 anslo vekst på 25%-100%. Nevroradiolog 2 anslo vekst på >200%. MR-beskrivelsen anslo uendret vekst. Volumsegmentering viste en vekst på >200% (relativ vekst på 661%)*

### **4.2.2 RANO-måling**

Vi har også sammenlignet graden av samsvar mellom de to nevreradiologene ved vekstvurdering gjort ved RANO-måling. Vi forventet også her at vi ville finne variasjon mellom nevreradiologenes vekstmålinger, men med noe høyere grad av enighet sammenlignet med visuell radiologisk bedømming, siden man her faktisk gjør kvantitative målinger på svulstene. Ved kategorisering etter RANO-kriterienes responskategorier forventet vi også her at resultatene ville vise tilsynelatende økt grad av enighet mellom nevreradiologene.

Ved bruk av prosentvise kategorier var grad av enighet mellom nevreradiologene 72%, med en Kappa-verdi på 0,60 (tabell 5). Uenigheten ble spredt utover alle vekstkategorier. Vi ser blant annet at nevreradiolog 1 og 2 anslo reduksjon mellom -50% og 0% i henholdsvis 8 og 11 tumorer, men at bare 6 av disse var de samme. Nevreradiolog 1 anslo vekst mellom 25% og 100% i 30 svulster, nevreradiolog 2 i 31 svulster, og de var enige i 21 av disse. Ved bruk av RANO-kriterienes responskategorier ser vi at enigheten øker til 84% (80 av 95 svulster), som tilsvarer en Kappa-verdi på 0,68 (tabell 6). Kappa-verdien endres altså fra 0,60 til 0,68, noe som begge tilsvarer ”moderat” grad av enighet. Vi ser igjen en økning i enighet ved bruk av RANO-kriteriene sin inndeling etter responskategorier. Dette kan oppfattes som en svakhet ved RANO-kriteriene, da de deler inn i responskategorier som er relativt vide (for eksempel går stabil sykdom fra 50% reduksjon av tumorvolumet til +25% økning av tumorvolumet).

### **4.2.3 Visuell radiologisk bedømming eller RANO-måling?**

Inter-rater variabiliteten ved visuell radiologisk bedømming og ved RANO-måling varierer med hvilke intervaller man setter for veksten. Dersom man fokuserer på nøyaktigheten av vekstvurderingene kan man argumentere for at man bør bruke smalere intervaller som vi har gjort ved datainnsamlingen (tabell 3 og 5), da dette vil gi et mer detaljert bilde på samsvaret i målingene mellom nevreradiologene og at man ser tydeligere spredningen i resultatene. Dersom man derimot ønsker å fokusere på hvordan variabiliteten er ved bruk av de retningslinjene som finnes for bruk i kliniske studier i dag, bør man bruke intervallene etter RANO-kriteriene (tabell 4 og 6). Ved å bruke våre kategorier fra datainnsamlingen økte enigheten fra en Kappa-verdi på 0,38 ved visuell bedømming til 0,60 ved to-dimensjonale RANO-målinger (tabell 3 og 5). Imidlertid ser vi at man oppnår nesten identiske enighet når man bruker RANO sine intervaller for responskategorier og sammenligner visuell radiologisk

bedømming med todimensjonelle målinger etter RANO-kriteriene, med en Kappa-verdi på henholdsvis 0,72 og 0,68 (tabell 4 og 6).

Uavhengig av kategoriseringsmetode er det relativt mange pasienter nevroradiologene ikke er enige om, dette til tross for bruk av RANO-måling hvor man kunne forventet høyere grad av enighet (tabell 5 og 6). Våre resultater samsvarer med tidligere studier, og viser at todimensjonale mål gir høy grad av variabilitet, noe som svekker reproduserbarheten [22, 23].

### **4.3 Visuell radiologisk bedømming vs. volumsegmentering**

Vi har sammenlignet vekstvurdering gjort ved visuell radiologisk bedømming med gullstandard volumsegmentering. Som forventet fant vi moderat grad av samsvar. Nevroradiolog 1 og 2 har henholdsvis 62% og 63% samsvar med volumsegmentering ved hjelp av visuell radiologisk bedømming (tabell 7). Selv om nevroradiologene oppnår omtrent lik prosentvis grad av enighet, omhandler dette ikke de samme svulstene, som vi kan se i tabell 4 med Kappa-verdi 0,72. Vi har også sett nærmere på metodens evne til å anslå progresjon. Ved visuell radiologisk bedømming anslo nevroradiolog 1 og 2 progresjon i henholdsvis 43 og 44 svulster (tabell 7). Volumsegmenteringen angav vekst i 67 tilfeller, nevroradiolog 1 og 2 hadde angitt i henholdsvis 40 og 41 av disse. For nevroradiolog 1 gir dette en sensitivitet på 0,60 og en spesifisitet på 0,89. For nevroradiolog 2 gir dette en sensitivitet på 0,61 og en spesifisitet på 0,89. Dette fører til at nevroradiolog 1 og nevroradiolog 2 vil få henholdsvis 27 og 26 falske negative vekstvurderinger (henholdsvis 39% og 40%). En stor andel svulster vil altså ikke vurderes til å ha vekst selv om gullstandard angir det. Det er betydelig bedre spesifisitet enn sensitivitet. Begge nevroradiologene vil få 3 falske positive (11%), og dermed angi “vekst”, selv om gullstandard angir “ikke vekst”.

PPV for både nevroradiolog 1 og 2 er 0,93. Det vil si at en svulst som har blitt angitt som “vekst”, har hatt en reell vekst i henhold til gullstandard i 93% av tilfellene. NPV for nevroradiolog 1 og 2 er henholdsvis 0,48 og 0,49. Dette fører til at det er 48% og 49% sannsynlighet for at en svulst som er angitt som “ikke-vekst”, i realiteten ikke har vokst bedømt med gullstandard.

## **4.4 RANO-måling vs. volumsegmentering**

Vi har sammenlignet vekstvurdering gjort ved RANO-måling med gullstandarden volumsegmentering. Nevroradiolog 1 og 2 har henholdsvis 63% og 61% samsvar med volumsegmentering ved RANO-måling (tabell 10).

Som vi også så under visuell radiologisk bedømming, har nevreradiologene oppnådd svært lik grad av enighet ved RANO-måling sammenlignet med volumetrisk segmentering. Men heller ikke her omhandler enigheten de samme svulstene, som vi ser i tabell 6 der Kappa-verdien er 0,68.

Vi så også her på metodens evne til å anslå progresjon. Nevroradiolog 1 og 2 anslo progresjon ved RANO-måling i henholdsvis 49 og 48 pasienter (tabell 10). Volumsegmentering angav vekst i 67 tilfeller, nevreradiolog 1 og 2 var enige i henholdsvis 44 og 42 av disse. Dette gir en sensitivitet på 0,66 og en spesifisitet på 0,82 for nevreradiolog 1, og en sensitivitet på 0,63 og en spesifisiteten på 0,79 for nevreradiolog 2. Dette fører til at mange svulster ikke havner i riktig responskategori, sammenlignet med gullstandarden volumsegmentering. Vi ser for eksempel at ved å bruke grensen for sikker vekst ved volumsegmentering vil flere svulster bli oppfattet til å ha progresjon enn om man hadde brukt RANO-kriteriene der progresjonsgrensen er satt til 25% vekst. I vårt materiale var det 23 og 25 pasienter som ble oppfattet som “ikke vekst” ved bruk av RANO-kriteriene (henholdsvis nevreradiolog 1 og 2), men som ble kategorisert som “vekst” ved volumsegmentering. I tillegg vil 5 og 6 pasienter bli oppfattet som “vekst” ved bruk av RANO-kriteriene (henholdsvis nevreradiolog 1 og 2), men blir kategorisert som “ikke vekst” dersom man bruker volumsegmentering.

PPV for nevreradiolog 1 og 2 er henholdsvis 0,90 og 0,88. Dette tilsvarer en sannsynlighet på 90% og 88% for at en svulst som har blitt angitt som “vekst” har hatt en reell vekst. NPV for nevreradiolog 1 og 2 er henholdsvis 0,50 og 0,47. Dette gir en sannsynlighet på 50% og 47% for at en svulst som er angitt som “ikke-vekst”, i realiteten ikke har vokst.

## **4.5 Visuell radiologisk bedømming, RANO-måling eller Volumsegmentering?**

Når vi sammenligner visuell radiologisk bedømming og RANO-måling med volumsegmentering, ser vi at sensitiviteten, spesifisiteten, PPV og NPV blir omtrent den samme for begge nevreradiologene ved bruk av begge metodene. Sensitiviteten blir ca. 0,60, spesifisiteten ca. 0,85, PPV ca. 0,90 og NPV ca. 0,50. Dette betyr at ingen av metodene vil klare å identifisere vekst i mer enn 60 % av de svulstene som faktisk har hatt en reell økning i størrelse. Dersom en svulst først blir anslått til å ha progresjon, er det derimot 90 % sannsynlighet for at dette er reelt, ettersom PPV er 0,90. Med en spesifisitet på 0,85 betyr det at metodene vil identifisere 85 % av de svulstene som ikke har hatt en reel størrelsesøkning. Positiv prediktiv verdi er imidlertid avhengig av prevalens av vekst og i dette materialet var den høy. Ettersom sensitiviteten er relativt lav, blir også NPV lav, og det er bare 50 % sannsynlighet for at en svulst faktisk ikke har vokst dersom den har blitt angitt som «ikke-vekst». En vekstvurdering som sier «ikke-vekst» er derfor lite pålitelig, mens en vekstvurdering om angir «vekst» er derimot betydelig mer pålitelig.

Vi forventet en høyere grad av treffsikkerhet ved bruk av RANO-måling, som bruker manuelle kvantitative målinger, enn ved visuell bedømming. At dette ikke forekommer i våre resultater, kan skyldes inndelingen i vide responskategorier, som kan skjule spredning i resultatene i tabell 7 og 10. Dette så vi tydeligere under avsnittet «3.2 Inter-rater variabilitet» der tabellene med inter-rater variabilitet kategorisert etter RANO-kategoriene (tabell 4 og 6) gir et feilaktig inntrykk av at metodene er nesten like gode (nesten lik Kappa-verdi), mens tabellene hvor inter-rater variabilitet er kategorisert i flere grupper etter prosentvis endring (tabell 3 og 5) viser en større spredning i resultatene (og større forskjell i Kappa-verdi). Man kan derfor spekulere i at dersom man hadde brukt prosentvise kategorier i tabell 7 og 10, ville man sett en større grad av samsvar mellom RANO-måling og volumsegmentering, enn mellom visuell radiologisk bedømming og volumsegmentering. At visuell radiologisk bedømming og RANO-måling skulle være like gode til å identifisere tumorvekst var overraskende, og kan brukes til å argumentere for at man ikke trenger å vurdere endring i tumorstørrelse ved å måle, men heller gjøre en visuell bedømming, så lenge man skal kategorisere veksten i RANO sine vekstkategorier.



At man ikke oppnår høyere sensitivitet enn ca. 60 % ved RANO-måling sammenlignet med gullstandard volumsegmentering, betyr at nesten 40 % av tumorene som har vokst ikke blir identifisert ved RANO-måling. Dette tyder på at det finnes store svakheter ved denne typen vekstmåling, da todimensjonale mål av en tredimensjonal, diffust infiltrerende struktur gir upresise størrelsesestimater. En forbedret vekstvurdering vil kunne oppnås ved å innføre volumetrisk segmentering i både forskning og klinisk praksis. Det er innen forskning svært viktig med nøyaktige størrelsesestimater og mulighet for sammenligning av pasientpopulasjoner på tvers av sentra og studier. Et sensitivt og robust vekstmål vil dessuten øke statistisk styrke i studier som vurderer behandlingsrespons, noe som fører til at man ikke trenger å inkludere like mange pasienter. Progresjonsfri overlevelse er et mye brukt endepunkt i gliomstudier i dag og er i all hovedsak basert på RANO-kriteriene. Det vil derfor være fordelaktig å ta i bruk volumsegmentering i forskningsøyemed, særlig om teknologien utvikles videre og blir mindre ressurskrevende. Dersom volumsegmentering blir mindre ressurskrevende, vil det også kunne implementeres i klinisk praksis, spesielt i tilfeller der det har behandlingsmessig konsekvens. Ved å ta i bruk volumsegmentering kunne man kanskje ha bedret diagnostikken og optimalisert behandling som følge av økt nøyaktighet. Frem til volumsegmentering blir en mer standardisert metode, vil RANO-måling og visuell radiologisk bedømming kunne være en gylden middelvei i klinisk praksis for estimering av tumorstørrelse.

Hvilken metode for vekstvurdering det er hensiktsmessig å bruke på et individuelt grunnlag er likevel foreløpig noe usikkert, da dette avhenger av hvilken prognostisk verdi de ulike vekstestimatene har. Det er for eksempel ikke sikkert at pasientene havner i en annen behandlingsgruppe selv om estimatet av tumorstørrelse blir mer nøyaktig. Dette går dessverre utover rammene for denne oppgaven, da man då måtte ha sett på behandlingen og overlevelsen til disse 106 pasientene i lys av de ulike måtene å vurdere veksten på.



## **5. Konklusjon**

Målsettingen var å undersøke hvor god radiologisk vekstvurdering av glioblastomer ved bruk av visuell bedømming og RANO-målinger er sammenlignet med volumsegmentering (gullstandard). Vi ønsket også å undersøke grad av samsvar (inter-rater variabilitet) mellom nevrordiologene.

Vi fant moderat grad av enighet (Kappa-verdi ca. 0,70) mellom nevrordiologenes bedømming av vekst ved bruk av begge metoder, når veksten ble kategorisert i henhold til RANO-kriterienes vekstkategorier. Det var noe overraskende at det ikke var større forskjeller mellom de to metodene siden man gjør kvantitative målinger når man bruker RANO-kriteriene, men våre funn tyder på at dette kan skyldes at RANO-kategoriene er relativt vide. Denne svakheten ved RANO-kriteriene er verd å merke seg når man leser eller gjennomfører studier med progresjonsfri overlevelse som endepunkt. Man bør også utvise forsiktighet i radiologisk tolkning av vekst dersom denne ikke er målt volumetrisk, spesielt i tilfeller der vekstestimatet vil få behandlingsmessige konsekvenser.

Vi fant at både visuell radiologisk bedømming og RANO-målinger var unøyaktige estimater for tumorveksten når vi sammenlignet med volumsegmentering. Begge disse metodene hadde omtrent samme sensitivitet, spesifisitet, NPV og PPV når det ble utført av to nevrordiologer. Begge metodene hadde en sensitivitet på omtrent 60%, noe som vil si at 40 % av svulstene som faktisk har vokst ikke vil bli fanget opp ved visuell bedømming eller ved RANO-målinger. Dette underbygger behovet for volumetrisk segmentering dersom man nøyaktig skal estimere vekst av tumor.

Hvorvidt nøyaktig estimering av tumorvekst er fordelaktig på et individuelt grunnlag gjenstår enda å se, da man må utføre videre studier hvor man ser på hvordan dette eventuelt ville ha påvirket behandlingsvalg og korrelerer med overlevelse.



## 6. Referanser

1. Osborn, A.G., *Osborn's brain : imaging, pathology, and anatomy*. 2013, Salt Lake City, Utah: Amirsys.
2. Thakkar, J.P., et al., *Epidemiologic and molecular prognostic review of glioblastoma*. *Cancer Epidemiol Biomarkers Prev*, 2014. **23**(10): p. 1985-96.
3. Stupp, R., et al., *Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma*. *N Engl J Med*, 2005. **352**(10): p. 987-96.
4. Ronning, P.A., et al., *A population-based study on the effect of temozolomide in the treatment of glioblastoma multiforme*. *Neuro Oncol*, 2012. **14**(9): p. 1178-84.
5. Stummer, W., et al., *Extent of resection and survival in glioblastoma multiforme: identification of and adjustment for bias*. *Neurosurgery*, 2008. **62**(3): p. 564-76; discussion 564-76.
6. Storstein, A., et al., *Høygradige gliomer hos voksne*. *Tidsskrift for den Norske Laegeforening*, 2011. **131**(3): p. 238-241.
7. Stensjoen, A.L., et al., *Growth dynamics of untreated glioblastomas in vivo*. *Neuro Oncol*, 2015. **17**(10): p. 1402-11.
8. Wen, P.Y., et al., *Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group*. *J Clin Oncol*, 2010. **28**(11): p. 1963-72.
9. Shah, G.D., et al., *Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas*. *Neuro Oncol*, 2006. **8**(1): p. 38-46.
10. WHO, *Handbook for Reporting Results of Cancer Treatment*. 1979, Offset Publication No. 48. Geneva: World Health Organization.
11. Miller, A.B., et al., *Reporting results of cancer treatment*. *Cancer*, 1981. **47**(1): p. 207-14.
12. Sorensen, A.G., et al., *Response criteria for glioma*. *Nat Clin Pract Oncol*, 2008. **5**(11): p. 634-44.
13. Quant, E.C. and P.Y. Wen, *Response Assessment in Neuro-Oncology*. *Current Oncology Reports*, 2010. **13**(1): p. 50-56.
14. Porz, N., et al., *Multi-modal glioblastoma segmentation: man versus machine*. *PLoS One*, 2014. **9**(5): p. e96873.
15. Macdonald, D.R., et al., *Response criteria for phase II studies of supratentorial malignant glioma*. *J Clin Oncol*, 1990. **8**(7): p. 1277-80.
16. Brandsma, D., et al., *Clinical features, mechanisms, and management of pseudoprogression in malignant gliomas*. *Lancet Oncol*, 2008. **9**(5): p. 453-61.
17. Taal, W., et al., *Incidence of early pseudo-progression in a cohort of malignant glioma patients treated with chemoirradiation with temozolomide*. *Cancer*, 2008. **113**(2): p. 405-10.
18. Kothari, P.D., et al., *Longitudinal restriction spectrum imaging is resistant to pseudoresponse in patients with high-grade gliomas treated with bevacizumab*. *AJNR Am J Neuroradiol*, 2013. **34**(9): p. 1752-7.
19. Fu, P., et al., *Bevacizumab treatment for newly diagnosed glioblastoma: Systematic review and meta-analysis of clinical trials*. *Mol Clin Oncol*, 2016. **4**(5): p. 833-838.
20. Henson, J.W., S. Ulmer, and G.J. Harris, *Brain Tumor Imaging in Clinical Trials*. *American Journal of Neuroradiology*, 2008. **29**(3): p. 419-424.

21. Wang, M.Y., et al., *Measurement of tumor size in adult glioblastoma: classical cross-sectional criteria on 2D MRI or volumetric criteria on high resolution 3D MRI?* Eur J Radiol, 2012. **81**(9): p. 2370-4.
22. Provenzale, J.M., C. Ison, and D. DeLong, *Bidimensional measurements in brain tumors: assessment of interobserver variability.* AJR Am J Roentgenol, 2009. **193**(6): p. W515-22.
23. Provenzale, J.M. and M.C. Mancini, *Assessment of intra-observer variability in measurement of high-grade brain tumors.* J Neurooncol, 2012. **108**(3): p. 477-83.
24. Stensjoen, A.L., *Intra-observer Variability in Segmentation of Glioblastomas,* in *Hovedoppgave medisn.* 2014, NTNU: DMF.
25. Stensjoen, A.L., *Volumetric Segmentation of Glioblastomas and Evaluation of Growth Dynamics,* in *Forskerlinjeoppgave medisn.* 2015, NTNU: DMF.
26. McHugh, M.L., *Interrater reliability: the kappa statistic.* Biochem Med (Zagreb), 2012. **22**(3): p. 276-82.

# Appendix

## Appendix 1

### RADIOLOGISK VURDERING AV GBM VEKST Nevroradiolog 1/2

CASE NUMMER & INITIALER	FØDSELSNUMMER
-------------------------	---------------

#### Radiologisk bedømming

1. Ved kun å «scrolle» i de kontrastforsterkede T1-volumene, hvor stor anslår du at veksten er?

- Reduksjon i tumorvolum
- 0 -25 %
- 25-100 %
- 100 – 200 %
- > 200 %

2. Hva er målet på den største tumor diameteren på aksiale snitt, og diameteren perpendiuklært på den største diameteren (i samme aksiale snitt). Hvis det er flere lesjoner, mål også på disse, men bare på kontrastladende lesjoner :

Preoperativ scan	Diagnostisk scan
Lesjon 1 <input type="checkbox"/> Diameter 1 (mm) : <input type="checkbox"/> Diameter 2 (mm) :	Lesjon 1 <input type="checkbox"/> Diameter 1 (mm) : <input type="checkbox"/> Diameter 2 (mm) :
Lesjon 2 <input type="checkbox"/> Diameter 1 (mm) : <input type="checkbox"/> Diameter 2 (mm) :	Lesjon 2 <input type="checkbox"/> Diameter 1 (mm) : <input type="checkbox"/> Diameter 2 (mm) :
Lesjon 3 <input type="checkbox"/> Diameter 1 (mm) : <input type="checkbox"/> Diameter 2 (mm) :	Lesjon 3 <input type="checkbox"/> Diameter 1 (mm) : <input type="checkbox"/> Diameter 2 (mm) :

Eventuell kommentar : \_\_\_\_\_

## Appendix 2

CASE NUMMER & INITIALER	FØDSELSNUMMER
-------------------------	---------------

Radiologisk bedømming gjort ved operasjonstidspunktet
<p>1. Er tumorvekst vurdert?</p> <ul style="list-style-type: none"><li><input type="checkbox"/> Nei</li><li><input type="checkbox"/> Ja<ul style="list-style-type: none"><li><input type="checkbox"/> Reduksjon</li><li><input type="checkbox"/> Uendret</li><li><input type="checkbox"/> Vekst</li></ul></li></ul>
<p>2. Er vekstmål oppgitt?</p> <ul style="list-style-type: none"><li><input type="checkbox"/> I antall millimeter. Hvordan er det målt? _____</li><li><input type="checkbox"/> I relativ vekst (f.eks. dobling)? _____</li><li><input type="checkbox"/> Annet _____</li></ul>
<p>3. Dersom vekstmål er oppgitt, hvordan samsvarer dette med volumetri</p> <ul style="list-style-type: none"><li>a. <input type="checkbox"/> Ikke aktuelt</li><li>b. <input type="checkbox"/> Vekstmål oppgitt: _____</li><li>c. <input type="checkbox"/> Volumetrisk mål: _____</li></ul>