



**Den diagnostiske verdien av fire nevrodynamiske tester i undersøkelsen av
cervikal radikulopati – En litteraturstudie**

Ørjan Hagen Sørli

Masterprogram i helsefag – Klinisk masterstudium i manuellterapi for
fysioterapeuter. Institutt for global helse og samfunnsmedisin, Universitetet i
Bergen

MANT 395, 4.semester, 2022

Antall ord: 10 212

Forord

Arbeidet med denne masteroppgaven er nå slutført. Det har vært givende å få anledning til å fordype seg i eget fagfelt. Prosessen har vært svært berikende, samtidig som det har vært en lang og krevende vei å gå.

En stor takk til min veileder Jon Joensen, som har fulgt opp undertegnede på en eksemplarisk måte gjennom hele prosessen. Joensen er faglig svært dyktig, samt dypt engasjert i sitt fagfelt og arbeid. Han har vært lett tilgjengelig, både for tilbakemeldinger og veiledning.

Takk til bibliotekar ved UiB, som var svært behjelpelig under søkeprosessen til denne oppgaven.

En stor takk rettes til familie og venner, for oppmuntring og støtte gjennom de to siste årene.

Avslutningsvis vil jeg rette en stor takk til min forlovede Mari, for gode råd og konstruktive innspill, samt uerstattelig støtte gjennom arbeidet.

Ørjan Hagen Sørli

November 2022

Sammendrag

Tema

Nakkesmerter er utbredt i befolkningen med en ettårsprevalens mellom 30-50%. I Norge ser man at nakkesmerter er blant de vanligste årsakene til langtidssykemeldinger og uføretrygd. Herunder er cervikal radikulopati en vanlig nakkelidelse, som er forbundet med uttalte smerter, funksjonsnedsettelse og redusert arbeidsevne. Det mangler konsensus om den diagnostiske verdien av kliniske undersøkelser som anvendes for å bekrefte eller avkrefte diagnosen.

Problemstilling

Hvilken diagnostisk verdi har fire nevrodynamiske tester (ULNT) i undersøkelsen av cervikal radikulopati (CR)?

Metode

Litteraturstudie som har vurdert den diagnostiske verdien av ULNT i undersøkelsen av CR, sammenlignet med magnetresonanstomografi (MR) eller elektromyografi (EMG). Data og resultater er hentet fra tre ulike diagnostiske kohortstudier.

Resultat

Testene viser varierende diagnostisk verdi hva gjelder sensitivitet, spesifisitet og likelihood ratios, både for hver individuelle test og om man sammenligner resultatene på tvers av studiene. Om man skal benytte enkeltstående ULNT, er det ULNT 2a og ULNT 3 som viser høyest diagnostisk verdi. Om man utfører ULNT som et test-cluster, øker testenes diagnostiske verdi og anbefales fra *to* av de inkluderte studiene å benyttes for å bekrefte CR som diagnose. *Én* av de inkluderte studiene indikerer at 4/4 negative ULNT kan utelukke CR som diagnose.

Konklusjon

Denne litteraturstudien støtter tidligere forskning på området, som viser at det ikke foreligger entydig evidens om den diagnostiske verdien til ULNT i undersøkelsen av CR. Den samlede informasjonen fra anamnese, samt resultater fra nevrologisk undersøkelse og andre spesifikke kliniske undersøkelser, *inkludert* test-cluster av ULNT, bør være grunnlag for diagnosen CR. Ved klinisk mistanke om CR, kan diagnosen eventuelt bekreftes med MR eller EMG.

Nøkkelord

Cervikal radikulopati, nevrodynamiske tester, ULNT, sensitivitet, spesifisitet, likelihood ratio

Summary

Theme

Neck pain is widespread in the general population with a one-year prevalence of 30-50%. Neck pain is among the most common causes for long-term sick leave and disability in Norway. Cervical radiculopathy is a common neck pathology, and is associated with severe pain, functional impairments and reduced work capacity. There is a lack of consensus regarding the diagnostic value of the clinical tests used in the assessment of the diagnosis.

Research question

What is the diagnostic value of upper limb neurodynamic tests (ULNT) in the assessment of cervical radiculopathy (CR)?

Methods

A literature study who assessed the diagnostic value of ULNT in the assessment of CR, compared to magnetic resonance imaging (MRI) or electromyography (EMG). Data and results are concluded from three different diagnostic cohort studies.

Results

This literature study reveals that ULNT has varying diagnostic value in regards to sensitivity, specificity and likelihood ratios, both for each individual test and across the included studies. If you are to use individual ULNT, ULNT 2a and ULNT 3 shows the greatest diagnostic value. The combination of all ULNT applied as a test-cluster, seems to increase the diagnostic value. *Two* of the studies recommends the application of ULNT as a test-cluster to rule-in CR as a diagnosis, and *one* study recommends a ULNT test-cluster of 4/4 negative tests to rule-out CR.

Conclusion

The findings from this study seems to support previous research, indicating a lack of consensus regarding the diagnostic value of ULNT in the assessment of CR. The combined information from anamnesis, results from neurological examination and other clinical tests *including* a cluster of ULNT, should be the basis for the diagnosis of CR. The diagnosis can be confirmed with MRI or EMG, if clinically suspected.

Keywords

Cervical radiculopathy, upper limb neurodynamic tests, ULNT, sensitivity, specificity, likelihood ratios.

Begrepsavklaring

Ord	Forklaring
Index-test	Intervensjonen – den testen som blir evaluert (Tacconelli, 2010)
ULNT	Upper limb neurodynamic test (Koulidis et al., 2019)
Test-cluster	Man kombinerer flere enkelttester som et middel for å bedre den diagnostiske prosessen, når man ønsker å bekrefte eller avkrefte en patologi (Hegedus et al., 2015)
Sensitivitet	Angir hvor mange «sant syke» som fremviser et positivt testresultat (Jamtvedt, 2015)
Spesifisitet	Angir hvor mange «sant friske» som fremviser et negativt testresultat (Jamtvedt, 2015)
Validitet	I hvilken grad en test eller et instrument måler det den er ment til å måle (Thomas et al., 2011)
Likelihood ratio	Beskriver hvor mange ganger mer sannsynlig det er at en deltager med målsykdommen vil oppnå et bestemt testresultat, enn en deltager uten målsykdommen (Tacconelli, 2010)
Referansetest	Den beste testen som finnes på området (Jamtvedt, 2015)

Innholdsfortegnelse

1.0 Innledning	1
1.1 Bakgrunn for valg av tema	1
1.2 Tidligere forskning på området	2
1.3 Hensikt og problemstilling.....	3
1.3.1 Hensikt.....	3
1.3.2 Problemstilling.....	3
1.3.3 Avgrensing av problemstilling.....	3
2.0 Teori.....	4
2.1 Cervikal radikulopati.....	4
2.1.1 Etiologi og patoanatomi	4
2.2 Smertekarakteristika for CR.....	5
2.3 Diagnostikk – Undersøkelse av pasienter med CR.....	5
2.3.1 Presentasjon av ULNT	6
2.3.2 MR og EMG som referansetester	7
3.0 Metode	8
3.1 Valg av forskingsdesign	8
3.2 Utvalg.....	9
3.2.1 PICO-skjema	9
3.2.2 Inklusjonskriterier	9
3.2.3 Eksklusjonskriterier.....	10
3.2.4 Datainnsamling	10
3.3 Vurdering av studienes kvalitet	11
3.3.1 Validitet	11
3.3.2 Intern validitet	11
3.3.3 Ekstern validitet	12
3.3.4 Reliabilitet	13
3.4 Statistisk analyse.....	13
3.4.1 Sensitivitet, spesifisitet, negativ og positiv likelihood ratio som utfallsmål.....	15
3.5 Etske betraktninger.....	16
4.0 Resultater	16
4.1 Valg av studier	16
4.1.1 Flytskjema over inkluderte og ekskluderte studier.....	17
4.2 Presentasjon av ekskluderte og inkluderte artikler	18
4.3 Utdypende presentasjon av inkluderte studier	19
4.4 Utdyping av studienes resultater.....	21
4.4.1 Resultater studie 1	21

4.4.2 Resultater studie 2	21
4.4.3 Resultater studie 3	22
4.5 Sjekkliste for vurdering av en studie som tester en ny diagnostisk test	24
5.0 Diskusjon	26
5.1 Metodediskusjon av inkluderte studier	26
5.1.1 Intern validitet	26
5.1.2 Ekstern validitet	28
5.2 Resultatdiskusjon	30
5.2.1 Hva forteller resultatene?	30
5.2.2 Hva kan ha påvirket resultatene?	31
5.2.3 Resultatdiskusjon av test-clustere	34
5.3 Implikasjoner for praksis	35
5.4 Masteroppgavens styrker og svakheter	35
6.0 Konklusjon	38
Litteraturliste	39
Vedlegg	43
Vedlegg 1 – Søkelogg i PubMed	43
Vedlegg 2 – Søkestrategi i PubMed	45

1.0 Innledning

Temaet som belyses i denne masteroppgaven omhandler undersøkelse av pasienter med nakke- og/eller armsmerter. Herunder er det ønskelig å fremme kunnskapen om den diagnostiske verdien til nevrodynamiske tester for overekstremitetene, og i hvilken grad disse testene kan anvendes for å diagnostisere cervikal radikulopati. I dette kapittelet vil det først presenteres bakgrunn for valg av tema. Dernest følger tidligere forskning på området, før hensikt og problemstilling, samt avgrensning av problemstilling blir fremlagt.

1.1 Bakgrunn for valg av tema

Nakkesmerter er utbredt i den internasjonale befolkningen med en ettårsprevalens mellom 30-50%, avhengig av ulikheter i definisjon og metode (Hogg-Johnson et al., 2008). I Norge ser man at nakkesmerter sammen med andre muskel- og skjelettplager er blant de vanligste årsakene til langtidssykemeldinger og uføretrygd (Nygaard et al., 2010). Som en subgruppe av nakkesmerter finner man cervikal radikulopati, heretter CR. CR er en relativt vanlig neurologisk lidelse å finne i klinisk praksis, og er assosiert med uttalte smerter, funksjonsnedsettelse og psykiske plager. Data i henhold til prevalens og epidemiologi av CR er begrenset (Koulidis et al., 2019), men én studie har estimert en årlig prevalens på 107.3 per 100 000 for menn, og 63.5 per 100 000 for kvinner (Radhakrishnan et al., 1994). Det er gjennomgående høyere prevalens blant yrkesaktive, der man også finner at CR i stor grad bidrar til den sosioøkonomiske byrden ved arbeidsrelaterte muskel- og skjelettplager (Koulidis et al., 2019). Epidemiologiske studier har avdekket at C7-nerveroten er den hyppigst affiserte ved CR, etterfulgt av nerverot C6 og C8 (Iyer & Kim, 2016).

Det foreligger i dag ingen nasjonale eller internasjonale retningslinjer for undersøkelse og behandling av pasienter med nakkesmerter, herunder CR, tilsvarende det som for eksempel finnes for pasienter med lave korsryggsmerter (Nygaard et al., 2010). Med bakgrunn i dette kan man argumentere for at det er behov for enhetlige retningslinjer for undersøkelse og behandling for denne pasientgruppen. Det å kunne identifisere pasienters smerte og dysfunksjoner gjennom standardiserte tester kan være en viktig del av diagnostikk i fysioterapien, der man som kliniker er avhengig av å kunne stole på svarene testene gir. Det finnes en rekke kliniske tester man kan anvende i undersøkelsen av pasienter med nakke- og/eller armsmerter, hvor blant annet nevrodynamiske tester er anbefalt å utføre (Thoomes et al., 2018). Nevrodynamiske tester kalles også for nervestrekktester. På engelsk benevnes disse som Upper Limb Tension Tests/Upper Limb Neurodynamic tests (ULTT/ULNT), og vil

heretter benevnes *ULNT*. Disse testene involverer målrettede bevegelsessekvenser som har til hensikt å fremprovosere mekanosensitivitet i nervevevet, ved å gradvis tilføre økende tensjon på de komponentene av nervesystemet som undersøkes (Elvey, 1986). Testene man utfører skal avdekke sentrale funn, som skal lede frem mot en diagnose og et hensiktsmessig behandlingsregime (Littlewood & Pope, 2017).

I min praksis som fysioterapeut og manuellterapistudent har jeg møtt mange pasienter med nakke- og armsmerter. Jeg opplever at denne pasientgruppen utgjør en stor andel av min praksis og jeg ser at deres plager i stor grad påvirker deres hverdag, arbeidsevne og generelle livsførsel. I arbeidet med disse pasientene opplever jeg at det er vesentlig med riktig diagnose så tidlig som mulig i forløpet, for å kunne igangsette den mest hensiktsmessige behandlingen. Under manuellterapistudiets praksis har min veileder hatt inngående teoretisk og praktisk undervisning om klinisk evaluering av det perifere nervesystemet i undersøkelsen av pasienter med nakke- og armsmerter, herunder bruk og tolkning av *ULNT*. Dette vekket min interesse for å fordype meg i dette temaet, og utforske i hvilken grad vi som klinikere kan anvende og stole på *ULNT* i vår arbeidshverdag. På bakgrunn av dette søkte jeg i litteratur som kunne si noe om eksisterende kunnskap på området.

1.2 Tidligere forskning på området

ULNT sammen med en rekke andre kliniske tester kan anvendes for å diagnostisere CR (Thoomes et al., 2018), og det er fra tidligere utført forskning på området som blant annet innebærer å vurdere den diagnostiske kvaliteten av de ulike testene. Herunder foreligger det flere systematiske oversiktsartikler (Koulidis et al., 2019; Rubinstein et al., 2007; Thoomes et al., 2018). Koulidis et al. (2019) diskuterer om *ULNT* kan bidra til å oppdage perifere nevrologiske smerter ved å vurdere nevralt mekanosensitivitet, og dermed differensiere mellom nevrologiske smerter og smerter fra annet vev, som kan være viktig i utredningen av CR. En annen systematisk oversiktsartikkel som omhandler utredning av nakkesmerter, beskriver at en negativ *ULNT* av nervus medianus er ansett å være validert som en spesifikk test for å utelukke CR, sammenlignet med resultater fra MR¹ (Rubinstein et al., 2007). En nylig publisert systematisk oversikt statuerer at det finnes begrenset evidens for den diagnostiske nøyaktigheten av kliniske tester for CR (Thoomes et al., 2018). Rubinstein et al. (2007) gjorde en evaluering av Spurlings test, Valsalvas manøver, traksjon/distraksjon av nakke, samt *ULNT* i undersøkelsen av CR, og konkluderte med at det er behov for ytterligere forskning for å bedømme testenes diagnostiske verdi. Dette støttes av Koulidis et al. (2019),

¹ Magnetresonanstomografi

som skriver at selv om ULNT benyttes av ulike klinikere er deres diagnostiske egnethet enda ikke fullt etablert. Fra dette kan man tolke at vi ikke har enighet hva gjelder utredning av CR ved hjelp av ULNT. Disse faktorene, sammen med aspektene som fremkommer av *kapittel 1.1*, kan indikere et behov for en systematisk litteraturgjennomgang av studier som har vurdert den diagnostiske nøyaktigheten av ULNT i undersøkelsen av CR. Dette kan potensielt gi nye innsikter som kan være av nyttig verdi for fysioterapeuter, manuellterapeuter og annet helsepersonell i undersøkelsen av pasienter med nakke- og/eller armsmerter, som videre kan bidra til å bedre utredningen og oppfølgingen av disse pasientene både i første- og andrelinjetjenesten.

1.3 Hensikt og problemstilling

1.3.1 Hensikt

Hensikten er å vurdere den diagnostiske verdien av ULNT i undersøkelsen av CR, for å utforske hva slags informasjon testene gir og i hvilken grad man kan stole på testresultatene. Videre er det ønskelig å kartlegge om det å sette testene sammen til et test-cluster, øker undersøkelsens diagnostiske verdi.

1.3.2 Problemstilling

Hvilken diagnostisk verdi har nevrodynamiske tester i undersøkelsen av cervikal radikulopati?

1.3.3 Avgrensning av problemstilling

I denne oppgaven avgrenses problemstillingen som følger:

Nevrodynamiske tester omfatter i denne oppgaven fire ULNT for henholdsvis nervus medianus (ULNT 1 og ULNT 2a), nervus radialis (ULNT 2b) og nervus ulnaris (ULNT 3). Se beskrivelse av ULNT i *kapittel 2.3.1*, tabell I. *Diagnostisk verdi* vurderes mot referansetestene MR og/eller EMG². Utfallsmålene for vurdering av ULNT inkluderer sensitivitet, spesifisitet, positiv og negativ likelihood ratio, samt positiv og negativ prediktiv verdi* for denne oppgaven. Det vil kun inkluderes litteratur omhandlende CR. Litteratur som omhandler annen patologi *alene* der ULNT anvendes som undersøkelsesmodalitet, ved eksempelvis entrapment-nevropatier, skulderpatologi eller lignende, vil ikke inkluderes. Studier som undersøker CR *sammen* med annen patologi, kan inkluderes.

² Elektromyografi. *= Positiv og negativ prediktiv verdi ble anvendt i søkeprosessen, men ikke inkludert som utfallsmål i analysen. Se *kapittel 4.0* for begrunnelse.

2.0 Teori

I dette kapittelet vil det presenteres sentral teori omhandlende CR som diagnose. Herunder vil litteratur som kan bidra til å definere diagnosen fremlegges, før det presenteres teori om etiologi og patoanatomiske forhold. Deretter vil det redegjøres for smertekarakteristika som er typisk for diagnosen. Avslutningsvis presenteres teori omhandlende diagnostikk, herunder klinisk undersøkelse, beskrivelse av ULNT, samt MR og EMG som referansetester.

2.1 Cervikal radikulopati

For å forklare diagnosen CR og virkningsmekanismene som ligger til grunn for de ulike symptomene, må *radikulopati* først defineres. Ifølge Helseth et al. (2019) er radikulopati definert som en skade eller dysfunksjon av en spinal nerverot, der affeksjon av fremre rot kan føre til pareser i tilhørende myotom, mens affeksjon av bakre rot kan lede til sensibilitetsutfall i tilhørende dermatom, samt refererte smerter.

I litteraturen finnes det mange ulike definisjoner av CR, der man ikke har konsensus med veldefinerte kliniske kriterier for diagnosen (Thoomes et al., 2012). I denne oppgaven vil det tas utgangspunkt i definisjonen til Rubinstein et al. (2007): *Symptomer som stammer fra dysfunksjon i en spinalnerve fra nakken*. Den kliniske manifestasjonen er bred og symptombildet kan variere. Pasientene presenterer ofte med nakkesmerter, armsmerter eller begge deler (Iyer & Kim, 2016). Typiske symptomer kan være sensoriske og motoriske endringer, reduserte eller utslukkede reflekser, smerter og kombinasjoner av disse (Koulidis et al., 2019; Thoomes et al., 2018).

2.1.1 Etiologi og patoanatomi

I en nylig publisert systematisk oversiktsstudie er det foreslått at CR kan skyldes mekanisk kompresjon av en intervertebral disk og dens nerverot, eller andre plassoppfyllende lesjoner, samt inflammasjon av nerveroten (Thoomes et al., 2018). Patoanatomisk er de vanligste årsakene cervikal spondylose og ulike former for prolaps (Iyer & Kim, 2016). Man skal også være oppmerksom på at etiologien, i tillegg til kompressive virkningsmekanismer som ved for eksempel spondylose, også kan skyldes ikke-degenerative prosesser. Ikke-degenerative prosesser som kan lede til CR er blant annet foreslått å være infeksjon, redusert eller blokkert blodtilførsel til nerveroten, rotavulusjon, tumorer og demyelinering (Carette & Fehlings, 2005). Selv om de kompressive patoanatomiske forholdene er klart hyppigst forekommende, er det viktig å hensynta de ikke-degenerative prosessene som mulige forklaringer for en god differensialdiagnostisk utredning (Iyer & Kim, 2016). Herunder har Carette og Fehlings (2005) foreslått funn som kan indikere alvorlig underliggende patologi eller ikke-degenerative

prosesser å være; Lhermittes tegn, gangvansker, muskelsvakhet og atrofi i hender, samt sykdomshistorikk med blant annet feber, uforklarlig vekttap, immunsuppresjon og kjent cancer. Dette støttes av Nygaard et al. (2010) som beskriver traumer, malignitet, myelopati, infeksjoner og systemiske lidelser som potensielle røde flagg.

2.2 Smertekarakteristika for CR

Med bakgrunn i diagnosens anatomiske beliggenhet affiseres det perifere nervesystemet, og vil i så måte ofte medføre perifer nevropatisk smerte (PNS) (Koulidis et al., 2019). PNS kan defineres som smerte som stammer fra en lesjon eller sykdom i det somatosensoriske nervesystemet (Finnerup et al., 2016). PNS kan oppstå når en perifer nervestamme eller nerverot har blitt utsatt for skade, kompresjon, inflammasjon eller hypoksi, som ved CR, som resulterer i affeksjon av nervesystemet (Nee & Butler, 2006). Symptomene ved slike nevropatier kan kategoriseres som positive eller negative. *Positive* symptomer inkluderer smerter, parestesier, dysestesier, hyperalgesi og allodyni, som indikerer unormal eksitabilitet i nervesystemet. De *negative* symptomene som hypoestesi eller anestesi, samt pareser, indikerer redusert overføring av nerveimpulser (Woolf, 2004). De vanligste lidelsene som affiserer det perifere nervesystemet og kan medføre PNS er CR, i tillegg til entrapment-nevropatier, som carpal tunnel syndrom og cubital tunnel syndrom (Woolf, 2004).

2.3 Diagnostikk – Undersøkelse av pasienter med CR

CR er definert som en *klinisk* diagnose basert på anamnese og kliniske undersøkelser, som kan bekreftes ved eventuell billeddiagnostikk eller elektrofysiologiske målinger (Sharma et al., 2014). Den kliniske undersøkelsen baserer seg på anamnese og et utvalg av tester. Den kliniske undersøkelsen bør innefatte en fullstendig nevrologisk undersøkelse, med formål om å undersøke integritet og funksjon i nervesystemet med hensyn til overføring av afferente og efferente impulser (Baselgia et al., 2017). Flere pasienter med CR kan presentere normal nevrologisk undersøkelse hva gjelder reflekser, kraft og sensibilitet, men likevel oppleve PNS (Schmid et al., 2009). Derfor bør spesifikke kliniske tester tas i bruk. I en retningslinje fra Helsedirektoratet i 2014 henvises det til at en positiv Spurlings test, kompresjon/traksjon av nakken og Valsalvas manøver kan tyde på CR (Rubinstein et al., 2007). I følge Schmid et al. (2009) bør en grundig klinisk undersøkelse også omfatte en vurdering av nervesystemets mekaniske sensitivitet. Begrepet mekanosensitivitet beskriver nervevevets evne til å tolerere mekanisk kraft og strekk. Mekanosensitivitet kan være viktig å vurdere i diagnostiseringen av pasienter med nakke- og/eller armsmerter, da det kan indikere patologi av nevralt vev (Schmid et al., 2009). Det er her ULNT kan ha sin funksjon, ved å nettopp kunne identifisere

endringer av mekanosensitivitet i nervesystemet og dermed kunne oppdage og identifisere CR (Baselgia, 2017). En ULNT vil tillegge mekaniske krefter til deler av nervesystemet, men vil også belaste ikke-nevralt vev. Derfor kan smerteresponsen ved en ULNT være relatert til sensitivitet i *både* nevralt og ikke-nevralt vev. Herunder er smerteresponsen vurdert å være relatert til nevral mekanosensitivitet om symptomene endrer seg ved bevegelser av en kroppsdel som ligger fjernt fra smerteområdet, som enten belaster eller avlastar nervevevet (Nee et al., 2012). Denne prosedyren er beskrevet som *strukturell differensiering*. De belastende eller avlastende bevegelsene kan for eksempel være lateralfleksjon av nakke, fleksjon av albue eller slipp av håndleddsekstensjon, avhengig av hvor pasienten opplever smerteresponsen (Grondin et al., 2021). Se *tabell 1* for presentasjon av de fire ULNT hva gjelder utførelse, strukturell differensiering og tolkning av testresultat.

2.3.1 Presentasjon av ULNT

Herunder følger en forklaring av utførelsen av index-testene etter definisjon og prosedyre av Grondin et al. (2021) og Koulididis et al. (2019).

Tabell I – Utførelse, strukturell differensiering og tolkning av ULNT

Test	Utførelse	Tolkning av positivt resultat
ULNT 1 (Nervus medianus)	<ul style="list-style-type: none"> • Pasient i ryggliggende • Stabilisering av skulderbue • Skulderabduksjon • Ekstensjon av hånd/fingre • Supinasjon av underarm • Lateralrotasjon av skulderleddet • Albueekstensjon • Strukturell differensiering: Lateral fleksjon av cervikalcolumna eller slipp av håndleddsekstensjon 	En nevrodynamisk test er ansett som positiv om den reproducerer pasientens egne symptomer og hvis disse kan endres gjennom strukturell differensiering
ULNT 2a (Nervus medianus)	<ul style="list-style-type: none"> • Pasient i ryggliggende • Depresjon av skulderbue • Albueekstensjon • Lateralrotasjon av skulderledd og supinasjon av underarm • Ekstensjon av håndledd/fingre • Skulderabduksjon • Strukturell differensiering: Lateral fleksjon av cervikalcolumna, slipp av skulderdepresjon eller slipp av håndleddsekstensjon 	En nevrodynamisk test er ansett som positiv om den reproducerer pasientens egne symptomer og hvis disse kan endres gjennom strukturell differensiering

ULNT 2b (Nervus radialis)	<ul style="list-style-type: none"> • Pasient i ryggliggende • Depresjon av skulderbue • Albueekstensjon • Lateralrotasjon av skulderledd og pronasjon av underarm • Fleksjon av håndledd/fingre • Skulderabduksjon • Strukturell differensiering: Slipp av skulderdepresjon eller håndleddsflexjon 	En nevrodynamisk test er ansett som positiv om den reproducerer pasientens egne symptomer og hvis disse kan endres gjennom strukturell differensiering
ULNT 3 (Nervus ulnaris)	<ul style="list-style-type: none"> • Pasient i ryggliggende • Håndledds/fingerekstensjon • Pronasjon av underarm • Albuefleksjon • Lateralrotasjon av skulderledd • Depresjon av skulderbue • Skulderabduksjon • Strukturell differensiering: Lateralflexjon av cervikalcolumna, slipp av depresjon skulderbue eller slipp av håndleddsflexjon 	En nevrodynamisk test er ansett som positiv om den reproducerer pasientens egne symptomer og hvis disse kan endres gjennom strukturell differensiering

2.3.2 MR og EMG som referansetester

Når man vurderer diagnostiske målemetoder er det ønskelig at testen som undersøkes skal måles opp mot en referansetest. For å sikre at den nye målemetoden kvalitetssikres på best mulig måte, bør referansetesten som velges være den nåværende gullstandarden på området (Jamtvedt, 2015). Ifølge Nasjonal faglig retningslinje for bildediagnostikk ved ikke-traumatiske muskel- og skjelettlidelser fra Helsedirektoratet råder det enighet om at MR og EMG er nåværende gullstandard for undersøkelse av CR (Sharma et al., 2014).

I en systematisk litteraturgjennomgang fant man at EMG har en sensitivitet på 50-71% hos personer med nevrologiske eller radiologiske tegn på CR (So, 1999). En nyere retrospektiv studie finner at EMG har en sensitivitet og spesifisitet på henholdsvis 34% og 74%, ved en klinisk positiv CR (Govindarajan et al., 2013). Vedrørende MR så betraktes dette som en gullstandard undersøkelsesmetode for bløtvev og benmarg, herunder blant annet CR, da den evner til å skille mellom patologiske tilstander i disse vevene. MR er anbefalt for å påvise skiveprolaps og degenerative forandringer som kan påvirke nerverøtter (Sharma et al., 2014). MR har i en studie vist sensitivitet på 76% og en spesifisitet på 93% ved en klinisk positiv CR (Govindarajan et al., 2013). Likevel er det viktig å være oppmerksom på at ingen av disse metodene er fullstendig valide, da funnene fra undersøkelsene kan være asymptomatiske for pasienten (Hanchard et al., 2013).

3.0 Metode

I dette kapittelet vil først valg av forskingsdesign begrunnes. Deretter følger kriterier relatert til utvalg, med tilhørende PICO-skjema, inklusjons- og eksklusjonskriterier samt datainnsamling. Videre fremlegges kriterier vedrørende vurdering av de inkluderte studienes kvalitet, før statistisk analyse presenteres. Avslutningsvis følger etiske betraktninger rundt denne masteroppgavens metode.

3.1 Valg av forskingsdesign

Litteraturstudier er en egnet metode når man ønsker å samle inn og systematisere kunnskap som allerede finnes om et tema (Magnus & Bakketeig, 2000). I denne masteroppgaven vil systematisk litteraturstudie anvendes som studiedesign for å innhente og systematisere kunnskapen rundt den diagnostiske verdien av ULNT med hensyn til målsykdommen CR. Litteraturstudier er kjennetegnet ved å være metodiske, omfattende, transparente og reproduerbare. De involverer en systematisk søkeprosess for å oppdage all relevant informasjon som adresserer ett eller flere forskningsspørsmål, i tillegg til en systematisk presentasjon av karakteristika og funn fra det aktuelle søket (Jamtvedt, 2015; Jørgensen et al., 2016; Thomas et al., 2011). Den systematiske metodikken har som formål å minimere subjektivitet og bias. Bias er et engelsk uttrykk, som betyr skjevhet eller feilaktighet, ved at resultater eller slutninger avviker fra de virkelige forholdene som utforskes i studien (Jørgensen et al., 2016; Thomas et al., 2011). De beste og mest anvendelige systematiske oversiktene bruker litteraturen som er vurdert til å utvikle en ny teori eller for å evaluere eksisterende teorier og deres implikasjoner for praksis. Det er vesentlig at inklusjons- og eksklusjonskriterier i den systematiske oversikten kommer tydelig frem, slik at avgjørelsen om å inkludere eller ekskludere studier er tydelig for leseren eller andre forskere (Siddaway et al., 2019).

Temaet i denne oppgaven omhandler *diagnostikk*. Diagnostikk handler om hvordan man kan avgjøre om noen har et helseproblem. Under diagnostiseringen kan man anvende ulike måleinstrumenter, deriblant kliniske tester. For å vurdere kvaliteten av disse kliniske testene, forsøker man å finne ut av hvor godt testen skiller mellom syke og friske (Jamtvedt, 2015). Når man undersøker diagnostiske målemetoder er det foretrukne studiedesignet ifølge Tacconelli (2010) *diagnostiske kohortstudier* der index-testene, i dette tilfellet ULNT, blir sammenlignet med en referansetest. Herunder beskrives det at det er to hovedtyper av studiedesign som er å foretrekke for å undersøke nøyaktigheten av en test, henholdsvis «single-gate» og «two-gate». *Single-gate* er lignende til det som også blir kalt diagnostisk

kohortstudie og *two-gate* er lignende til case-kontroll studier. Disse studiedesignene skiller seg fra andre kohort- og case-kontrollstudier ved at de begge er tverrsnittstudier i sin struktur. Single-gate studier inkluderer deltagere med ukjent sykdomsstatus og sammenligner resultatene fra index-testen med resultatene fra en referansetest. Dermed blir dette svært representativt for hvordan testen vil bli brukt i klinisk praksis. Two-gate studiene sammenligner derimot resultatene fra index-testen hos en gruppe som allerede har en etablert diagnose, mot en referansegruppe som er beskrevet som «frisk», eller har en annen diagnose. Kjent status, med hensyn til målsykdommen, blir dermed brukt som referert standard. Dette designet blir mindre representativt, da det er lite sannsynlig at det dekker hele spekteret for hvordan testen vil bli anvendt i praksis (Tacconelli, 2010). Med dette lagt til grunn, vil det kun inkluderes «single-gate» studier (benevnes diagnostisk kohortstudie) for å besvare forskningsspørsmålet.

3.2 Utvalg

3.2.1 PICO-skjema

PICO er et verktøy for å gi struktur og klargjøring av spørsmålet for litteratursøk, samtidig som det hjelper til med utvelgelse og kritisk vurdering av litteraturen (Helsebiblioteket, 2016).

Tabell II – PICO-skjema

Person	Intervention	Comparison	Outcome
Personer med cervikal radikulopati eller symptomer på dette	Nevrodynamiske tester 1-3 (ULNT 1-3)	Elektromyografi (EMG) eller magnetresonanstomografi (MR)	Sensitivitet, spesifisitet, positiv og negativ likelihood ratio, positiv og negativ prediktiv verdi

3.2.2 Inklusjonskriterier

Inklusjonskriterier bestemmer omfanget av litteraturen som blir fanget inn (Magnus & Bakketeig, 2000). Deltagerne i studiene skal være over 18 år. Som beskrevet i *kapittel 3.1* skal også designet være diagnostisk kohortstudie, da dette designet er mest representativt for hvordan testene blir anvendt i klinisk praksis. Videre skal studiene vurdere den diagnostiske kvaliteten av ULNT, opp mot målsykdommen CR, for at testene skal vurderes på likt grunnlag mot den undersøkte diagnosen. I tillegg skal testene vurderes i forhold til en gullstandard referansetest, herunder MR eller EMG. Studiene skal vurdere index-testene etter minimum 3/6 utfallsmål, som beskrevet i PICO-skjemaet, for å vurdere testene basert på flere like kriterier. Minimum 2/4 ULNT må vurderes for at studiene kan bli inkludert, som et middel for å innhente tilstrekkelig mengde data på testene. Studiene skal ikke være publisert for mer enn 20 år siden, for inklusjon av den nyeste og mest oppdaterte litteraturen. Til slutt

kan det inkluderes litteratur på engelsk, norsk, svensk og dansk, da dette er språk jeg behersker akademisk.

Inklusjonskriteriene i søket er som følger:

- Studier ikke eldre enn 20 år
- Diagnostisk kohortstudie som studiedesign
- Deltagere i studien skal være over 18 år
- MR eller EMG som referert standard
- Vurdering av minst 3/6 utfallsmål
- Vurdering av minst 2/4 ULNT
- Engelsk, norsk, svensk eller dansk språk

3.2.3 Eksklusjonskriterier

Eksklusjonskriterier omhandler karakteristika i en studie som gjør at de ikke kan inkluderes i denne masteroppgaven, ved at man for eksempel kan forkaste artikler som ikke måler et fenomen på en bestemt måte (Magnus & Bakketeig, 2000).

Eksklusjonskriteriene er som følger:

- Andre studiedesign enn diagnostisk kohortstudie
- Studier som ikke anvender MR eller EMG som referert standard
- Andre underliggende sykdommer som: Reumatologi, sentralnervøse lidelser og systemiske sykdommer

3.2.4 Datainnsamling

Datainnsamlingen ble gjennomført som et systematisk litteratursøk. Etter samtale med veileder og bibliotekar ved UIB ble det fattet enighet om å utføre søket i databasen PubMed. Som en del av søkeprosessen ble PICO-skjema benyttet for å utarbeide søkeord og søkestrategi. Søkeprosessen foregikk i perioden 01.06.22 – 21.06.22. Søket ble først gjennomført med ulike kombinasjoner for å få en oversikt over datamaterialet som finnes. I søket ble det benyttet boolske operatører som OR og AND, samt trunkeringer og MESH-termer. OR utvidet søket ved å gi treff på litteratur som inneholder enten det ene eller det andre søkeordet (Jamtvedt et al., 2015), som øker kvaliteten i søket ved å inkludere nærliggende eller synonyme begreper. Termen AND ble benyttet for å inkludere alle aspektene ved problemstillingen og for å presisere søket (Thomas et al., 2011).

Det endelige søket ble utført 21.06.2022 i samarbeid med kyndig bibliotekar ved UIB. For fullstendig oversikt over valg av søkeord og søkelogg se *vedlegg 1* og *vedlegg 2*.

Etter endt søk ble studiene gjennomgått etter tittel, dernest tittel/abstrakt og til sist fulltekst, for å vurdere om studien oppfylte nevnte inklusjons- og eksklusjonskriterier. Det ble

utarbeidet et flytskjema over inklusjons- og eksklusjonsprosessen etter endt søk, før alle inkluderte studier ble sjekklistervurdert. For å kritisk vurdere studiene ble «*Sjekkliste for vurdering av en studie som tester en ny diagnostisk test*» anvendt, hentet fra Helsebiblioteket (2018).

3.3 Vurdering av studienes kvalitet

I prosessen med å vurdere studier er det flere faktorer som må vektlegges, der blant annet kravene til metode må være oppfylt. I det følgende kapitlet vil det presenteres begreper knyttet til kvalitetsvurdering av studiene. Nevnte begreper vil senere anvendes i diskusjonskapitlet.

3.3.1 Validitet

Validitet betyr *gyldighet* og fungerer som et krav til kvaliteten av målinger. Validitet beskriver i hvilken grad en test måler det den er faktisk ment for å måle. Avhengig av hensikten med studien finnes det ulike typer validitet (Thomas et al., 2011). To former for validitet er *intern* og *ekstern* validitet (Dahlum, 2015).

3.3.2 Intern validitet

Intern validitet handler om resultatene fra undersøkelsene eller studiene er til å stole på (Jamtvedt et al., 2015). For å sikre høy intern validitet, er det vesentlig at man har kontroll over mulige bias som kan gi systematiske skjevheter i resultatene (Dahlum, 2015; Jamtvedt et al., 2015). Bias er viktige interne forhold å være oppmerksom på, da de fleste undersøkelser inneholder dette av en eller annen grad. Om bias forekommer, vil det kunne påvirke studienes interne validitet og materialet er ikke det det utgir seg for å være (Jørgensen et al., 2016). Den interne validiteten av kohortstudier kan trues av ulike kilder til systematiske feil, men de kan grovt inndeles i seleksjonsbias og informasjonsbias (Jørgensen et al., 2016). *Seleksjonsbias* kan defineres som systematiske feil som følge av forskjeller i karakteristika i utvalget som inkluderes eller eventuelt ikke inkluderes i en studie (Thomas et al., 2011). I vurderingen av potensielle seleksjonsbias vil studienes interne forhold, herunder studiedesign, rekruttering, utvalg og blinding vurderes. *Informasjonsbias* indikerer usikkerhet på hva man måler, ved at informasjonen som anvendes i en studie enten blir målt, beskrevet eller tolket feilaktig (Jørgensen et al., 2016; Thomas et al., 2011). Ulikheter i utførelse og tolkning av index-testene og/eller referansetestene kan resultere i ulike estimater av testens diagnostiske kvalitet. Det er derfor viktig at prosedyrene til målemetodene som blir anvendt blir detaljert beskrevet (Tacconelli, 2010). Om dette ikke foreligger kan det føre til informasjonsbias og dermed svekke studiens interne validitet (Jørgensen et al., 2016).

3.2.2.1 Blinding

Blinding er en måte å kontrollere en trussel til intern validitet i en studie slik at resultatet ikke påvirkes (Thomas et al., 2011). Deltaker, helsepersonell, forsker (utfallsmåler) og den som gjør analysene er alle kandidater for bias og bør blindes om mulig (Thomas et al., 2011). Hvis enten deltagere eller behandler er uvitende om hvilken behandling som gis i studien, kalles det enkeltblinding. Det kalles dobbeltblinding hvis begge er uvitende om behandlingen (Jørgensen et al., 2016). Begrepene enkelt- og dobbeltblinding brukes likevel ikke konsistent og er tvetydige dersom det ikke oppgis hvilke personer som er blindet (Thomas et al., 2011). Dobbeltblinding er vanskelig å få til for de fleste behandlingsformer eller undersøkelsesmetoder fysioterapeuter tilbyr, men utfallsmåler kan som regel blindes (Jamtvedt, 2015).

3.3.3 Ekstern validitet

Ekstern validitet omhandler *generaliserbarheten* eller *overføringsverdien* av resultatene som kommer frem i studien. Spørsmålet om resultatene som stammer fra et begrenset utvalg i studien kan genereres til en større populasjon, står svært sentralt (Dahlum, 2015). For å oppnå det man kaller et representativt utvalg, skal ideelt sett utvalget i studier tilsvare sammensetningen i populasjonen. Dette vil si et pasientutvalg med variasjon av sentrale kjennetegn, der det er heterogenitet av personkarakteristika som kan finnes i pasientpopulasjonen, som for eksempel alder og kjønn (Johannessen et al., 2021). I følge Tacconelli (2010) kan et utvalg av deltagere som *ikke* ligner på populasjonen som testen skal brukes på i praksis, lede til skjevheter i tolkningene av testresultatene og overestimere testenes diagnostiske kvaliteter. Et utvalg som *er* representativt kan medføre at testenes diagnostiske kvaliteter blir vurdert uten bias i forhold til utvalget. Dette kan igjen medføre at utvalget er en mindre trussel for resultatene og øke generaliserbarheten til en større populasjon (Tacconelli, 2010). I følge Thomas et al. (2011) kan et stort utvalg i seg selv lede til generaliserbare resultater, ved å øke studiens «power», eller styrke. Det er vanskelig å konkret skulle definere hvor stort utvalget skal være i diagnostiske kohortstudier for å kunne generalisere resultatene til en større populasjon, men det finnes teori som kan gi gode indikasjoner. Thomas et al. (2011) konkretiserer forholdet mellom utvalgsstørrelse og styrke, der det statueres at en utvalgsstørrelse på $n=100$ vil gi styrke på .95, som nærmest vil gi en sikkerhet på å finne reelle funn som kan generaliseres til en større populasjon. Om utvalgsstørrelsen er mindre, vil det gi suksessivt synkende styrke og således svekket generaliserbarhet av resultatene.

3.3.4 Reliabilitet

En målemetode eller test vil aldri kunne være valid hvis den ikke er reliabel (Thomas et al., 2011). Reliabilitet omhandler i forskningssammenheng reproduserbarheten av en måling, der det er ønskelig at målemetoden eller testen gir samme resultat ved gjentatte målinger. Det er imidlertid ikke realistisk å få nøyaktig samme resultat ved uendelig gjentatte målinger, og en viss variasjon må alltid godtas (Beyer & Magnusson, 2003). Reliabilitet kan undersøkes på flere måter; Intertester-reliabilitet undersøkes ved at flere testere utfører samme test, der man vurderer til hvilken grad det er enighet mellom testerne. Intratester-reliabilitet undersøkes ved at en tester utfører en test flere ganger og vurderer testens reproduserbarhet (Thomas et al., 2011).

3.4 Statistisk analyse

De primære utfallsmålene av interesse når man analyserer studier som undersøker nøyaktigheten av diagnostiske tester, er dataene man kan kalkulere ut av en 2*2 tabell (Tacconelli, 2010). En 2*2 tabell beskriver forholdet mellom resultatene fra index-testen og referansetesten, ved en gitt diagnostisk terskel (punktet der resultatet er klassifisert som positivt eller negativt) (Tacconelli, 2010). Ved å anvende tabellen kan man kalkulere følgende utfallsmål, som er vanlig for å analysere kvaliteten av diagnostiske tester. Se *figur 1* for 2*2 tabell og *figur 2* for utdypning av 2*2 tabell:

		Reference standard	
		Positive	Negative
Index test	Positive	TP	FP
	Negative	FN	TN

Figur 1. – 2*2 tabell, hentet fra Systematic reviews: CRD's guidance for undertaking reviews in health care, av Tacconelli, E, 2010, s 115.

Sensitivitet – andelen av deltagere som har målsykdommen som har et positivt testresultat

$$\frac{SP}{SP + FN}$$

Spesifisitet – andelen av deltagere uten målsykdommen som har et negativt testresultat

$$\frac{SN}{SN + FP}$$

Overall accuracy – andelen av deltagere som ble korrekt klassifisert av testen

$$\frac{SP + SN}{SP + FN + FP + SN}$$

Positiv prediktiv verdi – sannsynligheten for sykdom blant deltagere med positivt testresultat

$$\frac{SP}{SP + FP}$$

Negativ prediktiv verdi – sannsynligheten for fravær av sykdom blant deltagere med negativt testresultat

$$\frac{SN}{SN + FN}$$

Likelihood ratios – beskriver hvor mange ganger mer sannsynlig det er at en deltager med målsykdommen vil oppnå et bestemt testresultat, enn en deltager uten målsykdommen

PLR

$$\left(\frac{SP}{SP + FN} \right)$$

$$\left(\frac{FP}{FP + FN} \right)$$

NLR

$$\left(\frac{FN}{SP + FN} \right)$$

$$\left(\frac{SN}{SN + FN} \right)$$

Figur 2 – Utdypning av 2*2 tabell. Hentet fra Systematic reviews: CRD's guidance for undertaking reviews in health care, av Tacconelli, E, 2010, s 115-116.

3.4.1 Sensitivitet, spesifisitet, negativ og positiv likelihood ratio som utfallsmål

For å kunne vurdere ULNT som diagnostisk redskap ved utredning av CR, er det nødvendig å definere i hvilken grad sensitivitet, spesifisitet, samt positiv og negativ likelihood ratio sier noe om den diagnostiske verdien. Herunder følger en gradering av nevnte utfallsmål, da det er disse utfallsmålene som blir anvendt til å vurdere resultatene i denne masteroppgaven.

Sensitivitet som utfallsmål rangerer fra 0.00-1.00, der en sensitivitet på 1.00 vil fange opp alle «sant syke» med et positivt testresultat og således være den optimale sensitiviteten av en test (Jamtvedt, 2015). Dette vil medføre svært høy diagnostisk verdi der alle med målsykdommen blir fanget opp. Dersom sensitiviteten av en test synker, vil det bli økende grad av feilklassifisering mellom «sant syke» og «sant friske» individer, og den diagnostiske verdien blir dermed suksessivt lavere i takt med synkende grad av sensitivitet. I denne oppgaven vil $S_n = 0.00-0.50$ vurderes til lav diagnostisk verdi, $S_n = 0.50-0.75$ vurderes til moderat diagnostisk verdi og $S_n = 0.75-1.00$ vurderes til høy diagnostisk verdi.

Spesifisitet rangerer også fra 0.00-1.00, der en spesifisitet på 1.00 vil fange opp alle «sant friske» individer med et negativt testresultat, og være den optimale spesifisiteten av en test, med svært høy diagnostisk verdi (Jamtvedt, 2015). På samme måte vil også synkende grad av spesifisitet, øke feilklassiferingen mellom «sant friske» og «sant syke», der den diagnostiske verdien faller i takt med spesifisiteten. Graderingen av den diagnostiske verdien av spesifisitet tolkes i denne oppgaven som: $S_p = 0.00-0.50$ vurderes til lav diagnostisk verdi, $S_p = 0.50-0.75$ vurderes til moderat diagnostisk verdi og $S_p = 0.75-1.00$ vurderes til høy diagnostisk verdi.

Likelihood ratios er et utfallsmål som kombinerer sensitivitets- og spesifisitetsdatene av en spesifikk test (Tacconelli, 2010). Positiv likelihood ratio (LR+) er den andelen av individer med en positiv test, som faktisk har målsykdommen. Negativ likelihood ratio (LR-) er den andelen av individer med en negativ test, som faktisk ikke har målsykdommen. LR+ over 10 og LR- mindre enn 0.1, er vurdert til å være overbevisende diagnostisk bevis (Tacconelli, 2010).

3.5 Etske betraktninger

Ettersom forskningsdesignet er en systematisk litteraturgjennomgang, er det begrenset med etiske betraktninger som må hensyntas. Eventuelle etiske betraktninger under gjennomføring av søkeprosessen ble diskutert og hensyntatt sammen med veileder og kyndig bibliotekar. Det ble ikke innhentet organisatorisk data eller pasientdata, som medfører at søknader til instanser som Regionale komiteer for medisinsk og helsefaglig forskningsetikk (REK) eller Norsk senter for forskningsdata (NSD) ikke var nødvendig. Av andre etiske hensyn etterstrebes det at forskningen utøves hederlig, uten noen form for plagiering, eller tolkninger uten tydelig referansebruk.

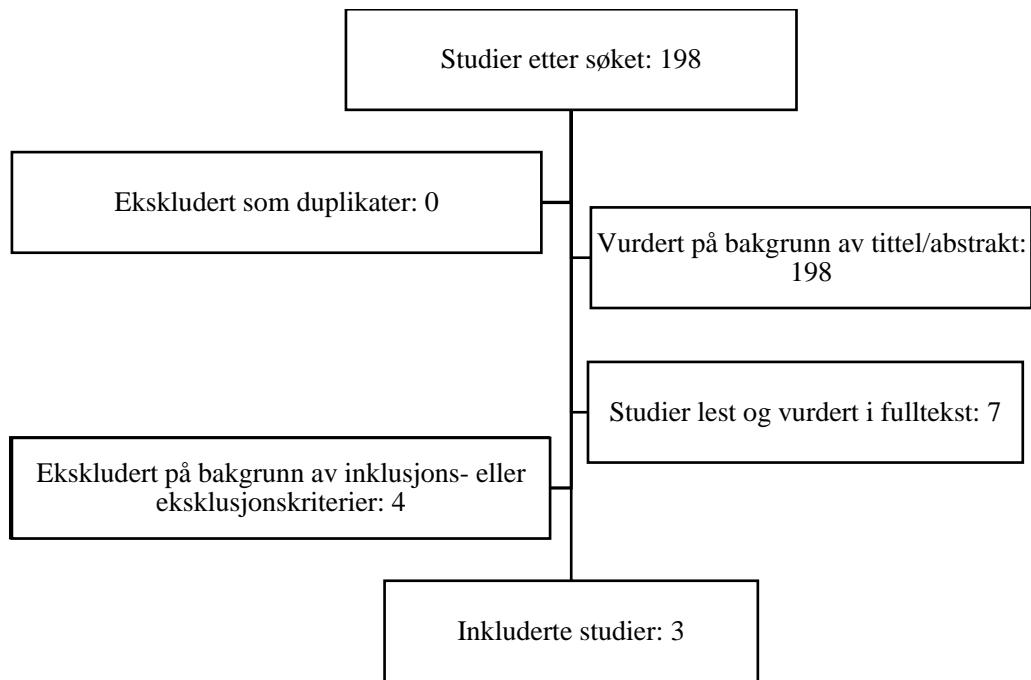
4.0 Resultater

I kapittel 4.1 og 4.2 blir valg av studier med flytskjema og en oversikt av ekskluderte og inkluderte studier presentert. Deretter vil de inkluderte studiene bli utdypende presentert, der tabell V i *kapittel 4.3* fremstiller de tre studienes viktigste innhold og resultater. De påfølgende tabellene gir en utdypende presentasjon av resultatene som er relevant for problemstillingen i denne litteraturstudien, basert på utfallsmålene sensitivitet, spesifisitet, samt positiv og negativ likelihood ratio. Deretter følger sjekklister vurdering av studienes kvalitet. For å diskutere utfallsmål som direkte måler testen diagnostiske verdi, var det ønskelig å sammenligne utfallsmål som ble brukt i alle de inkluderte studiene. Alle studiene har anvendt sensitivitet, spesifisitet, samt positiv og negativ likelihood ratio som utfallsmål, bortsett fra S2 som i stedet anvender positiv og negativ prediktiv verdi. For S2 har undertegnede regnet ut positiv og negativ likelihood ratio, etter formlene i *figur 2*. Dette for å skape et sammenligningsgrunnlag basert på like kriterier.

4.1 Valg av studier

Etter endt søk ble det funnet totalt 198 studier (se vedlegg 1). Deretter startet prosessen med å gjennomlese studiene for å fullføre inklusjonsprosessen. Totalt 191 studier ble ekskludert på bakgrunn av tittel og abstrakt. Etter denne prosessen var det totalt 7 studier som ble lest i fulltekst og vurdert i henhold til inklusjons- og eksklusjonskriteriene for denne oppgaven. Fire studier ble ekskludert, mens tre av studiene møtte de beskrevne kriteriene for inklusjon.

4.1.1 Flytskjema over inkluderte og ekskluderte studier



4.2 Presentasjon av ekskluderte og inkluderte artikler

Tabell III – Oversikt over ekskluderte studier etter lesning i fulltekst

Forfattere	Koulidis et al. (2019)	Lemeunier et al. (2017)	Nee and Butler (2006)	Thoomes et al. (2018)
Navn på studie	Diagnostic accuracy of upper limb neurodynamic tests for the assessment of peripheral neuropathic pain: A systematic review	Reliability and validity of clinical tests to assess the anatomical integrity of the cervical spine in adults with neck pain and its associated disorders: Part 1-A systematic review from the Cervical Assessment and Diagnosis Research Evaluation (CADRE) Collaboration	The validity of upper-limb neurodynamic tests for detecting peripheral neuropathic pain	Value of physical tests in diagnosing cervical radiculopathy: a systematic review
Publisert	Musculoskelet Sci Pract 2019 Vol. 40 Pages 21-33	Eur Spine J 2017 Vol. 26 Issue 9 Pages 2225-2241	J Orthop Sports Phys Ther 2012 Vol. 42 Issue 5 Pages 413-24	Spine J 2018 Vol. 18 Issue 1 Pages 179-189
Design	Systematisk oversiktsartikkel	Systematisk oversiktsartikkel	Systematisk oversiktsartikkel	Systematisk oversiktsartikkel
Inkludert	Nei	Nei	Nei	Nei
Kommentar	Møter ikke inklusjonskriteriet om studiedesign	Møter ikke inklusjonskriteriet om studiedesign	Møter ikke inklusjonskriteriet om studiedesign	Møter ikke inklusjonskriteriet om studiedesign

Tabell IV – Oversikt over inkluderte studier etter lesning i fulltekst

Forfattere	Wainner et al. (2003)	Apelby-Albrecht et al. (2013)	Grondin et al. (2021)
Navn på studie	Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy	Concordance of upper limb neurodynamic tests with medical examination and magnetic resonance imaging in patients with cervical radiculopathy: a diagnostic cohort study	Diagnostic accuracy of upper limb neurodynamic tests in the diagnosis of cervical radiculopathy
Publisert	Spine (Phila Pa) 2003 Vol. 28 Issue 1 Pages 52-62	J Manipulative Physiol Ther 2013 Vol. 36 Issue 9 Pages 626-32	Musculoskelet Sci Pract 2021 Vol. 55 Pages 102427
Design	Blindet, prospektiv diagnostisk teststudie	Diagnostisk kohortstudie	Diagnostisk kohortstudie
Referert standard	EMG	MR	MR
Inkludert	Ja	Ja	Ja

4.3 Utdypende presentasjon av inkluderte studier

For enklere lesning er navnene til de inkluderte studiene forkortet. Wainner et al. (2003)

navngis som *S1*, Apelby-Albrecht et al. (2013) navngis *S2* og Grondin et al. (2021) navngis som *S3*.

Tabell V – Fremstilling av de inkluderte studienes viktigste innhold og resultater

Studie	S1	S2	S3
Tittel	Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy	Concordance of upper limb neurodynamic tests with medical examination and magnetic resonance imaging in patients with cervical radiculopathy: a diagnostic cohort study	Diagnostic accuracy of upper limb neurodynamic tests in the diagnosis of cervical radiculopathy
Hensikt	Å vurdere nøyaktigheten av individuell klinisk undersøkelse og selvrapporing for diagnosen cervikal radikulopati, samt identifisere og vurdere nøyaktigheten av et optimalt test-cluster for cervikal radikulopati.	Å undersøke samsvar mellom ULNT og en utvalgt referert standard, bestående av medisinsk undersøkelse og MR, hos pasienter med cervikal radikulopati.	Å avgjøre den diagnostiske nøyaktigheten til enkeltstående og kombinerte ULNT (1, 2a,2b og 3) for cervikal radikulopati, ved å bruke funn fra tester som er like de som brukes i klinisk praksis.
Design	En blindet prospektiv diagnostisk test studie	Diagnostisk kohortstudie	Diagnostisk prospektiv kohortstudie
Inklusjonskriterier	-Mistenkt cervikal radikulopati (CR) eller carpal tunnel syndrom (CTS). -Kun pasienter som hadde funn som antydte disse patologiene ved elektromyografi ble vurdert for inklusjon.	-Nakkesmerter med opplevelse av armsmerter.	-18-65 års alder. - Armsmerter med eller uten nakkesmerter i minst 3 måneder. - Minimum 30mm og maksimum 80mm på Visual Analog Scale (VAS).
Eksklusjonskriterier	-Systemiske sykdommer som kan lede til perifere nevropatier -Bilaterale symptomer -Tidligere skade/sykdom av den symptomatiske armen som nedsetter funksjon i stor grad	- Multitraume som involverer ryggraden -Malignitet av ryggraden -Systemiske sykdommer som kan disponere for nevropatier - Generelle sykdommer (fysiske/psykiske) som	-Om deltageren ikke forsto Fransk -Traume av cervicalcolumna -Tidligere nakke- eller armkirurgi -Revmatiske sykdommer -Fibromyalgi -Diabetes -Gravide - Kardiovaskulær/nevrologisk/

	<p>-Ute av stand til å jobbe i mer enn 6 måneder på grunn av tilstanden</p> <p>-Tidligere kirurgi for tilstander som kan gi nakkesmerter eller CTS</p> <p>-Tidligere gjennomgått elektrofysiologiske målinger av symptomatisk arm, med mistanke om CR eller CTS</p> <p>- Mottatt sykepenger eller pågående rettssak</p>	kan påvirke testresultatet	<p>neoplastisk/psykiatrisk sykdom</p> <p>-Cervikal myelopati</p> <p>-Pyramidal eller ekstrapyramidal sykdom</p>
Utvalg	<p>N= 82</p> <p>Kvinner= 41</p> <p>Menn= 41</p> <p>Gjennomsnittlig alder= 45 år ± 12 år</p>	<p>Vurdert for inklusjon= 58</p> <p>Ekskludert= 7</p> <p>N= 51</p> <p>Kvinner= 27</p> <p>Menn= 24</p> <p>Gjennomsnittlig alder= 51år</p>	<p>Vurdert for inklusjon= 109</p> <p>Ekskludert= 25</p> <p>N= 85</p> <p>Kvinner= Ikke oppgitt</p> <p>Menn= Ikke oppgitt</p> <p>Gjennomsnittlig alder= 43.9</p>
Utfallsmål	<p>Sensitivitet</p> <p>Spesifisitet</p> <p>Positiv likelihood ratio</p> <p>Negativ likelihood ratio</p> <p>95% konfidensintervall</p> <p>Kappa (k)</p>	<p>Sensitivitet</p> <p>Spesifisitet</p> <p>Positiv likelihood ratio</p> <p>Negativ likelihood ratio</p> <p>ROC</p> <p>Kappa (k)</p>	<p>Sensitivitet</p> <p>Spesifisitet</p> <p>Positiv likelihood ratio</p> <p>Negativ likelihood ratio</p> <p>Post-test sannsynlighet med 95% konfidensintervall</p>
Referert standard	EMG	MR	MR
Blinding	<p>Utfallsmåler av de kliniske testene var blindet for EMG-resultatene. Det ble opprettet en standardisert prosedyre for teknisk utførelse av kliniske tester, som alle utfallsmålere fulgte. Testene ble utført 15-30 minutter etter EMG-undersøkelsen.</p>	<p>Pasientene ble ved inklusjon undersøkt av to erfarne nevrokirurger, som satte diagnoser CR basert på nevrologisk undersøkelse og MR-funn. Utfallsmålere som utførte index-testene, samt deltagerne i studien, var blindet for funnene til nevrokirurgene. Index-testene ble utført cirka 60 minutter etter den initiale undersøkelsen utført av nevrokirurgene. Det var totalt to utfallsmålere som utførte index-testene, etter en standardisert</p>	<p>Pasientene ble ved inklusjon undersøkt av en nevrokirurg med 15 års erfaring, som satte diagnosen CR eller annen diagnose, basert på klinisk undersøkelse og MR-funn. Utfallsmåler (fysioterapeut med 10 års erfaring på nakkesmerter og avansert kunnskap innen ortopedi) som utførte index-testene var blindet for disse funnene, samt anamnetiske opplysninger vedrørende deltagerne. Index-testene ble utført cirka 60 minutter etter den initiale undersøkelsen, og ingen intervensjoner var tiltatt mellom disse to undersøkelsene.</p>

		prosedyre som ble utarbeidet i forkant.	
Oppsummering av relevante funn	Flere av de undersøkte testene ble vurdert til å ha akseptabel grad av nøyaktighet for diagnosen cervikal radikulopati, sammenlignet med referert standard. Et sparsommelig test-cluster bestående av enkelttester fra den kliniske undersøkelsen produserte større post-test sannsynlighet, enn noen enkeltstående test alene.	Upper limb neurodynamic test, utført i et cluster av erfarne manuelle terapeuter, viser betydelig samsvar med funn fra nevrologisk undersøkelse og MR-funn, utført av nevrokirurger. Resultatene indikerer viktigheten av å inkludere ULNT for å komplementere den kliniske undersøkelsen hos pasienter med cervikal radikulopati	Resultatene støtter tidligere forskning som indikerer at bruken av en enkeltstående ULNT for å inkludere eller ekskludere CR, ikke er anbefalt. Når de brukes i kombinasjon, har testene høyere klinisk nytteverdi. Når alle ULNT er negative, kan CR ekskluderes. Når $\frac{3}{4}$ ULNT er positive, kan CR inkluderes. Således, anbefales det kun å bruke alle ULNT kombinert, som et test-cluster.

4.4 Utdyping av studienes resultater

Herunder følger en utdyping av studienes resultater som er relevante for oppgavens problemstilling, som vil si presentasjon av utfallsmålene sensitivitet, spesifisitet, positiv og negativ likelihood ratio.

4.4.1 Resultater studie 1

Herunder følger resultatene fra S1 presentert i tabellformat. Kun resultater som omhandler ULNT er inkludert, med bakgrunn i oppgavens problemstilling.

Tabell VI – Resultater S1

Test	Sn (95 KI)	Sp (95 KI)	LR- (95 KI)	LR+ (95 KI)
ULNT 2a	0.97 (0.90–1.0)	0.22 (0.12–0.33)	0.12 (0.01–1.9)	1.3 (1.1–1.5)
ULNT 2b	0.72 (0.52–0.93)	0.33 (0.21–0.45)	0.85 (0.37–1.9)	1.1 (0.77–1.5)

Sn = Sensitivitet =, Sp = Spesifisitet =, LR- = Negativ likelihood ratio, LR+ = Positiv likelihood ratio, 95 KI = 95% konfidensiellintervall

4.4.2 Resultater studie 2

Herunder følger resultatene fra S2 presentert i tabellformat. Den første raden i tabellen er resultater fra test-clusteret, mens de påfølgende radene er resultater fra hver enkelt ULNT.

Tabell VII – Resultater S2

Test	Sn (95 KI)	Sp (95 KI)	LR+*	LR-**
ULNT - cluster	0.97 (0.85-1.00)	0.69 (0.41-0.89)	3.12	0.28
ULNT 1	0.83 (0.66-0.93)	0.75 (0.48-0.93)	3.32	0.30
ULNT 2a	0.66 (0.48-0.81)	0.75 (0.48-0.93)	2.64	0.37
ULNT 2b	0.43 (0.26-0.61)	0.75 (0.48-0.93)	1.72	0.58
ULNT 3	0.71 (0.54-0.85)	0.87 (0.62-0.98)	5.46	0.18

*Sn = Sensitivitet =, Sp = Spesifisitet =, LR- = Negativ likelihood ratio, LR+ = Positiv likelihood ratio, 95 KI = 95% konfidensiellintervall * = Sensitivitet / (1 - Spesifisitet). ** = (1 - Spesifisitet) / Sensitivitet.*

4.4.3 Resultater studie 3

S3 gjør i likhet med S2 kalkulering av diagnostisk nøyaktighet både for hver enkelt ULNT og vurderer nøyaktigheten når testene blir utført som et cluster. *Tabell VIII* viser funn etter utfallsmål for de enkeltstående ULNT, mens *tabell IX* beskriver funn når testene er utført som et cluster.

Tabell VIII – Diagnostisk nøyaktighet av enkeltstående ULNT

Test	Sn	Sp	LR+ (95 KI)	LR- (95 KI)
ULNT 1	0.59 (0.38 - 0.77.)	0.75 (0.62 - 0.86)	2.46 (1.41 – 4.27)	0.54 (0.33 – 0.87)
ULNT 2a	0.70 (0.49 - 0.86)	0.72 (0.59 - 0.83)	2.55 (1.57 – 4.14)	0.41 (0.22 – 0.75)
ULNT 2b	0.55 (0.35 - 0.74)	0.75 (0.62 - 0.86)	2.30 (1.30 – 4.06)	0.59 (0.38 – 0.92)
ULNT 3	0.40 (0.22 - 0.61)	0.93 (0.83 - 0.98)	5.91 (2.07 - 16.87)	0.64 (0.46 – 0.88)

Sn = Sensitivitet =, Sp = Spesifisitet =, LR- = Negativ likelihood ratio, LR+ = Positiv likelihood ratio, 95 KI = 95% konfidensiellintervall

Tabell IX – Diagnostisk nøyaktighet av test-cluster

Test	Sn	Sp	LR+ (95 KI)	LR- (95 KI)
1/4 positive	0.96 (0.81 - 0.99)	0.46 (0.33 - 0.60)	1.80 (1.40 - 2.32)	0.08 (0.01 - 0.56)
2/4 positive	0.85 (0.66 - 0.95)	0.74 (0.60 - 0.84)	3.29 (2.07 - 5.23)	0.20 (0.08 - 0.50)
3/4 positive	0.44 (0.25 - 0.64)	0.96 (0.88 - 0.99)	12.89 (3.10 - 53.62)	0.58 (0.41 - 0.81)
4/4 positive	0.11 (0.23 - 0.29)	1.0 (0.93 - 1.0)	Inf.	0.89 (0.78 - 1.02)

Sn = Sensitivitet =, Sp = Spesifisitet =, LR- = Negativ likelihood ratio, LR+ = Positiv likelihood ratio, 95 KI = 95% konfidensiellintervall, Inf. = Infinity (uendelig).

4.5 Sjekkliste for vurdering av en studie som tester en ny diagnostisk test

Sjekkliste for vurdering av en studie som tester en ny diagnostisk test	S1	S2	S3
1. Er formålet med studien klart formulert?	Ja	Ja	Ja
2. Var det en hensiktsmessig sammenlikning mellom den nye testen og en egnet referansetest?	Ja, Index-test versus EMG.	Ja, Index-test versus MR.	Ja, Index-test versus MR.
3. Ble både den nye testen og referansetesten brukt på alle pasientene?	Ja	Ja	Ja
4. Ble det utført en uavhengig, blindet sammenlikning av den nye testen og referansetesten?	Ja, utfallsmåler blindet for EMG resultater.	Ja, utfallsmåler blindet for MR-funn.	Ja, utfallsmåler blindet for MR-funn.
5. Ble testen utprøvd på et pasientutvalg som ligner tilstrekkelig på populasjonen testen skal brukes på?	Usikkert, se <i>kapittel 5.1.1</i>	Usikkert, se <i>kapittel 5.1.1</i>	Usikkert, se <i>kapittel 5.1.1</i>
6. Er testprosedyrene detaljert nok beskrevet til at de kan gjentas andre steder?	Ja	Ja	Ja

7. Hva er resultatene?	<i>Se kapittel 4.0 - Resultater</i>	<i>Se kapittel 4.0 - Resultater</i>	<i>Se kapittel 4.0 - Resultater</i>
8. Hvor presise er resultatene?	<i>Se kapittel 5.0 - Diskusjon</i>	<i>Se kapittel 5.0 - Diskusjon</i>	<i>Se kapittel 5.0 - Diskusjon</i>
9. Er resultatene relevante for dine pasienter?	Ja	Ja	Ja
10. Kan testen benyttes på ditt arbeidssted?	Ja	Ja	Ja
11. Kan testen benyttes på dine pasienter?	Ja	Ja	Ja
12. Vil dine pasienter ha nytte av denne testen?	<i>Ja, se kapittel 6.0</i>	<i>Ja, se kapittel 6.0</i>	<i>Ja, se kapittel 6.0</i>

5.0 Diskusjon

Det følgende kapittelet er inndelt i en metode- og resultatdiskusjon. I metodediskusjonen drøftes de inkluderte studienes interne og eksterne validitet. Resultatdiskusjonen omhandler resultatene, og hva som kan ha påvirket disse. Videre i oppgaven vil det diskuteres om test-clustere kan bedre den diagnostiske verdien av ULNT og hva slags implikasjoner funnene fra denne litteraturstudien gir for praksis. Avslutningsvis redegjøres det for styrker og svakheter i denne masteroppgaven.

5.1 Metodediskusjon av inkluderte studier

5.1.1 Intern validitet

5.1.1.1 Studiedesign og rekruttering

Det er særlig interessant å diskutere eventuelle seleksjonsbias, når man skal vurdere studienes design og rekrutteringsprosess. Om rekrutteringsprosessen ikke er utført på en tilfredsstillende måte i henhold til studiedesignet, kan dette svekke studienes interne validitet (Tacconelli, 2010).

S1 rekrutterte deltagere mellom desember 1998 og april 2000, fra fire ulike medisinske sentre i USA. Pasientene var henvist til elektrofysiologisk avdeling, med mistanke om CR eller carpal tunnel syndrom (CTS). Kun de pasientene som viste tegn til CR eller CTS ved elektrofysiologiske målinger, ble vurdert som kvalifisert til å delta i S1. Referansetestene ble utført av sertifisert personell, før index-testene ble utført av to ulike utfallsmålere etter standardisert prosedyre, 15-30 minutter etter at referansetestene ble utført. S1 inkluderte dermed deltagere som hadde sykehistorie, symptomer og elektrofysiologiske funn som antydte CR eller CTS, *før* index-testene ble utført. En slik rekruttering møter ikke kriteriene til single-gate designet. Dette fordi S1 inkluderer deltagere og etablerer et utvalg med *kjent* sykdomsstatus, *før* de utfører og sammenligner resultater fra referanse- og index-tester. I tillegg kan det å vurdere index-testene ut ifra to antatte patologier, ikke dekke hele spekteret for hvordan testene vil bli anvendt for diagnostikk i klinisk praksis, som kan være en svakhet i studiedesignet (Tacconelli, 2010). Denne formen for seleksjonsbias kan videre gi systematiske feil i tolkningen av resultatene fra studien (Jørgensen et al., 2016), som vil diskuteres nærmere i resultatdiskusjonen.

S2 rekrutterte deltagere i perioden november 2007 til februar 2008. Samtlige deltagere ble henvist av fastlege til nevrokirurg. Alle pasienter som ble henvist hadde utført MR i løpet av de siste 6 måneder. Inklusjonskriteriet var nakkesmerter med opplevelse av armsmerter. Den valgte undersøkelsen i S2 var basert på anamnese, klinisk undersøkelse og referansetest med

MR-funn, som ble utført av to spesialister i nevrokirurgi. Denne prosessen ble standardisert før oppstart av studien. Index-testene ble utført cirka 60 minutter etter endt referanseundersøkelse av to forskjellige utfallsmålere. Utførelsen av index-testene ble også standardisert før oppstart av studien. Deltagerne i S3 ble henvist fra sin fastlege, eller spesialist, til nevrokirurgisk avdeling. Rekrutteringsprosessen fant sted mellom september 2017 til september 2019. Alle pasientene hadde mistenkt nakkelidelse. For å bli inkludert i studien måtte deltagerne være mellom 18-65 år og ha armsmerter med eller uten nakkesmerter med en varighet på minst 3 måneder. Index-testene ble utført av én enkelt utfallsmåler. På én side, kan det tolkes at rekrutteringsprosessen i S2 og S3 også legger føringer for at det inkluderes et utvalg med kjent sykdomsstatus. Likevel vurderes det at denne formen for seleksjonsbias er mindre fremtredende i S2 og S3, sammenlignet med S1. Dette fordi man ønsker å etablere et utvalg som faktisk er representativt i forhold til utvalget som index-testene skal anvendes på i klinisk praksis, jamfør definisjonen og typiske symptomkarakteristika av CR (Jamtvedt, 2015; Rubinstein et al., 2007). Dette kan være en metodisk styrke i både S2 og S3.

De beskrevne seleksjonsbiasene kan være en svakhet i studienes metode, ved at de øker risikoen for systematiske skjevheter av resultatene og dermed svekker den interne validiteten. Dette fordi seleksjonsbias hvor «alle» deltakere har den aktuelle diagnosen kan føre med seg at det blir få i gruppene med «negative funn». Dette fører igjen til små tall i negativ-funn boksene i 2x2 tabellen og dermed svekkes utregningsgrunnlaget for spesifisitet, og på den annen side vil utregningsgrunnlaget for sensitivitet styrkes. Dette kan samlet sett lede til feilestimering av testenes diagnostiske verdi. Det vurderes at denne formen for bias er av klart størst betydning i S1, med bakgrunn i den overnevnte diskusjonen. I følge Thomas et al. (2011) finnes det virkemidler som kan anvendes for å redusere konsekvensene av seleksjonsbias, der blinding står sentralt.

5.1.1.2 Blinding

Når studiene omhandler diagnostikk blir det særlig viktig at utfallsmåler fra de ulike målemetodene blir blindet for hverandres funn (Tacconelli, 2010). Samtlige inkluderte studier i denne oppgaven beskriver at det ble gjort tiltak i forhold til blinding. I både S1, S2 og S3 ble referansetester og index-tester utført av ulike utfallsmålere, som var blindet for hverandres resultater. Dette er en form for dobbel-blinding der utfallsmålerne fra begge undersøkelsesmetodene er blindet for hverandres funn (Jørgensen et al., 2016). Dette vil styrke studiens interne validitet, ved å redusere faren for at resultatene påvirkes som en

konsekvens av manglende blinding (Thomas et al., 2011). I tillegg var utfallsmålerne som utførte index-testene blindet for målsykdommen, amnestiske opplysninger, andre kliniske undersøkelser eller funn fra EMG eller MR. Dette kan medføre at eventuelle seleksjonsbias fra rekrutteringsprosessen ble av mindre betydning, og kan således være en styrke til den interne validiteten av både S1, S2 og S3 (Thomas et al., 2011).

5.1.1.3 Dokumentasjon og utførelse av tester

S1 navngir hvilke kliniske undersøkelser de har tatt i bruk som sine index-tester og refererer videre til et vedlegg for nærmere beskrivelse av testene. Her dokumenteres det tydelig for hvordan index-testene har blitt utført. Studien dokumenter også godt for hvordan referansetestene har blitt utført, der protokollene er svært detaljert og ettertrykkelig dokumentert flytende i teksten. S2 beskriver utførelsen av de inkluderte index-testene i tabellformat. Referansetestene blir beskrevet fortløpende i det metodiske kapittelet. S3 beskriver index-testene fortløpende i tekst, og henviser til primærkilde i henhold til teknisk utførelse av testene. De beskriver i tillegg utførelsen som punktformat, med bilder av hver enkelt sekvens i testen utførelse. S3 beskriver også tydelig hvilke kriterier som ble vektlagt for å etablere diagnosen CR under referansetestene. Samlet ser man at alle inkluderte studier dokumenterer index-tester og kriteriene for referansetestene nøyaktig. Utførelsen av testene er standardisert og veldokumentert. Dette vil ifølge Jørgensen et al. (2016) øke sannsynligheten for at man måler det man ønsker å måle og vil dermed kunne redusere risikoen for informasjonsbias og således styrke studienes interne validitet.

Videre ser man at S1 og S2 har to ulike utfallsmålere, mens S3 har én enkelt utfallsmåler for index-testene. Av dette kan man tolke at intertester-reliabiliteten er høyere i S1 og S2, sammenliknet med S3, der tolkning av index-testene beror på vurderingen hos én enkelt person i S3. Det fremkommer heller ikke av S3 hvorvidt utfallsmåleren utførte index-testene flere ganger, som gir usikkerhet rundt intratester-reliabiliteten av index-testene. Dette kan være en metodisk svakhet i S3, hvor man kan diskutere om resultatene i forskningen er på bakgrunn av så objektive funn som mulig, eller er et resultat av utfallsmåleren subjektive holdninger, forventninger og meninger.

5.1.2 Ekstern validitet

Ekstern validitet handler, som nevnt i metodekapitlet, om generaliserbarheten av resultatene til en studie (Thomas et al., 2011).

S1 kan etter data fra Thomas et al. (2011) vurderes å ha et tilfredsstillende stort utvalg med 82 deltagere. Videre inkluderte studien deltagere fra fire ulike medisinske sentre, som kan sørge

for et bredt utvalg gjennom spredning i rekrutteringsprosessen. Kjønnfordelingen var henholdsvis 41 kvinner og 41 menn, med en alder på 45 år (± 12 år). Deltagerne kan med bakgrunn i disse kriteriene anses som et heterogent utvalg (Johannessen et al., 2021). Dette, sammen med utvalgsstørrelsen, kan bidra til å øke studiens styrke og således generaliserbarheten av resultatene til en større populasjon (Thomas et al., 2011). Likevel ser man som tidligere diskutert at S1 snevret inn inklusjonen av deltagere på bakgrunn av to antatte patologier. En slik seleksjonsbias gjør utvalget mer homogent (Johannessen et al., 2021) og kan svekke studiens eksterne validitet der resultatene i mindre grad vil kunne generaliseres til en større og mer heterogen populasjon der eksempelvis andre patologier kan være til stede (Thomas et al., 2011).

S2 hadde et utvalg på 52 personer, bestående av 27 kvinner og 24 menn, med en alder på 25-67 år. Studien inkluderte deltagere fortløpende ettersom de ble henvist til nevrokirurgisk avdeling fra sin fastlege. Utvalgsstørrelsen i S2 kan implisere at resultatene fra studiene er mindre generaliserbare enn i S1, grunnet svekket styrke (Thomas et al., 2011). På den annen side kan den jevne fordelingen mellom kvinner og menn, bred aldersfordeling, samt inkluderingsprosedyren vedrørende henvisning fra ulike fastleger styrke studiens eksterne validitet, da disse parameterne kan gi et heterogent utvalg (Johannessen et al., 2021). I tillegg hadde S2 «nakkesmerter med opplevelser av armsmerter» som inklusjonskriterie, noe Koulidis et al. (2019) statuerer å være en av de vanligste symptombeskrivelsene hos pasientene med CR. Utvalget i S3 er på totalt 85 personer, med en gjennomsnittsalder på 43.9 år. Det oppgis ikke variasjon mellom kvinner og menn. Dette kan således være et representativt utvalg i forhold til størrelse, men mangelen på informasjon rundt kjønn gjør at det blir vanskelig å generalisere funnene på bakgrunn av kjønn. Inkluderingsprosedyren rundt rekruttering av deltagere forøvrig, er lik den i S2. S3 har også rekruttert deltagere med antatt nakkepatologi og har et inklusjonskriterie som sier «armsmerter med eller uten nakkesmerter med en varighet på minst 3 måneder». Inklusjonskriteriet om nakke- og armsmerter i S2 og S3 vil kunne bidra til at det rekrutteres deltagere som har symptomer som ligner på de personene som testene sannsynligvis vil bli anvendt på i klinisk praksis, som vil styrke studiens eksterne validitet. Et slikt inklusjonskriterie kan dog øke risikoen for det etableres et mer homogent utvalg, som gjør at testenes utføres på bakgrunn av antatt patologi, som kanskje ikke dekker hele spekteret for hvordan testen vil anvendes i klinisk praksis (Johannessen et al., 2021; Thomas et al., 2011). Likevel vurderes dette samlet sett som en styrke for den eksterne validiteten av S2 og S3.

5.2 Resultatdiskusjon

5.2.1 Hva forteller resultatene?

Herunder følger en vurdering av index-testenes diagnostiske verdi, basert på *samlede* resultater fra de inkluderte studiene, fra laveste til høyeste verdi. Testene vil bli vurdert til å ha *lav*, *moderat* eller *høy* diagnostisk verdi blant utfallsmålene sensitivitet, spesifisitet, samt positiv og negativ likelihood ratio, som forklart i *kapittel 3.4.1*.

ULNT 1 viser følgende variasjon av utfallsmål blant de inkluderte studiene; Sn= 0.59-0.83. Sp= 0.75. LR+ = 2.46-3.32. LR- = 0.54-0.30. ULNT 1 tolkes dermed til å ha moderat til høy diagnostisk verdi i henhold til Sn og Sp. Data vedrørende LR+ og LR- er ikke tilfredsstillende og tolkes til lav diagnostisk verdi.

ULNT 2a viser følgende variasjon av utfallsmål blant de inkluderte studiene; Sn= 0.66-0.97. Sp= 0.22-0.75. LR+ = 1.3-2.55. LR- = 0.41-0.12. ULNT 2a tolkes til å moderat til høy diagnostisk verdi hva gjelder Sn og Sp. LR+ og LR- tolkes til lav diagnostisk verdi.

ULNT 2b viser følgende utfallsmål; Sn= 0.43-0.72. Sp= 0.33-0.75. LR+ = 1.1-2.30. LR- = 0.85-0.58. ULNT 2b tolkes til lav til moderat diagnostisk verdi vedrørende Sn og Sp. LR + og LR – tolkes til å ha lav diagnostisk verdi.

ULNT 3 viser følgende utfallsmål; Sn= 0.4-0.71. Sp= 0.87-0.93. LR+ = 5.46-5.91. LR- = 0.64-0.18. ULNT 3 tolkes til å ha lav til moderat diagnostisk verdi i henhold Sn. Det vurderes at Sp er av høy diagnostisk verdi. LR + og LR – tolkes til lav diagnostisk verdi.

5.2.1.1 Fremstilling av de enkeltstående ULNT med høyest diagnostisk verdi

Med bakgrunn i problemstillingen vil de enkeltstående ULNT som er vurdert til høyest diagnostisk verdi og som videre gir klinikerer mest informasjon ved undersøkelse av CR fremstilles i *tabell X*.

Tabell X – Spredning i sensitivitet, spesifisitet, positiv og negativ likelihood ratio for ULNT 2a og ULNT 3

Test	ULNT 2a (n.medianus)	ULNT 3 (n.ulnaris)
Sensitivitet	(0.66-0.97)	(0.4-0.71)
Spesifisitet	(0.22-0.75)	(0.87-0.93)
Positiv likelihood ratio	(1.3-2.55)	(5.46-5.91)
Negativ likelihood ratio	(0.41-0.12)	(0.64-0.18)

Disse to ULNT viser samlet høyest diagnostisk verdi, som indikerer at anvendelsen av disse testene i størst grad vil kunne diskriminere mellom syke og friske, og dermed hjelpe klinikeren til å bekrefte eller avkrefte CR (Tacconelli, 2010). Basert på disse resultatene vil ULNT 2a best kunne *bekrefte* CR ved et positivt testresultat, med $S_n = 0.66-0.97$ og $LR_+ = 1.3-2.55$, mens ULNT 3 vil i størst grad kunne *avkrefte* CR ved et negativt testresultat med $S_p = 0.87-0.93$, samt $LR_- = 0.64-0.18$. Imidlertid er verdiene for LR_+ og LR_- ikke av overbevisende diagnostisk bevis i noen av studiene. Selv om disse ULNT viser moderat til høy diagnostisk verdi for S_n og S_p , vil verdier for LR_+ og LR_- medføre usikkerhet på om et positivt eller negativt testresultat vil gi ytterligere informasjon om sannsynligheten av sykdomsstatus og dermed begrense den diagnostiske verdien av ULNT 2a og ULNT 3.

5.2.2 Hva kan ha påvirket resultatene?

I det følgende kapitlet vil rekrutteringsprosessen og utvalg, blinding, samt valg, beskrivelse og tolkning av index-tester og referansetest diskuteres. Jeg har forsøkt å drøfte hvordan disse faktorene kan ha påvirket resultatene. Diskusjonen vil fortløpende sammenligne og drøfte resultater på tvers av studiene under hvert enkelt tema.

5.2.2.1 Rekrutteringsprosess og utvalg

Som diskutert i *kapittel 5.1.1*, ser man at alle inkluderte studier til en viss grad har seleksjonsbias i sin rekrutteringsprosess. Dette kan lede til feilestimer av index-testenes diagnostiske verdi. *S1* tolkes å ha seleksjonsbias i størst grad, da de inkluderte deltagere som hadde sykehistorie, symptomer og elektrofysiologiske funn som antydte CR eller CTS, *før* index-testene ble utført. Om man da sammenligner funnene fra eksempelvis ULNT 2a på tvers av studiene, ser man at *S1* angir $S_n = 0.97$ og $S_p = 0.22$. *S2* angir $S_n = 0.66$ og $S_p = 0.75$, mens *S3* angir $S_n = 0.70$ og $S_p = 0.73$ for den samme index-testen. Her kan man diskutere om de svært gode sensitivitetsresultatene for ULNT 2a fra *S1*, kan være et overestimat fordi utvalget bestod av personer som allerede var antatt å ha CR. Sannsynligheten for at testen gir positivt utslag hos dette utvalget vil da øke, og testen kan tolkes som svært sensitiv om funn fra MR eller EMG kan bekrefte diagnosen. Man ser også at S_p i *S1* er betydelig lavere estimert enn i *S2* og *S3*. Dette kan ha bakgrunn i seleksjonsbias som medfører få «sant friske» i *S1* som gir en underestimert av S_p . Disse betraktningene kan styrke vurderingen om at rekrutteringsprosessen i *S2* og *S3* etablerte et mer representativt utvalg, i henhold til heterogenitet og ukjent sykdomsstatus, som diskutert i *kapittel 5.1.1*. Dette kan medføre et større antall «sant friske» i *S2* og *S3* som sannsynlig vil føre til at flertallet av friske ender i kvadranten «sanne negative» i 2*2 tabellen, og gi «sannere» spesifisitetsberegninger for disse

studiene. Ut ifra dette kan man diskutere om noen av forskjellene i resultatene på tvers av studiene, kan tilskrives ulikheter i utvalget.

5.2.2.2 Blinding

Det er viktig at funnene er et resultat av forskning og ikke av forskerens subjektive holdninger (Johannessen et al., 2021). Forventninger kan ha stor betydning i forskning, fordi både deltakers og utfallsmålers adferd kan påvirkes av egne forventninger og dermed representere en mulig feilkilde for resultatet (Svartdal, 2019). I dette øyemed bør utfallsmålerene alltid blindes, både fra funn fra tester som er inkludert i studien og fra eventuelle tilleggsopplysninger om deltageren (Tacconelli, 2010). Samtlige inkluderte studier har utarbeidet gode rutiner for blinding, der alle utfallsmålere både fra index-tester og referansetest er blindet for hverandres funn. Utfallsmåler fra index-test er også blindet for tilleggsopplysninger om deltagerne i samtlige studier. Dette kan lede til at resultatene i mindre grad blir påvirket av utfallsmålerenes subjektive holdning, men blir et direkte resultat av forskningen (Johannessen et al., 2021). Dette vil som tidligere omtalt styrke studiene ytterligere og redusere sannsynligheten for at resultatene påvirkes som følge av mangel på blinding (Jamtvedt, 2015).

5.2.2.3 Beskrivelse av testprosedyre

Som tidligere beskrevet er det viktig at prosedyrene fra både index-tester og referansetest blir utført likt på alle deltagere, og at dokumentasjonen av disse kommer tydelig frem (Tacconelli, 2010). Studiene har utarbeidet en standardisert og tydelig beskrevet protokoll for utførelse av index-tester og referansetester, som et middel for å objektivere og standardisere undersøkelsesmetodene. Dette vurderes som en styrke for testenes reliabilitet, samtidig som det vil redusere risikoen for informasjonsbias, som samlet sett kan redusere sannsynligheten for feilestimer i resultatene av testenes diagnostiske kvaliteter.

5.2.2.4 Tidsrom mellom utførelse av index-test og referansetest

Det er vesentlig at det ikke går for lang tid mellom utførelsen av index-test og referansetest. Dette for å unngå endringer av sykdomsstatus som følge av et tilstrekkelig tidsrom, som kan være et bias i testmetoden (Hanchard et al., 2013; Tacconelli, 2010). Dette tidsrommet kan være vanskelig å definere, men Hanchard et al. (2013) stadfester at referansetesten bør finne sted innen en måned etter den kliniske undersøkelsen. Et annet alternativ er å ta utgangspunkt i hvor lenge deltagerne har hatt symptomer og utføre referansetesten innen dette tidsrommet (Hanchard et al., 2013). Dersom en deltager eksempelvis har hatt symptomer i to uker, bør ikke referansetesten utføres senere enn to uker etter index-testene. Samtlige inkluderte studier

utfører index-tester og referansetester innenfor en tidsramme på 15-60 minutter. Dette er en metodisk styrke, som vil redusere eventuelle bias i henhold til undersøkelsesmetoder og medføre at resultatene sannsynligvis ikke påvirkes på bakgrunn av endringer i sykdomsstatus (Hanchard et al., 2013).

5.2.2.5 Tolkning av index-tester

I forhold til hvordan studiene definerer en positiv test, så finnes det ulikheter som kan ha innflytelse på resultatene. *S1* definerer testen som positiv om *én* av de følgende kriteriene blir møtt; 1) Reproduksjon av pasientens symptomer. 2) Sidedforskjeller (>10 grader) i albueekstensjon (ULNT 2a), eller håndleddsflexjon (ULNT 2b), når alle bevegelsessekvenser er fullført. 3) På symptomatisk side: kontralateral lateralfleksjon av nakken øker symptomer, eller ipsilateral lateralfleksjon av nakken reduserer symptomer.

S2 definerte ULNT som positiv om *alle tre* følgende kriterier ble møtt; 1) Reproduksjon av nevropatisk smerte i nakke og arm (brennende eller lynaktig smerte, prikkete sensasjoner, i henhold til dermatomalt mønster om det er nerverotspatologi). 2) Økte eller reduserte symptomer med strukturell differensiering. 3) Forskjeller i smertefull radiering mellom høyre og venstre side.

S3 definerte ULNT som positiv om *begge* av de følgende kriteriene ble møtt; 1) Full eller delvis reproduksjon av en gjenkjennelig symptomatisk klage av arm og/eller nakkesmerte (smerte eller dysestesi som inkluderer brennende eller lynaktig smerte, eller prikkete sensasjoner). 2) Strukturell differensiering: Når en slik gjenkjennbar klage (punkt 1) ble fremprovosert, ble det utført en strukturell differensiering mellom nevrogene og ikke-nevrogene kilder ved å utføre sensitiverende bevegelser fjernt fra smerten (ipsilateral eller kontralateral lateralfleksjon av nakken, flexjon eller ekstensjon av albue eller håndledd, eller elevasjon av skulderbuen).

Her ser man at alle inkluderte studier vektlegger reproduksjon av kjent smerte og strukturell differensiering som viktige kriterier for å bedømme testen som positiv. En ulikhet er at *S1* kun krever at *én* av kriteriene må oppfylles for at testens skal tolkes som positiv, mens *S2* og *S3* krever at *samtlig*e kriterier er til stede for et positivt testresultat. Dette kan således gi ulikheter i resultatene mellom studiene. Herunder ser man av tabell VI at *S1* har svært gode sensitivitetsdata for sine ULTT A og B, mens *S2* og *S3* har mer sparsommelige tall vedrørende sensitivitet (tabell VII og VIII). Dette kan dels forklares ved at kriteriene for et positivt testresultat er strengere og mer omfattende i *S2* og *S3*, sammenlignet med *S1*. Det er

dermed *lavere terskel* for et *positivt testresultat* i *S1*, som vil kunne overestimere testenes egenskaper i forhold til sensitivitet. På den annen side kan de mer omfattende kravene i *S2* og *S3* medføre en *høyere terskel* for et *positivt resultat* og således kunne estimere mer nøyaktige egenskaper i forhold til sensitivitet.

Om man på den annen side vurderer dataene hva gjelder spesifisitet, observerer man en motsatt trend. Her viser tall fra *S1* at ULNT 2a og ULNT 2b har *lav spesifisitet*, mens tall fra *S2* og *S3* viser *høyere spesifisitet* blant de undersøkte ULNT. Her vil de mer omfattende kriteriene i *S2* og *S3* kunne medføre at man fanger opp en større andel av de som *faktisk er friske*, og dermed *estimere «sannere» verdier* for spesifisitet. Derimot kan de mer «liberale» kriteriene i *S1* vurderes å medføre mange *falske positive* testresultater, som videre gir testene *lavere* spesifisitet.

5.2.3 Resultatdiskusjon av test-clustere

Som et middel for å bedre den diagnostiske verdien kan man kombinere flere enkelttester, når man ønsker å bekrefte eller avkrefte en patologi. Dette kalles et test-cluster (Hegedus et al., 2015). *S2* og *S3* har undersøkt om anvendelsen av test-cluster øker den diagnostiske verdien av ULNT i undersøkelsen av CR. Resultater fra *S2* viser at kombinasjonen av alle fire ULNT gir $S_n = 0.97$, og $S_p = 0.62$. LR_+ og LR_- var på henholdsvis 0.87 og 0.92. Sammenlignet med resultater fra samme studie der testene blir vurdert hver for seg, ser man at sensitiviteten øker, mens spesifisiteten reduseres når testene utføres som et cluster. Dette indikerer at clusteret i større grad er anvendelig for å bekrefte CR, enn for å utelukke. Tall vedrørende LR_+ og LR_- impliserer at et positivt eller negativt testresultat ikke vil gi ytterligere informasjon om sannsynligheten for om pasienten har CR eller ikke, og kan ifølge Tacconelli (2010) dermed svekke test-clusterets diagnostiske verdi.

Resultater fra *S3* omhandlende diagnostiske utfallsmål for test-clusteret er oppsummert i *tabell IX*. *S3* statuerer at ULNT har større nytteverdi i diagnostiseringen av CR, når testene anvendes som et test-cluster. De angir at 3/4 positive tester gir grunnlag for å bekrefte CR som diagnose. Om man her analyserer dataene fra *tabell IX*, ser man at 3/4 positive tester gir følgende utfallsmål: $S_n = 44.44$, $S_p = 96.55$, $LR_+ = 12.89$ og $LR_- = 0.58$. Her ser man at selv om sensitiviteten kun er 44.44, så er LR_+ svært tilfredsstillende (Tacconelli, 2010), og impliserer dermed at 3/4 positive tester er av overbevisende diagnostisk verdi for å bekrefte CR som diagnose. I tillegg vil en $S_p = 96.55$ medføre at clusteret er svært nyttig til å ekskludere de «sant friske» og dermed i stor grad inkludere de «sant syke» som tester positivt.

Videre viser resultatene fra S3 at et test-cluster med 1 av 4 positive tester gir $LR^- = 0.08$, der forfatterne indikerer at et test-cluster med 4 av 4 negative tester dermed kan utelukke CR.

5.3 Implikasjoner for praksis

Resultatene fra de inkluderte studiene peker i retning av at ULNT *kan* ha en diagnostisk verdi i undersøkelsen av pasienter med nakke- og eller armsmerter, når man ønsker å bekrefte eller avkrefte CR som diagnose. De er enkle å anvende på arbeidststedet, kan ta liten tid og gi tilleggsinformasjon til anamnese og andre kliniske undersøkelser. Klinikere må være oppmerksom på at de fire ulike ULNT fremholder varierende diagnostisk verdi, hva gjelder sensitivitet, spesifisitet og likelihood ratios. Dette indikerer at testresultatene bør tolkes med forsiktighet. Funn fra denne masteroppgaven viser at enkeltstående ULNT (2a og 3) kan fremholde god sensitivitet og spesifisitet, men er likevel ikke av overbevisende diagnostisk verdi til å bekrefte eller avkrefte CR, grunnet begrensede verdier for LR^+ og LR^- . Om man derimot anvender ULNT som et test-cluster øker den diagnostiske verdien, der resultater fra to av de inkluderte studiene viser overbevisende diagnostisk verdi både til å bekrefte og avkrefte diagnosen. Det anbefales derfor at klinikere anvender ULNT som test-cluster i undersøkelsen av pasienter med nakke- og/eller armsmerter, når man ønsker å bekrefte eller avkrefte CR som diagnose. ULNT bør også vurderes sammen med funn fra anamnese og eventuelle andre kliniske undersøkelser, for en grundig diagnostisk utredning.

5.4 Masteroppgavens styrker og svakheter

I prosessen med å utarbeide problemstilling, var det ønskelig å kartlegge verdien av kliniske tester for CR og herunder finne ut av hvilken informasjon klinikeren får av ULNT for denne diagnosen. Det ville samtidig undersøkes om det var måter å kombinere disse testene på, som et middel for å øke den diagnostiske verdien. I de første søkene ble det kun kombinert søkeord som omhandlet ”Cervikal radikulopati”, sammen med ulike synonymer for ”Clinical tests”. Deretter ble søkestrategien utvidet, ved å inkludere flere termer i søket. Jeg gikk inn i litteraturen og fant ut av hvilke ord som ofte ble brukt for å beskrive CR. De neste søkene inneholdt flere synonymer for CR og for kliniske tester (se søkelogg, vedlegg 1). Dette resulterte i flere treff, men samtidig åpnet det for studier som også vurderte annen nakkepatologi enn den som er aktuell for min problemstilling, samt variasjoner av studiedesign og referansetester. Derfor ble det tillagt flere søkeord for å inkludere studier som vurderte diagnostisk nøyaktighet og ønskelig referansetest, som et middel for å skape et mer presist søk. Dette resulterte i et større antall relevante studier. For å minimalisere ulikhetene i de inkluderte studiene, slik at problemstillingen kunne besvares på best mulig måte, ble det

utarbeidet inklusjons- og eksklusjonskriterier som studiene måtte tilfredsstill. Se kapittel 3.2.2 og 3.2.3. Ved å sette disse kriteriene risikerte jeg å gå glipp av annen relevant litteratur. Samtidig økte det sannsynligheten for at jeg fant relevant litteratur som kunne besvare min problemstilling ut fra like kriterier.

I tillegg kan en svakhet i denne litteraturstudien være at det kun ble inkludert diagnostiske kohortstudier, som dermed ekskluderer andre relevante studier av annet studiedesign. Dog var det ønskelig å inkludere studier av likt design for å bedre kunne vurdere og sammenligne resultater på tvers av studiene. I tillegg viser det seg at resultatene som stammer fra diagnostiske kohortstudier er svært representative for klinisk praksis (Tacconelli, 2010), og kan sørge for at det ble innhentet gode data for å besvare problemstillingen.

En annen faktor som kan svekke denne oppgavens metode er min manglende erfaring med systematiske litteratursøk og bruk av databaser. For å minimere konsekvensene av dette, ble det endelige søket gjennomført i samarbeid med bibliotekar ved UiB. Her ble søkeord, søkemotorer og databaser diskutert og systematisk søkt i. Søkelogg, samt valg av database og inkluderte studier ble også vurdert sammen med veileder. I tillegg har jeg konferert med medstudenter og praksisveileder for tips og råd. Jeg anvendte litteraturen for å finne ut av hvilke databaser som var mest anvendelige og kunne gi mest informasjon om diagnostikk, før jeg diskuterte med veileder og bibliotekar. Med bakgrunn i dette, ble det endelige søket utført i PubMed. Det faktum at PubMed ble anvendt som eneste database er også en klar svakhet ved min metode. Epistemonikos og Cochrane Library er eksempler på databaser som er aktuelle ved forskning på diagnostikk (Jamtvedt, 2015), og ved å utelate disse kan jeg ha gått glipp av relevant litteratur som kunne besvart min problemstilling. Samtidig er PubMed en database som dekker hele helsefeltet, er anbefalt ved spørsmål om diagnostikk, har over 34 millioner referanser og inkluderer også databasen MEDLINE (Jamtvedt, 2015).

”Sjekkliste for vurdering av en studie som tester en ny diagnostisk test” (Helsebiblioteket, 2018) ble anvendt for å kritisk vurdere de inkluderte studiene. Jeg har liten erfaring i kritisk vurdering av studier som kan være en faktor som har påvirket analysen. Jeg kan potensielt ha vurdert studiene annerledes enn hva andre med mer erfaring ville gjort. For å styrke den kritiske vurderingen av studiene spurte jeg en kollega om en uavhengig vurdering, der vi i etterkant kom til enighet.

For å bedre kunne sammenligne resultatene på tvers av studiene, valgte jeg å inkludere ulike utfallsmål. Jeg gikk til litteraturen og sporet opp hvilke utfallsmål som er vanlig for å beskrive

kvaliteten av diagnostiske tester. Disse var blant annet sensitivitet, spesifisitet, positiv og negativ likelihood ratio, samt positiv og negativ prediktiv verdi. Minimum 3/6 av disse utfallsmålene måtte beskrives i studiene om de skulle inkluderes i denne litteraturstudien. Jeg har begrenset erfaring hva gjelder vurdering av diagnostiske tester og bruken av disse statistiske begrepene, og må derfor være åpen for eventuelle feiltolkninger av resultatene. Likevel har jeg støttet meg til faglitteratur for å tolke resultatene på best mulig måte.

Flere måleinstrumenter blir anvendt for å diagnostisere nakkepatologi, men det virker som at det er konsensus for at MR eller EMG er den nåværende gullstandard for diagnostisering av CR (Sharma et al., 2014). Dermed inkluderte jeg kun studier som vurderte de kliniske testene for CR, opp mot MR eller EMG, for å imøtekomme kravet om gullstandard. Dog kan dette ha resultert i at jeg ekskluderte annen litteratur som er relevant for min problemstilling, men som undersøkte fenomenet med en annen referansetest.

6.0 Konklusjon

Denne litteraturstudien har svart på følgende problemstilling:

«Hvilken diagnostisk verdi har nevrodynamiske tester i undersøkelsen av cervikal radikulopati?»

Funn fra denne litteraturstudien indikerer at ULNT 2a for nervus medianus og ULNT 3 for nervus ulnaris er de *enkelstående* ULNT som viser høyest diagnostisk verdi i undersøkelsen av CR. Disse to ULNT fremholder moderat til høy diagnostisk verdi for sensitivitet og spesifisitet, men er likevel ikke av overbevisende diagnostisk verdi grunnet begrenset positiv og negativ likelihood ratio. Resultater fra *to* av de inkluderte studiene viser at ULNT utført som test-cluster medfører høyere sensitivitet, og *én* studie viser tilfredsstillende positiv likelihood ratio, som således øker den diagnostiske verdien til å bekrefte CR som diagnose. Resultater fra *én* studie indikerer at 4 av 4 negative ULNT kan utelukke CR, med bakgrunn i svært tilfredsstillende diagnostisk verdi for LR-. Det indikerer at man *kan* velge å benytte enkeltstående ULNT (2a og 3) i undersøkelsen av denne pasientgruppen, men testsvarene man får vil ikke være av overbevisende diagnostisk verdi. I klinisk praksis anbefales det derfor å utføre ULNT som test-cluster når man ønsker å bekrefte eller avkrefte CR som diagnose, da dette øker den diagnostiske verdien til testene.

Det er viktig være oppmerksom på at denne masteroppgaven er basert på et snevert utvalg av studier, som gjør at man ikke kan trekke klare og bastante slutninger. Det ser ut til at funnene fra denne masteroppgaven støtter tidligere forskning på området, som viser at det ikke foreligger enighet om den diagnostiske verdien av ULNT i undersøkelsen av CR. De inkluderte studiene fremholder varierende resultater og har metodiske svakheter. En generell problemstilling ved forskning på CR er at det ikke foreligger konsensus med veldefinerte kliniske kriterier for diagnosen, som i dette tilfellet kan problematisere valideringen av testene. Det vurderes som sentralt at man utfører en grundig undersøkelse av pasienter med nakke- og/eller armsmerter, der man baserer en eventuell diagnose på bakgrunn av flere kriterier. Herunder bør den samlede informasjonen fra anamnese, samt resultater fra nevrologisk undersøkelse og andre spesifikke kliniske undersøkelser, *inkludert* test-cluster av ULNT, være grunnlag for diagnosen CR. Ved klinisk mistanke om CR, kan diagnosen eventuelt bekrefte med MR eller EMG. Det vurderes som nødvendig med flere studier som inkluderer større og representative utvalg, med høyere metodisk kvalitet, for å sikre et forskningsgrunnlag som kan anvendes og implementeres i klinisk praksis.

Litteraturliste

- Apelby-Albrecht, M., Andersson, L., Kleiva, I. W., Kvåle, K., Skillgate, E., & Josephson, A. (2013). Concordance of upper limb neurodynamic tests with medical examination and magnetic resonance imaging in patients with cervical radiculopathy: a diagnostic cohort study. *J Manipulative Physiol Ther*, *36*(9), 626-632.
<https://doi.org/10.1016/j.jmpt.2013.07.007>
- Baselgia, L. T., Bennett, D. L., Silbiger, R. M., & Schmid, A. B. (2017). Negative Neurodynamic Tests Do Not Exclude Neural Dysfunction in Patients With Entrapment Neuropathies. *Archives of Physical Medicine and Rehabilitation*, *98*(3), 480-486. <https://doi.org/https://doi.org/10.1016/j.apmr.2016.06.019>
- Carette, S., & Fehlings, M. G. (2005). Clinical practice. Cervical radiculopathy. *N Engl J Med*, *353*(4), 392-399. <https://doi.org/10.1056/NEJMcp043887>
- Elvey, R. L. (1986). Treatment of Arm Pain Associated with Abnormal Brachial Plexus Tension. *Aust J Physiother*, *32*(4), 225-230. [https://doi.org/10.1016/S0004-9514\(14\)60655-3](https://doi.org/10.1016/S0004-9514(14)60655-3)
- Finnerup, N. B., Haroutounian, S., Kamerman, P., Baron, R., Bennett, D. L. H., Bouhassira, D., Cruccu, G., Freeman, R., Hansson, P., Nurmikko, T., Raja, S. N., Rice, A. S. C., Serra, J., Smith, B. H., Treede, R.-D., & Jensen, T. S. (2016). Neuropathic pain: an updated grading system for research and clinical practice. *PAIN*, *157*(8), 1599-1606.
<https://doi.org/10.1097/j.pain.0000000000000492>
- Govindarajan, R., Kolb, C., & Salgado, E. (2013). Sensitivity and Specificity of MRI and EMG in Diagnosing Clinically Evident Cervical Radiculopathy: A Retrospective Study (P02.224). *Neurology*, *80*(7 Supplement), P02.224.
http://n.neurology.org/content/80/7_Supplement/P02.224.abstract
- Grondin, F., Cook, C., Hall, T., Maillard, O., Perdrix, Y., & Freppel, S. (2021). Diagnostic accuracy of upper limb neurodynamic tests in the diagnosis of cervical radiculopathy. *Musculoskeletal Science and Practice*, *55*, 102427.
<https://doi.org/https://doi.org/10.1016/j.msksp.2021.102427>
- Hanchard, N. C., Lenza, M., Handoll, H. H., & Takwoingi, Y. (2013). Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement. *Cochrane Database Syst Rev*, *2013*(4), Cd007427.
<https://doi.org/10.1002/14651858.CD007427.pub2>

- Hegedus, E. J., Cook, C., Lewis, J., Wright, A., & Park, J. Y. (2015). Combining orthopedic special tests to improve diagnosis of shoulder pathology. *Phys Ther Sport*, 16(2), 87-92. <https://doi.org/10.1016/j.ptsp.2014.08.001>
- Helsebiblioteket. (2018). *Sjekkliste for vurdering av en studie som tester en ny diagnostisk test*. Helsebiblioteket.no. Retrieved 21.06.2022 from <https://www.helsebiblioteket.no/innhold/artikler/kunnskapsbasert-praksis/kunnskapsbasertpraksis.no#4kritisk-vurdering>
- Helseth, E., Rootwelt, T., & Harbo, H. F. (2019). *Nevrologi og nevrokirurgi : fra barn til voksen* (7. ed.). Fagbokforlaget.
- Hogg-Johnson, S., van der Velde, G., Carroll, L. J., Holm, L. W., Cassidy, J. D., Guzman, J., Côté, P., Haldeman, S., Ammendolia, C., Carragee, E., Hurwitz, E., Nordin, M., & Peloso, P. (2008). The burden and determinants of neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976)*, 33(4 Suppl), S39-51. <https://doi.org/10.1097/BRS.0b013e31816454c8>
- Iyer, S., & Kim, H. J. (2016). Cervical radiculopathy. *Curr Rev Musculoskelet Med*, 9(3), 272-280. <https://doi.org/10.1007/s12178-016-9349-4>
- Jamtvedt, G. (2015). *Kunnskapsbasert fysioterapi : metoder og arbeidsmåter* (2. ed.). Gyldendal akademisk.
- Johannessen, A., Christoffersen, L., & Tufte, P. A. (2021). *Introduksjon til samfunnsvitenskapelig metode* (6. ed.). Abstrakt forlag.
- Jørgensen, T., Christensen, E., & Linneberg, A. (2016). *Klinisk forskningsmetode : en grundbog* (4. ed.). Munksgaard.
- Koulidis, K., Veremis, Y., Anderson, C., & Heneghan, N. R. (2019). Diagnostic accuracy of upper limb neurodynamic tests for the assessment of peripheral neuropathic pain: A systematic review. *Musculoskeletal Science and Practice*, 40, 21-33. <https://doi.org/https://doi.org/10.1016/j.msksp.2019.01.001>
- Lemeunier, N., da Silva-Oolup, S., Chow, N., Southerst, D., Carroll, L., Wong, J. J., Shearer, H., Mastragostino, P., Cox, J., Côté, E., Murnaghan, K., Sutton, D., & Côté, P. (2017). Reliability and validity of clinical tests to assess the anatomical integrity of the cervical spine in adults with neck pain and its associated disorders: Part 1-A systematic review from the Cervical Assessment and Diagnosis Research Evaluation (CADRE) Collaboration. *Eur Spine J*, 26(9), 2225-2241. <https://doi.org/10.1007/s00586-017-5153-0>

- Magnus, P., & Bakketeig, L. S. (2000). *Prosjektarbeid i helsefagene*. Gyldendal akademisk.
- Nee, R. J., & Butler, D. (2006). Management of peripheral neuropathic pain: Integrating neurobiology, neurodynamics, and clinical evidence. *Physical Therapy in Sport*, 7(1), 36-49. <https://doi.org/10.1016/j.ptsp.2005.10.002>
- Nee, R. J., Jull, G. A., Vicenzino, B., & Coppiters, M. W. (2012). The validity of upper-limb neurodynamic tests for detecting peripheral neuropathic pain. *J Orthop Sports Phys Ther*, 42(5), 413-424. <https://doi.org/10.2519/jospt.2012.3988>
- Nygaard, Ø., Rø, M., Andersen, T., & Zwart, J.-A. (2010). Nakkesmerter med og uten nerverotsaffeksjon. *Tidsskrift for den Norske Lægeforening*, 130(22), 2252-2255. <https://doi.org/10.4045/tidsskr.09.1285>
- Radhakrishnan, K., Litchy, W. J., O'Fallon, W. M., & Kurland, L. T. (1994). Epidemiology of cervical radiculopathy. A population-based study from Rochester, Minnesota, 1976 through 1990. *Brain*, 117 (Pt 2), 325-335. <https://doi.org/10.1093/brain/117.2.325>
- Rubinstein, S. M., Pool, J. J. M., van Tulder, M. W., Riphagen, I. I., & de Vet, H. C. W. (2007). A systematic review of the diagnostic accuracy of provocative tests of the neck for diagnosing cervical radiculopathy. *European Spine Journal*, 16(3), 307-319. <https://doi.org/10.1007/s00586-006-0225-6>
- Schmid, A. B., Brunner, F., Luomajoki, H., Held, U., Bachmann, L. M., Künzer, S., & Coppiters, M. W. (2009). Reliability of clinical tests to evaluate nerve function and mechanosensitivity of the upper limb peripheral nervous system. *BMC Musculoskeletal Disord*, 10, 11. <https://doi.org/10.1186/1471-2474-10-11>
- Sharma, S., Bjørnara, B. T., Robinson, H. S., Hjemly, H., Hammerlund, K., Myrseth, L. E., Glasø, M., Juel, N. G., Bjørner, T., & Kongshavn, T. (2014). *Nasjonalt faglig retningslinje for bildediagnostikk ved ikke-traumatiske muskel- og skjelettlidelser: anbefalinger for primærhelsetjenesten*. Helsedirektoratet.
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annu Rev Psychol*, 70, 747-770. <https://doi.org/10.1146/annurev-psych-010418-102803>
- So, Y. T. (1999). Guidelines in electrodiagnostic medicine. Practice parameter for needle electromyographic evaluation of patients with suspected cervical radiculopathy. *Muscle Nerve Suppl*, 8, S209-221.
- Svartdal, F. (2019). forventning - psykologi. In *Store norske leksikon*.

- Tacconelli, E. (2010). Book: Systematic reviews: CRD's guidance for undertaking reviews in health care. *The Lancet Infectious Diseases*, 10(4), 226.
- Thomas, J. R., Silverman, S. J., & Nelson, J. K. (2011). *Research methods in physical activity* (6th ed.). Human Kinetics.
- Thoomes, E. J., Scholten-Peeters, G. G. M., de Boer, A. J., Olsthoorn, R. A., Verkerk, K., Lin, C., & Verhagen, A. P. (2012). Lack of uniform diagnostic criteria for cervical radiculopathy in conservative intervention studies: a systematic review. *European Spine Journal*, 21(8), 1459-1470. <https://doi.org/10.1007/s00586-012-2297-9>
- Thoomes, E. J., van Geest, S., van der Windt, D. A., Falla, D., Verhagen, A. P., Koes, B. W., Thoomes-de Graaf, M., Kuijper, B., Scholten-Peeters, W. G. M., & Vleggeert-Lankamp, C. L. (2018). Value of physical tests in diagnosing cervical radiculopathy: a systematic review. *Spine J*, 18(1), 179-189. <https://doi.org/10.1016/j.spinee.2017.08.241>
- Wainner, R. S., Fritz, J. M., Irrgang, J. J., Boninger, M. L., Delitto, A., & Allison, S. (2003). Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine (Phila Pa 1976)*, 28(1), 52-62. <https://doi.org/10.1097/00007632-200301010-00014>
- Woolf, C. J. (2004). Dissecting out mechanisms responsible for peripheral neuropathic pain: Implications for diagnosis and therapy. *Life Sciences*, 74(21), 2605-2610. <https://doi.org/https://doi.org/10.1016/j.lfs.2004.01.003>

Vedlegg

Vedlegg 1 – Søkelogg i PubMed

Dato: 21.06.2022

Søkenummer	Beskrivelse av søket
#1	Search: cervical vertebra*[Title/Abstract] OR cervical disk*[Title/Abstract] OR cervical disc*[Title/Abstract] OR neuralgia[Title/Abstract] OR distal neuropathic pain*[Title/Abstract] OR peripheral neuropathic pain*[Title/Abstract] OR peripheral entrapment neuropath*[Title/Abstract] OR distal entrapment neuropath*[Title/Abstract] OR nerve compression syndrome*[Title/Abstract] OR radicular pain*[Title/Abstract] OR radiculit*[Title/Abstract] OR radiculopath*[Title/Abstract] 37,477 07:21:56
#2	Search: "Radiculopathy"[Mesh] OR "Cervical Vertebrae"[Mesh] OR "Nerve Compression Syndromes"[Mesh:NoExp] 57,035 07:23:02
#3 (#1 OR #2)	Search: (cervical vertebra*[Title/Abstract] OR cervical disk*[Title/Abstract] OR cervical disc*[Title/Abstract] OR neuralgia[Title/Abstract] OR distal neuropathic pain*[Title/Abstract] OR peripheral neuropathic pain*[Title/Abstract] OR peripheral entrapment neuropath*[Title/Abstract] OR distal entrapment neuropath*[Title/Abstract] OR nerve compression syndrome*[Title/Abstract] OR radicular pain*[Title/Abstract] OR radiculit*[Title/Abstract] OR radiculopath*[Title/Abstract]) OR ("Radiculopathy"[Mesh] OR "Cervical Vertebrae"[Mesh] OR "Nerve Compression Syndromes"[Mesh:NoExp]) 82,334 07:23:11
#4	Search: Upper limb neurodynamic test*[Title/Abstract] OR upper limb tension test*[Title/Abstract] OR neural provocation test*[Title/Abstract] OR neuro exam*[Title/Abstract] OR neurologic exam*[Title/Abstract] OR neurological exam*[Title/Abstract] 18,136 07:24:17
#5	Search: "Neurologic Examination"[Mesh:NoExp] 27,940 07:24:35
#6 (#4 OR #5)	Search: (Upper limb neurodynamic test*[Title/Abstract] OR upper limb tension test*[Title/Abstract] OR neural provocation test*[Title/Abstract] OR neuro exam*[Title/Abstract] OR neurologic exam*[Title/Abstract] OR neurological exam*[Title/Abstract]) OR ("Neurologic Examination"[Mesh:NoExp]) 42,954 07:24:43
#7	Search: Diagnostic accuracy[Title/Abstract] OR diagnostic precision[Title/Abstract] OR sensitivity[Title/Abstract] OR specificity[Title/Abstract] OR valid*[Title/Abstract] OR predictive value*[Title/Abstract] OR reference standard*[Title/Abstract] OR diagnostic reference level*[Title/Abstract] OR likelihood*[Title/Abstract] 2,206,344 08:43:32

#8	<p>Search: "Sensitivity and Specificity"[Mesh] OR "Reference Standards"[Mesh:NoExp] OR "Diagnostic Reference Levels"[Mesh] OR "Likelihood Functions"[Mesh] 695,205 08:44:07</p>
#9 (#7 OR #8)	<p>Search: (Diagnostic accuracy[Title/Abstract] OR diagnostic precision[Title/Abstract] OR sensitivity[Title/Abstract] OR specificity[Title/Abstract] OR valid*[Title/Abstract] OR predictive value*[Title/Abstract] OR reference standard*[Title/Abstract] OR diagnostic reference level*[Title/Abstract] OR likelihood*[Title/Abstract]) OR ("Sensitivity and Specificity"[Mesh] OR "Reference Standards"[Mesh:NoExp] OR "Diagnostic Reference Levels"[Mesh] OR "Likelihood Functions"[Mesh]) 2,591,259 08:44:16</p>
#10 (#3 AND #6 AND #9)	<p>Search: (((cervical vertebra*[Title/Abstract] OR cervical disk*[Title/Abstract] OR cervical disc*[Title/Abstract] OR neuralgia[Title/Abstract] OR distal neuropathic pain*[Title/Abstract] OR peripheral neuropathic pain*[Title/Abstract] OR peripheral entrapment neuropath*[Title/Abstract] OR distal entrapment neuropath*[Title/Abstract] OR nerve compression syndrome*[Title/Abstract] OR radicular pain*[Title/Abstract] OR radiculit*[Title/Abstract] OR radiculopath*[Title/Abstract]) OR ("Radiculopathy"[Mesh] OR "Cervical Vertebrae"[Mesh] OR "Nerve Compression Syndromes"[Mesh:NoExp])) AND ((Upper limb neurodynamic test*[Title/Abstract] OR upper limb tension test*[Title/Abstract] OR neural provocation test*[Title/Abstract] OR neuro exam*[Title/Abstract] OR neurologic exam*[Title/Abstract] OR neurological exam*[Title/Abstract]) OR ("Neurologic Examination"[Mesh:NoExp]))) AND ((Diagnostic accuracy[Title/Abstract] OR diagnostic precision[Title/Abstract] OR sensitivity[Title/Abstract] OR specificity[Title/Abstract] OR valid*[Title/Abstract] OR predictive value*[Title/Abstract] OR reference standard*[Title/Abstract] OR diagnostic reference level*[Title/Abstract] OR likelihood*[Title/Abstract]) OR ("Sensitivity and Specificity"[Mesh] OR "Reference Standards"[Mesh:NoExp] OR "Diagnostic Reference Levels"[Mesh] OR "Likelihood Functions"[Mesh])) 198 08:49:53</p>

Vedlegg 2 – Søkestrategi i PubMed

Søkeelement	Cervikal radikulopati	Nevrodynamiske tester	Diagnostisk nøyaktighet
Fritekst (Tittel & abstract)	cervical vertebra* OR cervical disk* OR cervical disc* OR neuralgia OR distal neuropathic pain* OR peripheral neuropathic pain* OR peripheral entrapment neuropath* OR distal entrapment neuropath* OR nerve compression syndrome* OR radicular pain* OR radiculit* OR radiculopath*	Upper limb neurodynamic test* OR upper limb tension test* OR neural provocation test* OR neuro exam* OR neurologic exam* OR neurological exam*	Diagnostic accuracy OR diagnostic precision OR sensitivity OR specificity OR valid* OR predictive value* OR reference standard* OR diagnostic reference level* OR likelihood*
MeSH	"Radiculopathy"[Mesh] OR "Cervical Vertebrae"[Mesh] OR "Nerve Compression Syndromes"[Mesh:NoExp]	"Neurologic Examination"[Mesh:NoExp]	"Sensitivity and Specificity"[Mesh] OR "Reference Standards"[Mesh:NoExp] OR "Diagnostic Reference Levels"[Mesh] OR "Likelihood Functions"[Mesh]