

Masteroppgave

Masterprogram i fysioterapi
Mai, 2020

Kan Start Back Screening Tool predikere funksjonsbegrensinger hos eldre som oppsøker primærhelsetjenesten for behandling av sine ryggmerter?

- En prospektiv kohortstudie med 3 måneders oppfølging

Kandidatnavn: Solveig Flugstad

Kandidatnummer: 403

Emnekode: MAFYS5900_1

Antall ord: 16 644.

Fakultet for helsevitenskap

FORORD

Årene ved masterutdanningen ved Oslo Metropolitan University har vært lærerike og utviklende. Arbeidet med denne masteroppgaven har til tider vært krevende og intenst, men også veldig spennende og givende. Dyktige og engasjerte lærere og medelever har en stor del av æren for dette!

Jeg vil først og fremst rette en stor takk til mine veiledere Kjersti Storheim og Ørjan Nesse Vigdal. Dere har vært engasjerte i masterstudien min fra første dag. Dere har også vært tilgjengelige for møter, diskusjon og veiledning når jeg har hatt behov for det. Med deres erfaring og faglige tyngde har dere gitt meg utallige konstruktive tilbakemeldinger og bidratt til en rekke nyttige faglige diskusjoner. Jeg er veldig takknemlig for innsatsen dere har lagt ned!

Jeg vil også takke medstudentene mine som har vært viktige både for faglig utvikling og trivsel gjennom masterforløpet. Flere av dere har også bidratt med faglige innspill på masteroppgaven. Til slutt vil jeg takke prosjektgruppen i BACE-N, både for datasettet jeg har benyttet i masterstudien, inkludering i forskningsgruppen og nyttige innspill. En takk rettes også til fysiofondet for tildeling av utdanningsstipend.

Solveig.

Sammendrag

Bakgrunn: Psykososiale faktorer er identifisert som spesielt viktige i utviklingen fra akutte til langvarige ryggsmarter. Start Back Screening Tool (SBT) er et screeninginstrument som har til hensikt å klassifisere personer i lav- medium- eller høy risikogruppe på bakgrunn av fysiske og psykososiale prognostiske faktorer for langvarige korsryggsmarter.

Formål: Formålet med masterstudien var å undersøke SBTs prediktive validitet hos eldre ≥ 55 år som oppsøkte primærhelsetjenesten for en ny episode med ryggsmarter.

Materiale og metode: Studien er en prospektiv kohortstudie med 3 måneders oppfølging (n=299). En cut-off-verdi på $\geq 7/24$ poeng på Roland-Morris Disability Questionnaire (RMDQ) ble benyttet som utfallsmål. Det ble gjennomført analyser av sensitivitet, spesifisitet, samt positiv og negativ likelihood ratio for risikogruppe lav/medium mot høy og lav mot medium/høy (anbefalte cut-off-verdier), samt for risikogruppe lav mot høy (identifisert ved hjelp av den optimale cut-off-verdien). Diskriminerende evner og optimal cut-off-verdi på SBT ble undersøkt ved hjelp av arealet under Receiver Operating Characteristics Curve (ROC).

Resultater: SBT klassifiserte 65 %, 27 % og 7% i henholdsvis lav-, medium- og høy risikogruppe ved baseline. For risikogruppe lav/medium mot høy ble det funnet en sensitivitet på 0.11 og en spesifisitet på 0.99 mens positiv likelihood ratio og negativ likelihood ratio var henholdsvis 11 og 0.9. For risikogruppe lav mot medium/ høy ble det funnet en sensitivitet på 0.43, en spesifisitet på 0.78 og en positiv og negativ likelihood ratio på henholdsvis 1.95 og 0.73. Arealet under ROC-kurven var 0.68 (95 % KI: 0.61, 0.75). Optimal cut-off-verdi var $\geq 3/9$ poeng på SBT. Dette ga en sensitivitet og spesifisitet på henholdsvis 0.64 og 0.60 og en positiv og negativ likelihood ratio på henholdsvis 1.35 og 0.60.

Konklusjon: Det var kun høyriskogruppen på SBT (ved bruk av anbefalte cut-off-verdier) som hadde evne til å predikere et dårlig utfall etter tre måneder. SBTs samlede evne til å diskriminere mellom et godt og dårlig utfall var lav, men på grensen til moderat. Antall studiedeltagere som ble klassifisert i høyriskogruppen var imidlertid få og det er behov for mer forskning for å kunne anbefale screening med SBT hos eldre personer med ryggsmarter i klinikk og forskning.

Abstract

Background: Psychosocial factors have been identified as particularly important in the development from acute to long-term back pain. The Start Back Screening Tool (SBT) is a screening instrument which aim is to classify people into low- medium- or high-risk groups based on prognostic physical and psychosocial factors for long-term low back pain.

Objective: The aim of this master's study was to examine the predictive validity of the SBT among older adults at age of ≥ 55 who sought primary care with a new episode of back pain.

Material and methods: The master's study is a prospective cohort study with 3 months follow-up (n= 299). A cut-off value of $\geq 7/24$ points on the Roland-Morris Disability Questionnaire (RMDQ) was used as the outcome measure. Analyses of sensitivity, specificity, as well as positive and negative likelihood ratio were performed using the low/medium against high and the low against medium/high-risk group (recommended cut-off values). These analyses were also performed using the low against high risk group (identified using the optimal cut-off value). The SBTs discriminative abilities and the optimal cut-off value on the SBT were examined using the Area under the Receiver Operating Characteristics curve (AUC).

Results: At baseline 65 %, 27 % and 7% were considered as low-, medium- and high risk, respectively. Using the cut-off low/medium versus high risk group the sensitivity was 0.11 and the specificity 0.99, while the positive and negative likelihood ratio were 11 and 0.90, respectively. Using the cut-off low versus medium/high risk group the sensitivity was 0.43 and the specificity 0.78, and the positive and negative likelihood ratio were 1.95 and 0.73, respectively. The AUC was 0.68 (95 % KI: 0.61, 0.75). The optimal cut-off value on the SBT was $\geq 3/9$ points. Using this cut-off value, the sensitivity was 0.64 and the specificity was 0.60 and the positive and negative likelihood ratio were 1.35 and 0.60, respectively.

Conclusion: Only the high-risk group (using the recommended cut-off-value) was able to predict a poor outcome in the master's study. The SBTs overall discriminative performance was low, but close to moderate. However, there was a low number of study participants classified in the high-risk group. More research is needed to recommend screening with SBT in older adults with back pain in clinic and research.

Vedlegg:

Vedlegg 1: 2*2 krysstabeller som danner grunnlaget for utregning av sensitivitet, spesifisitet og likelihood ratio

Vedlegg 2: ROC-analyse med tabell som danner grunnlaget for identifisering av optimal cut-off-verdi

Vedlegg 3: Informert samtykkeskjema

Vedlegg 4: The Keele STarT Back Screening Tool – norsk versjon

Vedlegg 5: Roland-Morris Disability Questionnaire – norsk versjon

Vedlegg 6: Godkjenning fra regional etisk komite

Liste over forkortelser:

SBT: STarT Back Screening Tool

RMDQ: Roland-Morris Disability Questionnaire

BBQ: Back Beliefs Questionnaire

FES-I: Falls Efficacy Scale

CES-D: Center of Epidemiologic studies depression scale

FABQ: Fear Avoidance and beliefs Questionnaire

PCS: Pain Catastrophising Scale

NRS: Numerisk smerteskala

KMI: Kroppsmasseindeks

KI: Konfidensintervall

Liste over tabeller og figurer:

Tabell 1. Studier av prognose hos eldre med ryggmerter

Tabell 2: 2*2 krysstabeller for analyse av sensitivitet, spesifisitet og likelihood ratioer.

Tabell 3: Baselinekarakteristika.

Tabell 4: Psykososiale risikofaktorer assosiert med langvarige ryggmerter, smerteintensitet og funksjonsbegrensninger fordelt på risikogrupper på SBT ved baseline.

Tabell 5: Krysstabeller basert på anbefalte cut-off-verdier på SBT

Tabell 6: Sensitivitet, spesifisitet likelihood ratioer for å identifisere personer med et dårlig utfall etter 3 måneder ved bruk av SBTs anbefalte cut-off-verdier.

Tabell 7: Krysstabell basert på optimal cut-off-verdi på SBT (≥ 3 poeng).

Tabell 8: Sensitivitet, spesifisitet, samt negativ og positiv likelihood ratioer ved bruk av optimal cut-off-verdi på SBT.

Tabell 9: Viser et utvalg av resultater fra studier av SBTs prediktive validitet med lignende utfallsmål, oppfølgingstid eller klinisk setting.

Figur 1: Flytskjema for inkluderte deltakere

Figur 2: Box-plot av median på RMDQ etter 3 måneder fordelt på ulike risikogrupper på SBT ved baseline.

Figur 3: Diskriminerende evner for totalscore på SBT når ≥ 7 poeng på RMDQ ble benyttet som utfallsmål.

Figur 4: Prosentandel av studiedeltagerne et dårlig utfall etter 3 måneder fordelt på lav og høy risikogruppe identifisert ved optimal cut-off-verdi på SBT.

Innholdsfortegnelse

1.0 INNLEDNING	9
1.1 Hensikt med masterstudien og forskningsspørsmål	10
1.2 Prosjektorganisasjon	11
1.3 Begreper og definisjoner	11
<i>1.3.1 Definisjoner, karakteristikk og bruk av begrepet ryggsmarter</i>	11
<i>1.3.2 Andre begrepsdefinisjoner</i>	12
2.0 TEORI OG TIDLIGERE FORSKNING	14
2.1 Forståelsen av ryggsmarter i et historisk og biopsykososialt perspektiv	14
2.2 Epidemiologi og befolkningsvekst	14
2.3 Samfunnskostnader	15
2.4 Etiologi, undersøkelse og diagnostisering	15
<i>2.4.1 Risikofaktorer for første episode korsryggsmarter</i>	17
2.5 Behandling	18
2.6 Prognose	18
2.7 Screeninginstrument for prognostiske faktorer	21
<i>2.7.1 Ørebro screeninginstrument for muskel- og skjelett smerter versus SBT</i>	21
<i>2.7.2 Start Back Screening Tool</i>	22
<i>2.7.3 Start Back Screening Tools prediktive validitet</i>	23
2.8 Studiens rasjonale og forskningshypoteser	24
<i>2.8.1 Forskningshypoteser</i>	24
3.0 METODE	26
3.1 Studiedesign	26
<i>3.1.1 Inklusjons- og eksklusjonskriterier</i>	26
<i>3.1.2 Datainnsamling</i>	27
3.2 Screeninginstrument	27
<i>3.2.1 SBTs psykometriske måleegenskaper</i>	28
3.3 Utfallsmål	29
<i>3.3.1 RMDQs psykometriske måleegenskaper</i>	29
3.4 Oppfølging og frafall	30
3.5 Statistisk analyse	30
<i>3.5.1 Bakgrunnsvariabler</i>	30
<i>3.5.2 Imputering, sensitivitetstester og andre statistiske tester</i>	31

3.5.3	<i>Sensitivitet og spesifisitet</i>	32
3.5.4	<i>Positiv og negativ likelihood ratio</i>	33
3.5.5	<i>Arealet under Receiver Operating Characteristics Curve</i>	33
3.6	Etikk	34
4.0	RESULTATER	35
4.1	Baselinekarakteristikker	35
4.2	Klinisk forløp 3 måneder etter baseline	39
4.3	Sensitivitet, spesifisitet og likelihood ratioer ved anbefalte cut-off-verdier	40
4.4	Diskriminerende evner målt med arealet under ROC-kurve	42
4.4.1	<i>Sensitivitet spesifisitet og likelihood ratioer ved bruk av optimal cut-off verdi på SBT</i>	43
5.0	DISKUSJON	46
5.1.1	Metodediskusjon	46
5.1.1	<i>Studiedesign</i>	46
5.1.2	<i>Datainnsamling</i>	46
5.1.3	<i>Utvalg og generaliserbarhet</i>	47
5.1.4	<i>Utfallsmål</i>	49
5.1.5	<i>Oppfølgingstid</i>	52
5.1.6	<i>Frafall</i>	53
5.1.7	<i>Analysemetoder</i>	54
5.2	Resultatdiskusjon	57
5.2.1	<i>Resultatene sett opp mot forskningshypotesene</i>	57
5.2.2	<i>Sammenligning av resultater og tidligere forskning</i>	59
5.2.3	<i>Resultater sammenlignet med andre studier av Start Back Screening Tool</i>	60
5.3	Kliniske implikasjoner hos eldre med ryggmerter	67
5.4	Videre forskning	70
6.0	KONKLUSJON	71
7.0	REFERANSELISTE	72

1.0 INNLEDNING

Ryggsmerter er den hyppigste muskel- og skjelettlidelsen og forårsaker flest år levd med funksjonsbegrensninger på verdensbasis (1, 2). Ryggsmerter er også den muskel- og skjelettlidelsen som utgjør størst samfunnskostnader i Norge i form av sykefravær og trygdeytelser. I Norge opplever 60-80 % ryggsmerter i løpet av livet (3). Studier viser at forekomsten av ryggsmerter øker med alderen, for så å flate ut ved 60-70års-alderen (4-7). Samtidig vil andelen eldre i befolkningen øke betydelig de neste tiårene (1).

En rekke fysiske og psykososiale prognostiske faktorer er assosiert med utviklingen fra første episode ryggsmerter til langvarige ryggsmerter. I denne sammenheng regnes prognostiske psykososiale faktorer som spesielt viktige (8). Det har derfor blitt utviklet en rekke screeningsinstrumenter for å kunne identifisere prognostiske psykososiale faktorer tidlig i et behandlingsforløp. På denne måten kan klinikerne i en tidlig fase planlegge og tilpasse behandlingsstrategier for å forebygge utvikling av residiverende eller vedvarende ryggsmerter (9, 10).

«The Keele STarT Back Screening Tool» (SBT) er et screeninginstrument som ble utviklet for å identifisere modifierbare prognostiske fysiske og psykososiale faktorer hos personer med uspesifikke korsryggsmerter i primærhelsetjenesten. Formålet med å benytte SBT er å klassifisere pasienter i ulike risikogrupper for langvarige ryggsmerter, som igjen indikerer ulike behandlingstilnærminger (11). Stratifisert behandling basert på SBT har vist bedre behandlingseffekt på smerte og funksjon og høyere kostnadseffektivitet enn annen retningslinjebasert behandling hos personer med korsryggsmerter (12-15). Videre har SBT vist gode evner til å predikere dårlige ryggrelaterte utfall hos personer i yrkesaktiv alder, spesielt i primærhelsetjenesten (11, 16-32).

I lys av en stadig eldre befolkningssammensetning er det behov for å vurdere tiltak for å bedre behandlings- og kostnadseffektiviteten av ryggbehandling også hos eldre. Kun en studie har undersøkt SBTs prediktive evner hos eldre med ryggsmerter. Studien fant at SBTs risikoklassifisering predikerte funksjonsbegrensninger etter 6 måneder hos eldre ≥ 65 år. Forskerne fant imidlertid også at SBT ikke hadde evne til å diskriminere mellom dårlige funksjonsutfall i medium- og høyrisikogruppen i deres utvalg. Norge har et annet språk, helsesystem og muligens andre sosiokulturelle normer knyttet til ryggsmertene enn USA (25).

Det vil derfor være viktig å kartlegge SBTs prediktive evner hos eldre med ryggsmarter også i Norge. På den måten kan klinikerne planlegge og tilpasse målrettede behandlingsstrategier til den enkelte som søker behandling for ryggsmarter for å forebygge utvikling og opprettholdelse av ryggsmarter. Dette er også i tråd med samhandlingsreformen, hvor økt kvalitet i helsetjenestene og fokus på forebyggende helsearbeid hos eldre er nevnt som vesentlig for å kunne yte bærekraftige helsetjenester i fremtiden (33).

1.1 Hensikt med masterstudien og forskningsspørsmål

Hensikten med masterstudien var å undersøke SBTs prediktive validitet i form av sensitivitet, spesifisitet, likelihood ratio og diskriminerende evner hos eldre personer som oppsøkte primærhelsetjenesten for behandling av sine ryggsmarter. For å undersøke dette skulle følgende forskningsspørsmål besvares:

1. Hva er Start Back Screening Tools evne til å predikere et dårlig utfall etter 3 måneder hos eldre personer i primærhelsetjenesten?
2. Hva er Start Back Screening Tools evne til å diskriminere mellom et godt og dårlig utfall 3 måneder etter baseline?
3. Videre; Hvilken cut-off-verdi på Start Back Screening Tool er optimal for å skille mellom et godt og dårlig utfall etter 3 måneder i det foreliggende datamaterialet av eldre som oppsøker primærhelsetjenesten for behandling av sine ryggsmarter?

Det er fremsatt hypoteser for hvert forskningsspørsmål i kap. 2.8.1.

1.2 Prosjektorganisasjon

Studien er et mastergradsprosjekt tilknyttet master i fysioterapi ved Oslo Metropolitan University. Masterstudien er videre tilknyttet den norske armen av en internasjonal prospektiv kohortstudie, BACE – Back Complaints in Elders. BACE-studien er et samarbeidsprosjekt som springer ut fra et internasjonalt konsortium som er ledet av professor Chris Maher (Australia) og Bart Koes (Nederland). Den norske armen (BACE-N) ble etablert i 2015 og ledes av Professor Margareth Grotle. Den norske studien har 2 års oppfølgingstid. Det overordnede formålet med BACE-studien er å få økt kunnskap om det kliniske forløpet, faktorer som påvirker ryggsmertene og funksjonsproblemene hos eldre personer ≥ 55 år som oppsøker primærhelsetjenesten med en ny episode av rygg smerter. Datamaterialet fra de ulike samarbeidsnasjonene skal på sikt inngå i en felles database som gir mulighet til å studere rygg smerter hos eldre på tvers av flere land. I denne masterstudien er hovedveileder og forskningsansvarlig for prosjektet professor Kjersti Storheim, mens medveileder er doktorgradsstipendiat Ørjan Nesse Vigdal.

1.3 Begreper og definisjoner

1.3.1 Definisjoner, karakteristikk og bruk av begrepet rygg smerter

Det internasjonale diagnoseregisteret: I den internasjonale statistiske klassifikasjonen av sykdommer og helseproblemer (ICD-10) klassifiseres rygg lidelser som lidelser som har opphav fra sakral- til oksipito-atlanto-aksialregionen. Grovt klassifiseres deformerende rygg lidelser i diagnosekode (M40-43) og lidelser i rygg søylen (M45-49) slik som ankyloserende spondylitt (M45) eller spondylose (M47). Diagnosekoden rygg smerter (M54) tilhører siste hovedklassifikasjon, andre rygg lidelser (M50-54). Diagnosekoden rygg smerter tar innunder seg flere diagnosekoder, deriblant lumbago med og uten isjialgi, rygg smerte i thorkalsegmentet, uspesifisert rygg smerte og annen spesifisert rygg smerte (34).

Definisjoner og begreper i litteraturen: Det eksisterer en rekke ulike definisjoner på rygg smerter i litteraturen og det ser ut til å være lite konsensus rundt hvilke definisjoner og begreper som skal benyttes. Den store majoriteten av forskningslitteraturen omhandler korsrygg smerter, på engelsk «low back pain». I norske kliniske retningslinjer er

korsryggsmerter definert ut fra anatomisk utbredelse, som smerter i området mellom nederste ribbein og glutealfoldene med eller uten utstråling til underekstremitetene (8).

Korsryggsmerter som ikke er assosiert med alvorlig underliggende patologi er i litteraturen også benevnt som «uspesifikke», «mekaniske», eller «muskulære» korsryggsmerter (9, 35).

Klassifisering ut fra varighet: I litteraturen klassifiseres ryggsmerter ofte på bakgrunn av varighet siden den akutte episoden inntraff. Korsryggsmerter betegnes ofte som akutte ved varighet under 3 måneder, subakutte mellom 6 og 12 uker og langvarige når smerteepisoden har vart i mer enn 3 måneder (8). I senere tid er det i større grad gått bort fra denne definisjonen, da tilbakevendende episoder er vanlig. Korsryggsmerter er derfor ansett som en langvarig tilstand, med et fluktuerende forløp og ikke en serie med urelaterte episoder (2).

1.3.2 Andre begrepsdefinisjoner

Eldre: Defineres i masterstudien som personer ≥ 55 år. Når det refereres til studier av «eldre» er disse også basert på utvalg med personer som er ≥ 55 år.

Prediktiv validitet: Er sammen med samtidig validitet en undergruppe av begrepsvaliditet. Prediktiv validitet viser til hvorvidt score på et måleinstrument predikerer score på gullstandarden i fremtiden. Samtidig validitet viser derimot til hvorvidt måleinstrumentet samsvarer med score på gullstandarden når de blir målt på samme tidspunkt (36, s.159).

Prognostisk faktor: Dette er en spesifikk variabel hos en person som har prediktiv verdi for et negativt utfall (37), som i denne studien er langvarige funksjonsbegrensende ryggsmerter. Det er imidlertid viktig å skille mellom risikofaktorer for første episode med ryggsmerter og prognostiske faktorer for residiverende eller vedvarende ryggsmerter, som forteller om prognosen. Prognostiske faktorer kan eksempelvis være demografiske faktorer, pasientkarakteristikker og fysiologiske eller psykososiale faktorer. Prognostiske faktorer interagerer ikke med eventuelle behandlingstiltak og skiller seg fra effektmedierende- eller modererende faktorer, som interagerer med behandlingstiltak og påvirker dets effekt (38). En variabel kan imidlertid være både prognostisk for et dårlig utfall og mediere/moderere behandlingseffekt, slik som for eksempel negative forventninger om bedring av ryggsmerter.

Cut-off-verdi: Viser til den verdien på en ordinal kategorisk eller en kontinuerlig skala som deler utvalget i to. Cut-off-verdien indikerer i denne studien skillet mellom «et dårlig utfall» eller «et godt utfall».

2.0 TEORI OG TIDLIGERE FORSKNING

2.1 Forståelsen av ryggmerter i et historisk og biopsykososialt perspektiv

Det har skjedd et paradigmeskifte innen helsesektoren i vestlige land, fra en biomedisinsk forståelse av helse på 1800- og 1900-tallet til dagens biopsykososiale forståelse av helse. Georg Engel mente at psykologiske faktorer slik som tanker og følelser, og sosiale faktorer slik som familie, kultur og religion ikke nødvendigvis kan forklares i fysiologiske lover, som er utgangspunktet i den biomedisinske modellen. Videre fremhevet Engel at biologiske, psykologiske og sosiale faktorer er dynamiske og gjensidig forbundet med hverandre (39).

Gordon Wadell tilpasset den biopsykososiale forståelsesmodellen til personer med ryggmerter (40, 41). Wadell påpekte at ryggmerter er et symptom og ikke en sykdom (42, s. 2). Det biopsykososiale perspektivet medfører derfor at ryggmerter ikke nødvendigvis skal forstås som forårsaket av vevsskade, men som en plage som påvirkes av fysiologiske, psykologiske og sosiale faktorer.

2.2 Epidemiologi og befolkningsvekst

Korsryggmerter er den hyppigste muskel og-skjelettlidelsen og forårsaker flest år levd med funksjonsbegrensninger verden over (1). I Norge står ryggmerter for 11.6 % av alle år levd med funksjonsbegrensninger, et tall som er høyere enn i verden for øvrig (43). Internasjonalt er det estimert at gjennomsnittlig punktprevalens for aktivitetsbegrensende korsryggmerter >1 dag er 11.9 %, mens månedsprevalensen er 23.2 %. (7). I Norge angir de fleste studier at mellom 60-80 % vil oppleve ryggmerter en eller flere ganger i løpet av livet (3).

Det finnes få epidemiologiske studier av ryggmerter hos personer ≥ 55 år isolert (44). Eksisterende studier rapporterer imidlertid at ryggmerter er en minst like hyppig årsak til funksjonsbegrensninger hos eldre (45, 46). Flere studier rapporterer at forekomsten øker med alderen og flater ut rundt 60-70års-alderen (4-7). Videre er forekomsten av alvorlig funksjonsbegrensende ryggmerter rapportert å øke etter 65-årsalderen (5). I Norge er det også vist at langvarige muskel- og skjelettsmerter øker med alderen (47). Samtidig forventes det i utviklingsland en betydelig økning i antallet eldre de neste 10-årene (1). I perioden 2005

til 2015 økte den forventede levealderen på 82 år med 22 måneder i Norge (48, 49). Andelen eldre over 67 år er forventet og fordobles fram mot 2050 (33). Det er altså stor variasjon i rapportering om forekomsten av ryggsmarter, som igjen vanskeliggjør sammenligning på tvers av studier og gjør at resultater fra systematiske oversiktsartikler må tolkes med forsiktighet (7). Noe av årsaken kan være at studier av ryggsmarter gjennomgående har stor variasjon i blant annet definisjonen av ryggsmarter, karakteristikk ved populasjonen som studeres (5, 35, 45, 50) samt prevalensperiode (7).

2.3 Samfunnskostnader

Ryggsmarter er også en av de tilstandene som utgjør størst samfunnskostnader i Vesten (51). Samfunnskostnadene er i stor grad knyttet til bruk av helsetjenester, sykefravær og trygdeytelser. I Norge doblet samfunnskostnadene knyttet til muskel og-skjelettplager seg i perioden 2004-2009. I 2009 utgjorde kostnadene knyttet til muskel og skjelettplager mellom 69 og 73 milliarder kroner (3). Videre kontaktet i underkant av 1.2 millioner fastlegen for behandling av muskel- og skjelettlidelser i 2015. Den største andelen av disse var middelaldrende menn og eldre kvinner (52). Ryggsmarter sto for 3.6 % av alle fastlegekonsultasjoner i 2018 (53), mens muskel og- skjelettlidelser utgjorde 31 % av alle legemeldte sykefraværstilfeller i 2019 (54). I utgangen av 2015 utgjorde ryggglidelser 11 % av alle uføretilfeller (55). Samtidig var andelen legemeldt sykefravær uavhengig av diagnosegruppe størst i alderen 55-64 år (56). Selv om kostnadene knyttet til muskel- og skjelettplager har steget kraftig på 2000-tallet har kostandene knyttet til ryggsmarter blitt mindre (57). Dette kan ha årsak i at behandlere vektlegger opprettholdelse av normal aktivitet og retur til arbeid for personer med ryggsmarter (3).

2.4 Etiologi, undersøkelse og diagnostisering

Det er en rekke biologiske og psykososiale faktorer som kan forårsake ryggsmarter (2). Kliniske retningslinjer verden over har derfor anbefalt å benytte den diagnostiske triaden i anamnesen og den kliniske undersøkelsen (8, 58). Formålet er å ekskludere spesifikk patologi, radikulopati eller alvorlig underliggende patologi som årsak til ryggsmertene (8, 9). Siden

årsaksforholdet ofte er sammensatt er det også viktig å kartlegge grønne og gule flagg som gir informasjon om prognosen for bedring (8) og dermed vil være viktig for hvilken oppfølging som blir gitt.

80-90 % av korsryggsmertene i den generelle befolkningen defineres som uspesifikke, altså uten kjent patoanatomisk årsak (8). Det er en rekke strukturer i og rundt ryggspylen som er innervert med sensoriske nerver og som derfor kan være biologiske årsaker ved uspesifikke rygg smerter (9). Dette inkluderer eksempelvis mellomvirvelskiver, fasettledd, muskulatur, og senefester. Radikulopater utgjør 5-10 % av alle korsrygg smerter (8). Opp mot 90 % av radikulopater er forårsaket av prolaps av en mellomvirvelskive (59), mens spinal stenose er en annen vanlig årsak (3).

Alvorlig underliggende patologi kartlegges ved bruk av røde flagg og utgjør 1-5 % av de som rapporterer rygg smerter i den generelle befolkningen. Av alvorlige underliggende årsaker er fraktur, neoplasme, inflammatorisk og infeksjons rygg lidelse, nyrestein, cauda equina syndrom og aortaaneurisme mest vanlig (8). Den diagnostiske nøyaktigheten til røde flagg er imidlertid vist å være begrenset (60). Det har også tidlig bildediagnostikk hos eldre (61). Det er rapportert at minst 80 % av personer med akutte korsrygg smerter har minst ett rødt flagg (62). Bildediagnostikk er derfor kun anbefalt dersom det foreligger en klar indikasjon på alvorlig underliggende patologi, da det ikke er evidens for at identifisering av røde flagg bedrer prognosen til rygg pasienter (2). En naturlig årsak til dette er at degenerative forandringer forekommer hyppig også hos den asymptotiske delen av befolkningen, spesielt med økende alder. Det er eksempelvis rapportert at degenerasjon av mellomvirvelskiven eksisterer hos 37 % av 20-åringene og 96 % av 80-åringene uten symptomer på rygg smerter (63, 64). En annen studie av eldre rapporterte at det ikke var noen årsakssammenheng mellom lav beintetthet og rygg smerter hos personer over 55 år (65).

Forekomsten av komorbide tilstander og ulike aldersrelaterte tilstander slik som osteoporose, artrose og spinal stenose er vist å øke med alderen (63, 66, 67). Sarkopeni er en annen aldersrelatert tilstand som innebærer nedsatt muskelmasse og styrke, og ofte også nedsatt funksjon i daglige aktiviteter (68, 69). Det er rapportert høyere forekomst av rygg smerter med årsak i alvorlig underliggende patologi hos eldre enn i den yrkesaktive delen av befolkningen (5, 63). I en nederlandsk BACE-studie ble 43 % diagnostisert med en spesifikk årsak til smertene, hvorav osteoporotiske frakturer utgjorde 26 %. Videre ble 6 % diagnostisert med

alvorlig underliggende patologi, hvorav vertebrale frakturer utgjorde 5 % (63). Disse studiene tyder derfor på at aldersrelaterte tilstander bidrar til at det er en annen fordeling av årsaker til ryggmerter hos eldre enn i den generelle befolkningen. Aldersrelaterte tilstander ser også ut til å påvirke det kliniske forløpet av ryggmerter hos eldre.

2.4.1 Risikofaktorer for første episode korsryggmerter

Parreira et al. (70) gir i sitt umbrella review en oversikt over studier som rapporterer om risikofaktorer for første episode med korsryggmerter eller isjiassmerter hos personer over 18 år. Forfatterne rapporterte motstridende resultater med hensyn til om type kjønn er en risikofaktor for utvikling av korsryggmerter. Videre ble lavere sosioøkonomisk status og økende alder rapportert som risikofaktorer (71, 72). En rekke faktorer knyttet til dårlig generell helse ble også rapportert. Dette inkluderte blant annet faktorer som tidligere eller nåværende røyker (72), et overdrevent alkoholinntak (73), alvorlige søvnproblemer (71) og overvekt (72, 74).

Ulike systematiske oversiktsartikler basert på den yrkesaktive delen av befolkningen har også vist at faktorer relatert til fysisk stress øker risikoen for å utvikle korsryggmerter. Det er blant annet vist et positivt dose-respons forhold mellom tunge løft og risikoen for ryggmerter. Frekvensen av tunge løft er også av betydning (75). Andre arbeidsrelaterte prognostiske faktorer som er rapportert er å kjøre bil over to timer (72), å sitte på kne over 15 min, å gjøre knebøy over 15 min, å jobbe med hendene over skuldrene, å jobbe med over 60 grader fleksjon av trunkus i over 5 min, samt å stå eller gå over lengre tidsperioder (71). En rekke fysiske risikofaktorer som ikke nevnes her er også rapportert (70). Samtidig kan det se ut til at psykososiale faktorer er mindre dokumentert som risikofaktorer for utvikling av første episode ryggmerter enn som prognostiske faktorer.

2.5 Behandling

Det har vært forsket på en rekke behandlingstiltak for korsryggsmerter med og uten radierende beinsmerter i den generelle befolkningen. (58, 76). Trening alene eller i kombinasjon med pasientinformasjon har vist størst effekt på å forebygge nye episoder med korsryggsmerter. Disse tiltakene er anbefalt som grunnbehandling ved korsryggsmerter (76). Det eksisterer lite forskning på den eldre aldersgruppen isolert (77). Fysisk aktivitet har imidlertid vist en positiv effekt på smerter og funksjonsbegrensninger også hos eldre (78). Ved langvarige korsryggsmerter er også veiledet trening og kognitiv atferdsterapi anbefalt som grunnbehandling (76).

Dersom det skal være grunnlag for å benytte et screeninginstrument i klinisk praksis må det ikke bare kunne predikere prognose. Aktuelle behandlingstiltak må også redusere risikoen for at et negativt klinisk utfall oppstår (79, s. 119, 120). Studier har vist at behandlingsstrategier tilpasset SBTs risikoklassifisering er mer effektive for å øke funksjonsnivået i daglige aktiviteter og har høyere kostnadseffektivitet enn behandlingstilnæringer som ikke er basert på SBTs risikoklassifisering (12-15). Dette fordi en unngår overbehandling spesielt i lavrisikogruppen, samt for lite og spesifikk behandling av høyrisikogruppen (13, 80). I en nyere studie ble det derimot funnet at klinikere i primærhelsetjenesten benyttet SBT på omlag halvparten av sine pasienter, men at dette ikke påvirket hvilken behandling som ble gitt. SBTs risikostratifisering påvirket derfor hverken pasientutfall eller bruk av helsetjenester (81). Videre er det satt i gang en ny studie for å undersøke behandlingseffektivitet og kostnadseffektivitet ved bruk av SBT i primærhelsetjenesten i USA (82).

2.6 Prognose

Prognose omhandler å kunne forutsi det sannsynlige forløpet og utfallet av en sykdom eller tilstand (79, s. 138). I dette kapitlet redegjøres det for forekomsten av gode eller dårlige kliniske utfall ved ryggsmerter. En systematisk oversiktsartikkel rapporterte at mellom 22 % og 77 % av personer med korsryggsmerter opplevde residiv innen et år. Forfatterne konkluderte med at 33 % residiv i løpet av ett år var det beste estimatet (83). Hos personer med en smertevarighet under 3 uker er det rapportert at de fleste opplevde rask bedring med

hensyn til smerte- og funksjonsbegrensninger, med en gjennomsnittlig reduksjon på 58 % innen en måned. Etter dette rapporterte studiedeltagerne lite bedring i smerte- og funksjonsbegrensninger frem til 3 måneder. (84, 85). Lignende resultater ble rapportert i en oversiktsartikkel som sammenlignet personer med en smertevarighet på inntil 12 uker og >12 uker. Begge gruppene opplevde en markant bedring med hensyn til smerte og funksjon de første 6 ukene og deretter lite bedring frem til ett år. (86).

Det er også hos eldre rapportert at den største endringen i smerte og funksjon skjer de første 3 månedene etter en ny episode ryggmerter. I studier av eldre med ryggmerter av ulik varighet er det rapportert at mellom 58 % og 62 % har et dårlig utfall med hensyn til smerte og funksjonsrelaterte utfall etter 3 måneder (se tabell 1) (87-89). Videre hadde mellom 47 % og 68 % har et dårlig utfall etter ett år (65, 88-90). En studie rapporterte også at mer enn halvparten av studiedeltagerne hadde residiverende eller vedvarende ryggmerter etter 5 år (89). Det ser altså ut til at det er en høyere andel eldre som rapporterer dårlige smerte- og funksjonsrelaterte utfall, sammenlignet med personer med korsryggmerter i den generelle befolkningen.

Tabell 1. Oppsummering av studier av prognose som det er referert til i kapittel 2.6. Studiene er gjennomført på eldre personer (≥ 55 år) med ryggsmarter i primærhelsetjenesten.

Studie	Klinisk setting	Definisjon av dårlig utfall + oppfølgingstidspunkt	%andel m/dårlig utfall
Jarvik et al. (93)	Primærh.tj., USA ≥ 65 år Oppsøkt primh.tj. siste 3 uker pga RS	RMDQ (0-24), < 30 % bedring. Oppfølgingstp.: 2 år	55-58 %
Deyo et al. (90)	Primærh.tj./sekh.tj., USA ≥ 65 år. Pga en ny episode RS	RMDQ (0-24) <30 % bedring Oppfølgingstp.: 1 år	59 %
Rundell et al. (92)	Primærh.tj., USA ≥ 65 år. Oppsøkt primh.tj. pga ny episode RS	RMDQ (0-24) >4 Oppfølgingstp.: 1 år	60 %
Rundell et al. (88)	Primærh.tj., USA ≥ 65 år. Oppsøkt primh.tj. pga ny episode RS	RMDQ (0-24) <30 % bedring Oppfølgingstp.: 3 mnd, 1 år	3 mnd: 58 % 1 år: 55 %
Scheele et al. (87)	Primær.h.tj., Nederland >55 år. Oppsøkt fastlege pga ny episode RS	GPE (1-7 likertskala) Dårlig utfall: «Noe bedre», «som før», «noe verre», «mye verre» eller «verre enn noen gang» Oppfølgingstp: 6 uker, 3 mnd.	6 uker: 64 % 3 mnd: 61 %
Kim et al. (65)	Primærh.tj. Nederland >55 år, oppsøkt fastlege pga RS	NRS (0-10) Relativ vedvarende RS: < 30 % bedring. Absolutt vedvarende RS: >1 poeng Oppfølgingstp.: 1 år	Relativ vedvarende RS: 47 % Absolutt vedvarende RS: 68 %
van Der Gaag et al. (89)	Primærh.tj. Nederland >55 år, oppsøkt fastlege pga RS	GPE (1-7 likertskala) Dårlig utfall: «Noe bedre», «som før», «noe verre», «mye verre» eller «verre enn noen gang». Oppfølgingstp: 3 mnd, 5 år.	3 mnd: 62 % 5 år: 57 %

Studiene er gjennomført i primærhelsetjenesten dersom ikke annet er oppgitt. Forkortelser: Primærh.tj. = primærhelsetjenesten. Sek.htj.= sekundærhelsetjenesten. Oppfølgingstp.= oppfølgingstidspunkt. RS= ryggsmarter. NRS= numerisk smerteskala. RMDQ= Roland Morris Disability Questionnaire. GPE= Global Percieved Effect. Pga= på grunn av. M= med.

2.7 Screeninginstrument for prognostiske faktorer

Hensikten med screeninginstrumenter er å identifisere uoppdaget sykdom, presymptomatisk sykdom eller prognostiske faktorer for sykdom (79, s. 132). Da det er vist at klinikere er lite samstemte med hensyn til hvem som har forøket risiko for langvarige plager og ikke (103) virker det hensiktsmessig å benytte et screeninginstrument i kartleggingen av ryggpasienter. Personer med ryggmerter er en heterogen populasjon som trolig nyttiggjør seg ulike behandlingsalternativer og det har derfor blitt utviklet en rekke screeninginstrumenter for å bidra i den kliniske beslutningstagningen (104-110).

Noen screeninginstrumenter klassifiserer på bakgrunn av underliggende mekanisme for smertene slik som anatomiske, patologiske eller psykologiske faktorer, for så å rette behandlingen mot disse faktorene. Flere screeninginstrumenter er også laget for å identifisere pasienter som kan nyttiggjøre seg spesifikke behandlingsmodaliteter slik som eksempelvis manuelle teknikker eller spesifikk trening (104, 105). Enkelte screeninginstrumenter stratifiserer pasientene på bakgrunn av prognostiske faktorer for vedvarende funksjonsnedsettende ryggmerter (9). Dette inkluderer SBT og Ørebro screeningskjema for muskel- og skjelettsmerter. Disse screeninginstrumentene er mest benyttet i Norge, og er anbefalt i nyere kliniske retningslinjer (9).

2.7.1 Ørebro screeninginstrument for muskel- og skjelett smerter versus SBT

Ørebro screeningskjema for muskel- og skjelettsmerter ble utviklet for screening av prognostiske psykososiale faktorer for langvarig sykefravær hos personer med nakke- eller ryggmerter. Screeninginstrumentet består av 21 spørsmål som retter seg mot sinnsstemning, holdninger til arbeid, samt tanker, overbevisninger og atferd knyttet til ryggsmertene (110). I senere tid har det også blitt utviklet en kortversjon av screeningsinstrumentet som har vist høy korrelasjon med originalversjonen og tilnærmet like gode prediktive evner (111). Fullversjonen av Ørebro screeningskjema for muskel- og skjelettsmerter er oversatt til norsk av Grotle og Vøllestad i 2001 og har demonstrert akseptable måleegenskaper med hensyn til test-retest-reliabilitet og intern konsistens hos personer med korsryggmerter (112).

Ørebro screeningskjema for muskel- og skjelettsmerter kartlegger prognostiske faktorer for nedsatt arbeidsevne grunnet smerter lokalisert til rygg, nakke eller skuldre. Screeninginstrumentet kartlegger arbeidssituasjon, sykefravær, smertebilde, funksjonsbegrensninger, smertemestring, angst, depresjon, samt frykt- og unngåelsestanker (110). Videre har screeninginstrumentet vist seg fordelaktig sammenlignet med SBT når det gjelder å identifisere livsstilsfaktorer som kan være prognostiske for negative ryggrelaterte utfall (113). SBT gir derimot prognostisk informasjon om funksjonsnedsettende ryggmerter i daglige aktiviteter. SBT retter seg dermed i større grad mot befolkningen ≥ 55 år, siden en stor del av disse vil være ute av arbeidslivet. Videre er det en styrke med SBT sammenlignet med Ørebro screeningskjema for muskel- og skjelettsmerter at det både kan predikere prognose og samtidig indikerer ulike behandlingsstrategier. SBT inneholder færre delspørsmål enn fullversjonen av Ørebro screeninginstrument for muskel- og skjelettsmerter og tar kortere tid å gjennomføre. Det er også hevdet at SBT er bedre egnet til å predikere funksjonsutfall enn Ørebro screeningskjema for muskel- og skjelettsmerter (114).

2.7.2 Start Back Screening Tool

Start Back Screening Tool baserer seg på den biopsykososiale modellen og har i løpet av få år blitt et av de mest brukte screeninginstrumentene for pasienter med korsryggmerter (9). Screeninginstrumentet ble utviklet for å identifisere modifiserbare prognostiske faktorer hos personer med korsryggmerter, og samtidig klassifisere pasienter i ulike risikogrupper med tilhørende behandlingsstrategi. I utgangspunktet skulle dette hjelpe fastlegenes første valg av behandlingsstrategi i primærhelsetjenesten (11).

Basert på risikogrupperingen anbefales det at lavrisikogruppen skal motta 1-2 oppfølgingstimer av sin fastlege for informasjon og råd basert på kliniske retningslinjer. Medium risikogruppe blir i tillegg henvist til fysioterapeut for retningslinjebaserte tiltak. Høyriskogruppen mottar samme tiltak som lav- og medium risikogruppen og skal i tillegg henvises for en psykologisk tilnærming basert på prinsipper i kognitiv atferdsterapi. Behandlingen skal rettes mot pasientenes risikoprofil (13).

Dersom SBTs risikoklassifisering viser til gode prediktive evner også hos den eldre delen av befolkningen kan risikogrupperingen brukes til å planlegge og styre behandling.

Behandlingsmetodene vil dermed kunne gi større effekt ved at de benyttes hos personer som i større grad kan nyttiggjøre seg dem. Samtidig vil SBT også kunne gi økt grunnlag for å individualisere ulike behandlingsmetoder, slik som for eksempel kognitive teknikker til den enkelte. Dette kan argumenteres for å være spesielt viktig hos eldre, som er en gruppe som ofte blir underbehandlet eller feilbehandlet (51). Samtidig er det i en nordisk studie vist at helsepersonell kommuniserer dårligere med denne pasientgruppen (115).

2.7.3 Start Back Screening Tools prediktive validitet

SBT har demonstrert gode prediktive evner hos pasienter i yrkesaktiv alder i primærhelsetjenesten. Pasientene i de ulike studiene hadde ryggsmarter både med og uten utstrålende beinsmerter og ulik grad av symptomvarighet (11, 16-30). SBT har imidlertid rapportert dårligere prediktive evner hos pasienter som følges opp av kiropraktor sammenlignet med pasienter som følges opp av fysioterapeut eller fastlege (24, 116-119). Videre har SBT demonstrert begrensede prediktive evner i sekundærhelsetjenesten (120-123). To studier har likevel demonstrert prediktiv nytteverdi også i sekundærhelsetjenesten (31, 32). Videre har SBT demonstrert dårligere prediktive evner benyttet på personer med under 2 ukers symptomvarighet (21, 116).

To studier har rapportert at endringscoren på SBT mellom baseline og henholdsvis 4 og 5 uker har bedre prediktiv verdi enn baselinescoren alene (17, 20). Videre har SBT demonstrert like god evne til å predikere funksjonsutfall som Ørebro screeningskjema for muskel- og skjelettsmerter og bedre prediktive evner enn Vermont Disability Prediction Questionnaire (18). To studier har også rapportert at SBTs psykososiale subscore har like god eller bedre evne til å predikere dårlig utfall sammenlignet med totalscoren (16, 26). Selv om det er gjennomført en rekke studier for å kartlegge SBTs prediktive validitet er det så vidt oppgaveforfatteren kjent kun én studie som har vurdert SBTs prediktive evner hos personer ≥ 55 år isolert. Forfatterne rapporterte at SBT predikerte funksjonsnedsettelse hos personer ≥ 65 år etter 6 måneder. SBT kunne imidlertid ikke diskriminere mellom personer med et dårlig utfall i henholdsvis medium og høy risikogruppe etter 6 måneder (25).

Det er også utviklet to kortversjoner av SBT på henholdsvis 5 og 6 spørsmål (124). Videre er det utviklet en generisk versjon ved navn Start MSK Tool (125), som er tilpasset bruk på muskel- og skjelettplager uavhengig av smerteregion. Den generiske versjonen har demonstrert gode prediktive evner hos personer med smerte lokalisert til fem ulike kroppsregioner (126).

2.8 Studiens rasjonale og forskningshypoteser

Ryggsmerter er den lidelsen som forårsaker flest år levd med funksjonsbegrensninger (1). I tillegg til store helsemessige konsekvenser for den enkelte medfører ryggsmerter også store samfunnskostnader og bruk av helsetjenester (3). Ryggsmerter forstås på bakgrunn av et komplekst samspill mellom biologiske, psykologiske og sosiale faktorer. Psykososiale prognostiske faktorer er ansett som spesielt viktige for utvikling av langvarige ryggsmerter og det har derfor vært blitt utviklet en rekke screeninginstrumenter for å fange opp disse. SBT er anbefalt i nyere kliniske retningslinjer (9) og har vist gode prediktive evner benyttet på den generelle befolkningen i primærhelsetjenesten (11, 16-32). Det er imidlertid ikke gjennomført studier av prediktiv validitet i Norge og kun én studie av eldre (25). Da økende alder er identifisert som en prognostisk faktor for langvarige ryggsmerter (88, 91-93), og gjennomsnittsalderen i befolkningen vil øke i årene fremover (48, 49) vil det være viktig å kartlegge SBTs prediktive validitet også i denne aldersgruppen. For å besvare forskningsspørsmålene gjengitt i kapittel 1.1 ble derfor følgende forskningshypoteser fremsatt:

2.8.1 Forskningshypoteser

1. Start Back Screening Tools evne til å predikere et dårlig utfall etter 3 måneder er moderat for sensitivitet og spesifisitet, og lav for positiv og negativ likelihood ratio (Dette ble forventet på bakgrunn av funn gjort i studien av Suri et al. (25)).

2. Start Back Screening Tool har lave til moderate evner til å diskriminere mellom et godt og dårlig utfall etter 3 måneder. (Dette ble forventet på bakgrunn av studiene til Kongsted et al. (119) og Morsø et al. (22)).

3. Den optimale cut-off-verdien for å skille mellom et godt og dårlig utfall i dette utvalget vil være lavere enn den optimale cut-off-verdien som ble identifisert av Beneciuk et al. (127) (≥ 5 poeng på SBTs totalscore). Dette ble forventet da det er vist at det er flere som scorer lavt på totalscoren på SBT som har et dårlig utfall i aldersgruppen ≥ 65 år (25) enn i den yrkesaktive delen av befolkningen (127).

3.0 METODE

3.1 Studiedesign

Datamaterialet som ble benyttet i masterstudien tilhører BACE-Nstudien. Alle studiedeltagerne inkludert i BACE-Nstudien mottok oppfølgingsspørreskjemaer elektronisk eller per post etter 3, 6, 12 og 24 måneder. I masterstudien ble det kun benyttet datamateriale som var innsamlet ved baseline og 3 måneders oppfølging. Pasientene ble rekruttert fra praksispoliklinikkene ved fysioterapiutdanningen ved OsloMet og i primærhelsetjenesten i Viken, Trøndelag og Vestlandet.

3.1.1 Inklusjons- og eksklusjonskriterier

I masterstudien ble samme inklusjon- og eksklusjonskriterier som i den norske BACE-Nstudien benyttet. Kvinner og menn ≥ 55 år som hadde kontaktet primærhelsetjenesten (lege, fysioterapeut eller kiropraktor) for behandling grunnet en ny episode av rygg smerter ble inkludert. Rygg smerter ble definert som smerter mellom scapula og første sakrale vertebra. En ny episode med rygg smerter ble definert som at studiedeltagerne ikke hadde kontaktet primærhelsetjenesten for ryggplagene i løpet av de siste 6 månedene. Studiedeltagerne kunne heller ikke ha oppsøkt en annen primærbehandler de siste 6 månedene. Personer med kognitive vansker, manglende skriftlige og muntlige norskkunnskaper, samt større synsproblemer som hindret dem i å besvare spørreskjemaene ble ekskludert. Videre ble personer som ikke kunne fullføre den fysiske undersøkelsen som inngår i testbatteriet for BACE-Nstudien på grunn av store mobilitetsproblemer (for eksempel rullestolbrukere) ekskludert. Mistenkt underliggende patologi eller aldersrelaterte plager som osteoporotiske brudd eller spinal stenose var ikke eksklusjonskriterier. Et spesifikt inklusjonskriterium for masterstudien var at studiedeltagerne måtte ha svart fullstendig på SBT ved baseline.

3.1.2 Datainnsamling

I masterstudien ble datamateriale fra BACE-Nstudien som var samlet inn mellom april 2015 og oktober 2018 benyttet. Baselinekartleggingen skulle optimalt sett skje før oppstart av behandlingen, og senest etter 3-4 uker. Ved baselinekartleggingen besvarte studiedeltagerne først standardiserte spørreskjema elektronisk på Infopad. Deretter ble det gjennomført en standardisert poliklinisk baselinekartlegging bestående av en klinisk undersøkelse og fysiske funksjonstester. Resultatene fra den fysiske funksjonsundersøkelsen ble ikke benyttet til å presentere utvalget i masterstudien. Totalt tok baselinekartleggingen cirka 1.5 timer per studiedeltager. I etterkant av den kliniske undersøkelsen besvarte studiedeltagerne også to spørreskjemaer på egenhånd som ikke kunne besvares elektronisk på Infopad. Disse spørreskjemaene ble deretter returnert til forskningskoordinatoren i BACE-N elektronisk eller per post. Forskningskoordinatoren dobbeltsjekket inkluderte pasienter i etterkant, ut fra inklusjon- og eksklusjonskriteriene. I masterstudien ble det kun benyttet et selektert utvalg av datamaterialet som kunne bidra til å belyse problemstillingene og beskrive pasientmaterialet (RMDQ, SBT, samt demografiske bakgrunnsvariabler). Masterstudenter, doktorgradsstipendiater, samt ulike behandlere deltok i testingen ved baseline. Oppgaveforfatteren var ikke involvert i innsamling av datamateriale til masterstudien.

3.2 Screeninginstrument

SBT er et selvrapportert screeningsinstrument som inneholder 9 delspørsmål. Disse omhandler modifiserbare fysiske og psykososiale prognostiske faktorer for vedvarende rygg smerter. Spørsmål 1-8 formuleres som påstander med dikotome svaralternativer som besvares med enig eller uenig (gir 0 eller 1 poeng). Spørsmål 9 ble besvares på en 5-punkts likertskala. Dette spørsmålet gir også 0 eller 1 poeng, henholdsvis ut fra svaralternativ 1-3 og 4-5. Spørsmål 1-4 danner den fysiske subscoren, og omhandler spørsmål om referert beinsmerte, nakke- eller skuld smerter, vanskeligheter med gange og vanskeligheter med påkledning. Spørsmål 5-9 danner en psykososial subscore fra 0-5 poeng. Spørsmålene omhandler frykt for fysisk aktivitet, angstsymptomer, smertekatastrofering, depressive symptomer og hvor plagsomme rygg smertene oppleves. Spørsmål 1-9 gir en totalscore på 0-9 poeng. Pasienter med en totalscore på ≤ 3 poeng ble kategorisert i lavrisikogruppen. Personer med en totalscore på over >3 poeng, men en psykososial subscore <4 poeng ble kategorisert i

medium risikogruppe. Personer med en psykososial subscore ≥ 4 poeng ble uansett totalscore kategorisert i høyrisikogruppen (11, 128). I denne studien ble den norske versjonen som er oversatt av Storheim og Grotle i 2012 benyttet (129) (vedlegg 4).

3.2.1 SBTs psykometriske måleegenskaper

SBT har blitt oversatt til en rekke land (124) og har vist valide psykometriske egenskaper benyttet på personer i over 18 år både i primærhelsetjenesten (113, 130-140) og i sekundærhelsetjenesten (141-143). I en nyere systematisk oversiktsartikkel ble det rapportert at tilnærmet alle de oversatte versjonene har vurdert SBTs innholdsvaliditet som perfekt (144). Høyrisikogruppen har vist signifikant høyere score på variabler som omhandler depresjon- og angstsymptomer, frykt og unngåelsestanker, samt negative livsstilsfaktorer sammenlignet med de andre risikogrupperne (113). Videre har en overvekt av studiene rapportert at SBTs totalscore har moderat til høy korrelasjon målt med spearman eller pearson korrelasjonskoeffisient ($r/\rho = 0.45-0.8$) med både fullversjonen og kortversjonen av Ørebro screeningskjema for muskel- og skjelettmerter (29, 134, 136, 137, 142, 145, 146). Det ble likevel påpekt at flere av studiene som undersøkte SBTs begrepsvaliditet ikke hadde prespesifisert hypoteser som sammenheng, som kan ha ført til systematiske skjevheter (144). Seks studier rapporterte at det ikke eksisterte tak- eller gulveffekter (11, 27, 135, 137, 139-141), mens en studie rapporterte om en gulveffekt på 23 % (19). Samlet viser studiene at SBT har demonstrert valide måleegenskaper på tvers av ulike land, kulturer og helsesystemer.

SBT har også demonstrert reliable måleegenskaper ved bruk både i sekundær- og primærhelsetjenesten (11, 19, 132, 133, 137, 140, 143, 146-149). En overvekt av studiene rapporterte om akseptabel intern konsistens målt med cronbachs alfa >0.70 (11, 19, 132, 133, 137, 147, 148). Mens en studie rapporterte verdier rett i underkant av 0.70 (135). Videre rapporterte en rekke studier akseptabel relativ målefeil, målt med ICC-verdier >0.70 (henholdsvis mellom 0.67 og 0.93) (134, 137, 139, 141, 147). Intra-rater-reliabiliteten har i flertallet av studiene blitt rapportert å være akseptabel eller svært god målt med kappaverdier mellom 0.70 – 0.80 (11, 141, 148, 149), mens noen studier også har rapportert kappa-verdier under 0.7 (29, 103, 134, 135, 142, 146). Absolutt reliabilitet målt med standard error of measurement har i en studie blitt rapportert å være 1.9 % (133). Smallest detectable change

for totalscore på SBT har i en norsk studie fra primærhelsetjenesten blitt rapportert å være 1.60 poeng for totalscore og 1.35 poeng for psykososial subscore (139).

3.3 Utfallsmål

RMDQ ble benyttet som utfallsmål og var dermed «gullstandard» i de videre analysene. RMDQ er et selvrapportert kartleggingsinstrument utviklet med bakgrunn i et lengre spørreskjema ved navn Sickness Impact Profile i 1983 for bruk på pasienter med korsryggsmarter (150). RMDQ inneholder 24 spørsmål, hvor 20 spørsmål omhandler funksjonsbegrensninger i daglige aktiviteter relatert til personens korsryggsmarter. De resterende 4 spørsmålene omhandler, smerte, søvn, appetitt og emosjonelle funksjoner. Hvert spørsmål er gir en score på 0 eller 1 poeng og en sumscore mellom 0 og 24 poeng. Null poeng indikerer ingen funksjonsbegrensninger og 24 poeng alvorlige funksjonsbegrensninger. RMDQ tar om lag fem minutter å gjennomføre (150). I denne studien ble den norske versjonen (vedlegg 5) som er oversatt av Grotle og Vøllestad i 2001 benyttet.

I masterstudien ble RMDQ statistisk sett behandlet som et kategorisk utfallsmål. I originalstudien av Hill et al. (11) og i studien av Suri et al. (25), som inkluderte subanalyser av personer ≥ 65 år, ble cut-off-verdi for å indikere et «godt utfall» eller «god selvrapportert funksjon» etter 3 måneder satt til ≥ 7 poeng på RMDQ. Denne cut-off-verdien ble derfor også benyttet i masterstudiens analyser.

3.3.1 RMDQs psykometriske måleegenskaper

RMDQs psykometriske egenskaper har blitt testet på personer over 18 år med korsryggsmarter av ulik varighet og har demonstrert valide, reliable og responsive måleegenskaper. RMDQ har demonstrert god innholdsvaliditet med tanke på forståelighet, men har begrensninger med hensyn til unidimensjonalitet (151, 152). Videre har det ikke blitt rapportert om tak- eller gulveffekter (153). RMDQ har vist høy korrelasjon med andre kartleggingsinstrumenter for selvrapporterte funksjonsbegrensninger (154). RMDQ har også

demonstrert høy til svært høy intern konsistens (155) og høy test-retest-reliabilitet ved baseline og opp til 6 uker etter behandlingsoppstart (154). Absolutt målefeil er i ulike studier rapportert mellom 1.4 og 3.7 poeng i studier med ulikt intervall mellom test og retest. I tillegg har RMDQ demonstrert akseptable egenskaper med hensyn til responsivitet (156). Det er anbefalt å benytte en score på 5 poeng for minste klinisk viktige endring (157). Den norske versjonen har også demonstrert akseptable egenskaper med hensyn til begrepsvaliditet og reliabilitet og hos personer med både akutte og langvarige ryggmerter (158).

3.4 Oppfølging og frafall

Etter inklusjon i BACE-Nstudien ble studiedeltagerne fulgt opp som normalt av sine behandlere. Det vil si at de fikk behandling etter eget behov og ønske i samarbeid med helsepersonell. Behandlende helsepersonell mottok ikke informasjon om testresultater fra baselineundersøkelsen fra prosjektgruppen i BACE-N. Videre fikk masterstudenten tildelt sitt datasett før alle studiedeltagerne som var inkludert i datamaterialet ved baseline hadde besvart RMDQ etter 3 måneder. Disse studiedeltagerne tilhører derfor gruppen med studiedeltakere ikke har besvart RMDQ etter 3 måneder og dermed ikke kunne inkluderes i analysene av prediktiv validitet i masterstudien (figur 1). Det var syttifem (25 %) av studiedeltagerne som ikke hadde besvart RMDQ etter 3 måneder. Tjueni (9.7 %) av disse hadde besvart RMDQ etter 3 måneder i etterkant av at masterstudenten fikk tildelt sitt datasett (data ikke tilgjengelig for masterstudenten). Av de resterende 46 studiedeltagerne som ikke kunne inkluderes i analysene hadde 3 personer også droppet ut av BACE-N-studien. En person oppga at årsaken til frafall var helsemessig begrunnet.

3.5 Statistisk analyse

3.5.1 Bakgrunnsvariabler

Statistikkprogrammet IBM SPSS versjon 25 ble benyttet for å analysere datamaterialet. For å beskrive bakgrunnsvariablene ble deskriptiv statistikk benyttet. Sosiodemografiske bakgrunnsvariabler som ble kartlagt var alder, kjønn, morsmål, utdanningsnivå, om

studiedeltagerne hadde lønnet arbeid, røyking, kroppsmasseindeks (159) og sivilstatus (87, 160). Kliniske variabler som ble benyttet i masterprosjektet var smertevarighet, smerteintensitet (161), utstrålende ryggmerter til bein og tidligere ryggmerter. Psykososiale variabler som ble kartlagt var frykt- og unngåelsestanker og atferd (162), tanker og holdninger til ryggsmertene (163), depresjonssymptomer (164), smertekatastrofering (165) frykt for å falle (166) og den enkelte studiedeltakers forventning om bedring av ryggsmertene tre måneder frem i tid (87, 94). Andre kliniske variabler som ble kartlagt var søvnproblemer (167), andre sykdommer (168), fallhistorikk, selvrappoert osteoporosediagnose (92), medikamentbruk for ryggsmertene (160) og behandler som ble oppsøkt for inkludering i studien. For mer detaljer rundt bakgrunnsvariablene som ble kartlagt, se tabell 3 og 4.

3.5.2 Imputering, sensitivitetstester og andre statistiske tester

For datamaterialet som i utgangspunktet var tilgjengelig for masterstudenten (n= 308) ble eventuelle forskjeller mellom studiedeltagerne med manglende verdier på SBT og studiedeltagerne med fullstendig besvarelse på SBT studert deskriptivt. Det ble ikke funnet noen forskjeller med hensyn til bakgrunnsvariablene. Studiedeltagerne med manglende verdier på SBT ble deretter ekskludert fra studien (se figur 1). Studiedeltagerne med inntil 25 % manglende verdier på RMDQ etter 3 måneder fikk imputert de manglende verdiene med gjennomsnittsverdien til resten av studiedeltagerne på de aktuelle spørsmålene. Eventuelle forskjeller mellom gruppen som hadde imputerte verdier på RMDQ etter 3 måneder, gruppen som ikke hadde besvart RMDQ etter 3 måneder og gruppen med fullstendig besvarelse på RMDQ etter 3 måneder ble først studert deskriptivt. Dersom det var mistanke om statistisk signifikante forskjeller ble det gjennomført Mann Whitney U-test mellom to og to grupper eller Fishers exact test, på henholdsvis kontinuerlige og kategoriske variabler. Det var ingen studiedeltagere med manglende utfyllelse på RMDQ etter 3 måneder som hadde over 25 % manglende verdier. Signifikanstesting på RMDQ ble gjennomført på alle bakgrunnsvariabler (ved baseline) oppgitt i tabell 3 og 4. Videre ble ikke de enkelte bakgrunnsvariablene (ved baseline) signifikanttestet for forskjeller mellom studiedeltagere med manglende verdier, deltagere som ikke hadde besvart de ulike baselinevariablene og de med fullstendig besvarelse.

Eventuelle forskjeller i RMDQ-score etter 3 måneder mellom de ulike risikogruppene ble analysert ved hjelp av Mann Whitney U-test mellom to og to grupper. Det ble også undersøkt hvorvidt det var signifikante forskjeller i andelen studiedeltagere som hadde og ikke hadde en reduksjon i RMDQ-scoren mellom baseline og 3 måneder. Wilcoxon test ble benyttet fordi differansen i RMDQ-scoren mellom baseline og tre måneder ikke var normalfordelt. I tillegg ble eventuelle forskjeller i andel studiedeltagere med et dårlig utfall i de ulike risikogruppe identifisert med SBT ved baseline (både ved bruk av anbefalt og optimal cut-off-verdi) analysert ved hjelp av kji-kvadrat-test. Videre var alle de psykososiale variablene (inkludert RMDQ og NRS, se tabell 4) ikke-normalfordelte på minst en av risikogruppene. For å bedre sammenligningsgrunnlaget på tvers av risikogruppene ble derfor alle de kontinuerlige variablene oppgitt med median og range. Mann Whitney U-test ble benyttet mellom to og to grupper for å undersøke om det var statistisk signifikante forskjeller i score mellom risikogruppene på de ulike bakgrunnsvariablene. Fishers exact test ble benyttet på den kategoriske variabelen forventning om bedring etter 3 måneder (FOB 3 mnd) (tabell 4).

3.5.3 Sensitivitet og spesifisitet

Sensitivitet- og spesifisitetsverdier regnes ut ved å sammenligne resultatene fra en dikotom test, med resultatene fra en dikotom gullstandard (169). Utregningen skjer ved hjelp av 2×2 krysstabeller (Se tabell 2, vedlegg 1). SBTs risikogrupeer ble derfor dikotomisert i to analysesett for å kunne gjennomføre sensitivitet- og spesifisitetsanalyser. Analysesettene var lav/medium mot høyrisikogruppe og lav risikogruppe mot medium/høyrisikogruppe. Sensitivitet omhandler testens evne til å korrekt klassifisere de med en sykdom eller «et klinisk dårlig utfall». (170). En tests spesifisitet viser til dets evne til å korrekt klassifisere de som ikke har en sykdom eller «et klinisk godt utfall» (171). Sensitivitet- og spesifisitetsverdier oppgis i prosent. 1 tilsvarer 100 %, altså perfekt evne til å klassifisere pasientene i korrekt risikogruppe (170).

3.5.4 Positiv og negativ likelihood ratio

Positiv og negativ likelihood ratio er andre nøyaktighetsmål som gir et tall på hvor mange ganger mer eller mindre sannsynlig det er at en person med en sykdom har et gitt testresultat enn en person som ikke har tilstanden. Likelihood ratio angis med et tall over eller under 1, som indikerer henholdsvis høyere eller lavere sannsynlighet (172). Positiv likelihood ratio ble regnet ut ved å dividere sensitivitet på (1-spesifisitet). Negativ likelihood ratio ble regnet ut ved å dividere (1-sensitivitet) på spesifisitet. En positiv likelihood ratio over 10 er svært høy, mens en negativ likelihood ratio under 0.1 nærmest utelukker muligheten for at personen har sykdommen (173). En positiv likelihood ratio mellom 5 og 10 og en negativ likelihood ratio mellom 0.1 og 0.2 er regnes som moderat. Positiv og negativ likelihood ratio mellom henholdsvis 2-5 og 0.5- 0.2 er lave, og kan gi små (men klinisk relevante) skifter i post-testodds. Likelihood ratioer mellom 1 og 2 og 0.5 og 1 gir imidlertid sjelden klinisk nytteverdi (174).

3.5.5 Arealet under Receiver Operating Characteristics Curve

Receiver under Operating Characteristics Curve (ROC-kurven) er en grafisk metode for å undersøke en tests samlede evne til å diskriminere mellom personer med og uten en sykdom (175). Arealet under kurven angis med ett tall mellom 0.5 og 1, hvor 0.5 betyr at en test har 50 % evne til å diskriminere mellom god eller dårlig selvpoplevd funksjon, mens 1 indikerer en perfekt diskriminerende evne. Et areal under ROC-kurven mellom 0.5 – 0.7 indikerer lave diskriminerende evner, 0.7-0.9 moderate diskriminerende evner, mens areal over 0.9 indikerer høye diskriminerende evner (176). Resultatene fra ROC-kurven ble også brukt til å identifisere den optimale cut-off-verdien. Dette representerer den verdien som har best evne til å identifisere både de med god og dårlig selvrapportert funksjon etter 3 måneder. Den optimale cut-off-verdien ble videre benyttet til å lage nye 2*2 krysstabeller for utregning av sensitivitet, spesifisitet, samt positiv og negativ likelihood ratio.

3.6 Etikk

Innsamling og lagring av data fra BACE-N-studien er godkjent av norsk senter for datalagring (NSD). Masterstudien er godkjent av regional etisk komité (REK referansenummer: 2018:1424, vedlegg 6). Studiedeltagerne ble forelagt informasjon om studiens formål og oppfølgingsperiode og ga et informert skriftlig samtykke før inklusjon i BACE-Nstudien. Det ble gitt informasjon om at studien ikke vil påvirke videre behandling av ryggplagene (vedlegg 3), og at det ikke ville påvirke videre oppfølging hos primærbehandler dersom studiedeltaker trakk seg underveis. Direkte personidentifiserende opplysninger i masterstudentens datamateriale ble erstattet med et referansenummer som viste til en adskilt navneliste som ble oppbevart av prosjektleder. Datamaterialet ble derfor oppbevart i henhold til datatilsynets retningslinjer (177). Studien forøvrig vurderes også å være i tråd med Helsinkideklarasjonens etiske retningslinjer (178).

4.0 RESULTATER

4.1 Baselinekarakteristikker

Det var 299 studiedeltakere som oppfylte kriteriene for inklusjon i studien. Etter 3 måneder var det 224 deltakere (75 %) som oppfylte kriteriene som var satt for fullstendig besvarelse på RMDQ, og dermed ble inkludert til videre analyser. Det var ingen statistisk signifikante forskjeller mellom gruppen som hadde fullstendig besvarelse på RMDQ etter 3 måneder, gruppen som hadde imputerte verdier, og heller ikke gruppen som ikke hadde besvart RMDQ på samme tidspunkt på noen av bakgrunnsvariablene. SBT klassifiserte 65 %, 27 % og 7 % til henholdsvis lav-, medium- og høy risikogruppe ved baseline. Median alder var 67 år og 55 % av studiedeltagerne hadde en smertevarighet på ≤ 4 uker. Median totalscore på SBT var 3 poeng ved baseline (tabell 3). Det fullstendige flytskjemaet er presentert i figur 1. Baselinekarakteristikker er presentert i tabell 3 og 4.

Figur 1: Flytdiagram som illustrerer antall studiedeltagere i masterstudien ved baseline, antallet studiedeltagere som ble ekskludert ved baseline og antall studiedeltagere som hadde fullstendig besvarelse på RMDQ etter 3 måneder. Anbefalte cut-off-verdier på SBT ved baseline er illustrert ved dikotomisering høy mot lav/medium risikogruppe (blå farge) og medium/høy mot lav risikogruppe (grønn farge).



Forkortelser: RMDQ= Roland Morris Disability Questionaire. SBT= Start Back Screening Tool. Lav= lavrisikogruppe. Med.= medium risikogruppe. Høy= høyrisikogruppen. Mnd.= måneder. FBR= Funksjonsbegrensninger som følge av ryggsmarter kartlagt med RMDQ. Har FBR= ≥ 7 poeng på RMDQ. Har ikke FBR= < 7 poeng på RMDQ.

* Antall studiedeltakere med imputerte verdier på RMDQ, fordelt på risikokategori (dikotomisert) på SBT:
 SBT baseline høy: Imputert score på RMDQ 3 mnd, n= 1
 SBT baseline lav/med.: Imputert score på RMDQ 3 mnd, n= 14
 SBT baseline med./høy: Imputert score på RMDQ 3 mnd, n= 8
 SBT baseline lav: Imputert score på RMDQ 3 mnd, n= 7

Tabell 3: Baselinekarakteristikker for de inkluderte studiedeltagerne undersøkt med deskriptiv statistikk.

VARIABEL, n= 299 (100%)	
Alder (n= 299), median (range)	67 (55 - 89)
Kjønn (n= 298), kvinne, n (%)	166 (55.5)
Sivilstatus (n= 299), gift/samboer, n (%)	232 (77.6)
Morsmål (n= 296), norsk, n (%)	284 (95.0)
KMI (kg/m²), (n= 288), n (%)	
18.5-24.9	86 (28.8)
>25	202 (67.6)
Utdanningsnivå (n= 297), har høyere utdanning, n (%)	124 (41.5)
Lønnet arbeid (n= 276), ja, n (%)	116 (38.8)
Røyking (n= 295), ja, n (%)	42 (14.0)
Tidligere ryggmerter (n= 296), ja, n (%)	284 (95.0)
Varighet av ryggmerter (n= 261), dager, n (%)	
≤4 uker	164 (54.8)
5 uker-3 mnd	38 (12.7)
>3 mnd	59 (19.7)
Utstrålende ryggmerter til bein siste uke (n= 294), ja, n (%)	128 (42.8)
Bruker medisiner for ryggsmertene (n= 288), ja, n (%)	124 (41.5)
Søvnproblemer pga ryggsmertene siste måned (n= 293), ja, n (%)	175 (58.6)
Fallhistorikk siste 6 uker (n= 294), ja, n (%)	47 (15.7)
Andre sykdommer (n= 233), >2 andre sykdommer, n (%)	153 (51.2)
Osteoporose, selvrapportert (n= 294), ja, n (%)	20 (6.7)
Rekruiterende profesjon (n= 287), n (%)	
Lege	51 (17.1)
Fysioterapeut	122 (40.8)
Kiropraktor	114 (38.1)
SBT, totalscore (0-9), (n= 299), median (range)	3 (0 - 9)
SBT, psykososial subscore (0-5), (n= 299), median (range)	1 (0 - 5)

Forkortelser og forklaringer: SBT= Start Back Screening Tool (fra 0-9, hvor høyere score indikerer høyere tilstedeværelse av fysiske og psykososiale prognostiske faktorer for langvarige korsryggplager). Psykososial subscore (fra 0-5, hvor høyere score indikerer høyere tilstedeværelse av psykososiale prognostiske faktorer). KMI= kroppsmasseindeks (kilogram/meter²). Søvnproblemer siste måned (spørsmål 5 i Pittsburgh Sleep Quality Index). Andre sykdommer (The Self Administered Comorbidity Questionnaire). Pga= på grunn av.

I tabell 4 er psykososiale prognostiske faktorer, samt smerteintensitet og funksjonsbegrensninger fordelt på risikogrupper på SBT presentert. Tabell 4 viser at median score på RMDQ og de fleste andre variablene øker i samsvar med høyere risikogruppe på SBT. Forskjellene i score mellom de tre risikogruppene var statistisk signifikant ($p < 0.01$) på de fleste psykososiale variablene. På variablene FABQ fys.akt., RMDQ, CES-D og PCS ser en også klinisk relevante forskjeller. Høyrisikogruppen hadde på disse variablene mellom 2 og 3 ganger høyere median score enn lavrisikogruppen. På variablene NRS og BBQ var det imidlertid ingen signifikante forskjeller i score mellom medium- og høyrisikogruppen. For variabelen forventning om bedring etter 3 måneder (FOB 3 mnd) var det en statistisk signifikant forskjell i prosentandel med ulike forventninger om bedring i alle de tre risikogruppene. Det var imidlertid ikke en konsekvent sammenheng med hensyn til personer som hadde negative og positive forventninger til ryggsmertene 3 måneder frem i tid og hvilken risikogruppe de tilhørte. Det var likevel 33.3 % av dem som hadde forventninger om at ryggsmertene ville bli verre som tilhørte høyrisikogruppen, mens det ikke var noen i lavrisikogruppen. Samtidig var det 4 % av dem som trodde ryggsmertene ville bli helt bra 3 måneder frem i tid som tilhørte lavrisikogruppen sammenlignet 68 % i høyrisikogruppen.

Tabell 4: Psykososiale prognostiske faktorer (inkludert RMDQ og NRS) fordelt på risikogrupper på Start Back Screening Tool ved baseline. Variablene er kartlagt med deskriptiv statistikk.

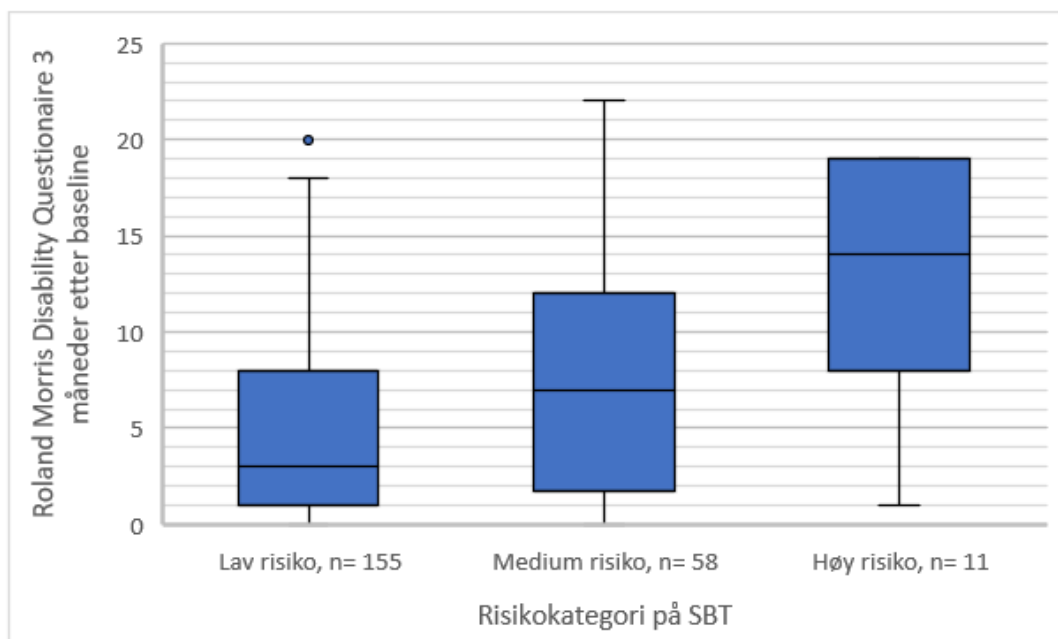
VARIABEL	Lav risikogruppe, n= 195 (65%)	Medium risikogruppe, n= 82 (27 %)	Høy risikogruppe, n= 22 (7 %)
NRS (0-10), (n= 292), median (range)	5 (0 - 10)	7 (1 - 10)	7 (3 - 10)
FABQ, fys.akt. (0-30), (n= 299), median (range)	8 (0 - 24)	12 (0 - 24)	15 (5 - 24)
RMDQ (0-24), (n= 299), median (range)	6 (0 - 20)	12 (4 - 20)	17.5 (5 - 21)
FES-I (16-64), (n= 287), median (range)	19 (16 - 52)	23 (16 - 48)	27 (17 - 53)
CES-D (0-60), (n= 284), median (range)	5 (0 - 32)	11 (0 - 44)	18 (8 - 43)
PCS (0-52), (n= 292), median (range)	7 (0 - 35)	13 (0 - 37)	26 (5 - 48)
BBQ (45-9), (n= 282), median (range)	32 (15 - 45)	27 (11 - 42)	20.5 (9 - 42)
FOB 3 mnd, (n=298), n (%)			
Helt bra	51 (68.0)	21 (28.0)	3 (4.0)
Mye bedre	91 (61.5)	46 (31.1)	11 (7.4)
Ingen forandring	53 (73.6)	12 (16.7)	7 (9.7)
Mye verre	0 (0.0)	2 (66.7)	1 (33.3)
Forkortelser og forklaring av score: NRS= Numerisk smerteskala (Smerteintensitet uken før baselinekartlegging). FABQ= Fear Avoidance and Beliefs Questionnaire – fys.akt= fysisk aktivitet subskala (fra 0 ingen frykt og unngåelsestanker – 30). RMDQ= Roland Morris Disability Questionnaire (fra 0 ingen funksjonsbegrensinger – alvorlige funksjonsbegrensere). FES-I= Falls Efficacy Scale (16= Ingen frykt for å falle – 64= alvorlig frykt for å falle). CES-D= Center of Epidemiologic studies Depression Scale (fra 0 = ingen depressiv symptomatologi – 60 høy depressive symptomatologi). PCS: Pain Catastrophising Scale (fra 0 ingen smertekatastrofering – 52) Back Beliefs Questionnaire (lavere score indikerer større grad av maladaptive eller pessimistiske tanker om ryggsmertene). FOB 3 mnd: Forventning om bedring av ryggsmertene 3 måneder frem i tid fra tidspunktet på baselinekartleggingen.			

4.2 Klinisk forløp 3 måneder etter baseline

Samlet hadde 41.5 % (n= 93) av studiedeltagerne et dårlig klinisk utfall etter 3 måneder. Det var likevel signifikant flere studiedeltagere som fikk redusert sin score på RMDQ etter 3 måneder ($p < 0.01$). Median score på RMDQ ved baseline for hele utvalget var 9 poeng (range 0-21), mens median score etter 3 måneder var 4 poeng (range 0-22). Andelen pasienter med et dårlig utfall etter tre måneder økte i henhold til risikogruppene. I lav-, medium-, og høyriskogruppen hadde henholdsvis 53 av 155 (34 %), 30 av 58 (52 %) og 10 av 11 (91 %) studiedeltakere ett dårlig utfall etter 3 måneder. Forskjellene var statistisk signifikante med en p-verdi på < 0.01 mellom alle de tre risikogruppene. Forskjellene vurderes også som klinisk relevante.

I Figur 2 vises et box-plot av median og interkvartilrange på RMDQ etter 3 måneder fordelt på de ulike risikogruppene på SBT ved baseline. Det var en statistisk signifikant forskjell i RMDQ-score etter 3 måneder mellom alle risikogruppene. Lav mot medium risikogruppe ($p=0.03$), medium mot høy risikogruppe ($p=0.01$) og lav mot høy risikogruppe ($p<0.01$). Høyere risikogruppe på SBT indikerte derfor høyere score på RMDQ etter 3 måneder.

Figur 2: Box- plot av median score på RMDQ etter 3 måneder fordelt på risikogrupper på SBT ved baseline.



4.3 Sensitivitet, spesifisitet og likelihood ratioer ved anbefalte cut-off-verdier

Tabell 5 viser krysstabellene som ble benyttet som grunnlag for utregning av sensitivitet, spesifisitet og likelihood ratioer. Resultatene i tabell 6 viser at SBT hadde svært høy spesifisitet og til gjengjeld svært lav sensitivitet ved bruk av cut-off lav/medium mot høy risikogruppe på SBT. Sensitiviteten var fortsatt lav med bruk av cut-off lav mot medium/høy risikogruppe. Spesifisiteten var derimot 0.78. Når det gjelder estimatene av positiv likelihood

ratio var tallene vesentlig høyere når cut-off-verdien lav/medium mot høy risikogruppe ble benyttet enn når cut-off lav mot medium/høy risikogruppe ble benyttet. Det var 11 ganger mer sannsynlig at personer som havnet i høyrisikogruppen ved baseline hadde et dårlig utfall etter 3 måneder sammenlignet med personer i lav- og medium risikogruppe.

Tabell 5: Krysstabeller basert på anbefalte cut-off-verdier for de ulike risikogruppene på SBT og en cut-off for et dårlig utfall på ≥ 7 poeng på RMDQ. Krysstabellene danner grunnlaget for utregning av sensitivitet, spesifisitet og likelihood ratio i tabell 7. Tallene i krysstabellen er oppgitt i antall (n). Totalt antall i krysstabellen: n= 224.

	≥ 7 på RMDQ	< 7 på RMDQ		≥ 7 på RMDQ	< 7 på RMDQ
SBT Medium/høy risikogruppe	40	29	SBT Høy risikogruppe	10	1
SBT Lav risikogruppe	53	102	SBT Lav/medium risikogruppe	83	130

Forkortelser: RMDQ= Roland Morris Disability Questionnaire. SBT= Start Back Screening Tool.

Tabell 6: Sensitivitet, spesifisitet, samt negativ og positiv likelihood ratio for å identifisere personer med et dårlig utfall (≥ 7 poeng på Roland-Morris Disability Questionnaire) etter 3 måneder ved bruk av de anbefalte cut-off-verdiene på SBT. (Totalt antall deltagere i analysene: n=224)

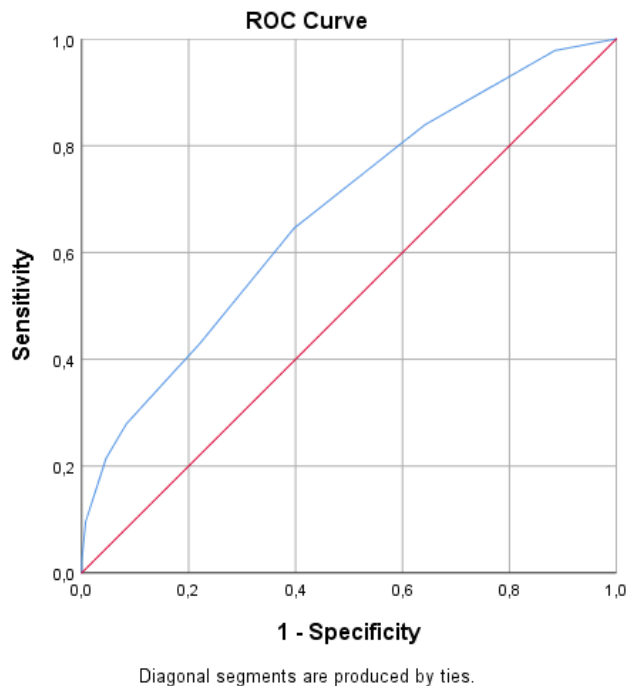
	Sensitivitet	Spesifisitet	+ LR	-LR
Med./høy versus Lav risikogruppe	0.43	0.78	1.95	0.73
Høy versus Lav/med. risikogruppe	0.11	0.99	11	0.90

Forkortelser: +LR= Positiv likelihood ratio. -LR= Negativ likelihood ratio. Med. = medium risikogruppe

4.4 Diskriminerende evner målt med arealet under ROC-kurve

I figur 3 illustreres SBTs evne til å diskriminere mellom et godt og et dårlig utfall etter 3 måneder ved hjelp av arealet under ROC-kurven. Når ≥ 7 poeng på RMDQ ble benyttet som cut-off-verdi var arealet under ROC-kurven 0.68 (95 % KI: 0.61, 0.75. $p < 0.01$).

Figur 3: Diskriminerende evner for totalscore på Start Back Screening Tool når ≥ 7 poeng på Roland-Morris Disability Questionnaire ble benyttet som utfallsmål (n=224). Analysert og illustrert ved hjelp av arealet under Receiver Operating Characteristics Curve.



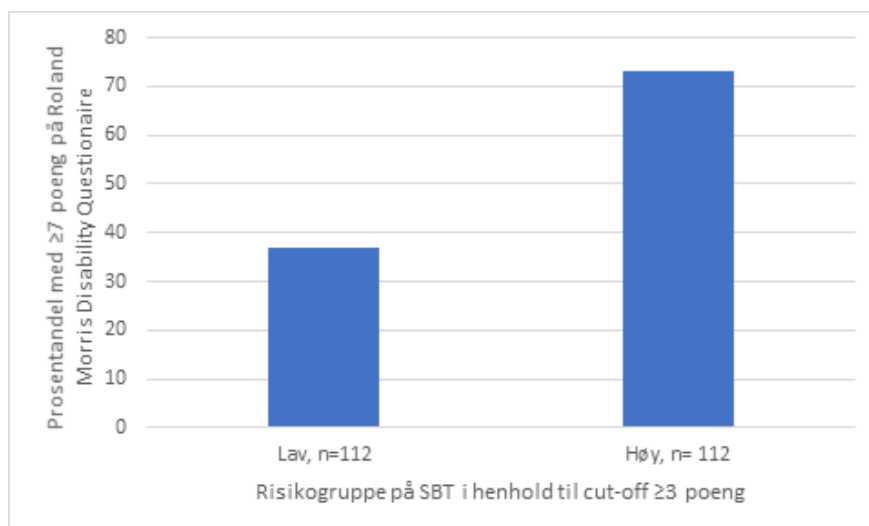
4.4.1 Sensitivitet spesifisitet og likelihood ratioer ved bruk av optimal cut-off verdi på SBT

Ut fra ROC-analysen (se vedlegg 2) ble det funnet at den optimale cut-off-verdien for å diskriminere mellom et godt og dårlig utfall i dette utvalget var ≥ 3 poeng. Studiedeltagerne med en baselinescore på SBT på ≥ 3 poeng hadde en median RMDQ-score etter 3 måneder på 7 poeng (range 0-22). Studiedeltagerne som hadde < 3 poeng på SBT hadde en median RMDQ-score etter 3 måneder på 2 poeng (range 0-20).

Basert på den optimale cut-off-verdien (≥ 3 poeng på SBT) som ble identifisert i ROC-analysen ble studiedeltagerne fordelt i to nye risikogrupper, henholdsvis lav og høy risikogruppe (Tabell 7). Det var en statistisk signifikant høyere andel med et dårlig utfall i

høyrisikogruppen (73.2 %) enn i lavrisikogruppen (36.9 %), ($p < 0.01$) (Figur 4).

Figur 4: Prosentandel av studiedeltagerne med et dårlig utfall etter 3 måneder fordelt på lav og høy risikogruppe på Start Back Screening Tool (SBT) ved baseline, basert på den optimale cut-off-verdien (≥ 3 poeng) ($n=224$).



Basert på den optimale cut-off-verdien som ble identifisert i ROC-analysen ble det også laget nye 2*2 krysstabeller for beregning av sensitivitet, spesifisitet, positiv og negativ likelihood ratio (tabell 7). Av tabell 8 ser man at sensitiviteten var høyere når den optimale cut-off-verdien ble benyttet enn når de anbefalte cut-off-verdiene på SBT ble benyttet, mens spesifisiteten var lavere (tabell 6). Videre var positiv likelihood ratio dårligere og negativ likelihood ratio bedre når den optimale cut-off-verdien ble benyttet enn når den anbefalte cut-off-verdien lav/medium mot høy risikogruppe på SBT ble benyttet.

Tabell 7: Krysstabell basert på optimal cut-off-verdi på Start Back Screening Tool (≥ 3 poeng) og cut-off-verdi for et dårlig utfall på RMDQ på ≥ 7 poeng. Krysstabellen danner grunnlaget for utregning av sensitivitet, spesifisitet og likelihood ratio. Tallene i krysstabellen er oppgitt i antall (n), totalt antall i krysstabellen: $n=224$.

	≥ 7 på RMDQ	< 7 på RMDQ
SBT Høy risikogr. (≥ 3)	60	52
SBT Lav risikogr. (< 3)	33	79

Forkortelser: RMDQ: Roland-Morris Disability Questionnaire. SBT: Start Back screening Tool. Risikogr.: risikogruppe.

Tabell 8: Sensitivitet, spesifisitet, samt negativ og positiv likelihood ratio for å identifisere personer med et dårlig utfall (≥ 7 poeng på Roland-Morris Disability Questionnaire) etter 3 måneder ved bruk av optimal cut-off-verdi (≥ 3 poeng) på Start Back Screening Tool (Totalt antall: n= 224).

	Sensitivitet	Spesifisitet	+ LR	-LR
Lav risikogr. versus Høy risikogr.	0.64	0.60	1.35	0.60

Forkortelser: +LR: Positiv likelihood ratio. -LR: Negativ Likelihood ratio. Risikogr.: risikogruppe

5.0 DISKUSJON

Formålet med denne studien var å kartlegge SBTs evne til å predikere et dårlig utfall etter 3 måneder hos eldre som oppsøkte primærhelsetjenesten for behandling av sine ryggmerter. Et av hovedfunnene i studien var at det kun var høyrisikogruppen på SBT som hadde prediktiv verdi for funksjonsutfall etter 3 måneder. SBT hadde svært høy spesifisitet og positiv likelihood ratio ved bruk av cut-off lav/medium mot høy risikogruppe på SBT. Videre hadde SBT ingen klinisk nytteverdi i å predikere funksjonsutfall ved bruk av cut-off-verdien lav mot medium/høy risikogruppe. Sensitiviteten og den negative likelihood ratioen var lav både ved bruk av lav/medium mot høy risikogruppe og lav mot medium/høy risikogruppe. SBTs samlede evne til å diskriminere mellom et godt og dårlig utfall etter 3 måneder var lav, men på grensen til moderat. Videre var den optimale cut-off-verdien på SBT i dette utvalget ≥ 3 av 9 poeng. Verdiene for likelihood ratio ved bruk av optimal cut-off-verdi gav ingen klinisk nytteverdi (174).

5.1 Metodediskusjon

5.1.1 Studiedesign

Hensikt og fremgangsmåte brukt ved screeninginstrumenter er sammenlignbare med prognostiske tester hvor det kreves et kohortdesign. Disse må generelt være longitudinelle for å fange opp insidensen av sykdom (169, 79, s. 138). Et prospektivt kohortdesign ble derfor benyttet i masterstudien. En annen fordel med et prospektivt design er at sjansen for «recall bias» er mindre enn ved et retrospektivt design (179, s. 17).

5.1.2 Datainnsamling

Forfatteren av masterstudien var ikke inkludert i innsamling av data til masterstudien og kan derfor ikke ha påvirket scoringene ved baseline. Oppgaveforfatteren bidro imidlertid til datainnsamlingen ved baseline på et senere tidspunkt og fikk dermed førstehåndserfaring med, og kjennskap til datainnsamlingsprosessen. Flere av de andre klinikerne som bidro i

datainnsamlingen var imidlertid både tester og behandler. Dette kan ha ført til en systematisk skjevhet ved at studiedeltagerne følte en forpliktelse eller et ønske om å rapportere bedring av symptomer (179, s. 17). Samtidig kan testpersonell som også var behandler ha fått mer kunnskap om den enkelte pasienten, som igjen kan ha påvirket behandlingen som ble gitt. Gjennomføringen av de kliniske undersøkelsene og besvarelse av spørreskjemapakken ved baseline tok cirka 1.5 timer. Det kan tenkes at studiedeltagerne med størst smerter og funksjonsbegrensninger hadde utfordringer med å gjennomføre hele baselineundersøkelsen. Omfanget av baselineundersøkelsen kan ha påvirket svarene og scoringer på spørreskjemaene. Ved oppfølging etter 3 måneder mottok studiedeltagerne spørreskjemapakken per e-post eller brevpost. Her var det kun en mindre del av spørreskjemapakken fra baseline som skulle besvares.

5.1.3 Utvalg og generaliserbarhet

En utfordring for generaliserbarheten i masterstudien er at selv om det er ønskelig at behandlere som rekrutterer pasienter skulle inkludere alle aktuelle pasienter fortløpende, så har man ingen kontroll over om det var noen som ikke ble spurt om å delta i BACE-studien. Det eksisterer ingen oversikt hverken over hvem som ble spurt og takket nei til studien, eller hvem som ikke ble spurt i det hele tatt. Dette kan ha påvirket hvilke studiedeltakere som er representert i masterstudien og kan dermed være et problem for studiens eksterne validitet, til tross for at studiedeltagerne i masterstudien er sammenlignbare med tilsvarende kohorter fra andre studier.

Masterstudien benyttet samme inklusjon- og eksklusjonskriterier som BACE-N studien. I tillegg ble personer som ikke hadde fullstendig besvarelse på SBT ved baseline ekskludert. Eldre personer ble i masterstudien definert som personer ≥ 55 år. WHO definerer eldre i den vestlige verden som personer >60 år. Dette med utgangspunkt i personkarakteristikker som pensjonsalder, kronologisk alder og reduksjon i funksjonelle evner (180). En forskergruppe i USA har etablert et lignende studiekonsept som BACE, ved navn «The Back pain Outcomes using Longitudinal Data» (BOLD). Studiene benyttet lignende inklusjon- og eksklusjonskriterier som BACE-konsortiet. En vesentlig forskjell er at alders-cut-off-verdien for inklusjon ble satt på ≥ 65 år. I BOLD-studiene og i studien av SBTs prediktive validitet av

Suri et al. (25) ble det også gjennomført analyser hvor aldersgrensen for inklusjon var på ≥ 65 år. Cut-off ≥ 55 år ble likevel benyttet i masterstudien da BACE-konsortiet med allerede publiserte studier benytter denne cut-off-verdien (63, 87, 89, 94, 160, 181), samt for å oppnå best mulig statistisk styrke.

Utvalget i denne masterstudien er sammenlignbart med BACE-studier fra Nederland med hensyn til baselinevariabler som blant annet alder, kjønn og andel personer med ryggsmerte med varighet over tre måneder (63, 87, 89, 181). BOLD-studien av Jarvik et al. (93) hadde høyere gjennomsnittsalder (73.7 ± 6.8) og en noe høyere andel kvinner (65 %) enn masterstudien. I likhet med andre BACE- og BOLD-studier krever masterstudien at studiedeltagerne ikke har oppsøkt primærhelsetjenesten for sine ryggsmerte de siste 6 månedene. I den generelle befolkningen er det demonstrert forskjeller i baselinekarakteristikker hos personer som søker hjelp og de som ikke søker hjelp for sine ryggsmerte. Det er rapportert at kvinner, personer som oppgir høyere smerteintensitet eller funksjonsnedsettelse, større grad av mental engstelse, smerterelaterte arbeidsbegrensninger eller lavere selvrappert helse oftere oppsøker hjelp hos helsepersonell (182, 183). Dette betyr at resultatene i masterstudien ikke nødvendigvis er overførbare hos personer ≥ 55 år som hyppig oppsøker behandler på grunn av ryggsmerte.

Kun studiedeltagerne med fullstendig besvarelse på SBT ved baseline ble inkludert i masterstudien (figur 1). Disse kriteriene har også blitt benyttet i to tidligere studier (26, 184). Studiedeltagerne med manglende verdier kunne også fått erstattet disse med imputerte verdier, ved å imputere med gjennomsnittsverdien til de resterende studiedeltagerne. En alternativ løsning kunne vært å la totalscoren bli gjeldene selv med manglende verdier, og erstatte manglende verdier med 0 for «uenig». Dette ble vurdert å kunne gi unøyaktige resultater da et manglende svar like gjerne kunne vært 1 for «enig». Samtidig er det kun 9 spørsmål på SBT, og et manglende svar kunne derfor fått stor betydning for hvilken risikogruppe studiedeltageren havnet i. Det var kun 9 av 308 (2.9 %) studiedeltagere med manglende verdier på SBT ved baseline. Denne prosentandelen var relativt liten og ble vurdert å ikke ha stor betydning for den statistiske styrken. Studiedeltagerne med manglende verdier på SBT ved baseline skilte seg heller ikke fra studiedeltagerne med fullstendig besvarelse og ble derfor ekskludert fra studien. Et annet moment som bør problematiseres i denne sammenheng, omhandler de studiedeltagerne som scoret 0 poeng på RMDQ ved

baseline. Ifølge scoringsprosedyren indikerer dette at de ikke har ryggrelaterte funksjonsbegrensninger (150). Disse studiedeltagerne ble likevel inkludert i masterstudien, da dette heller ikke er benyttet som et eksklusjonskriterium i andre studier av SBTs prediktive validitet med RMDQ som utfallsmål (11, 22, 25, 119).

Masterstudien er gjennomført i primærhelsetjenesten og resultatene er derfor ikke nødvendigvis overførbare til eksempelvis spesialisthelsetjenesten. Samtidig har også tidligere studier vist at SBT har dårligere prediktive evner i sekundærhelsetjenesten enn i primærhelsetjenesten (120-123). Studiedeltagerne ble rekruttert fra kiropraktorklinikker, fysioterapiinstitutter og fastleger, noe som øker masterstudiens overføringsverdi til ulike kliniske settinger i primærhelsetjenesten. I tillegg er pasientene rekruttert fra ulike områder i Norge, både større byer og mer landlige områder. 41.5 % hadde høyere utdanning, mens 57 % hadde grunnskole og/eller videregående utdanning. Dette gjør at utvalget trolig er representativt for personer med ulik sosioøkonomisk status.

5.1.4 Utfallsmål

Prediktiv validitet er en undergruppe av kriterievaliditet, som viser til hvorvidt score på måleinstrumentet predikerer score på gullstandarden i fremtiden (36, s. 159). Det har derfor avgjørende betydning for slike studiers interne validitet at gullstandarden som benyttes ligger nærmest mulig den perfekte gullstandarden. RMDQ og Oswestry Disability Index er de to mest benyttede utfallsmålene for å kartlegge ryggrelaterte funksjonsbegrensninger ved korsryggsmerter (50). Det foreligger internasjonal konsensus om å benytte et av disse utfallsmålene i populasjoner med korsryggsmerter (185). Ved bruk i primærhelsetjenesten er det argumentert for å benytte RMDQ fremfor Oswestry Disability Index da en kan forvente mildere funksjonsbegrensninger i denne populasjonen sammenlignet med eksempelvis sekundærhelsetjenesten (154, 186).

RMDQ inneholder ulike aspekter av funksjon. Disse inkluderer funksjonsbegrensninger i daglige aktiviteter som følge av personens korsryggsmerter, samt smerte, søvn, appetitt og emosjonelle funksjoner. RMDQ var opprinnelig ansett som et unidimensjonalt

kartleggingsinstrument som måler fysisk funksjon i daglige aktiviteter. Det er senere argumentert av flere for at RMDQ bør forstås som et multidimensjonalt kartleggingsinstrument, da innholdsvaliditeten er funnet å være utilstrekkelig (151, 152). Andre screeningsinstrumenter inneholder også andre aspekter av funksjon i dagliglivet. Eksempelvis inneholder Oswestry Disability Index i motsetning til RMDQ også spørsmål knyttet til sosialt liv og seksualliv (158). I tolkningen av resultatene er det derfor viktig å ha i mente at RMDQ-scoren er en sum av et begrenset utvalg dimensjoner av funksjonsbegrensninger knyttet til ryggsmertene.

Det er ingen etablert cut-off-verdi på RMDQ for å skille mellom personer med et godt og et dårlig utfall i form av ryggrelaterte funksjonsbegrensninger. Hill et al. (11) etablerte Start Back Screening Tool med utgangspunkt i en absolutt cut-off-verdi for et dårlig utfall på ≥ 7 poeng på RMDQ. Denne cut-off-verdien er også benyttet i andre studier av SBTs prediktive validitet (to studier har gjort om RMDQ-skalaen til en skala fra 0-100) (22, 25, 121, 187). Ulempen med å benytte en absolutt score som cut-off-verdi er at en person som eksempelvis går fra 24 poeng på RMDQ ved baseline til 7 poeng etter tre måneder sannsynligvis også vil oppnå et dårlig utfall, til tross for svært stor prosentvis endring. Samtidig vil en person som går fra 1 poeng på RMDQ ved baseline til 6 poeng etter 3 måneder oppnå et godt utfall.

På grunn av utfordringene med en absolutt score som utfallsmål er det argumentert for å benytte en cut-off-verdi basert på en absolutt endringsscore på 5 poeng som indikerer minste klinisk viktige endring på RMDQ (157). Dette vil dermed kunne forenkle fortolkningen av hvem som har et godt utfall og ikke. Samme forskergruppe foreslår alternativt å benytte en relativ endringsscore på $\geq 30\%$ for å indikere et godt utfall. Fordelen med denne metoden er at den tar høyde for baselinescoren slik at alle (bortsett fra dem med en RMDQ-score på 0 ved baseline) har mulighet til å oppnå et godt utfall. På den andre siden er det også utfordringer med å benytte relativ endringsscore som cut-off-verdi fordi denne er sensitiv for studiedeltagernes varians i score ved baseline. I motsetning til når en absolutt endringsscore eller en absolutt score benyttes som cut-off-verdi vil det kreves flere poengs endring desto høyere sumscore personen har ved baseline (188). En person med 24 poeng ved baseline må ha 8 poeng reduksjon på RMDQ etter 3 måneder for å oppnå $\geq 30\%$ bedring. Samtidig trenger en person med 5 poeng ved baseline kun 2 poeng reduksjon på RMDQ etter tre måneder for å oppnå samme utfall. Selv om 2 poeng reduksjon i score på RMDQ i dette tilfellet vil indikere

et godt utfall ved bruk av prosentvis endringsscore, så er en endring på 2 poeng lavere enn både absolutt målefeil og minste klinisk viktige endring (156, 157). En reduksjon i score på under 5 poeng, både ved bruk av absolutt cut-off-verdi og prosentvis endringsscore kan altså være et uttrykk for målefeil og ikke en reell endring i funksjon hos pasienten.

For pasientene kan det også være vanskelig å fortolke prosentvis endring sammenlignet med en absolutt score. Dette kan eksemplifiseres med at en person i høyrisikogruppen vil ha høy risiko for å oppnå <30 % bedring av funksjonsbegrensningene, sammenlignet med å ha høy risiko for vedvarende funksjonsbegrensninger. Samtidig har en norsk studie vist at absolutt score som utfallsmål hadde like gode evner til å indikere bedring ved rygg smerter som relativ endringsscore i prosent (189). Videre var median score på RMDQ ved baseline i masterstudien sammenlignbar originalstudien av Hill et al. (11). Den absolutte cut-off-verdien på ≥ 7 på RMDQ ble også valgt for å bedre sammenligningsgrunnlaget med studien av Suri et al. (25) som har gjennomført en subanalyse av SBTs prediktive validitet hos personer ≥ 65 år. I tillegg vil bruk av en annen cut-off-verdi på RMDQ enn i originalstudien gi SBT andre forutsetninger for å kunne predikere et dårlig utfall. En cut-off-verdi på ≥ 7 poeng på RMDQ ble derfor benyttet i masterstudien.

Valg av cut-off-verdi avgjør hva som defineres som et dårlig utfall. Cut-off-verdien for et dårlig utfall vil derfor også påvirke estimatene for sensitivitet, spesifisitet, likelihood ratioer og diskriminerende evner (190). I masterstudien satte cut-off-verdien et distinkt skille mellom de med langvarige funksjonsbegrensende rygg smerter og de som ikke hadde det. En kunne testet hvor robust valg av cut-off var ved å sammenligne cut-off ≥ 7 poeng på RMDQ etter tre måneder med et annet utfallsmål. Ut fra variablene som var tilgjengelig i BACE-materialet ved 3 måneder kunne en eksempelvis brukt Global Percieved Effect eller numerisk smerteskala, som begge er blant de mest benyttede utfallsmålene i ryggforskning (50). Global Percieved Effect er en likertskala fra 1-7 (eventuelt 1-5) som måler selvopplevd endring (191). Cut-off for et dårlig utfall kunne da ha blitt satt ved svaralternativ 6 og 7, «mye verre» eller «verre enn noen gang før». En validering av resultatene ved bruk av et annet utfallsmål ble ikke gjennomført da masterstudien allerede inneholder mange analyser og ytterligere analyser ville gjort masterstudien omfattende sett opp mot de tidsrammene som var satt.

5.1.5 Oppfølgingstid

For å fange opp insidensen av sykdom, og dermed kunne si noe om prognose må studiedeltagerne følges opp en viss tid (79, s. 138). Det kan derfor være at en oppfølgingstid på 3 måneder var for kort til å fange opp alle som fikk residiv av ryggsmertene. En oppfølgingstid på 3 måneder ble likevel benyttet for å bedre sammenligningsgrunnlaget med flere andre studier (22, 26, 117, 119). Samtidig ble det vurdert som interessant å beregne en prediksjon av utfall som ikke var for langt frem i tid. Dette fordi det kan forventes at den største forandringen i smerte og funksjon skjer i løpet av de første tre månedene etter en ny episode med rygg smerter (84, 85, 89). Det ble også vurdert som viktig med et relativt tidlig oppfølgingstidspunkt da det er gunstig å kunne gi rett oppfølging til rett person på et tidligst mulig tidspunkt etter en ny episode rygg smerter.

Det er likevel problematisk å indikere et godt eller dårlig utfall ved rygg smerter på ett bestemt tidspunkt. Dette fordi rygg smerter ofte har et fluktuerende forløp (2, 192). Det vil si at selv om en studiedeltager har scoret seg selv <7 poeng på RMDQ etter 3 måneder kan det være at denne scoren er ≥ 7 poeng, eksempelvis uken før eller uken etter besvarelsen. En annen faktor som ofte er problematisk i prospektive studier er et fenomen som kalles «regression towards the mean» (193). Dette handler om at folk mest sannsynlig oppsøker behandler for hjelp når plagene har blitt verre, og kanskje også er på sitt verste. Derfor har spesielt studiedeltagere med ekstremverdier ved baseline høy sannsynlighet for å score nærmere gjennomsnittet ved oppfølgingstidspunktet. Da rygg smerter har et naturlig fluktuerende forløp kan det tenkes at denne typen systematisk skjevhet er spesielt aktuell her. Problemene som er forespeilet i dette avsnittet indikerer at selv om personer defineres til et godt utfall etter tre måneder, er dette ingen garanti for senere residiv av rygg smertene.

Det er vist at tidsperioden mellom første konsultasjon for rygg smertene og tidspunktet en gjennomfører screening av risikoprofil er av betydning. En studie viste at endringsscoren fra baselinekartleggingen med SBT til andre screening med SBT etter 4 uker hadde bedre til å predikere smerteintensitet og funksjonsbegrensinger etter 6 måneder enn baselinescoren alene (17). I en annen studie av Newell et al. (118) var det kun to dager mellom måletidspunktene for SBT og over en tredel av pasientene endret risikogruppe. Ut fra disse studiene kan det virke mer hensiktsmessig å gjennomføre screening ved hjelp av SBT en periode etter pasientens første konsultasjon for rygg smertene. Problemet med å vente med å benytte SBT i

kartleggingen er at en mister verdifull tid til å kunne gjennomføre riktige tiltak for pasienter med økt risiko for langvarige rygg smerter på et tidlig tidspunkt. Studiene reiser likevel spørsmål om hvorvidt en også bør benytte SBT på et senere tidspunkt i forløpet og eventuelt korrigere videre tiltak dersom pasienten har endret risikoprofil.

5.1.6 Frafall

Etter 3 måneder var det 75 av 299 (25%) studiedeltagere som ikke hadde besvart RMDQ. Av de gjenværende studiedeltagerne på oppfølgingstidspunktet 3 måneder (n=224) var det 15 (6.7 %) personer som fikk imputerte verdier på RMDQ. Blant studiedeltagerne med manglende besvarelse på RMDQ etter 3 måneder, som ikke hadde besvart RMDQ i etterkant av at masterstudenten fikk tildelt datamaterialet var det kun en studiedeltager som oppga årsak til frafall. Det er derfor ikke mulig å vite om frafallet er tilfeldig eller systematisk. Dersom frafallet var systematisk, kan det ha påvirket studiens generaliserbarhet. Signifikanstesting viste imidlertid ingen forskjeller mellom studiedeltagere med fullstendig besvarelse, inntil 25 % manglende verdier og de som ikke hadde besvart RMDQ etter 3 måneder. Dette er en styrke med tanke på overføringsverdien til masterstudien.

Andelen studiedeltakere som ikke hadde besvart RMDQ etter 3 måneder i masterstudien var vesentlig høyere enn hva som har blitt rapportert med hensyn til frafall i studien av Suri et al. (25) som undersøkte SBTs prediktive validitet og hadde 6 måneders oppfølging (9 %). I andre studier av SBTs prediktive validitet med en oppfølgingstid på 3 måneder har det blitt rapportert om en lignende andel som ikke hadde besvart spørreskjema for studienes utfallsmål (registrert som frafall i disse studiene) som i masterstudien. Morsø et al. (22) hadde 28 % og 14 % frafall i sine to kohorter, Kongsted et al. (119) hadde 21 %, Toh et al. (26) hadde 33 %, mens Field og Newell (117) hadde mellom 30 % og 40 % frafall i de ulike risikogrupperne. For å bevare den statistiske styrken best mulig ble det derfor valgt å imputere de manglende verdiene med gjennomsnittsverdien til de resterende studiedeltagerne. Grensen for imputering ble satt ved maksimalt 25 % manglende verdier, da denne grensen har blitt benyttet tidligere av Hill et al. (145). Samtidig er en annen imputeringsmetode, såkalt multipl imputering regnet for å være en mer valid imputeringsmetode (194). Denne metoden ble imidlertid vurdert for avansert for en masterstudie. Videre ble det ikke signifikantestet med hensyn til

forskjeller mellom studiedeltagerne med manglende og fullstendig besvarelse på de ulike bakgrunnsvariablene ved baseline. Dette fordi bakgrunnsvariablene ikke ble benyttet i videre analyser og fordi det var en relativt liten andel studiedeltagere med manglende verdier på de ulike bakgrunnsvariablene. Eventuell imputering på bakgrunnsvariablene ville derfor ikke gitt en vesentlig forskjell på sumscorene sammenlignet med ikke-imputerte sumscorer.

Ifølge COSMIN-retningslinjene kreves en utvalgsstørrelse på minst 100 personer ekskludert dem med manglende verdier ved validitetstesting (195). Til tross for at det var 25 % som ikke hadde besvart RMDQ etter 3 måneder i masterstudien var det et relativt stort antall studiedeltagere som inngikk i analysene både ved baseline og etter tre måneder. Den statistiske styrken ble likevel redusert da utvalget ble splittet basert på de anbefalte cut-off-verdiene på SBT til totalt 4 grupper ved baseline. Hver gruppe ble videre dikotomisert i henhold til et godt eller dårlig utfall etter tre måneder. Dette ga totalt 8 undergrupper i analysene av sensitivitet, spesifisitet og likelihood ratio, og andelen studiedeltagere i høyrisikogruppen med et godt og et dårlig utfall ble da relativt liten. Et større utvalg kunne trolig økt den statistiske styrken og gitt et smalere konfidensintervall rundt SBTs diskriminerende evner. Samtidig vil et større utvalg oftest redusere sjansen for å begå både type 1- og type 2-feil i analysene. Ved type 1-feil forkaster man 0-hypotesen feilaktig, mens ved type 2-feil beholder man 0-hypotesen når den skulle vært forkastet (171, s. 207). For å redusere faren for type 1-feil er imidlertid det viktigste tiltaket og øke signifikansnivået (196).

5.1.7 Analysemetoder

Analyser av sensitivitet, spesifisitet og likelihood ratio må gjennomføres ved krysstabellanalyser og det kreves derfor at både screeninginstrumentet og utfallsmålet blir dikotomisert. For SBT betydde dette at henholdsvis høy og lav risikogruppe ble isolert i de to analysesettene. Dette gjorde at det ikke ble gjennomført analyser av de prediktive evnene til medium risikogruppe isolert. Ulempen med å kategorisere kontinuerlige data er at en mister informasjon og taper statistisk styrke (197). Det er trolig også større sannsynlighet for å korrekt indikere om en person har høy eller lav risiko for langvarige funksjonsbegrensende ryggmerter dersom personen scorer veldig høyt og altså har tilstedeværelse av mange prognostiske faktorer som kartlegges med SBT eller scorer veldig lavt på SBT (tilnærmet

fravær av prognostiske faktorer kartlagt med SBT). Dette betyr at grenseområdet mellom cut-off-verdien for et «godt utfall» og et «dårlig utfall» på RMDQ i denne studien ikke nødvendigvis gir så mye informasjon. En fordel med dikotome utfallsmål er derimot at de er lettere å fortolke enn kontinuerlige.

Analysen av arealet under ROC-kurven ga informasjon om SBTs diskriminerende evner uavhengig av cut-off-verdi på de ulike risikogrupperne på SBT. Hensikten med SBT er at de ulike risikogrupperne indikerer ulike behandlingstilnærminger. Sånn sett kan det argumenteres for at nytteverdien synker når ROC-analysene blir basert på en kontinuerlig skala av totalscoren på SBT, og en optimal cut-off-verdi kun definerer høy og lav risikogruppe. Samtidig er høyrisikogruppen på SBT kun et resultat av den psykososiale subscoren (11). Det var likevel ønskelig å identifisere optimal cut-off-verdi på SBT for å diskriminere mellom et godt og et dårlig utfall i dette utvalget for å se om den optimale cut-off-verdien ga bedre prediktive evner i form av sensitivitet, spesifisitet og likelihood ratio enn de anbefalte cut-off-verdiene i masterstudien utvalg. I tillegg var det ønskelig å kunne sammenligne resultatene med studien av Beneciuk et al. (127).

Det ble ikke justert for konfunderende variabler i masterstudien. Resultatene må derfor tolkes med forbehold fordi det kan være andre faktorer en risikogruppe på SBT som med større presisjon kan predikere hvilket utfall studiedeltagerne fikk. Dersom utfallsmålet er kontinuerlig kan en multipl regresjonsanalyse benyttes for å kartlegge hvor stor del variasjonen i score på utfallsmålet som kan tilskrives henholdsvis SBT og ulike bakgrunnsvariabler (171, s. 104). Beneciuk et al. (16) fant at baselinescoren på Oswestry Disability Index forklarte mellom 38.3 % og 38.5 % av variasjonen i score på Oswestry Disability Index etter 6 måneder. SBTs totalscore og psykososiale subscore forklarte ytterligere 2.5 % og 4.5 % av variasjonen.

Ved dikotome utfallsmål slik som i masterstudien må en benytte logistisk regresjon dersom en skal justere for betydningen av konfunderende variabler (171, s. 168). Morsø et al. (22) justerte for klinisk setting (fysioterapeut eller fastlege) og endring på variabler som måler fenomener som også er inkludert i SBTs psykososiale subscore. De fant høyere odds for et dårlig funksjonsutfall hos personer i medium- og høyrisikogruppen i de justerte analysene sammenlignet med i de ujusterte analysene. Høyrisikogruppen hadde en justert odds ratio på

15.85 (95 % KI: 5.22, 48.17) for et dårlig utfall definert som ≥ 30 av 100 på RMDQ sammenlignet med lavrisikogruppen. Morsø et al. (121) justerte for smerteintensitet ved baseline og varighet av ryggsmertene og fant derimot at medium- og høyriskogruppen hadde lavere odds for et dårlig utfall i de justerte analysene enn i de ujusterte analysene. I de justerte analysene var oddsen for et dårlig utfall for personer i høyriskogruppen 3.31 (95% KI: 2.18, 5.03) sammenlignet med for personer i lavrisikogruppen. Det ble ikke gjennomført regresjonsanalyser for å kartlegge betydningen av eventuelt konfunderende variabler i masterstudien da undervisning i regresjonsmodeller ikke inngår i masterstudiet. Omfanget av masterprosjektet ble også vurdert å være tilstrekkelig med det som inngår.

SBT var opprinnelig designet som et verktøy for å iverksette stratifiserte behandlingstilnæringer basert på tilstedeværelse av prognostiske faktorer (13). I masterstudien var det imidlertid ønskelig å se om SBT kunne predikere funksjonsbegrensninger. Behandlerne i masterstudien gjennomførte ikke målrettet behandling basert på SBTs risikoklassifisering. Likevel er det sannsynlig at behandlingen som ble gitt mellom baseline og 3 måneder kan ha adressert noen av de prognostiske faktorene i SBT. Behandlingen som ble gitt er uansett en konfunderende faktor i masterstudien som kan ha påvirket studiedeltagernes utfall etter 3 måneder og dermed SBTs prediktive evner i ulik retning.

Trolig var både mengde og type behandling relativt ulik blant annet avhengig av hvilken type behandler som ble oppsøkt. Eksempelvis har personer med høy sosioøkonomisk status som mottok stratifisert behandling basert på SBT vist 2.5 ganger lavere odds for et dårlig utfall enn personer med lav sosioøkonomisk status som mottok retningslinjebasert behandling (37). I en annen studie ble det funnet at grad av funksjonsbegrensninger ved baseline i interaksjon med risikogruppe på SBT påvirket behandlingsutfall ved fysioterapi. For alle risikogruppene hadde de med høyere funksjonsbegrensninger ved baseline høyere odds for bedring av ryggsmertene. Denne effekten var imidlertid mindre uttalt desto høyere risikogruppe studiedeltagerne befant seg i. Andre variabler som påvirket SBTs prediktive evner var å være mørkhudet, å ha medisinsk forsikring og å tidligere ha oppsøkt fysioterapeut for behandling (28). Dersom behandlingen som ble gitt i masterstudien hadde dårlig effekt på studiedeltagernes funksjonsbegrensninger vil dette trolig ha påvirket SBTs prediktive og diskriminerende egenskaper negativt. Dette fordi behandling med negativ effekt trolig vil ha

ført til flere falske negative testresultater. Samtidig vil en behandling med god effekt kunne ha ført til falske positive resultater, og SBTs sensitivitet vil dermed bli lavere. Det er uansett ønskelig at behandlingen som gis skal påvirke pasientens prognose positivt, men dette vil nødvendigvis påvirke et screeninginstruments prediktive evner negativt.

SBT er utformet på et gruppenivå (11). Det ble ikke testet hvorvidt enkeltspørsmål har større betydning for SBTs diskriminerende evner enn andre. Dette kunne blitt gjort ved å gjennomføre en ROC-analyse for hvert enkelt delspørsmål på SBT. Dette er gjennomført for SBTs diskriminerende evner ved baseline i originalstudien (11) og av Morsø et al. (143). Dersom en hadde funnet at enkeltspørsmål skiller seg ut med hensyn til å kunne diskriminere mellom et godt og dårlig utfall, kunne dette vært nyttig kunnskap for en kliniker å ta med i en totalvurdering, spesielt i tilfeller hvor pasienter scorer på grensen mellom to ulike risikogrupper.

5.2 Resultatdiskusjon

5.2.1 Resultatene sett opp mot forskningshypotesene

1. Start Back Screening Tools evne til å predikere et dårlig utfall etter 3 måneder er moderat for sensitivitet og spesifisitet, og lav for positiv og negativ likelihood ratio (Dette ble forventet på bakgrunn av funn gjort i studien av Suri et al. (25)).

Denne hypotesen ble delvis bekreftet. Det var kun cut-off lav/medium mot høy risikogruppe som hadde klinisk nytteverdi. Ved bruk av cut-off lav/medium mot høy risikogruppe hadde SBT svært lav sensitivitet og til gjengjeld svært høy spesifisitet. Positiv likelihood ratio var svært høy, mens negativ likelihood ratio var lav. Da både spesifisiteten og den positive likelihood ratioen var svært høy vurderes høyrisikogruppen å ha moderate evner til å predikere funksjonsutfall. Ved bruk av cut-off lav- mot medium/høy risikogruppe var sensitiviteten også lav, mens spesifisiteten var moderat. Videre var det ved bruk av cut-off lav mot medium/høy risikogruppe en stor andel både falske positive og falske negative testresultater. Denne cut-off-verdien hadde derfor ingen klinisk nytteverdi.

Positiv og negativ likelihood ratio baserer seg både på sensitivitet og spesifisitet, og det var derfor ikke overaskende at disse estimatene heller ikke hadde klinisk nytteverdi ved bruk av cut-off lav mot medium/høy risikogruppe. Den positive likelihood ratioen var da 1.95, mens den negative likelihood ratioen var 0.73. Den positive likelihood ratioen lå imidlertid opp mot grensen for å indikere klinisk relevans (2 og 0.5 for henholdsvis positiv og negativ likelihood ratio) (174). Negativ likelihood ratio ved bruk av cut-off lav/medium mot høy risikogruppe gav ingen klinisk nytteverdi, mens den positive likelihood ratioen var 11. Dette betyr at å havne i høyrisikogruppen ved baseline økte sannsynligheten for et dårlig utfall etter tre måneder betraktelig sammenlignet med å havne i lav/medium risikogruppe. En ser likevel at antallet studiedeltagere i høyrisikogruppen kun var 11. Videre ble 1 av disse identifisert som falsk positiv. Dersom ytterligere en studiedeltager hadde blitt identifisert som falsk positiv ville dette fått stor betydning for estimatene av likelihood ratio (Se tabell 6). Spesifisiteten ble ikke påvirket nevneverdig da det relativt sett var et stort antall studiedeltagere som ble klassifisert som ekte negative. Det kan derfor være at den svært høye positive likelihood ratioen som ble identifisert når høyrisikogruppen ble isolert kan forklares av lav statistisk styrke.

2) Start Back Screening Tool har lave til moderate evner til å diskriminere mellom et godt og dårlig utfall etter 3 måneder. (Dette ble forventet på bakgrunn av studiene til Kongsted et al. (119) og Morsø et al. (22)).

Denne hypotesen ble bekreftet. Arealet under ROC-kurven var 0.68 (95 % KI 0.61, 0.75) Dette nådde imidlertid nesten grensen for å indikere moderate diskriminerende evner på 0.7 (176), og en ser at øvre del av konfidensintervallet strekker seg opp til 0.75. Arealet under ROC-kurven ble beregnet på bakgrunn av andelen personer med et positivt og negativt testresultat, da «ekte positive» utgjør y-aksen og «falske positive» x-aksen. Resultatene indikerer derfor at SBTs totalscore ikke har klinisk nytteverdi i å diskriminere mellom et godt og et dårlig utfall hos personer ≥ 55 år som oppsøker primærhelsetjenesten med ryggmerter. Dette er naturligvis gitt at en benytter SBTs totalscore, samme utfallsmål, cut-off-verdi og metodiske forhold som i masterstudien.

3) Den optimale cut-off-verdien for å skille mellom et godt og dårlig utfall i dette utvalget vil være lavere enn den optimale cut-off-verdien som ble identifisert av Beneciuk et al. (127) (≥ 5 poeng på SBTs totalscore). Dette ble forventet da det er vist at det er flere som scorer lavt på totalscoren på SBT som har et dårlig utfall i aldersgruppen ≥ 65 år (25) enn i den yrkesaktive delen av befolkningen (127).

Denne hypotesen ble bekreftet. Den optimale cut-off-verdien for å skille mellom et godt og et dårlig utfall i masterstudien var ≥ 3 poeng på SBT. Den optimale cut-off-verdien er avhengig av fordelingen av totalscore på SBT i utvalget og andelen med et dårlig utfall. Studien av Suri et al. (25) med en subanalyse av personer ≥ 65 år identifiserte en større andel med et dårlig utfall blant dem med lavere totalscore på SBT (27 % og 71 % i henholdsvis lav og medium risikogruppe) enn Beneciuk et al. (127). Masterstudien identifiserte også en større andel med et dårlig utfall blant dem med lav totalscore på SBT (34 % og 52 % i henholdsvis lav og medium risikogruppe) enn Beneciuk et al. (8 % og 39 % med et dårlig utfall i henholdsvis lav og medium risikogruppe) (127). Det kan derfor være at høyere alder hos studiedeltagerne i masterstudien enn i studien av Beneciuk et al. (127) er en årsak til den ulike andelen med et dårlig utfall.

Den optimale cut-off-verdien er også avhengig av hvilket dikotomt utfallsmål som benyttes. Beneciuk et al. (127) benyttet en kombinasjon av RMDQ og skala for smerterelatert psykisk stress og uhensiktsmessige mestringsstrategier som utfallsmål. Forskjellene i utfallsmål kan derfor også være en årsak til at optimal cut-off-verdi er lavere i masterstudien enn i studien av Beneciuk et al. (127). En annen årsak kan være at Beneciuk et al. (127) undersøker samtidig validitet. Det er logisk å anta at det en høyere andel studiedeltagere med et dårlig utfall ved baseline enn etter 3 måneder og at dette igjen vil kunne påvirke den optimale cut-off-verdien.

5.2.2 Sammenligning av resultater og tidligere forskning

I flere enkeltstudier av eldre med rygg smerter har det blitt rapportert at psykososiale symptomer som høy grad av depressive symptomer, symptomer på angst og liten tro på bedring av ryggsmertene har en negativ innvirkning på prognosen (87, 88, 93, 94). SBT

demonstrerte i masterstudien lave diskriminerende evner, imidlertid helt på grensen til hva som tolkes som moderate diskriminerende evner. To studier har også rapportert lavere diskriminerende evner enn masterstudien, begge med et areal under ROC-kurven på under 0.57 (118, 119). Fire studier har rapportert moderate diskriminerende evner (22, 23, 187, 198), mens originalstudien av Hill et al. (11) rapporterte høye diskriminerende evner i en populasjon hvor gjennomsnittsalderen var 45 ± 10 år. Disse studiene kan derfor tyde på at det også er andre psykososiale, fysiske eller demografiske faktorer som har betydning for prognosen hos eldre med ryggsmarter. Enkeltstudier peker eksempelvis på at faktorer som øker med alderen har betydning for prognosen ved ryggsmarter. Dette inkluderer blant annet et større antall komorbide plager (65, 87, 90, 92), kneartrose, lumbal spinal stenose og >2 fall siste 3 uker før studieinkludering (92). I tillegg er økende alder også blant den eldste aldersgruppen rapportert å være en prognostisk faktor i flere studier (88, 92, 93). Dette kan være med på å forklare hvorfor SBT hadde lavere diskriminerende evner i masterstudien enn i studier av den yrkesaktive befolkningen.

Ulike populasjoner, utfallsmål og oppfølgingstid vanskeliggjør en direkte sammenligning mellom SBTs prediktive validitet i masterstudien og andre studier. Alle studiene av SBTs prediktive validitet er gjennomført på personer ≥ 18 år, med unntak av studien av Suri et al. (25) som har gjennomført en subanalyse på personer ≥ 65 år. Flere andre studier i primærhelsetjenesten benyttet også en RMDQ-score på ≥ 7 poeng (eller tilsvarende på en 0-100-skala) som utfallsmål (11, 22, 25, 187). Av disse er det kun Morsø et al. (22) som også har en oppfølgingstid på tre måneder. Tabell 9 refererer til de ulike studiene gjennomført i primærhelsetjenesten som blir sammenlignet med masterstudien.

5.2.3 Resultater sammenlignet med andre studier av Start Back Screening Tool

Hill et al. (11), Beneciuk et al. (16), Toh et al. (26) og Karstens et al. (198) var de eneste av sammenligningsstudiene av prediktiv validitet fra primærhelsetjenesten som oppga middelveier for totalscore og psykososial subscore ved baseline for hele kohorten samlet. Median totalscore på SBT og psykososial subscore i masterstudien var cirka poenget lavere enn gjennomsnittlig totalscore og psykososial subscore ved baseline i studiene av Hill et al. (11), Beneciuk et al. (16) og Toh et al. (26). Gjennomsnittlig totalscore og subscore i studien

av Hill et al. (11) var henholdsvis 3.83 ± 2.3 poeng og 1.82 ± 1.5 poeng. I studien av Karstens et al. (198) var totalscore og subscore nesten nøyaktig den samme som i masterstudien.

I masterstudien var det 65 %, 27 % og 7 % i henholdsvis lav-, medium- og høyrisikogruppen ved baseline. Det er vesentlig flere i lavrisikogruppen og færre i høyrisikogruppen i masterstudien enn majoriteten av sammenligningsstudiene (11, 20, 22, 24-26, 28, 117, 118, 184, 187). Andelen i medium risikogruppe ved baseline var i de aller fleste av disse studiene kun noe lavere enn i masterstudien. Et unntak var studien av Matsudaira et al. (19) hvor det var en større andel studiedeltagere i lavrisikogruppen (78 %), færre i medium risikogruppe (14 %) og omtrent nøyaktig samme andel studiedeltagere i høyrisikogruppen (8 %) sammenlignet med masterstudien. Videre ble det i en populasjonsstudie av Haglund et al. (30) funnet lignende resultater som i masterstudien. Utvalget besto av personer som hadde hatt korsryggsmertner siste uken. SBT identifiserte 80 %, 18 % og 2 % i henholdsvis lav-, medium- og høyrisikogruppen ved baseline. Noe av forklaringen på at de fleste i denne studien tilhørte lavrisikogruppen, og svært få tilhørte høyrisikogruppen kan være at studien inkluderte en større andel eldre studiedeltagere (personer mellom 40 og 70 år), eller at studien hadde få inkluderte studiedeltagere. En annen nærliggende forklaring er at studien inkluderte personer uavhengig av om de søkte hjelp for korsryggsmertene eller ikke. Karstens et al. (198) hadde også en fordeling i risikogrupperne ved baseline som var forholdsvis lik som i masterstudien. Medium risikogruppe skal identifisere personer med overvekt av fysiske risikofaktorer, mens høyrisikogruppen skal identifisere personer med tilstedeværelse av psykososiale faktorer som kan hindre bedring av ryggsmertene (11). På bakgrunn av resultatene fra majoriteten av studiene som er presentert i dette avsnittet kan det derfor se ut til at studiedeltagere i andre studier har større tilstedeværelse av både fysiske og spesielt psykososiale prognostiske faktorer enn studiedeltagerne i masterstudien.

Disse forskjellene i tilstedeværelse av fysiske og psykososiale prognostiske faktorer kan forklares ved forskjeller i utvalget. Masterstudien inkluderte personer ≥ 55 år. Selv om dette også inkluderte en del personer i yrkesaktiv alder, så inkluderte sammenligningsstudiene stort sett kun personer mellom 18 og 65 år. Videre var det forskjeller i rekrutteringsmåte samt inklusjons- og eksklusjonskriterier. Masterstudien inkluderte kun personer som ikke hadde oppsøkt primærbehandler siste halvår, dette kriteriet hadde ikke de andre studiene. De fleste studiene inkluderte kun personer med ulike definisjoner av uspesifikke korsryggsmertner (11, 22, 23, 25, 117) eller ekskluderte personer med spesifikke årsaker til ryggsmertene slik som

spinal stenose, vertebral fraktur, skivedegenerasjon og osteoporose (24, 25). Videre var det i masterstudien 56 % av studiedeltagerne med en smertevarighet på under 4 uker. Andelen i studiene som oppgav dette var mellom 17 % og 44 % (11, 22, 25), og mellom 37 % og 62 % avhengig av risikogruppe i studiene av (117, 118). Dette kan bety at masterstudien i større grad fanget opp tilfeller av ryggsmarter med kort smertevarighet, og at det derfor var mindre grad av fysiske og psykososiale prognostiske faktorer tilstede.

SBTs sensitivitet og spesifisitet kan sammenlignes med andre studier fra primærhelsetjenesten. I England hadde SBT lignende spesifisitet ved bruk av henholdsvis cut-off lav/medium mot høy risikogruppe og lav mot medium/høy risikogruppe (94.6 % og 65.4 %), men høyere sensitivitet (henholdsvis 39.6 % og 80.1 %) enn i masterstudien (11). I australsk primærhelsetjeneste var også sensitiviteten høyere med bruk av de anbefalte cut-off-verdiene. Her var imidlertid sensitiviteten vesentlig høyere ved bruk av lav/medium mot høy risikogruppe (86 %), mens spesifisiteten var vesentlig lavere (37 %). For cut-off lav mot medium/høy risikogruppe var både sensitiviteten (40 %) og spesifisiteten (76 %) tilnærmet lik som i masterstudien (187). I studien av Suri et al. (25) fra USA hadde cut-off lav/medium mot høy risikogruppe også sammenlignbare resultater med masterstudien. Spesifisiteten var noe lavere (91.8 %), mens sensitiviteten til gjengjeld var høyere (34.2 %) enn i masterstudien. For cut-off lav mot medium/høy risikogruppe var sensitiviteten vesentlig bedre (81.0 %) og spesifisiteten dårligere (64.6 %). Dette samsvarer med studiene fra England og Australia. Samlet indikerer resultatene fra masterstudien at SBT hadde høyere spesifisitet enn hva som er rapportert tidligere og samtidig lavere sensitivitet.

Ved bruk av de anbefalte cut-off-verdiene var spesifisiteten høyere i masterstudien enn i sammenligningsstudiene når høyrisikogruppen ble isolert. Dette betyr at det var færre falske positive testresultater i masterstudien. Videre var sensitiviteten gjennomgående dårligere i med bruk av begge cut-off-verdiene enn i sammenligningsstudiene. Dette betyr at det var en høyere andel falske negative testresultater i masterstudien. Dette peker på at formålet med å benytte SBT ikke stemmer for masterstudiens utvalg. Det kan være ulike årsaker til at såpass mange i lavrisikogruppen i masterstudien hadde et dårlig utfall. Flere studier har vist at SBT har dårligere prediktive evner i populasjoner med kortere varighet av ryggsmertene (21, 116). Det kan også være at psykososiale faktorer ikke er like viktige som prognostiske faktorer hos eldre eller at det er andre prognostiske faktorer som er like viktige hos eldre som ikke blir

kartlagt med SBT. I denne sammenheng vil det være nærliggende å kartlegge aldersrelaterte faktorer som tidligere er rapportert å ha prognostisk verdi nærmere (87, 92).

Resultatene for positiv og negativ likelihood ratio er også sammenlignbare med resultater fra primærhelsetjenesten (Tabell 9). I England var positiv og negativ likelihood ratio ved bruk av cut-off lav/medium mot høy risikogruppe 5.51 og 0.74 (11). Med tilsvarende bruk av cut-off-verdi i Australia var den positive likelihood ratioen 1.67 og den negative 0.79 (187). I USA var positiv og negativ likelihood ratio ved bruk av samme cut-off-verdi henholdsvis 4.16 og 0.72 (25). Av disse resultatene ser en derfor at den positive likelihood ratioen i masterstudien på 11 var betydelig høyere og den negative likelihood ratioen på 0.9 lavere enn hva som er rapportert i sammenligningsstudiene. Det var heller ikke i noen av sammenligningsstudiene rapportert om klinisk relevante resultater for negativ likelihood ratio ved bruk av denne cut-off-verdien. Årsakene til forskjeller i likelihood ratio kan være de samme som er nevnt for sensitivitet og spesifisitet. Videre rapporterte en nyere studie fra den nederlandske armen av BACE at den største fremgangen med hensyn til funksjonsbegrensninger hos eldre med ryggsmerte skjedde de første 3 månedene etter baseline, men at det også var noe fremgang frem til 6 måneder (89). Det største problemet i masterstudien er at det eksisterer en høy andel falske negative, og det kan derfor være at en lengre oppfølgingstid ville gjort at SBT demonstrerte bedre prediktive evner. Samtidig kan forskjeller i klinisk setting med hensyn til land og hvilke profesjoner studiedeltagerne har oppsøkt påvirke SBTs nøyaktighet.

Sensitivitets- og spesifisitetsverdiene som ble identifisert ved bruk av optimal cut-off-verdi var lave og verdiene for likelihood ratio hadde ingen klinisk nytteverdi i masterstudien. Det eksisterer få andre studier som har sett på prediktiv validitet for optimal cut-off-verdi på SBT. I studien av Beneciuk et al. (127) ga en optimal cut-off-verdi på ≥ 5 en sensitivitet og spesifisitet på 89 % og 82 % og en positiv og negativ likelihood ratio på henholdsvis 4.98 og 0.13. Disse resultatene er vesentlig bedre enn resultatene i masterstudien. Videre rapporterte en studie av en modifisert generisk versjon av SBT at optimal cut-off-verdi for å predikere fysisk helse etter 6 måneder var ≥ 5 i ryggpopulasjonen. Denne cut-off-verdien hadde en sensitivitet og spesifisitet på henholdsvis 71 % og 65 %. Dataene i denne studien er imidlertid heller ikke direkte sammenlignbare med masterstudien da denne studien benyttet en modifisert versjon av SBT (126). Morsø et al. (121) gjennomførte post-hoc-analyser hvor de undersøkte om endring av cut-off-verdi på de ulike risikogruppene påvirket SBTs evne til å predikere funksjonsbegrensninger. Forskerne i denne studien fant at endring av cut-off-verdi

på risikogruppene ikke påvirket SBTs prediktive evner i særlig grad. Forfatterne oppga ingen optimal cut-off-verdi for et dårlig utfall. Den optimale cut-off-verdien i masterstudien var derfor lavere enn i andre sammenlignbare studier.

Tabell 9: Viser et utvalg av resultater fra studier av SBTs prediktive validitet med lignende utfallsmål, oppfølgingstid eller klinisk setting.

Studie	Klinisk setting og alder	Definisjon av utfall m/cut-off-verdi, oppfølgings-tidspunkt (er)	%fordeling i lav/med./høy RG m/dårlig utfall	Sensitivitet (%)	Spesifisitet (%)	Likelihood ratio	AUC (95 % KI)	OR for dårlig utfall (justerte), lav/med./Høy RG (95 % KI)
Hill et al. (11)	Primærh.tj England 18-59 år	RMDQ (0-24) ≥ 7 , 6 mnd	17/53/78	Lav/med.- høy: 39.6 Lav-med./ høy: 80.1	Lav/med.- høy: 94.6 Lav-med./ høy: 65.4	Lav/med.- høy: 5.51 - LR: 0.74 Lav -med./høy: 2.32 - LR: 0.30	0.90 (0.88, 0.93)	
Forsbrand et al. (184)	Primærh.tj Sverige 18-67 år RS og/eller NS	EQ-5D (-0.59-1.0) < 0.6 Median 13 mnd	4/11/36	Lav/med.- høy: 38.5 Lav - med./høy: 84.6	Lav/med.- høy: 91.5 Lav - med./høy: 46.7	Lav/med.- høy: 4.53 - LR: 0.67 Lav - med./høy: 1.59 - LR: 0.33	0.73 (0.61, 0.84)	1/ 1.84 (0.51, 6.51)/ 6.16 (1.50, 25.26)
Field og Newell (117)	Primærh.tj England, > 18 år	PGIC (1-7) ≥ 6 , («mye verre», «veldig mye verre»), 90 dager	12/14/9	94	23			19.3 (3.0, 29.0)/ 3.0 (1.3, 7.0)
Morsø et al. (22)	Primærh.tj Danmark 18-65 år	RMDQ (0-100) ≥ 30 , 3 mnd	24/57/64				0.71 (0.66, 0.77)	1/ 8.16 (3.44, 19.30)/ 15.85 (5.22, 48.17)
Newell et al. (118)	Primærh.tj England > 16 år	PGIC (1-7) ≥ 6 , («mye bedre», «veldig mye bedre»), 90 dager					0.53 (0.45, 0.61)	

Pagé et al. (23)	Primærh.tj Frankrike 18-60 år	ODI ≥24 % av baselinescore, 6 mnd		Lav- med/høy: 66.7	Lav- med/høy: 77.5	0.84 (0.69, 1.00)
Kongsted et al. (119)	Primærh.tj Danmark 18-65 år	RMDQ (0-100), >8, 3 mnd				0.59 (0.55, 0.63) <i>Forventet kort forløp: +LR: 0.73. -LR: 1.5 Forventet forlengt forløp: +LR: 1.4. - LR: 0.83 Forventet langvarig forløp: + LR: 2.0. -LR: 0.94</i>
Karstens et al. (198)	Primærh.tj Tyskland >18 år	CPGS (0-100), ≥grad 3 og 4, 12 mnd				0.79 (0.73, 0.85)
Kendell et al. (187)	Primærh.tj Australia 18-70 år	RMDQ (0-24), ≥7 1 år	12/38/33	Lav- med./høy: 86 Lav/med. - høy: 40	Lav-med./ høy:37 Lav/med. - høy: 76	0.71 (0.64, 0.77) Lav/med. -høy: +LR: 1.67. -LR: 0.79 Lav -med./høy: +LR: 1.36. -LR: 0.39
Katzan et al. (28)	Primærh.tj USA ≥18 år	MDQ (0-100) ≥10 poeng reduksjon i score, Opp mot 1 år				Odds for bedring: 1/0.61 (0.47, 0.80/0.48 (0.36, 0.64)
Suri et al. (25)	Primærh.tj USA ≥ 18 år	RMDQ (0-24) ≥7, 6 mnd	≥ 65 år: 27/71/79	Lav/med. - høy: 34.2 Lav- med./ høy: 81.0	Lav/med.- høy: 91.8 Lav- med./ høy: 64.6	Lav/med. -høy: +LR: 4.16. -LR: 0.72 Lav- med./høy: +LR: 2.29. -LR 0.29
Toh et al. (26)	Primærh.tj Singapore >20 år	NRS (0-10) ≥5, 3 mnd				1/1.29 (0.66, 2.49)/ 3.29 (1.49, 7.24)

Forkortelser og forklaringer: Primærh.tj. = primærhelsetjenesten. NRS= numerisk smerteskala. RMDQ= Roland Morris Disability Questionnaire. ODI: Oswestry Disability Index. GPE= Global Perceived Effect. CPGS = Chronic Pain Grade Scale. MDQ= Modified low back pain disability questionnaire. EQ-5d: EuroQol 5-dimension Questionnaire. OR= Odds ratio. Odds ratio gjelder for et dårlig utfall dersom ikke annet er oppgitt. -LR= Negativ likelihood ratio. +LR= Positiv likelihood ratio. AUC= Areal under Receiver Operating Characteristics Curve. Lav/med.= lav/medium risikogruppe. Med./høy= medium/høy risikogruppe. Mnd.= måneder. RG= risikogruppe. KS= korsryggsmerter. RS: ryggsmerter. NS: nakkesmerter. Mod.= modifisert. KI: Konfidensintervall.

5.3 Kliniske implikasjoner hos eldre med ryggmerter

I masterstudien ble det gjennomført analyser av prediktiv validitet for optimal og anbefalte cut-off-verdier på SBT. Til tross for at det på bakgrunn av masterstudien ikke kan anbefales å benytte den optimale cut-off-verdien fremfor de anbefalte cut-off-verdiene kan det i ulike situasjoner og settinger være flere faktorer som avgjør hvilken cut-off-verdi en benytter seg av. Dette vil diskuteres videre i delkapitlet.

Fordelen med bruk av optimal cut-off-verdi er at den har en bedre samlet evne til å identifisere både et godt og et dårlig utfall enn de anbefalte cut-off-verdiene. Likelihood ratioen ved bruk av optimal cut-off-verdi i masterstudien indikerte imidlertid ingen klinisk nytteverdi (174). Dette taler for å benytte de anbefalte cut-off-verdiene, som hadde svært god spesifisitet og positiv likelihood ratio når høyriskogruppen ble isolert mot lav/medium risikogruppe. Selv om nærmest alle som havnet i høyriskogruppen ved baseline hadde et dårlig utfall, var det også mange i lav- og medium risikogruppe som hadde et dårlig utfall (mange falske positive). Dette gjorde at den negative likelihood ratioen ga lite informasjon, da personer i lav og medium risikogruppe ikke hadde betydelig lavere sannsynlighet for et dårlig utfall enn personer i høyriskogruppen. Hvorvidt en skal benytte den optimale eller de anbefalte cut-off-verdiene avhenger også av hvilke egenskaper ved screeningsinstrumentet en vektlegger. Dette kan variere utfra om screeninginstrumentet skal vurdere risiko hos den enkelte pasient eller benyttes på gruppenivå. Det viktigste formålet med å benytte SBT er å identifisere personer i høyriskogruppen som kan nyttiggjøre seg en kognitiv intervensjon. Det var kun høyriskogruppen som hadde en prediktiv nytteverdi ved screening med SBT i masterstudien. Høy mot lav/medium risikogruppe demonstrerte nærmest perfekt spesifisitet. Det er derfor svært lite sannsynlig at en person som havnet i høyriskogruppen ved baseline hadde et godt utfall etter 3 måneder. På grunn av lav statistisk styrke i høyriskogruppen må resultatene fra masterstudien imidlertid etterprøves av andre før screening med SBT kan anbefales til bruk i klinisk praksis på eldre personer med ryggmerter.

SBT ble utviklet med den hensikt at ulike behandlingsstrategier skal følge de ulike risikogrupperne. Cut-off-verdien for høyriskogruppen som skal motta mer ressurskrevende kognitive behandlingstiltak og ble derfor satt høyt (11). Siden utvalget splittes i to ved bruk av den optimale cut-off-verdien, vil dette sannsynligvis fange opp flere personer med et dårlig utfall enn ved bruk av de anbefalte cut-off-verdiene. Det er en betydelig andel av dem som

hadde et dårlig utfall som havnet i lav eller medium risikogruppe ved baseline ved bruk av de anbefalte cut-off-verdiene. Det er derfor ikke trolig at deres behandlingsbehov vil adresseres selv om de havner i høyrisikogruppen ved bruk av optimal cut-off-verdi (høyrisikogruppen skal motta kognitiv intervensjon (11)). Videre ble den optimale cut-off-verdien basert på totalscoren på SBT. Denne inkluderer psykososiale aspekt, men ved en cut-off-verdi på ≥ 3 poeng er det like sannsynlig at det var fysiske prognostiske faktorer som førte til at studiedeltagerne havnet i høyrisikogruppen. Siden utvalget splittes i to ved bruk av optimal cut-off-verdi falt medium risikogruppe bort. Derfor er det sannsynlig at bruk av optimal cut-off-verdi vil føre til overbehandling av personer som ikke har behov for en kognitiv intervensjon, da andelen falske positive vil bli høyere. Siden svært få i den eldre kohorten i masterstudien med et dårlig utfall har tilstedeværelse av mange psykososiale prognostiske faktorer og trolig ikke vil nyttiggjøre seg en kognitiv intervensjon reises spørsmålet om hvilken behandling disse bør få. Dette faller likevel utenfor masterstudiens rammer og diskutere.

Ved analyser av optimal cut-off-verdi kan en i stedet for å velge den cut-off-verdien med best balanse mellom sensitivitet og spesifisitet, prioritere ett av parameterne fremfor det andre. Det eksisterer ingen fasit på om sensitivitet eller spesifisitet er viktigst for screeninginstrumenter i ryggpopulasjonen. Siden formålet med å screene etter prognostiske psykososiale faktorer er å identifisere de som har risiko for vedvarende problemer fremfor å sette en klinisk diagnose er det argumentert for at sensitiviteten er viktigere enn spesifisiteten (199). Ved å prioritere høy sensitivitet vil SBT være effektivt i å utelukke et dårlig klinisk utfall dersom pasienten havner i lav eller medium risikogruppe på SBT ved baseline (9). Det vil altså være svært sannsynlig at høyrisikogruppen fanger opp de pasientene som har behov for videre kartlegging og prioritering av ressurser. Dersom det er små konsekvenser ved å predikere feil prognose (79, s. 120) kan det argumenteres for å prioritere spesifisitet fremfor sensitivitet. Konsekvensene av langvarige ryggsmarter kan være sykefravær, nedsatt funksjon i daglige aktiviteter eller psykiske plager som angst og depresjon (9). Dersom det ikke eksisterer effektive behandlingstiltak eller eksisterende behandling er lite kostnadseffektiv (79, s. 120) kan det også argumenteres for å prioritere spesifisitet. Risikostratifisert behandling basert på SBT har demonstrert større behandlingseffekter og høyere kostnadseffektivitet sammenlignet med annen retningslinjebasert behandling, men potensielle kostnader ved en stor andel falske positive kan være høye (12-14). Så langt oppgaveforfatteren er kjent, har det ikke blitt

gjennomført studier av behandling rettet mot prognostiske psykososiale faktorer hos eldre. Dersom en skulle ha prioritert sensitivitet eller spesifisitet er det ideelt sett hensiktsmessig med høy sensitivitet for å fange opp alle som har behov for mer omfattende kartlegging eller behandling. Likevel er behandling rettet mot høyrisikogruppen ressurskrevende og det er usikkert hvorvidt det har en positiv behandlingseffekt og er kostnadseffektivt hos eldre. Det kan derfor ikke anbefales å prioritere hverken sensitivitet eller spesifisitet fremfor den andre.

I denne studien forteller sensitivitet og spesifisitet hvor stor andel av personer med et godt og dårlig utfall på SBT som ble identifisert i den høyeste og laveste risikogruppen. SBTs sensitivitet og spesifisitet gir dermed nyttig informasjon for klinikerne, men kun på et gruppenivå. Likelihood ratio utnytter derimot informasjonen som en test gir på et gruppenivå slik at den kan overføres til den enkelte pasient (169). Ved å bruke pre-test-odds for en «sykdom» kan en med likelihood ratio også regne ut hvor mye et gitt test-resultat endrer sannsynligheten for sykdom hos den enkelte pasienten (176)¹.

SBT skal ikke danne grunnlag for den kliniske beslutningstakingen alene, og er kun ment som et supplement til anamnesen og den kliniske undersøkelsen (11). En mulighet er å kombinere bruk av ulike kartleggingsinstrumenter. En studie av Haglund et al. (30) fant at man ved å benytte et kartleggingsinstrument som inneholdt spørsmål om smerteintensitet og utbredelse av smerten i kombinasjon med SBT hadde større sjanse for å identifisere personer med høy risiko for langvarige plager. Dette fordi smerteintensitet og smerteutbredelse er prognostiske faktorer som ikke kartlegges med bruk av SBT. Klinikerne kan eksempelvis også inkorporere tidligere sykehistorie og behandlingshistorikk, pasientenes preferanser og forventninger, samt livsomstendigheter i sin risikovurdering (103). Selv om SBT alene ikke kan anbefales for den eldre populasjonen på bakgrunn av masterstudien kan det være at SBT i kombinasjon med andre kartleggingsinstrumenter eller klinikernes risikovurdering ville demonstrert bedre prediktive evner.

¹ Pretest-odds (andel med et godt/andel med et dårlig utfall i hele kohorten) * likelihood ratio= posttest-odds

5.4 Videre forskning

SBT er oversatt til norsk, men det bør gjennomføres en fullstendig testing av psykometriske egenskaper. Dette bør gjennomføres for å sikre spørreskjemaets interne validitet. Videre har det blitt satt i gang flere studier av stratifisert behandling i primærhelsetjenesten. Dette gjelder for pasienter med korsryggmerter i Danmark (200), isjias-pasienter i Storbritannia (201), for pasienter i ulike helsesystemer i USA (202, 203), i Portugal (204) og for personer mellom 18 og 80 år i sekundærhelsetjenesten i Brasil (205). Hvis sistnevnte studie kan vise til overordnede behandlingsresultater og kostnadseffektivitet sammenlignet med retningslinjebasert behandling gir det grunnlag for å undersøke stratifisert behandling basert på SBT også hos eldre.

I denne studien var det kun høyrisikogruppen hadde klinisk nytteverdi i å predikere funksjonsutfall. Det trengs imidlertid flere studier av SBTs prediktive validitet som har større statistisk styrke i høyrisikogruppen. Dersom fremtidige studier kan vise til like gode eller bedre prediktive evner for høyrisikogruppen på SBT eller screeninginstrumentet samlet, kan det gi grunnlag for å anbefale å benytte SBT hos eldre. Videre har to studier fra primærhelsetjenesten rapportert at den psykososiale subscoren hadde minst like gode og bedre prediktive evner som totalscoren på SBT (16, 26). Med bakgrunn i dette kan det også være interessant å undersøke de samlede diskriminerende evnene til den psykososiale subskalaen på SBT isolert, hos eldre.

6.0 KONKLUSJON

Det var kun høyrisikogruppen på SBT som hadde evne til å predikere et dårlig utfall etter 3 måneder i masterstudien. Studiedeltagerne som scoret høyt på SBT ved inklusjon hadde større funksjonsbegrensninger 3 måneder etter. Screening med SBT hadde ingen prediktiv verdi for lavrisikogruppen. SBTs samlede diskriminerende evner for et godt eller dårlig utfall etter 3 måneder var lav, men på grensen til moderat. Den optimale cut-off-verdien på SBT i dette utvalget var ≥ 3 av 9 poeng, men dette tilførte ikke noe ytterligere med hensyn til SBTs prediktive evner. Samlet indikerer resultatene at SBT kan ha nytteverdi i å predikere dårlige kliniske utfall hos eldre personer som oppsøker primærhelsetjenesten på grunn av rygg smerter dersom de klassifiseres i høyrisikogruppen ved baseline. Antall studiedeltagere som ble klassifisert i høyrisikogruppen var imidlertid få og det er behov for mer forskning for å kunne anbefale screening med SBT hos eldre personer med rygg smerter i klinikk og forskning.

7.0 REFERANSELISTE

1. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2012;380(9859):2163-96.
2. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. *The Lancet*. 2018;391(10137):2356-67.
3. Lærum E, Brage S, Ihlebæk C, Johnsen K, Natvig B, Aas, E. Et muskel- og skjelettrengskap: Forekomst og kostnader knyttet til skader, sykdommer og plager i muskel- og skjelettsystemet. Oslo: FORMI, Klinikk for kirurgi og nevrofag, Oslo universitetssykehus; 2013 [Hentet 29. april 2020]. MST-rapport 1. Tilgjengelig fra: <https://fysio.no/Media/Files/Et-muskel-og-skjelettrengskap.-Forekomst-og-kostnader-knyttet-til-skader-sykdommer-og-plager-i-muskel-og-skjelettsystemet>
4. Hagen K, Svebak S, Zwart J-A. Incidence of Musculoskeletal Complaints in a Large Adult Norwegian County Population. The HUNT Study. *Spine*. 2006;31(18):2146-50.
5. Dionne CE, Dunn KM, Croft PR. Does back pain prevalence really decrease with increasing age? A systematic review. *Age Ageing*. 2006;35(3):229-34.
6. Meucci RD, Fassa AG, Faria NM. Prevalence of chronic low back pain: systematic review. *Rev Saude Publica*. 2015;49(73):1-9
7. Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, et al. A systematic review of the global prevalence of low back pain. *Arthritis and rheumatism*. 2012;64(6):2028-37.
8. Lærum E, Brox JI, Storheim K, Espeland A, Haldorsen E, Munch-Ellingsen J, et al. Korsryggsmerter - med og uten nerverotaffeksjon [Internett]. Oslo: Formidlingsenheten for muskel og skjelettlidelser, sosial- og helsedirektoratet; 2007 [Hentet 29.april 2020]. Tilgjengelig fra: <http://www.muskelskjeletthelse.no/wp-content/uploads/2016/06/Nasjonale-kliniske-retningslinjer-korsryggsmerter-2007-Fullversjon.pdf>
9. National Institute for Health and Care Excellence: Clinical Guidelines. Low Back Pain and Sciatica in Over 16s: Assessment and Management. Assessment and non-invasive treatments NICE guideline NG59 [Internett]. United Kingdom: National Guideline Centre; 2016 [Hentet 29. april 2020]. Tilgjengelig fra: https://www.ncbi.nlm.nih.gov/books/NBK401577/pdf/Bookshelf_NBK401577.pdf

10. Walsh NE, Brooks P, Hazes JM, Walsh RM, Dreinhofer K, Woolf AD, et al. Standards of care for acute and chronic musculoskeletal pain: the Bone and Joint Decade (2000-2010). *Arch Phys Med Rehabil.* 2008;89(9):1830-45.
11. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis Car Res.* 2008;59(5):632-41.
12. Bamford A, Nation A, Durrell S, Andronis L, Rule E, McLeod H. Implementing the Keele stratified care model for patients with low back pain: an observational impact study. *BMC Musculoskelet Disord.* 2017;18(1):66.
13. Hill JC, Whitehurst DGT, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): A randomised controlled trial. *The Lancet.* 2011;378(9802):1560-71.
14. Whitehurst DGT, Bryan S, Lewis M, Hay EM, Mullis R, Foster NE. Implementing stratified primary care management for low back pain: cost-utility analysis alongside a prospective, population-based, sequential comparison study. *Spine.* 2015;40(6):405-14.
15. Foster NE, Mullis R, Hill JC, Lewis M, Whitehurst DG, Doyle C, et al. Effect of stratified care for low back pain in family practice (IMPACT Back): a prospective population-based sequential comparison. *Ann Fam Med.* 2014;12(2):102-11.
16. Beneciuk JM, Bishop MD, Fritz JM, Robinson ME, Asal NR, Nisenzon AN, et al. The STarT back screening tool and individual psychological measures: evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. *Phys Ther.* 2013;93(3):321-33.
17. Beneciuk JM, Fritz JM, George SZ. The STarT Back Screening Tool for prediction of 6-month clinical outcomes: relevance of change patterns in outpatient physical therapy settings. *J Orthop Sports Phys Ther.* 2014;44(9):656-64.
18. Karran EL, McAuley JH, Traeger AC, Hillier SL, Grabherr L, Russek LN, et al. Can screening instruments accurately determine poor outcome risk in adults with recent onset low back pain? A systematic review and meta-analysis. *BMC Med.* 2017;15(1):13.
19. Matsudaira K, Oka H, Kikuchi N, Haga Y, Sawada T, Tanaka S. The Japanese version of the STarT Back Tool predicts 6-month clinical outcomes of low back pain. *J Orthop Sci.* 2017;22(2):224-9.
20. Medeiros FC, Costa LOP, Added MAN, Salomao EC, Costa L. Longitudinal Monitoring of Patients With Chronic Low Back Pain During Physical Therapy Treatment Using the STarT Back Screening Tool. *J Orthop Sports Phys Ther.* 2017;47(5):314-23.

21. Morsø L, Kongsted A, Hestbaek L, Kent P. The prognostic ability of the STarT Back Tool was affected by episode duration. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2016;25(3):936-44.
22. Morsø L, Kent P, Albert H, Hill J, Kongsted A, Manniche C. The predictive and external validity of the STarT Back Tool in Danish primary care. *Eur Spine J.* 2013;22(8):1859-67.
23. Pagé I, Abboud J, J. OS, Laurencelle L, Descarreaux M. Chronic low back pain clinical outcomes present higher associations with the STarT Back Screening Tool than with physiologic measures: a 12-month cohort study. *BMC Musculoskelet Disord.* 2015;16:201
24. Riis A, Rathleff MS, Jensen CE, Jensen MB. Predictive ability of the start back tool: an ancillary analysis of a low back pain trial from Danish general practice. *BMC Musculoskelet Disord.* 2017;18(1):360.
25. Suri P, Delaney K, Rundell SD, Cherkin DC. Predictive Validity of the STarT Back Tool for Risk of Persistent Disabling Back Pain in a U.S Primary Care Setting. *Arch Phys Med Rehabil.* 2018; 99(8):1533-39
26. Toh I, Chong H-C, Suet-Ching Liaw J, Pua Y-H. Evaluation of the STarT Back Screening Tool for Prediction of Low Back Pain Intensity in an Outpatient Physical Therapy Setting. *J Orthop Sports Phys Ther.* 2017;47(4):261-67.
27. Bier JD, Ostelo R, van Hooff ML, Koes BW, Verhagen AP. Validity and Reproducibility of the STarT Back Tool (Dutch Version) in Patients With Low Back Pain in Primary Care Settings. *Phys Ther.* 2017;97(5):561-70.
28. Katzan IL, Thompson NR, George SZ, Passek S, Frost F, Stilphen M. The use of STarT back screening tool to predict functional disability outcomes in patients receiving physical therapy for low back pain. *The Spine Journal.* 2019;19(4):645-54.
29. Forsbrand M, Grahn B, Hill JC, Petersson IF, Sennehed CP, Stigmar K. Comparison of the Swedish STarT Back Screening Tool and the Short Form of the Orebro Musculoskeletal Pain Screening Questionnaire in patients with acute or subacute back and neck pain. *BMC Musculoskelet Disord.* 2017;18(1):89.
30. Haglund E, Bremander A, Bergman S. The StarT back screening tool and a pain mannequin improve triage in individuals with low back pain at risk of a worse prognosis - A population based cohort study. *BMC Musculoskelet Disord.* 2019;20(460):1-9.

31. Storm L, Rousing R, Andersen MO, Carreon LY. Usefulness of the STarT Back Screening Tool to predict pain problems after lumbar spine surgery. *Dan Med J.* 2018;65(12): 1-4.
32. Tan CIC, Liaw JSC, Jiang B, Pothiawala SE, Li H, Leong MKF. Predicting outcomes of acute low back pain patients in emergency department: A prospective observational cohort study. *Medicine (Baltimore).* 2018;97(26):e11247:1-9.
33. St.meld. nr. 47 (2008–2009). Samhandlingsreformen. Rett behandling – på rett sted – til rett tid. Oslo: Det kongelige helse- og omsorgsdepartement; 2009.
34. Direktoratet for e-helse. Den internasjonale statistiske klassifikasjonen av sykdommer og beslektede helseproblemer 2019 [Internett]. Oslo; Direktoratet for e-helse; 2019. Tilgjengelig fra: <https://ehelse.no/kodeverk/kodeverket-icd-10-og-icd-11#Last%20ned%20ICD-10%202020>
35. Ihlebæk C, Lærum E. Plager flest - koster mest : muskel-skjelettlidelser i Norge. Oslo: Nasjonalt ryggnettverk, Forskningsenheten/Formidlingsenheten; 2004. Rapport 2004:1.
36. DeVet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: United Kingdom: Cambridge University Press; 2011.*
37. Beneciuk JM, Hill JC, Campbell P, Afolabi E, George SZ, Dunn KM, et al. Identifying Treatment Effect Modifiers in the STarT Back Trial: A Secondary Analysis. *J Pain.* 2017;18(1):54-65.
38. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and Moderators of Treatment Effects in Randomized Clinical Trials. *Arch Gen Psychiatry.* 2002;59(10):877-83.
39. Borrell-Carrió F, Suchman AL, Epstein RM. The biopsychosocial model 25 years later: principles, practice, and scientific inquiry. *Ann Fam Med.* 2004;2(6):576-82.
40. Waddell G. 1987 Volvo award in clinical sciences. A new clinical model for the treatment of low-back pain. *Spine (Phila Pa 1976).* 1987;12(7):632-44.
41. Waddell G. Biopsychosocial analysis of low back pain. *Baillieres Clin Rheumatol.* 1992;6(3):523-57.
42. Waddell G. *The back pain revolution.* 2. utg. Edinburgh: Churchill Livingstone; 2004.
43. Institute for Health Metrics and Evaluation (IHME) [Internett]. GBD Compare Data Visualization. Seattle. WA: IHME, University of Washington [08. november 2018, hentet 29. April 2020]. Tilgjengelig fra: <http://vizhub.healthdata.org/gbd-compare>.

44. Scheele J, Luijsterburg PA, Ferreira ML, Maher CG, Pereira L, Peul WC, et al. Back complaints in the elders (BACE); design of cohort studies in primary care: an international consortium. *BMC Musculoskelet Disord*. 2011;12:193.
45. Bressler BH, Keyes JW, Rochon AP, Badley AE. The Prevalence of Low Back Pain in the Elderly: A Systematic Review of the Literature. *Spine*. 1999;24(17):1813-19.
46. Hicks GE, Gaines JM, Shardell M, Simonsick EM. Associations of back and leg pain with health status and functional capacity of older adults: findings from the retirement community back pain study. *Arthritis Rheum*. 2008;59(9):1306-13.
47. Kinge JM, Knudsen AK, Skirbekk V, Vollset SE. Musculoskeletal disorders in Norway: prevalence of chronicity and use of primary and specialist health care services. *BMC Musculoskelet Disord*. 2015;16(1):75.
48. Knudsen AK, Tollånes MC, Haaland ØA, Kinge JM, Skirbekk V, Vollset SE. Sykdomsbyrde i Norge 2015: Resultater fra Global Burden of Diseases, Injuries, and Risk Factors Study 2015 (GBD 2015) [Internett]. Bergen/Oslo: Folkehelseinstituttet; 2017 [Hentet 29.04.2020]. Rapport 2017. Tilgjengelig fra: https://www.fhi.no/globalassets/dokumenterfiler/rapporter/2015/sykdomsbyrde_i_norge_2015.pdf
49. Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1459-544.
50. Froud R, Patel S, Rajendran D, Bright P, Bjørkli T, Buchbinder R, et al. A Systematic Review of Outcome Measures Use, Analytical Approaches, Reporting Methods, and Publication Volume by Year in Low Back Pain Trials Published between 1980 and 2012: Respice, adspice, et prospice. *PloS one*. 2016;11(10):e0164573.
51. Breivik H, Eisenberg E, O'brien T. The individual and societal burden of chronic pain in Europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. *BMC Public Health*. 2013;13(1):1229.
52. Mundal A. Flest til fastlegen på grunn av muskel- og skjelettlidelser [Internett]. Norge: Statistisk sentralbyrå; 3. juli 2018 [Hentet 29. april 2020]. Tilgjengelig fra: <https://www.ssb.no/helse/artikler-og-publikasjoner/flest-til-fastlegen-pa-grunn-av-muskel-og-skjelettlidelser>
53. Statistisk sentralbyrå. Almennlegetjenesten, 2018. Antall konsultasjoner hos fastlege fordelt etter diagnosegrupper, absolutte tall og prosent [Internett]. Norge: Statistisk

- sentralbyrå; 2018 [Hentet 29. april 2020]. Tilgjengelig fra:
<https://www.ssb.no/helse/statistikker/fastlegetj/aar/2019-06-13?fane=tabell&sort=nummer&tabell=389590>
54. Nav.no. SYFRA560 Legemeldt sykefravær etter diagnose og bosted. Kvartal. Oslo: NAV; 2020 [Hentet 29. april 2020].
55. Ellingsen J. Diagnoser uføretrygd. Utviklingen i uførediagnoser per 31. desember 2015 [Internett]. Norge: NAV; 27. juni 2019 [Hentet 30. april 2020]. Tilgjengelig fra:
<https://www.nav.no/no/nav-og-samfunn/statistikk/aap-nedsatt-arbeidsevne-og-uforetrygd-statistikk/uforetrygd/diagnoser-uforetrygd>
56. Sundell T. Utviklingen i sykefraværet, 4. kvartal 2019 [Internett]. Oslo: NAV; 11. mars 2020 [Hentet 30. april 2020]. Tilgjengelig fra: <https://www.nav.no/no/nav-og-samfunn/statistikk/sykefravar-statistikk/sykefravar>
57. Brage S, Ihlebæk C, Natvig B, Bruusgaard D. Muskel- og skjelettlidelser som årsak til sykefravær og uføreytelser. Tidsskrift for Den norske legeforening. 2010;130(23): 2369-70.
58. Oliveira C, Maher C, Pinto R, Traeger A, Lin CC, Chenot J-F, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J. 2018 Nov;27(11):2791-2803.
59. Koes BW, van Tulder MW, Peul WC. Diagnosis and treatment of sciatica. BMJ. 2007;334(7607):1313-17.
60. Henschke N, Maher CG, Ostelo R, de Vet HCW, Macaskill P, Irwig L. Red flags to screen for malignancy in patients with low-back pain. Cochrane Database Syst Rev. 2013;(2): CD008686.
61. Jarvik JG, Gold LS, Comstock BA, et al. Association of early imaging for back pain with clinical outcomes in older adults. Jama. 2015;313(11):1143-53.
62. Henschke N, Maher CG, Refshauge KM, Herbert RD, Cumming RG, Bleasel J, et al. Prevalence of and screening for serious spinal pathology in patients presenting to primary care settings with acute low back pain. Arthritis Rheum. 2009;60(10):3072-80.
63. Enthoven WT, Geuze J, Scheele J, Bierma-Zeinstra SM, Bueving HJ, Bohnen AM, et al. Prevalence and "Red Flags" Regarding Specified Causes of Back Pain in Older Adults Presenting in General Practice. Phys Ther. 2016;96(3):305-12.
64. Brinjikji W, Luetmer PH, Comstock B, Bresnahan BW, Chen LE, Deyo RA, et al. Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. AJNR Am J Neuroradiol. 2015;36(4):811-16.

65. Kim JH, Koes BW, Enthoven WTM, Bierma-Zeinstra SMA, Luijsterburg PAJ. No association between low bone quality and back pain in older adults: A cohort study. *J Back Musculoskelet Rehabil.* 2018;31(3):541-7.
66. De Schepper ITE, Damen BJJ, Van Meurs ZJ, Ginai WA, Popham MM, Hofman MA, et al. The Association Between Lumbar Disc Degeneration and Low Back Pain: The Influence of Age, Gender, and Individual Radiographic Features. *Spine.* 2010;35(5):531-6.
67. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet.* 2012;380(9836):37-43.
68. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, Boirie Y, Cederholm T, Landi F, et al. Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. *Age Ageing.* 2010;39(4):412-23.
69. Janssen I, Heymsfield SB, Wang ZM, Ross R. Skeletal muscle mass and distribution in 468 men and women aged 18-88 yr. *J Appl Physiol (1985).* 2000;89(1):81-8.
70. Parreira P, Maher CG, Steffens D, Hancock MJ, Ferreira ML. Risk factors for low back pain and sciatica: an umbrella review. *Spine J.* 2018;18(9):1715-21.
71. Taylor JB, Goode AP, George SZ, Cook CE. Incidence and risk factors for first-time incident low back pain: a systematic review and meta-analysis. *Spine J.* 2014;14(10):2299-319.
72. Cook CE, Taylor J, Wright A, Milosavljevic S, Goode A, Whitford M. Risk factors for first time incidence sciatica: a systematic review. *Physiother Res Int.* 2014;19(2):65-78.
73. Ferreira PH, Pinheiro MB, Machado GC, Ferreira ML. Is alcohol intake associated with low back pain? A systematic review of observational studies. *Man Ther.* 2013;18(3):183-90.
74. Dario AB, Ferreira ML, Refshauge KM, Lima TS, Ordoñana JR, Ferreira PH. The relationship between obesity, low back pain, and lumbar disc degeneration when genetics and the environment are considered: a systematic review of twin studies. *Spine J.* 2015;15(5):1106-17.
75. Coenen P, Gouttebarga V, van Der Burght ASAM, van Dieën JH, Frings-Dresen MHW, van Der Beek AJ, et al. The effect of lifting during work on low back pain: a health impact assessment based on a meta-analysis. *Occup Environ Med.* 2014;71(12):871-7.
76. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet.* 2018;391(10137):2368-83.

77. Paeck T, Ferreira ML, Sun C, Lin CW, Tiedemann A, Maher CG. Are older adults missing from low back pain clinical trials? A systematic review and meta-analysis. *Arthritis Care Res (Hoboken)*. 2014;66(8):1220-6.
78. Vadala G, Russo F, De Salvatore S, Cortina G, Albo E, Papalia R, et al. Physical Activity for the Treatment of Chronic Low Back Pain in Elderly Patients: A Systematic Review. *J Clin Med*. 2020;9(4):1023.
79. Newman TB, Kohn MA. Evidence-based diagnosis. New York: Cambridge University Press; 2009.
80. Bier JD, Sandee-Geurts JJW, Ostelo RWJG, Koes BW, Verhagen AP. Can Primary Care for Back and/or Neck Pain in the Netherlands Benefit From Stratification for Risk Groups According to the STarT Back Tool Classification? *Arch Phys Med Rehabil*. 2018;99(1):65-71.
81. Cherkin D, Balderson B, Wellman R, Hsu C, Sherman KJ, Evers SC, et al. Effect of Low Back Pain Risk-Stratification Strategy on Patient Outcomes and Care Processes: the MATCH Randomized Trial in Primary Care. *J Gen Intern Med*. 2018. 33(8):1324-36.
82. Delitto A, Patterson CG, Stevans JM, Brennan GP, Wegener ST, Morrisette DC, et al. Study protocol for targeted interventions to prevent chronic low back pain in high-risk patients: A multi-site pragmatic cluster randomized controlled trial (TARGET Trial). *Contemp Clin Trials*. 2019;82:66-76.
83. da Silva T, Mills K, Brown BT, Herbert RD, Maher CG, Hancock MJ. Risk of Recurrence of Low Back Pain: A Systematic Review. *J Orthop Sports Phys Ther*. 2017;47(5):305-13.
84. Pengel LHM, Herbert RD, Maher CG, Refshauge KM. Acute low back pain: systematic review of its prognosis. *BMJ*. 2003;327(7410):323.
85. Itz CJ, Geurts JW, Kleef M, Nelemans P. Clinical course of non-specific low back pain: A systematic review of prospective cohort studies set in primary care. 2013. *Eur J Pain*. 2013;17(1):5-15.
86. da C Menezes Costa L, Maher CG, Hancock MJ, McAuley JH, Herbert RD, Costa LOP. The prognosis of acute and persistent low-back pain: a meta-analysis. *CMAJ*. 2012;184(11):E613-E24.
87. Scheele TMJ, Enthoven MAW, Bierma-Zeinstra CS, Peul WW, Van Tulder MM, Bohnen YA, et al. Course and prognosis of older back pain patients in general practice: A prospective cohort study. *Pain*. 2013;154(6):951-7.

88. Rundell SD, Sherman KJ, Heagerty PJ, Mock CN, Jarvik JG. The Clinical Course of Pain and Function in Older Adults with a New Primary Care Visit for Back Pain. *J Am Geriatr Soc.* 2015;63(3):524-30.
89. van Der Gaag WH, Enthoven WTM, Luijsterburg PAJ, van Rijckevorsel-Scheele J, Bierma-Zeinstra SMA, Bohnen AM, et al. Natural History of Back Pain in Older Adults over Five Years. *JABFM.* 2019;32(6):781-89.
90. Deyo RA, Bryan M, Comstock BA, Turner JA, Heagerty P, Friedly J, et al. Trajectories of Symptoms and Function in Older Adults With Low Back Disorders. *Spine (Phila Pa 1976).* 2015;40(17):1352-62.
91. Artus M, Campbell P, Mallen CD, Dunn KM, van der Windt DA. Generic prognostic factors for musculoskeletal pain in primary care: a systematic review. *BMJ open.* 2017;7(1):e012901.
92. Rundell SD, Sherman KJ, Heagerty PJ, Mock CN, Dettori NJ, Comstock BA, et al. Predictors of Persistent Disability and Back Pain in Older Adults with a New Episode of Care for Back Pain. *Pain Med.* 2017;18(6):1049-62.
93. Jarvik JG, Gold LS, Tan K, Friedly JL, Nedeljkovic SS, Comstock BA, et al. Long-term outcomes of a large, prospective observational cohort of older adults with back pain. *Spine J.* 2018;18(9):1540-51.
94. Enthoven WT, Koes BW, Bierma-Zeinstra SMA, Bueving HJ, Bohnen AM, Peul WC, et al. Defining trajectories in older adults with back pain presenting in general practice. *Age Ageing.* 2016;45(6):878-83.
95. Wertli MM, Eugster R, Held U, Steurer J, Kofmehl R, Weiser S. Catastrophizing—a prognostic factor for outcome in patients with low back pain: a systematic review. *Spine J.* 2014;14(11):2639-57.
96. Wertli MM, Burgstaller JM, Weiser S, Steurer J, Kofmehl R, Held U. Influence of catastrophizing on treatment outcome in patients with nonspecific low back pain: a systematic review. *Spine (Phila Pa 1976).* 2014;39(3):263-73.
97. Ramond A, Bouton C, Richard I, Roquelaure Y, Baufreton C, Legrand E, et al. Psychosocial risk factors for chronic low back pain in primary care—a systematic review. *Family Pract.* 2011;28(1):12-21.
98. Hayden JA, Wilson MN, Riley RD, Iles R, Pincus T, Ogilvie R. Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor review. *Cochrane Database Syst Rev.* 2019;2019(11):1-126.

99. Pinheiro MB, Ferreira ML, Refshauge K, Maher CG, Ordonana JR, Andrade TB, et al. Symptoms of depression as a prognostic factor for low back pain: a systematic review. *Spine J.* 2016;16(1):105-16.
100. Ashworth J, Konstantinou K, Dunn K. Prognostic factors in non-surgically treated sciatica: A systematic review. *BMC Musculoskelet Disord.* 2011;12(1):208.
101. Wertli MM, Rasmussen-Barr E, Weiser S, Bachmann LM, Brunner F. The role of fear avoidance beliefs as a prognostic factor for outcome in patients with nonspecific low back pain: a systematic review. *Spine J.* 2014;14(5):816-36.
102. Riley RD, Hayden J, Steyerberg E, Moons K, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: *PLoS Med.* 2013;10(2): e1001380.
103. Hill JC, Vohora K, Dunn KM, Main CJ, Hay EM. Comparing the STarT back screening tool's subgroup allocation of individual patients with that of independent clinical experts. *Clin J Pain.* 2010;26(9):783-7.
104. Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Ann Intern Med.* 2004;141(12):920-8.
105. Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary Development of a Clinical Prediction Rule for Determining Which Patients With Low Back Pain Will Respond to a Stabilization Exercise Program. *Arch Phys Med Rehabil.* 2005;86(9):1753-62.
106. Truchon M, Côté D. Predictive validity of the Chronic Pain Coping Inventory in subacute low back pain. *Pain.* 2005;116(3):205-12.
107. Duijts SFA, Kant IJ, Landeweerd JA, Swaen GMH. Prediction of sickness absence: development of a screening instrument. *Occup Environ Med.* 2006;63(8):564-9.
108. Neubauer E, Junge A, Pirron P, Seemann H, Schiltenwolf M. HKF-R 10 – Screening for predicting chronicity in acute low back pain (LBP): A prospective clinical trial. *Eur J Pain.* 2006;10(6):559-66.
109. Denison E, Åsenlöf P, Lindberg P. Self-efficacy, fear avoidance, and pain intensity as predictors of disability in subacute and chronic musculoskeletal pain patients in primary health care. *Pain.* 2004;111(3):245-52.
110. Linton SJ, Halldén K. Can We Screen for Problematic Back Pain? A Screening Questionnaire for Predicting Outcome in Acute and Subacute Back Pain. *Clin J Pain.* 1998;14(3):209-15.
111. Linton SJ, MacDonald S, Nicholas M. Development of a short form of the Örebro musculoskeletal pain screening questionnaire. *Spine.* 2011;36(22):1891-5.

112. Grotle KM, Vøllestad IN, Brox IJ. Screening for Yellow Flags in First-time Acute Low Back Pain: Reliability and Validity of a Norwegian Version of the Acute Low Back Pain Screening Questionnaire. *Clin J Pain*. 2006;22(5):458-67.
113. Simula AS, Ruokolainen O, Oura P, Lausmaa M, Holopainen R, Paukkunen M, et al. Association of STarT Back Tool and the short form of the Örebro Musculoskeletal Pain Screening Questionnaire with multidimensional risk factors. *Sci Rep*. 2020;10(1):290.
114. Lheureux A, Berquin A. Comparison between the STarT Back Screening Tool and the Örebro Musculoskeletal Pain Screening Questionnaire: Which tool for what purpose? A semi-systematic review. *Ann Phys Rehabil Med*. 2019;62(3):178-88.
115. Gulbrandsen BP, Madsen SH, Benth SJ, Lærum SE. Health care providers communicate less well with patients with chronic low back pain – A study of encounters at a back pain clinic in Denmark. *Pain*. 2010;150(3):458-61.
116. Khan Y. The STarT back tool in chiropractic practice: a narrative review. *Chiropr Man Therap*. 2017;25:11.
117. Field J, Newell D. Relationship between STarT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. *Chiropr Man Therap*. 2012;20(1):17.
118. Newell D, Field J, Pollard D. Using the STarT Back Tool: Does timing of stratification matter? *Man Ther*. 2015;20(4):533-9.
119. Kongsted A, Andersen CH, Hansen MM, Hestbaek L. Prediction of outcome in patients with low back pain – A prospective cohort study comparing clinicians' predictions with those of the Start Back Tool. *Man Ther*. 2016;21:120-7.
120. Friedman BW, Conway J, Campbell C, Bijur PE, Gallagher EJ. Pain one week after an ED visit for acute low back pain is associated with poor three month outcomes. *Acad Emerg Med*. 2018; 25(10):1138-45.
121. Morsø L, Kent P, Manniche C, Albert HB. The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *Eur Spine J*. 2014;23(1):120-8.
122. Medeiros FC, Costa LOP, Oliveira IS, Oshima RK, Costa LCM. The use of STarT BACK Screening Tool in emergency departments for patients with acute low back pain: a prospective inception cohort study. *Eur Spine J*. 2018; 27(11):2823-30.
123. Karran EL, Traeger AC, McAuley JH, Hillier SL, Yau YH, Moseley GL. The Value of Prognostic Screening for Patients With Low Back Pain in Secondary Care. *J. Pain*. 2017;18(6):673-86.

124. Keele University. STarT Back Translations [Internett]. United Kingdom: Keele University; [Hentet 30. april 2020]. Tilgjengelig fra: <https://startback.hfac.keele.ac.uk/training/resources/start-back-translations/>.
125. Keele University. Welcome to the Keele STarT MSK Tool© 2017 [Internett]. United Kingdom: Keele University; [Hentet 30. April 2020]. Tilgjengelig fra: <https://www.keele.ac.uk/startmsk/>.
126. Hill JC, Afolabi EK, Lewis M, Dunn KM, Roddy E, van der Windt DA, et al. Does a modified STarT Back Tool predict outcome with a broader group of musculoskeletal patients than back pain? A secondary analysis of cohort data. *BMJ open*. 2016;6(10):e012445.
127. Beneciuk JM, Robinson ME, George SZ. Subgrouping for patients with low back pain: a multidimensional approach incorporating cluster analysis and the STarT Back Screening Tool. *J Pain*. 2015;16(1):19-30.
128. Main CJ, Sowden G, Hill JC, Watson PJ, Hay EM. Integrating physical and psychological approaches to treatment in low back pain: the development and content of the STarT Back trial's 'high-risk' intervention (StarT Back; ISRCTN 37113406). *Physiotherapy*. 2012;98(2):110-6.
129. Storheim K, Grotle IM. The Keele STarT Back Screening Tool- norsk versjon. Oslo: FORMI, Universitetssykehuset Ullevål; 2012. Tilgjengelig fra: Se vedlegg 4.
130. Gusi N, del Pozo-Cruz B, Olivares PR, Hernandez-Mocholi M, Hill JC. The Spanish version of the "STarT Back Screening Tool" (SBST) in different subgroups. *Aten Primaria*. 2011;43(7):356-61.
131. Kongsted A, Johannesen E, Leboeuf-Yde C. Feasibility of the STarT back screening tool in chiropractic clinics: a cross-sectional study of patients with low back pain. *Chiropr Man Therap*. 2011;19:10.
132. Azimi P, Shahzadi S, Azhari S, Montazeri A. A validation study of the Iranian version of STarT Back Screening Tool (SBST) in lumbar central canal stenosis patients. *J Orthop Sci*. 2014;19(2):213-7.
133. Pilz B, Vasconcelos RA, Marcondes FB, Lodovichi SS, Mello W, Grossi DB. The Brazilian version of STarT Back Screening Tool - translation, cross-cultural adaptation and reliability. *Braz J Phys Ther*. 2014;18(5):453-61.
134. Piironen S, Paananen M, Haapea M, Hupli M, Zitting P, Ryyanen K, et al. Transcultural adaption and psychometric properties of the STarT Back Screening Tool among Finnish low back pain patients. *Eur Spine J*. 2016;25(1):287-95.

135. Karstens S, Krug K, Hill JC, Stock C, Steinhäuser J, Szecsenyi J, et al. Validation of the German version of the STarT-Back Tool (STarT-G): a cohort study with patients from primary care practices. *BMC Musculoskelet Disord*. 2015;16:346.
136. Betten C, Sandell C, Hill JC, Gutke A. Cross-cultural adaptation and validation of the Swedish STarT Back Screening Tool. *Eur J Physiother*. 2015;17(1):29-36.
137. Bruyere O, Demoulin M, Beaudart C, Hill JC, Maquet D, Genevay S, et al. Validity and reliability of the French version of the STarT Back screening tool for patients with low back pain. *Spine (Phila Pa 1976)*. 2014;39(2):E123-8.
138. Matsudaira K, Oka H, Kikuchi N, Haga Y, Sawada T, Tanaka S. Psychometric Properties of the Japanese Version of the STarT Back Tool in Patients with Low Back Pain. *PloS one*. 2016;11(3):e0152019.
139. Robinson HS, Dagfinrud H. Reliability and screening ability of the StarT Back screening tool in patients with low back pain in physiotherapy practice, a cohort study. *BMC Musculoskelet Disord*. 2017;18(1):232.
140. Aebischer B, Hill JC, Hilfiker R, Karstens S. German Translation and Cross-Cultural Adaptation of the STarT Back Screening Tool. *PloS one*. 2015;10(7):e0132068.
141. Luan S, Min Y, Li G, Lin C, Li X, Wu S, et al. Cross-cultural adaptation, reliability, and validity of the Chinese version of the STarT Back Screening Tool in patients with low back pain. *Spine (Phila Pa 1976)*. 2014;39(16):E974-9.
142. Unsgaard-Tondel M, Kregnes IG, Nilsen TIL, Marchand GH, Askim T. Risk classification of patients referred to secondary care for low back pain. *BMC Musculoskelet Disord*. 2018;19(1):166.
143. Morsø L, Albert H, Kent P, Manniche C, Hill J. Translation and discriminative validation of the STarT Back Screening Tool into Danish. *Eur Spine J*. 2011;20(12):2166-73.
144. Al Zoubi FM, Eilayyan O, Mayo NE, Bussières AE. Evaluation of Cross-Cultural Adaptation and Measurement Properties of STarT Back Screening Tool: A Systematic Review. *J Manipulative Physiol Ther*. 2017;40(8):558-72.
145. Hill JC, Dunn KM, Main CJ, Hay EM. Subgrouping low back pain: a comparison of the STarT Back Tool with the Orebro Musculoskeletal Pain Screening Questionnaire. *Eur J Pain*. 2010;14(1):83-9.
146. Fuhro FF, Fagundes FR, Manzoni AC, Costa LO, Cabral CM. Orebro Musculoskeletal Pain Screening Questionnaire Short-Form and STarT Back Screening Tool: Correlation and Agreement Analysis. *Spine (Phila Pa 1976)*. 2016;41(15):E931-6.

147. Abedi M, Manshadi FD, Khalkhali M, Mousavi SJ, Baghban AA, Montazeri A, et al. Translation and validation of the Persian version of the STarT Back Screening Tool in patients with nonspecific low back pain. *Man Ther.* 2015;20(6):850-4.
148. Raimundo A, Parraca J, Batalha N, Tomas-Carus P, Branco J, Hill J, et al. Portuguese translation, cross-cultural adaptation and reliability of the questionnaire <<Start Back Screening Tool>> (SBST). *Acta Reumatol Port.* 2017;42(1):38-46.
149. Pilz B, Vasconcelos RA, Teixeira PP, Mello W, Marcondes FB, Hill JC, et al. Construct and discriminant validity of STarT Back Screening Tool - Brazilian version. *Braz J Phys Ther.* 2017;21(1):69-73.
150. Roland M, Morris R. A study of the natural history of back pain. Part 1 Development of a Reliable and Sensitive Measure of Disability in Low-Back Pain. *Spine (Phila Pa 1976).* 1983;8(2):141-4.
151. Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *J Clin Epidemiol.* 2018;95:73-93.
152. Grotle IM, Brox KJ, Vøllestad KN. Functional Status and Disability Questionnaires: What Do They Assess?: A Systematic Review of Back-Specific Outcome Questionnaires. *Spine.* 2005;30(1):130-40.
153. Yao M, Zhu S, Tian ZR, Song YJ, Yang L, Wang YJ, et al. Cross-cultural adaptation of Roland-Morris Disability Questionnaire needs to assess the measurement properties: a systematic review. *J Clin Epidemiol.* 2018;99:113-22.
154. Davies CC, Nitz AJ. Psychometric properties of the Roland-Morris Disability Questionnaire compared to the Oswestry Disability Index: a systematic review. *Phys Ther Rev.* 2009;14(6):399-408.
155. Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine (Phila Pa 1976).* 2000;25(24):3115-24.
156. Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell P, Ostelo RW. Roland-Morris Disability Questionnaire and Oswestry Disability Index: Which Has Better Measurement Properties for Measuring Physical Functioning in Nonspecific Low Back Pain? Systematic Review and Meta-Analysis. *Phys Ther.* 2016;96(10):1620-37.
157. Ostelo WJGR, Deyo AR, Stratford MP, Waddell CG, Croft CP, Von Korff CM, et al. Interpreting Change Scores for Pain and Functional Status in Low Back Pain: Towards International Consensus Regarding Minimal Important Change. *Spine.* 2008;33(1):90-4.

158. Grotle M, Brox J, Vollestad N. Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. *J Rehabil Med.* 2003;35(5):241-7.
159. WHO. Body mass index – BMI [Internett]. Regional Office for Europe: World Health Organization; [Hentet 30. April 2020]. Tilgjengelig fra: <http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi>.
160. Scheele J, Enthoven WT, Bierma-Zeinstra SM, Peul WC, van Tulder MW, Bohnen AM, et al. Characteristics of older patients with back pain in general practice: BACE cohort study. *Eur J Pain.* 2014;18(2):279-87.
161. Von Korff PM, Jensen PM, Karoly PP. Assessing Global Pain Severity by Self-Report in Clinical and Health Services Research. *Spine.* 2000;25(24):3140-51.
162. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain.* 1993;52(2):157-68.
163. Symonds TL, Burton AK, Tillotson KM, Main CJ. Do Attitudes and Beliefs Influence Work Loss Due to Low Back Trouble?. *Occup Med (Lond).* 1996;46(1):25-32.
164. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Appl Psychol Meas.* 1977;1(3):385-401.
165. Sullivan MJL, Bishop SR. The pain catastrophizing scale: Development and validation. *Psychol Assess.* 1995;7(4):524-32.
166. Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C. Development and initial validation of the Falls Efficacy Scale-International (FES-I). *Age Ageing.* 2005;34(6):614-9.
167. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res.* 1989;28(2):193-213.
168. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum.* 2003;49(2):156-63.
169. Hulley SB. 12: Designing Studies of Medical Tests. I: Newman TB, Browner WS, Cummings SR, Hulley SB (red). *Designing Clinical research.* 4. utg. Philadelphia; Wolters Cluver, Lippincott Williams & Wilkins; 2013. S. 176-7.
170. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain.* 2008;8(6):221-23.

171. Pallant J. SPSS survival manual : a step by step guide to data analysis using SPSS. 4. utg. Maidenhead: McGraw-Hill Open University Press; 2010.
172. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ*. 2004;329(7458):168.
173. Akobeng AK. Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatr*. 2007;96(4):487-91.
174. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *Jama*. 1994;271(9):703-7.
175. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. *BMJ*. 1994;309(6948):188.
176. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med*. 2003;29(7):1043-51.
177. Datatilsynet. Lover og regler [Internett]. Datatilsynet. [Hentet 30.04.2020]. Tilgjengelig fra: <https://www.datatilsynet.no/regelverk-og-verktoy/lover-og-regler/>
178. World Medical Association. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects [Internett]. World Medical Association. [Oppdatert: 9. Juli 2018; hentet: 30.04.2020]. Tilgjengelig fra: <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>.
179. Herbert R, Jamtvedt G, Hagen KB, Mead J. 3 - What constitutes evidence? I: Herbert R, Jamtvedt G, Hagen KB, Mead J, red. *Practical Evidence-Based Physiotherapy* (2. utg). Edinburgh: Churchill Livingstone; 2011. S.17.
180. Verdens helseorganisasjon (WHO). Health statistics and information systems. Proposed working definition of an older person in Africa for the MDS Project [Internett]. WHO: 2002; [Hentet 30 april 2020]. Tilgjengelig fra: <https://www.who.int/healthinfo/survey/ageingdefnolder/en/>
181. Enthoven WTM, Scheele J, Bierma-Zeinstra SMA, Bueving HJ, Bohnen AM, Peul WC, et al. Analgesic Use in Older Adults with Back Pain: The BACE Study. *Pain Med*. 2014;15(10):1704-14.
182. Ferreira ML, Machado G, Latimer J, Maher C, Ferreira PH, Smeets RJ. Factors defining care-seeking in low back pain – A meta-analysis of population based surveys. *Eur J Pain*. 2010;14(7):747.e1-.e7.

183. Woodhouse A, Pape K, Romundstad PR, Vasseljen O. Health care contact following a new incident neck or low back pain episode in the general population; the HUNT study.(Clinical report). *BMC Health Serv Res.* 2016;16(81).
184. Forsbrand MH, Grahn B, Hill JC, Petersson IF, Post Sennehed C, Stigmar K. Can the STarT Back Tool predict health-related quality of life and work ability after an acute/subacute episode with back or neck pain? A psychometric validation study in primary care. *BMJ open.* 2018;8(12): e021748.
185. Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LOP, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain.* 2018;159(3):481-95.
186. Bombardier C. Outcome Assessments in the Evaluation of Treatment of Spinal Disorders: Summary and General Recommendations. *Spine.* 2000;25(24):3100-3.
187. Kendell M, Beales D, O'Sullivan P, Rabey M, Hill J, Smith A. The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *J Physiother.* 2018;64(2):107-13.
188. Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Med Res Methodol.* 2001;1:6.
189. Austevoll IM, Gjestad R, Brox JI, Solberg TK, Storheim K, Rekeland F, et al. The effectiveness of decompression alone compared with additional fusion for lumbar spinal stenosis with degenerative spondylolisthesis: a pragmatic comparative non-inferiority observational study from the Norwegian registry for Spine surgery. *Eur Spine J.* 2017;26(2):404-13.
190. Aggarwal R, Ranganathan P. Understanding diagnostic tests - Part 3: Receiver operating characteristic curves. *Perspect Clin Res.* 2018;9(3):145-8.
191. Kamper SJ, Ostelo RWJG, Knol DL, Maher CG, de Vet HCW, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin epidemiol.* 2010;63(7):760-66.
192. Kongsted A, Kent P, Axen I, Downie AS, Dunn KM. What have we learned from ten years of trajectory research in low back pain? *BMC Musculoskelet Disord.* 2016;17:220.
193. Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. *BMJ.* 2003;326(7398):1083-4.

194. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC Med Res Methodol.* 2017;17(1):162.
195. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. Amsterdam, Nederland: Department of Epidemiology and Biostatistics Amsterdam Public Health research institute; 2019. Rapport 2019.
196. Pedersen TR, Vollset SE. Legemiddelutprøving. I: Laake P, Hjartåker A, Thelle DS, Veierød MB. *Epidemiologiske og kliniske forskningsmetoder.* 1. utg. Oslo: Gyldendal Norsk Forlag AS; 2007. s. 275.
197. Svensson E, Hjartåker A, Laake P. Hva skal måles og hvordan? I: Laake P, Hjartåker A, Thelle DS, Veierød MB, red. *Epidemiologiske og kliniske forskningsmetoder.* Oslo: Gyldendahl Norsk Forlag AS; 2007. S. 54-60.
198. Karstens S, Krug K, Raspe H, Wunderlich M, Hochheim M, Joos S, et al. Prognostic ability of the German version of the STarT Back tool: analysis of 12-month follow-up data from a randomized controlled trial. *BMC Musculoskelet Disord.* 2019;20(1):94.
199. Nicholas MK, Linton SJ, Watson PJ, Main CJ. Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. *Phys Ther.* 2011;91(5):737-53.
200. Morsø L, Schiottz-Christensen B, Sondergaard J, Andersen NV, Pedersen F, Olsen KR, et al. The effectiveness of a stratified care model for non-specific low back pain in Danish primary care compared to current practice: study protocol of a randomised controlled trial. *Trials.* 2018;19(1):315.
201. Foster N, Konstantinou K, Beardmore R, Dunn K, Lewis M, Bartlam B, et al. Stratified care for patients with sciatica and suspected sciatica in primary care: the scopic trial protocol (scopic - sciatica outcomes in primary care). *Trials.* 2015;16(2):P156.
202. Rhon D. Validation of the STarT Back Screening Tool in the Military [Internett]. *Clinicaltrials.gov*; 2020 [Hentet 30. april 2020]. Tilgjengelig fra: <https://clinicaltrials.gov/ct2/show/NCT03127826?term=start+back+screening+tool&cond=Back+Pain&draw=2&rank=3>.
203. Henry S. Risk Stratification in Primary Care Combined With Stratified-specific Physical Therapy Care for Low Back Pain [Internett]. *Clinicaltrials.gov*; 2020 [Hentet 30 april 2020]. Tilgjengelig fra: <https://clinicaltrials.gov/ct2/show/NCT03065894?term=start+back+screening+tool&cond=Back+Pain&draw=2&rank=3>.

ck+Pain&draw=2&rank=7.

204. Cruz EB. Stratified Primary Care for Low Back Pain (SPLIT) [Internett].

Clinicaltrials.gov; 2020 [Hentet 30. April 2020] Tilgjengelig fra:

<https://clinicaltrials.gov/ct2/show/NCT04046874?term=start+back+screening+tool&cond=Back+Pain&draw=2&rank=6>.

205. Costa L. Implementation of the STarT Back Screening Tool [Internett].

Clinicaltrials.gov; 2020 [Hentet 30. April 2020]. Tilgjengelig fra:

<https://clinicaltrials.gov/ct2/show/NCT03671278?term=start+back+screening+tool&cond=Back+Pain&draw=2&rank=1>.

VEDLEGG

Vedlegg 1: **Tabell 2:** 2*2 krysstabeller for analyse av sensitivitet, spesifisitet og likelihood ratioer. Hentet fra Hulley et al. (169).

GULLSTANDARD FOR DÅRLIG UTFALL			
	≥ 7 RDQ	< 7 RDQ	Total
Positiv test	Ekte positiv a	Falsk positiv b	a+b
Negativ test	Falsk negativ c	Ekte negativ d	c+d
Total	a+c	b+d	a+b+c+d
	Sensitivitet= $a/(a+c)$	Spesifisitet= $d/(b+d)$	

Vedlegg 2: ROC-analyse med tabell som danner grunnlaget for identifisering av optimal cut-off-verdi

Coordinates of the Curve

Test Result Variable(s): The Keele STarT Back

Screening Tool

Positive if Greater Than or Equal To ^a	Sensitivity	1 - Specificity
-1,00000	1,000	1,000
,50000	,978	,885
1,50000	,839	,641
2,50000	,645	,397
3,50000	,430	,221
4,50000	,280	,084
5,50000	,215	,046
6,50000	,097	,008
7,50000	,022	,000
9,00000	,000	,000

The test result variable(s): The Keele STarT Back Screening Tool has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Vedlegg 3: Informert samtykkeskjema

Forespørsel om deltakelse i forskningsprosjektet

«BACE: en prospektiv kohort studie av eldre med rygg smerter som søker hjelp i primærhelsetjenesten»

Bakgrunn og hensikt

Dette er et spørsmål til deg om å delta i en forskningsstudie for å undersøke forløpet av rygg smerter hos eldre, og for å forstå hvorfor episoder av rygg smerter i noen tilfeller kan bli kroniske. Personer som er over 55 år, som tar kontakt til primærhelsetjenesten på grunn av en ny episode med rygg smerter, blir forespurt om deltagelse. Fakultet for helsefag ved Høgskolen i Oslo og Akershus er hovedansvarlig for prosjektet.

Hva innebærer studien?

Studien innebærer at du ved inklusjon svarer på et spørreskjema om helse og funksjon, og at du blir undersøkt av fysioterapeut eller kiropraktor. Du vil bli bedt om å svare på spørreskjemaet på et nettbrett, evt. på papir om du foretrekker det. Undersøkelsen vil foregå på en av BACE prosjektets teststasjoner, og vil innebære en undersøkelse av ryggmuskulaturen og test av rygg- og hoftbevegelighet, sensibilitet, reflekser, beinstyrke og balanse. Videre vil det bli sendt ut oppfølgingsspørreskjema etter 3, 6, 12 og 24 mnd. Oppfølgingsskjema vil sendes ut elektronisk til din epost-adresse, evt. per post om du foretrekker det.

Mulige fordeler og ulemper

Erfaring fra studien vil kunne bidra til bedre forståelse av rygg smerter hos eldre, noe som vil være av betydning for å bedre behandlingsforløpet ved disse plagene. Testene som utføres er spesifikke for forskningsformål og forventes ikke å føre til ubehag for deltagerne i studien.

Hva skjer med informasjonen om deg?

Informasjonen som registreres om deg skal kun brukes slik som beskrevet i hensikten med studien. All informasjon blir lagret på en sikker forskningsserver ved IT-avdelingen, HiOA. Direkte personidentifiserende opplysninger erstattes med et referansenummer som viser til en atskilt navneliste (koblingsnøkkel). Navnelisten/koblingsnøkkelen vil oppbevares av prosjektleder. Kun prosjektleder og forskningskoordinator vil ha tilgang til navnelisten/koblingsnøkkelen.

Resultater fra prosjektet vil formidles i form av artikler publisert i nasjonale og internasjonale tidsskrift, samt deler av materialet vil benyttes i masteroppgaver ved HiOA. Resultatene vil også presenteres ved norske og internasjonale konferanser. Forskning knyttet til dette prosjektet vil skje i samsvar med Helseforskningsloven. Prosjektperioden starter etter at alle godkjenninger foreligger og dataene for prosjektet vil bestå til februar 2025. Etter 1.mars 2025 vil alle direkte og indirekte personopplysninger i prosjektet bli anonymisert eller slettet.

Frivillig deltakelse

Det er frivillig å delta i studien. Du kan når som helst og uten å oppgi noen grunn trekke ditt samtykke til å delta i studien. Dette vil ikke få konsekvenser for din videre behandling. Dersom du ønsker å delta, undertegner du samtykkeerklæringen på siste side. Om du nå sier ja til å delta, kan du senere trekke tilbake ditt samtykke uten at det påvirker din øvrige behandling. Dersom du senere ønsker å trekke deg eller har spørsmål til studien, kan du kontakte prosjektleder Margreth Grotle, tlf. 90111172.

Samtykke til deltakelse i forskningsprosjektet «BACE: en prospektiv kohort studie av eldre med rygg smerter som søker hjelp i primærhelsetjenesten»

Navn		ID nummer:
Adresse		
Telefonnummer		
Epostadresse		

- Jeg ønsker å få oppfølgingsspørreskjemaene tilsendt per post
- Jeg gir tillatelse til at prosjektansvarlig eller andre medarbeidere tilknyttet BACE prosjektet kan innhente informasjon om ryggsmertene mine fra min fastlege ved en senere anledning (sett kryss)

Fastlege (navn og kontaktinformasjon):

- Jeg gir tillatelse til at prosjektansvarlig eller andre medarbeidere tilknyttet BACE prosjektet kan ta kontakt til pårørende, hvis de ikke får kontakt med meg ved oppfølging etter 3, 6, 12 og 24 måneder (sett kryss)

1. Pårørende (navn og kontaktinformasjon):

2. Pårørende (navn og kontaktinformasjon):

1) Jeg er villig til å delta i studien

(Signert av prosjektdeltaker, dato)

2) Jeg bekrefter å ha gitt informasjon om studien

(Signert, rolle i studien, dato)

Vedlegg 4: The Keele STarT Back Screening Tool – norsk versjon

Tenk på de **siste 2 ukene** og kryss av ditt svaralternativ til følgende spørsmål:

		UENIG	ENIG
		0	1
1	Ryggsmertene mine har spredt seg ned i benet/beina en eller flere ganger de siste 2 ukene	<input type="checkbox"/>	<input type="checkbox"/>
2	Jeg har hatt smerter i skulder eller nakke en eller flere ganger de siste 2 ukene	<input type="checkbox"/>	<input type="checkbox"/>
3	Jeg har kun gått korte distanser på grunn av ryggsmertene mine	<input type="checkbox"/>	<input type="checkbox"/>
4	De siste 2 ukene har jeg kledd meg langsommere enn vanlig på grunn av ryggsmertene	<input type="checkbox"/>	<input type="checkbox"/>
5	Det er egentlig ikke trygt for en person med min tilstand å være fysisk aktiv	<input type="checkbox"/>	<input type="checkbox"/>
6	Jeg har hatt bekymrede tanker store deler av tiden	<input type="checkbox"/>	<input type="checkbox"/>
7	Jeg opplever at ryggsmertene mine er forferdelige og aldri vil bli bedre	<input type="checkbox"/>	<input type="checkbox"/>
8	Generelt sett har jeg ikke gledet meg over ting som jeg pleier å glede meg over	<input type="checkbox"/>	<input type="checkbox"/>

9. Totalt sett, hvor **plagsomme** har ryggsmertene dine vært **de siste 2 ukene**?

Ikke i det hele tatt	Litt	Moderat	Veldig mye	Ekstremt
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
0	0	0	1	1

Total score (alle 9): _____

Del score (spørsmål 5-9): _____

Vedlegg 5: Roland-Morris Disability Questionnaire- norsk versjon

ROLAND-MORRIS SPØRRESKJEMA FOR FUNKSJONSBEGRENSNINGER

Pas.nr: _____

Dato for utfylling: _____

Når du har vondt i ryggen kan det være vanskelig å gjøre noen av de tingene du vanligvis gjør. Her er noen setninger folk har brukt for å beskrive seg selv når de har ryggsmarter. Når du leser dem kan det være at noen av dem skiller seg ut fordi de beskriver deg i dag. Når du leser en setning som beskriver deg i dag, sett ett kryss for ja i boksen til høyre, hvis ikke setningen passer så kryss av i nei-boksen.

	Ja	Nei
1 Jeg holder meg for det meste hjemme på grunn av ryggen.....	<input type="checkbox"/>	<input type="checkbox"/>
2 Jeg skifter stilling ofte for å forsøke å gjøre det behagelig for ryggen.....	<input type="checkbox"/>	<input type="checkbox"/>
3 På grunn av ryggen går jeg saktere enn vanlig.....	<input type="checkbox"/>	<input type="checkbox"/>
4 På grunn av ryggen gjør jeg ingen av de tingene jeg vanligvis gjør i huset.....	<input type="checkbox"/>	<input type="checkbox"/>
5 På grunn av ryggen bruker jeg gelenderet for å gå opp trapper.....	<input type="checkbox"/>	<input type="checkbox"/>
6 På grunn av ryggen legger jeg meg oftere ned for å hvile.....	<input type="checkbox"/>	<input type="checkbox"/>
7 På grunn av ryggen må jeg støtte meg på noe for å komme meg opp av en lenestol..	<input type="checkbox"/>	<input type="checkbox"/>
8 På grunn av ryggen forsøker jeg å få andre til å gjøre ting for meg.....	<input type="checkbox"/>	<input type="checkbox"/>
9 På grunn av ryggen kler jeg på meg saktere enn vanlig.....	<input type="checkbox"/>	<input type="checkbox"/>
10 På grunn av ryggen står jeg oppreist bare i korte stunder av gangen.....	<input type="checkbox"/>	<input type="checkbox"/>
11 På grunn av ryggen forsøker jeg å la være å bøye meg eller sette meg på kne.....	<input type="checkbox"/>	<input type="checkbox"/>
12 På grunn av ryggen synes jeg det er vanskelig å reise meg fra en stol.....	<input type="checkbox"/>	<input type="checkbox"/>
13 Jeg har vondt i ryggen nesten hele tiden.....	<input type="checkbox"/>	<input type="checkbox"/>
14 På grunn av ryggen synes jeg det er vanskelig å snu meg rundt i sengen.....	<input type="checkbox"/>	<input type="checkbox"/>
15 På grunn av ryggen har jeg dårlig matlyst.....	<input type="checkbox"/>	<input type="checkbox"/>
16 På grunn av ryggsmertene har jeg vanskeligheter med å ta på meg strømpene.....	<input type="checkbox"/>	<input type="checkbox"/>
17 På grunn av ryggsmertene kan jeg bare gå korte strekninger.....	<input type="checkbox"/>	<input type="checkbox"/>
18 På grunn av ryggen sover jeg dårligere.....	<input type="checkbox"/>	<input type="checkbox"/>
19 På grunn av ryggen får jeg hjelp av andre til å kle på meg.....	<input type="checkbox"/>	<input type="checkbox"/>
20 På grunn av ryggen sitter jeg det meste av dagen.....	<input type="checkbox"/>	<input type="checkbox"/>
21 På grunn av ryggen unngår jeg tunge jobber i huset.....	<input type="checkbox"/>	<input type="checkbox"/>
22 På grunn av ryggen er jeg mer irritabel og i dårlig humør ovenfor andre enn vanlig....	<input type="checkbox"/>	<input type="checkbox"/>
23 På grunn av ryggen går jeg saktere opp trapper enn vanlig.....	<input type="checkbox"/>	<input type="checkbox"/>
24 På grunn av ryggen holder jeg for det meste sengen.....	<input type="checkbox"/>	<input type="checkbox"/>

Utviklet av Roland, M. and Morris, R.
(1983). Oversatt av Grotle, M og
Vøllestad, NK (2001)

Vedlegg 6: Godkjenning fra regional etisk komite



Region: REK vest	Saksbehandler: Camilla Gjerstad	Telefon: 55978499	Vår dato: 05.09.2018	Vår referanse: 2018/1424/REK vest
			Deres dato: 12.06.2018	
			Vår referanse må oppgis ved alle henvendelser	

Kjersti Storheim
FORMI / Nevrokliviken

2018/1424 Kan Start Back Screening Tool predikere klinisk utfall hos ryggpasienter over 55 år?

Forskningsansvarlig: OsloMet - storbyuniversitetet
Prosjektleder: Kjersti Storheim

Vi viser til søknad om forhåndsgodkjenning av ovennevnte forskningsprosjekt. Søknaden ble behandlet av Regional komité for medisinsk og helsefaglig forskningsetikk (REK vest) i møtet 15.08.2018. Vurderingen er gjort med hjemmel i helseforskningsloven (hfl.) § 10.

Prosjektomtale (opprinnelig)

Dette masterprosjektet ønsker å undersøke om et internasjonalt anbefalt screeningskjema (Start Back Screening Tool) som har vist gode psykometriske egenskaper i yngre pasientpopulasjoner med korsryggsmarter har tilfredsstillende evne til å predikere klinisk utfall og å diskriminere mellom sub-grupper av pasienter også blant eldre ryggpasienter. Man vil benytte allerede innsamlede data fra forskningsprosjektet BACE-N (Back Pain in Elders-Norway). BACE er en internasjonal prospektiv kohortstudie som kartlegger ryggplager hos personer over 55 år i primærhelsetjenesten, den norske armen (BACE-N) ledes av professor Margreth Grothe ved OsloMet. Screeningskjemaets diskriminerende evner skal i dette masterprosjektet undersøkes ved hjelp av ROC-analyser, som også vil benyttes for å fastslå cut-offverdien for studiedeltakernes endring i funksjonsnivå. BACE-N er tidligere sendt til vurdering hos REK (vest) men ble klassifisert som å falle utenfor REKs' mandat (se vedlegg).

Vurdering

Relatert prosjekt

2014/1634 BACE: en prospektiv kohort studie av eldre med korsryggsmarter som søker hjelp i primærhelsetjenesten. REK vest vurderte i 2014 at studien falt utenfor helseforskningsloven.

Studiepopulasjonen

Deltakerne er personer som er 55 år eller eldre som har tatt kontakt med primærhelsetjenesten på grunn av en ny episode med ryggplager. Ca. 200 pasienter vil bli inkludert i studien.

Forsvarlighetsvurdering, samtykke er allerede innhentet

Det omsøkte masterprosjektet er en metodestudie som vil undersøke om et internasjonalt anbefalt screeningskjema (Start Back Screening Tool) har tilfredsstillende prediktiv validitet for å diskriminere mellom subgrupper av pasienter, også blant eldre ryggpasienter. Studien vil benytte opplysninger innhentet via spørreskjema om pasientens helse og funksjon (ved baseline samt oppfølgingsspørreskjema etter tre måneder). Samtykke fra deltakerne er allerede innhentet i et annet prosjekt (2014/1634). REK vest har ingen merknader til samtykkeskrivet og mener det er dekkende for analysene i studien.

Besøksadresse:
Armauer Hansens Hus (AHH),
Tverrfløy Nord, 2 etasje, Rom
281, Haukelandsveien 28

Telefon: 55975000
E-post: post@helseforskning.etikkom.no
Web: <http://helseforskning.etikkom.no/>

All post og e-post som inngår i saksbehandlingen, bes adressert til REK vest og ikke til enkelte personer

Kindly address all mail and e-mails to the Regional Ethics Committee, REK vest, not to individual staff

Prosjektslutt

Oppgitt prosjektslutt er 15.08.2021. I følge søknaden vil data i studien slettes ved prosjektslutt. REK vest har ingen merknader til dette.

Vedtak

REK vest godkjenner prosjektet i samsvar med forelagt søknad.

Vi gjør samtidig oppmerksom på at etter ny personopplysningslov må det også foreligge et behandlingsgrunnlag etter personvernforordningen. Det må forankres i egen institusjon.

Sluttmelding og søknad om prosjektendring

Prosjektleder skal sende sluttmelding til REK vest på eget skjema senest 15.02.2022, jf. hfl. §

12. Prosjektleder skal sende søknad om prosjektendring til REK vest dersom det skal gjøres vesentlige endringer i forhold til de opplysninger som er gitt i søknaden, jf. hfl. § 11.

Klageadgang

Du kan klage på komiteens vedtak, jf. forvaltningsloven § 28 flg. Klagen sendes til REK vest. Klagefristen er tre uker fra du mottar dette brevet. Dersom vedtaket opprettholdes av REK vest, sendes klagen videre til Den nasjonale forskningsetiske komité for medisin og helsefag for endelig vurdering.

Med vennlig hilsen

Marit Grønning
dr.med. professor
komitéleder

Camilla Gjerstad
rådgiver

Kopi til: mgrotle@oslomet.no