

Reliabilitet av den norske versjonen av Pain
Response to Activity and Position: En test-retest
studie.



Student: Monica Sneis

Kandidatnummer: 21

Manuellterapi teori: MANT395, 4.semester, 2022

Masterprogram i helsefag – Klinisk masterstudium i manuellterapi for
fysioterapeuter.

Institutt for global helse og samfunnsmedisin

Universitetet i Bergen

Antall ord: 9172

Innholdsfortegnelse;

Forord.....	5
Sammendrag	6
Abstract	7
Sentrale begreper.....	8
Forkortelser	9
Figurer og tabeller	10
1.0 Innledning.....	11
1.1 Presentasjon av problemområdet og bakgrunn for valg av temaet.....	11
2.0 Teori	12
2.1 Epidemiologi	12
2.2 Definisjoner og klassifisering.....	13
2.2.1 Korsryggmerter med nerverotsaffeksjon	13
2.2.2 Spinal stenose	14
2.2.3 Uspesifikke korsryggmerter.....	14
2.3 Metodestudier	15
2.3.1 Validitet	15
2.3.1 Reliabilitet	16
2.3.2 Forholdet mellom validitet og reliabilitet.....	18
2.3.3 Studiedesign	18
2.3.4 Statistiske metoder.....	19
2.3.5 Kappa.....	19
2.4 Tidligere forskning på området	21
2.4.1 Reliabilitet	22
2.4.2 Validitet	23
3.0 Hensikt og problemstilling	25
3.1 Hensikt	25
3.2 Problemstilling	25
4.0 Metode.....	26
4.1 Valg av forskningsdesign	26
4.2 Oversetelse av Pain Respons to Activity and Position Questionnaire.....	26
4.3 Pain Response to Activity and Position questionnaire	27
4.4 Utvalg	27
4.4.1 Inklusjonskriterier.....	27
4.4.2 Eksklusjonskriterier.....	28
4.5 Variabler.....	28

4.6 Datainnsamling.....	29
4.7 Analyse.....	29
4.8 Klassifisering etter algoritmen	30
4.9 Tolkning av kappa.....	30
4.10 Etikk	30
5.0 Resultat.....	31
5.1 Demografi.....	33
5.2 Kappa for hvert spørsmål	34
5.3 Kappa for hver deltaker	36
5.4 Variasjon i utvalget	38
5.5 Klassifisering etter algoritme	38
6.0 Diskusjon.....	41
6.1 Diskusjon av metode	41
6.2 Diskusjon opp mot tidligere forskning.....	44
6.2.1 Hvert enkelt spørsmål.....	45
6.2.2 Hver enkelt pasient	47
6.2.3 Variasjon i utvalget.....	47
6.2.4 Algoritmen.....	48
6.4 Klinisk relevans.....	49
6.5 Videre forskning.....	49
7.0 Konklusjon	51
Referanseliste:	52
Vedlegg:	55
Vedlegg 1: PRAP	55
Vedlegg 2: Algoritme	56
Vedlegg 3: Krysstabeller per spørsmål	57
Vedlegg 4: Krysstabeller per pasient	65
Vedlegg 5: Klassifisering ved test.	76
Vedlegg 6: Klassifisering ved retest.....	78

Forord

I forbindelse med masterstudiet har det vært mye diskusjoner om tolkning av, og hva de ulike kliniske testene og spørreskjemaene «virkelig» tester, og om vi som klinikere vil oppnå de samme resultatene fra gang til gang. På bakgrunn av dette ønsket jeg å sette med mer inn i begreper som reliabilitet og validitet, og hva dette forteller om de ulike testene og spørreskjemaene vi benytter i praksis.

Jeg fikk muligheten til å skrive min masteroppgave, om test-retest reliabilitet, i forbindelse med det norske BACE-prosjektet, som pågår ved forskningsgruppen for muskel- og skjeletthelse ved OsloMet.

BACE-N har som formål å kartlegge; klinisk forløp av smerter, funksjonsbegrensninger og helserelatert livskvalitet, hos pasienter over 55 år som tar kontakt med norsk primærhelsetjeneste på grunn av rygg smerter med eller uten utstrålende symptomer, og å kartlegge prognostiske faktorer for vedvarende smerter og funksjonsbegrensninger. Denne oppgave er et delprosjekt under BACE-studien der hensikten er å undersøke test-retest reliabilitet for PRAP.

Jeg ønsker å rette en stor takk til min veileder, professor emeritus Rolf Moe-Nilssen ved UIB, for gode råd og tilbakemeldinger underveis i arbeidet med masteroppgaven.

Samtidig ønsker jeg å takke professor Margreth Grotle, prosjektleder for BACE-N, for muligheten til å være med i dette prosjektet.

Drøbak 10.11.2022

Monica Sneis

Sammendrag

Bakgrunn: Pain Response to Activity and Position questionnaire (PRAP) er et pasient rapportert spørreskjema med 30 spørsmål. Formålet med skjemaet er å screene for fire ulike typer korsryggsmerter; spinal stenose, skivesykdom, spinalstenose med skivesykdom og uspesifikke/godartete korsryggsmerter. PRAP har blitt oversatt til norsk i forbindelse med BACE-N studien. Psykometriske egenskaper for den norske utgaven har ennå ikke blitt undersøkt.

Hensikt: Denne studien har til hensikt å undersøke test-retest reliabiliteten til den norske utgaven av PRAP.

Metode: En prospektiv kohortstudie med en test-retest design. Pasienter som søkte hjelp for korsryggsmerter i primærhelsetjenesten, hos fysioterapeuter, kiropraktorer og fastleger, ble rekruttert mellom april 2015 og februar 2020. Deskriptiv data ble benyttet for å beskrive deltakerne. Kappa ble benyttet i analysene av test-retest reliabilitet.

Resultater: Totalt 127 deltakere mottok PRAP både ved test og retest, 76 av disse ble ekskludert på grunn av manglende besvarelse på hele eller deler av skjemaet. Det er inkludert totalt er 51 deltakere i analysene. Gjennomsnittlig kappa for hvert enkelt spørsmål oppnådde moderat enighet ($K = 0.449$). Kappaverdiene på de ulike spørsmålene varierte fra svak enighet til betydelig enighet ($K = 0.144-0.660$). Gjennomsnittlig for hver enkelt deltaker oppnådde rimelig enighet ($K = 0.347$). Kappaverdiene varierte fra ingen enighet til betydelig enighet for hver enkelt pasient ($K = -0.084-0.745$). Det var ikke mulig å klassifisere flertallet av deltakerne til en av de ulike diagnosene hverken ved test eller retest, og var derfor ikke mulig å utføre en test-retest analyse for algoritmen.

Konklusjon: Den norske versjonen av PRAP viste akseptabel test retest reliabilitet, med moderat enighet, på de enkelte spørsmålene i skjemaet. Test retest reliabilitet for hver enkelt deltaker viste seg derimot ikke å være akseptabel, med kun rimelig enighet. Det var ikke mulig å gjennomføre en test retest analyse på klassifiserings algoritmen i PRAP.

Nøkkelord: Pain Response to Activity and Position questionnaire (PRAP), korsryggsmerter, screeningskjema, primærhelsetjenesten, eldre, test-retest, reliabilitet

Abstract

Background: Pain Response to Activity and Position (PRAP) is a thirty item patient-reported questionnaire which assists in diagnosing patients with low back pain. PRAP screens for four different diagnosis of low back pain; spinal stenosis, disk disease, disk disease with spinal stenosis and benign low back pain. PRAP has recently been translated into Norwegian, but its psychometric properties have not been evaluated.

Aim: The aim of this study is to assess the test-retest reliability of the Norwegian translation of PRAP.

Method: A prospective cohort study with a test-retest design. Participants seeking help for low back pain were recruited from primary health care among physiotherapist, chiropractors and physicians between April 2015 and February 2020. Descriptive data were used to describe the participants. Kappa statistics were used in the analyses of test-retest reliability.

Results: Total of 127 participants received PRAP at baseline and retest, 76 were excluded due to lack of non-response at retest or lack of response to some of the questions in PRAP. Fifty-one elderly over the age of 55 years participated in the study, mean age (SD) of 66.3 (8.4). Mean kappa for each question achieved a moderate agreement ($K = 0.449$). Kappa for each questions variation form fair agreement to substantial agreement ($K = 0.144-0.660$). Mean kappa for each participant achieved fair agreement ($K = 0.347$). Kappa for each participant had a variation form zero agreement to substantial agreement ($K = -0.084 - 0.745$). Classification of the participants to only one of the four diagnoses were difficult and made it impossible/difficult to perform a test-retest analysis.

Conclusion: The Norwegian version of PRAP showed acceptable test-retest reliability with moderate agreement for each of the questions. Test-retest reliability for each patient showed to be not acceptable, with reasonable agreement. Difficulties with classification made it impossible to perform a test-retest analysis of the algorithm.

Keywords: Pain Response to Activity and Position questionnaire (PRAP), low back pain, Screening questionnaire, Primary health care, elderly, clinical assessment.

Sentrale begreper

Korsryggsmerter - Korsryggsmerter er definert som smerter lokalisert mellom nederste ribbe og over nederste glutealfold, med eller uten radierende smerter til ben (Lærum, 2007).

Spinal stenose - Spinal stenose er en vanlig degenerativ lidelse i ryggspylen og oppstår når spinalkanalen blir trangere og gir en kompresjon av ryggmargen og/eller nerverøtter i korsryggen. En sentral stenose er en forsnævring av hele spinalkanalen, mens formlinal stenose er en forsnævring av åpningen der nerverøttene forlater spinalkanalen (Ferreira & de Luca, 2017).

Skivesykdom - Skivesykdom er en nedbrytning/ruptur av mellomvirvelskiven som ligger mellom to ryggvirvler. Tilstanden kan være alt fra asymptomatisk til intense smerter, og gir som regel ryggmerter og/eller nerve- og strålesmerter mot setet og beina

Uspesifikke/godartete korsryggsmerter - Diagnosen brukes når bestemt patoanatomisk årsak til smertene er ukjent/ikke kan fastslåes. Flere ulike strukturer kan være årsaken til uspesifikke korsryggsmerter, men det finnes ingen kliniske tester som kan fastslå dette (Maher et al., 2017).

Reliabilitet - “the degree to which the measurement is free from measurement error” (Mokkink et al., 2010b)

Forkortelser

PRAP – Pain Respons to Activity and Position questionnaire

BACE – BACk pain in Elders

BACE-N – BACk pain in Elders Norway

COSMIN – COnsensus-based Standards for the selection of health status Measurement
INstruments.

Figurer og tabeller

Figur 1: Illustrasjon av validitet og reliabilitet.

Figur 2: Formel for beregning av kappa

Figur 3: Tolkning av kappaestimer.

Figur 4: Flytskjema.

Figur 5: Scatterplot over kappa for hvert enkelt spørsmål.

Figur 6: Scatterplot over kappa for hver enkelt deltaker.

Figur 7a: Fordeling av deltakere i forhold til antall diagnoser ved test.

Figur 7b: Fordeling av deltakere i forhold til antall diagnoser ved retest.

Tabell 1: Krysstabell for spørsmålet «Føler du smerter i ryggen når du går mer enn 5 minutter»

Tabell 2: Oversikt over studier på PRAP.

Tabell 2: Test-retest reliabilitet for de ulike spørsmålene.

Tabell 4: Sensitivitet og spesifisitet for de ulike kategoriene av korsryggsmerter.

Tabell 5: Positiv og negativ verdi for de ulike kategoriene av korsryggsmerter.

Tabell 6: Sensibilitet, spesifisitet, positiv og negativ prediktiv verdi for utstrålende smerter i ben og smerter i ben ved gange.

Tabell 7: Antall dager mellom test og retest.

Tabell 8: Deskriptive data.

Tabell 9: Kappaestimat for hvert enkelt spørsmål.

Tabell 10: Gjennomsnittlig kappa for hvert enkelt spørsmål.

Tabell 11: Fortolkning av kappa for hvert enkelt spørsmål.

Tabell 12: Kappa for hver enkelt deltaker.

Tabell 13: Gjennomsnittlig kappa for hver enkelt deltaker.

Tabell 14: Fortolkning av kappa for hver enkelt deltaker.

Tabell 15: Klassifisering av deltakere i forhold til diagnosene.

1.0 Innledning

1.1 Presentasjon av problemområdet og bakgrunn for valg av temaet

Som manuellterapeut i primærhelsetjeneste møter vi mange pasienter med smerter i korsryggen. Dette er den hyppigste av alle muskel- og skjelettlidelser, og den mest vanlige muskel- og skjelettlidelsen i primærhelsetjenesten (Lærum, 2007). Globalt sett er korsryggsmerter den ledende årsaken til Years Lived with Disability (YLD) (Wu et al., 2020).

Som klinikere kan vi, foruten anamnese, klinisk undersøkelse og ulike tester, bruke ulike spørreskjema som en del av undersøkelse og behandling. Det finnes et utall av ulike typer spørreskjemaer som kan benyttes, fra generiske til sykdomsspesifikke. Et mye brukt screening skjema for plager i muskel- og skjelettsystemet, inkludert korsryggsmerter, er Ørebro screeningskjema for gule flagg. Dette er et skjema som kan vurdere prognosen for utvikling av langvarige plager i muskel- og skjelettsystemet (Linton & Hallden, 1998). Andre sykdomsspesifikke skjemaer er; Roland-Morris Disability Questionnaire (RMDQ) (Ostelo & de Vet, 2005) og Ostwestry Disability Indeks Index (ODI) (Fairbank & Pynsent, 2000), som begge vurderer ryggens funksjon.

Et spørreskjema som nylig er oversatt til norsk er; Pain Response to Activity and Position questionnaire (PRAP). Dette er et selvrappoterende spørreskjema utviklet for å screene for fire ulike typer korsryggsmerter; skivesykdom, spinal stenose, spinal stenose med skivesykdom og godartet/uspesifikk korsryggsmerter, basert på pasientenes rapportering av smerter i ulike posisjoner og aktiviteter (Roach et al. 1997b).

For å kunne stole på svaret i slike spørreskjemaer er det viktig at disse har god validitet og reliabilitet. Validiteten sier noe om skjemaet tester det fenomenet det er ment å teste (De Vet, 2011; Parikh et al., 2008). Reliabilitet sier noe om resultatene er reproduerbare, det vil si om man får det samme resultatet ved ulike målinger (De Vet, 2011; Oladimeji, 2015).

Oversettelse av undersøkelsesskjemaer til andre språk kan påvirke skjemaets psykometriske egenskaper. Det er derfor viktig at de psykometriske egenskapene til den norske versjonen av PRAP undersøkes nærmere.

I min masteroppgave ønsker jeg å undersøke test-retest reliabiliteten til den norske versjonen av PRAP.

2.0 Teori

2.1 Epidemiologi

Smerter i korsryggen er den vanligste av alle muskel- og skjelettlidelser, med en livstidsprevalens på mellom 60 - 85% (Krismer, 2007). Til enhver tid vil rundt 15% av den voksne delen av befolkningen ha korsryggsmerter (Lærum, 2007). Korsryggsmerter bidrar til 11% av antall år levd med funksjonsnedsettelse, og er dermed den største enkeltårsak til dette (Lærum, 2013). Det har vist seg at prevalens og kostnader knyttet til korsryggsmerter har vært økende i mange år (Hartvigsen et al., 2018; Wu et al., 2020).

Korsryggsmerter kan føre til store påkjenninger for den enkelte person og for samfunnet. For den enkelte dreier det seg hovedsakelig om konsekvenser i form av smerte og nedsatt fysisk funksjon så lenge ryggsmertene varer, men for de med vedvarende smerter vil det også kunne påvirke arbeid og den økonomiske situasjonen (Hoy et al., 2010). For samfunnet er det knyttet store kostnader til korsryggsmerter. I Norge er de årlige kostnadene på grunn av korsryggsmerter blitt estimert til 15 milliarder kroner, og 16% av langtids sykefravær og 16% av nye uførekrav har korsryggsmerter som primærdiagnose (Lærum, 2013). Denne gruppen er også storforbrukere av helsetjenester, og står for 5 - 10% av konsultasjonene hos allmennlegene og 27% av fysioterapibehandlingene (Werner & Indahl, 2005). Det har vist seg at 75% av kostnadene knyttet til korsryggsmerter kan føres tilbake til 5% av pasientene i denne gruppen (Kent & Keating, 2008).

Prevalensen av alvorlige korsryggsmerter øker også med økende alder (Dionne et al., 2006; Docking et al., 2011). Det er estimert at mellom 2010 og 2050 vil antall personer med korsryggsmerter fra 60 år og oppover øke med 56% i I-land (Grotle, 2020), noe som vil gi en betydelig økt byrde fra vedvarende korsryggsmerter. Til tross for dette har den eldre populasjonen ofte blitt ekskludert fra forskning på korsryggsmerter (Grotle et al., 2005; Paeck et al., 2014). På bakgrunn av dette ble den internasjonale BACK pain in Elders; BACE-studien etablert i 2008 (Scheele et al., 2011). Det ble i 2015 etablert en norsk arm, BACKc pain in Elders Norway (BACE-N), av den internasjonale studien. En pilot studie ble gjennomført i Norge fra mars 2015 til april 2017. Det ble ikke gjort store endringer i protokollen til BACE-N så rekruttering fortsatte mer eller mindre uforandret frem til februar 2020 (Grotle, 2020)

2.2 Definisjoner og klassifisering

Korsryggsmarter er definert som smerter lokalisert mellom nederste ribbe og over nederste glutealfold, med eller uten radierende smerter til ben (Lærum, 2007).

Korsryggsmarter kan klassifiseres ut ifra varighet, og deles inn i;

- Akutte korsryggsmarter, med varighet under 3 måneder.
- Langvarige korsryggsmarter, med varighet over 3 måneder.
- Subakutte korsryggsmarter, er en undergruppe av de akutte, med varighet mellom 6 og 12 uker (Lærum, 2007).

I likhet med for eksempel hodepine og svimmelhet er korsryggsmarter et symptom, og det kan være flere ulike årsaker til smertene.

Korsryggsmertene kan klassifiseres ut ifra om de er spesifikke eller uspesifikke. Spesifikke korsryggsmarter er karakterisert ved at det finnes en patologisk forklaring/årsak til smertene. Dette kan være korsryggsmarter med nerverotaffeksjon, oftest forårsaket av skiveprolaps eller spinal stenose. Korsryggsmarter med mulig alvorlig underliggende sykdom, som for eksempel infeksjon, cancer, fraktur eller forårsaket av nevrologisk akutt tilfelle/cauda equina (Lærum, 2007).

I motsetning til de spesifikke korsryggsmertene finner man ingen bakenforliggende årsak til de uspesifikke korsryggsmertene.

Uspesifikke korsryggsmarter utgjør 80 - 90%, mens nerverotsaffeksjon utgjør mellom 5 - 10% av tilfellene. Mulig alvorlig underliggende sykdom inkludert cauda equina utgjør kun 1 - 5% av tilfellene (Lærum, 2007).

2.2.1 Korsryggsmarter med nerverotsaffeksjon

Ved en nerverotsaffeksjon vil det være en anatomisk kompresjon av nerverot. Pasienter med skivesykdom/prolaps med affeksjon av nerverot forteller ofte om mer distinkte smerter tilsvarende ett eller flere dermatomer. Smertene i benet er ofte mer intense enn smertene i ryggen. De kan også ha nummenhet og parestesier i ett eller flere dermatomer, samt nedsatt styrke tilsvarende en eller flere lumbosakrale nerverøtter. Pasienter forteller også om økte

smertes når de sitter, samt opplevd reproduksjon av de utstrålende smertene ved hosting og nysing (Benzakour et al., 2019; Lærum, 2007).

2.2.2 Spinal stenose

Spinal stenose er en vanlig degenerativ lidelse i ryggspylen. Lidelsen kan gi store smerter og betydelig funksjonsnedsettelse. Spinal stenose oppstår når spinalkanalen blir trangere og gir en kompresjon av ryggmargen og/eller nerverøtter i korsryggen. Spinal stenose kan gi nerverotsaffeksjon eller vertebrogen claudicatio. En sentral stenose er en forsnevring av selve spinalkanalen, mens forinial stenose er en forsnevring av åpningen der nerverøttene forlater spinalkanalen (Ferreira & de Luca, 2017).

Pasienter med spinal stenose forteller om nedsatt gangdistanse på grunn av smerter i sete og/eller bein, ved sentral stenose er smertene i begge bein. Smertene er hovedsakelig over kne. Mange opplever også smerter i korsryggen. Det kan også komme parestesier eller nedsatt styrke i sete, lår eller legger (Deer et al., 2019; Jensen et al., 2021; Lærum, 2007). Smertene kommer ved aktiviteter og stillinger som gir ekstensjon i ryggspylen, som stående stilling eller ved gange, disse smertene blir langsomt bedre om man stopper opp. Smertene bedres ved fleksjon i ryggen, som for eksempel om pasienten bøyer seg fremover eller ved sitting (Deer et al., 2019; Lærum, 2007)

Spinal stenose er betraktet som den ryggglidelsen som gir mest funksjonsnedsettelse hos eldre pasienter og prevalensen av spinal stenose er estimert til å være 47% hos personer eldre en 60 år (Ferreira & de Luca, 2017).

2.2.3 Uspesifikke korsryggsmarter

Uspesifikke korsryggsmarter er den mest vanligste formen for korsryggsmarter, og denne diagnosen brukes når bestemt patoanatomisk årsak til smertene er ukjent/ikke kan fastslåes. Flere ulike strukturer kan være årsaken til uspesifikke korsryggsmarter, men det finnes ingen kliniske tester som kan fastslå dette (Maher et al., 2017).

Disse pasientene er i god almenntilstand. I tillegg til smerter i korsrygg kan de ha smerter i sete og lår. Smerteintensiteten varierer ofte, og smertene er som regel bedre i ro (Lærum, 2007).

Gjennom undersøkelse av pasientene er målet å avdekke de med spesifikke årsaker til ryggmerter, som for eksempel cancer, fraktur, cauda equina syndrom, radikulopati eller spinal stenose. De gjenværende tilfellene der man ikke har avdekket noen bakenforliggende årsak diagnostiseres da som uspesifikke korsryggsmerter (Maher et al., 2017).

2.3 Metodestudier

Måling og måleinstrumenter er en viktig del av både medisinsk forskning og klinisk arbeid. Målinger danner basis for diagnostisering, prognoser og evaluering av behandling. God kvalitet på måleinstrumentene er derfor viktig (De Vet, 2011).

Evaluering av måleinstrumenter en viktig del av forskningen, og målet med slike metodestudier/-forskning er å dokumentere og forbedre kvaliteten på måleinstrumentene (Carter, 2015; De Vet, 2011). COSMIN initiativet er et internasjonalt samarbeid blant ulike forskere med ekspertise i utvikling og evaluering av måleinstrumenter. Målet for denne gruppen er å forbedre kvalitet på metodestudier ved forbedring og standardisering av metodologi, samt utvikle praktiske verktøy for evaluering av ulike måleegenskaper (COSMIN, 2022). Det er blant annet utviklet sjekklister for; evaluering av metodisk kvalitet på studier av måleegenskaper (Mokkink et al., 2010a), vurdering av måleegenskaper på eksisterende pasient-rapporterte måleinstrumenter (PROMs) (Mokkink, 2019), samt en konsensus om terminologi, taksonomi og definisjon av måleegenskaper hos helserelaterte PROMs (Mokkink et al., 2010b).

Egenskaper ved pasient-rapporterte utfallsmål kan deles inn i; validitet, reliabilitet og responsiveness (Mokkink et al., 2010b).

2.3.1 Validitet

Validitet er definert som;

«the degree to which an instrument truly measures the construct(s) it purports to measure»

(Mokkink et al., 2010b)

Validiteten sier noe om gyldigheten til målingene, at vi måler vi det vi ønsker å måle (Carter, 2015; De Vet, 2011). En skiller hovedsakelig mellom tre ulike former for validitet; innholds validitet, som fokuserer på innholdet, kriterievaliditet, som benyttes hvor det finnes en gullstandard for fenomenet, og undersøker hvorvidt måleinstrumentet samsvarer med gullstandarden, samt begrepsvaliditet som benyttes det det ikke finnes noen gullstandard, men undersøker hvorvidt måleinstrumentet oppnår akseptable skår basert på eksisterende kunnskap om fenomenet (De Vet, 2011).

Ved test av validitet på diagnostiske tester måles sensitivitet, som er evnen en test har til å korrekt klassifisere en person med sykdom, og spesifisitet, som er evnen en test har til å korrekt klassifisere en person uten sykdom (Parikh et al., 2008). I klinikken vil man ved testing få falske positive og negative svar på tester. Vi kan ikke vite om personen vi tester er en sann eller falsk positiv/negativ. Positiv prediktiv verdi (PPV) forteller hvor stor prosentvis sjans for at testen en sann positiv. Negativ prediktiv verdi (NPV) forteller hvor stor prosentvis sjans for at testen er sann negativ. Jo høyere PPV/NPV verdi, jo nærmere gullstandardens prediktive evne har testen (Parikh et al., 2008).

2.3.1 Reliabilitet

I klinisk sammenheng er det viktig at diagnostiske-, funksjonstester og spørreskjemaer vi benytter er reliable, det vil si at den observerte endringen er en reell endring, og ikke et resultat av lav reliabilitet (De Vet, 2011).

Reliabilitet sier noe om i hvilken grad et spørreskjema, en test, en observasjon eller en hvilken som helst prosedyre gir det samme resultatet ved gjentatte målinger. Det er et uttrykk for graden av stabilitet i målingene. Andre begreper som benyttes om det samme er repeterbarhet eller reproduserbarhet (De Vet, 2011).

Reliabilitet er definert som;

“the degree to which the measurement is free from measurements error”

(Mokkink et al., 2010b)

Det er i tillegg til den generelle definisjonen av reliabilitet, en utvidet definisjon;

«The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions»

(De Vet, 2011)

Reliabilitet kan undersøkes på tre ulike måter; 1) Ved å undersøke om to eller flere ulike undersøkere/personer/testere oppnår det samme resultatet, dette kalles intertester reliabilitet, 2) om samme undersøker oppnår det samme resultatet ved to eller flere målinger, dette kalles intratester reliabilitet og 3) ved test-retest reliabilitet, der det undersøkes samsvaret for samme tester over tid (De Vet, 2011; Oladimeji, 2015). Intern konsistens er også et reliabilitetsmål, her blir ulike deler av samme måleinstrument sammenlignet (De Vet, 2011).

Test-retest reliabilitet er en vanlig måte å undersøke reliabiliteten til spørreskjemaer på. Det samme utvalget vil da besvare spørreskjemaet ved to eller flere tidspunkter (De Vet, 2011). For et spørreskjema vil et tegn på test-retest reliabilitet være i hvilken grad en persons respons vil forbli det samme over tid. To antagelser ligger til grunn for test-retest reliabilitet. Det er; 1) at fenomenet som undersøkes ikke endrer seg over tid, og 2) er at perioden mellom de ulike målingene er kort nok, men allikevel lang nok til at deltakerens hukommelse om resultatet av første måling påvirker måling nummer to (Oladimeji, 2015).

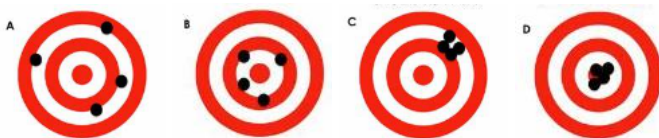
Klassisk test teori (CCT) er en strategi benyttet for å måle fenomener som ikke er observerbare. Informasjon om det ikke observerbare fenomenet fås ved å måle ulike egenskaper fenomenet manifesterer. CCT sier at den observerte målingen/-metoden består av en sann skår og ulike feilkomponenter/-kilder, og ved kvantitative målinger består resultatet av disse delene (De Vet, 2011). Ved undersøkelse av reliabilitet ønsker man å anslå disse målefeilene (feilkomponenter/-kilder) og hvordan man kan minimere disse.

Målefeil kan være systematiske eller tilfeldige. En systematisk målefeil vil påvirke gjennomsnittet av målingene enten i en positiv eller en retning. Tilfeldige målefeil omtales derimot som «støy» i målingene, og vil normalt ikke påvirke gjennomsnittet (De Vet, 2011). Tilfeldige målefeil kan være variasjoner hos deltakerne, hos personen som utfører målingene eller mellom personer som utfører målingene, manglete standardisering av test, variasjoner i

selve måleinstrumentet, konteksten målingene blir utført i eller instabilitet i selve fenomenet som blir målt (De Vet, 2011; Oladimeji, 2015).

2.3.2 Forholdet mellom validitet og reliabilitet

Forholdet mellom validitet og reliabilitet kan illustreres med en blink og noen dart piler (figur 1). Sentrum av blinken illustrerer fenomenet vi undersøker, mens dartpilene illustrerer antall tester/målinger og resultatet av disse. Jo nærmere sentrum blinken dartpilene treffer, jo høyere validitet, jo mer samlet alle dartpilene treffer, jo høyere reliabilitet (De Vet, 2011).



Figur 1: Illustrasjon av validitet og reliabilitet.

Figur 1A: Pilene treffer spredt og utenfor sentrum av blinken illustrerer både lav validitet og reliabilitet. Figur 1B: Pilene er nærmere sentrum, men fortsatt spredt illustrerer lav reliabilitet, men mulig valid resultat. Figur 1C: Pilene er treffer samlet viser og viser høy reliabilitet, men treffer ikke blink og viser lav validitet, og figur 1D: Pilene treffer både i sentrum og er samlet illustrerer både høy reliabilitet og validitet.

2.3.3 Studiedesign

Ved design av metodestudier på PROMs er det anbefalt å benytte COSMIN sjekkliste for studiedesign. Denne sjekklisten består av generelle anbefalinger for design av studier av måleegenskaper, samt spesifikke standarder for studier av de ulike egenskapene (Mokkink, 2019).

Ved undersøkelse av reliabilitet er det anbefalt minimum to målinger, at gjennomføringen av målingene er uavhengig, at pasienten/fenomenet er stabilt i måleperioden, at tiden mellom målingene er adekvat, at test situasjonene er lik/lignende og at analysene er gjennomført med et akseptabelt antall deltakere (Mokkink, 2019).

2.3.4 Statistiske metoder

Reliabilitet kan beregnes på to måter; absolutt og relativ reliabilitet. Relativ reliabilitet undersøker sammenhengen mellom to eller flere sett med repeterte målinger, mens absolutt reliabilitet undersøker variabiliteten mellom de ulike målingene (Carter, 2015).

Ved analyse av kontinuerlige variabler benyttes Intraclass Correlation Coefficient (ICC) og Standard Error of Measurement (SEM) for henholdsvis relativ og absolutt reliabilitet. Ved kategoriske data – nominal og ordinal - benyttes Cohens kappa for relativ reliabilitet. For absolutt reliabilitet finnes det derimot ikke metoder for kategoriske data, da disse dataene kun klassifiseres og rangeres. Det kan derimot undersøkes hvilken prosent av målingene som blir klassifisert i samme kategori, dette kalles «percentage of agreement» (De Vet, 2011).

2.3.5 Kappa

Som nevnt over benyttes Cohens Kappa for å undersøke overensstemmelse mellom kategoriske data, ved både nominale og ordinale data (De Vet, 2011). Ved nominaldata benyttes uvektet kappa, mens ved ordinaldata kan også vektet kappa benyttes (Elfving, 2015).

Kappa er et mål som korrigerer for grad av enighet som ville funnet sted ved tilfeldighet. Grad av tilfeldighet er den forventede enigheten (P_e) (De Vet, 2011).

Kappa beregnes ved å; dele observert enighet i utvalget (P_o) - forventet enighet i utvalget (P_e), på 1 minus forventet enigheten i utvalget (P_e) (figur 2).

$$kappa(\kappa) = \frac{P_o - P_e}{1 - P_e}$$

Figur 2: Formel for beregning av kappa.

Videre under vises et eksempel på beregning av kappa basert på krysstabell for observert enighet (tabell 1). I eksempelet er det benyttet krysstabellen til spørsmål «Føler du smerte i ryggen når du står oppreist i mer enn 5 minutter».

Tabell 1: Krysstabell for spørsmål «Føler du smerte i ryggen når du står oppreist i mer enn 5 minutter».

Test/retest	Ingen	Bedre	Samme	Verre	Ikke aktuelt	Totalt
Ingen	4	3	1	0	0	8
Bedre	2	1	5	1	0	9
Samme	1	3	11	1	0	15
Verre	1	2	4	9	2	18
Ikke aktuelt	0	0	0	0	0	0
Totalt	8	9	21	11	2	51

Beregning av observert enighet:

$$P_o = \frac{4+1+11+9}{51} = \frac{25}{51} = 0.491$$

Beregning av forventet enighet:

$$P_e = \frac{8}{51} \times \frac{8}{51} + \frac{9}{51} \times \frac{9}{51} + \frac{15}{51} \times \frac{21}{51} + \frac{18}{51} \times \frac{11}{51} + \frac{0}{51} \times \frac{2}{51} = 0.253$$

Beregning av kappa:

$$K = \frac{0.491 - 0.253}{1 - 0.253} = 0.31$$

Kappaverdien varierer fra -1 til 1, der en skår på 1 oppnås ved full enighet mellom de to målingene. Om verdien er 0 vil tolkes det som korrelasjonen er en tilfeldighet. Ved en verdi på under 0 vil det ikke være noen korrelasjon mellom målingene, dette er sjeldent relevant i praksis. Kappaverdiene fra 0 til 1 kan angis på en skala fra svak enighet til nesten perfekt enighet (figur 3) (De Vet, 2011).

Kappa		ICC	
0.80	Nesten perfekt	Utmerket	0.90
		God	
0.60	Betydelig	Moderat	0.50
0.40	Moderat	Dårlig	
0.20	Rimelig		
	Svak		

Figur 3: Tolkning av kappaverdier (De Vet, 2011).

Det anbefales å presentere krysstabeller med fordelingen av variablene sammen med kappaverdiene, dette fordi kappa-verdien blir påvirket av fordelingen mellom klassene, antall klasser og systematiske forskjeller mellom ulike testere. Kappas verdi er også avhengig av heterogeniteten i utvalget. Heterogeniteten i utvalget kan ses i krysstabellene. En jevn fordeling mellom de ulike klassene representerer et heterogent utvalg, mens et skjevfordelt utvalg, for eksempel nesten alle deltakerne er i samme klasse, peker i retning av et mer homogent utvalg. En mer skjevfordeling av variablene fører til en større brøkdel av tilfeldig enighet, noe som gir mindre rom for reell enighet. Selv om kappa-verdien teoretisk sett fortsatt kan nå 1, vil verdiene i praksis normalt være/bli lavere (De Vet, 2011).

2.4 Tidligere forskning på området

Det ble søkt i to databaser, Medline og Pedro, for å finne tidligere forskning på PRAP. Det ble benyttet søkeordene; «Pain Response to Activity and Position», PRAP, reliabilitet, test-retest og validitet.

Ved søk ble det kun funnet en artikkel som omhandlet PRAP, det var de Carvalho et al. (2019). I referanselisten til denne artikkelen ble det funnet ytterligere to artikler som omhandlet reliabilitet og validitet av PRAP (tabell 2).

Tabell 2: Tidligere publiserte artikler på reliabilitet og validitet av PRAP.

Forfatter	Artikkel	Reliabilitet	Validitet
Roach et al. 1997a	Test-retest Reliability of Patient Reports of low back pain.	Test-retest reliabilitet	
de Carvalho et al. 2019	Cross-cultural adaption and reliability of pain response to activity and position questionnaire.	Intra-rater, inter-rater reliabilitet	
Roach et al. 1997b	The sensitivity and Specificity of Pain Response to Activity and Position in Categorizing Patients with Low Back Pain		Sensitivitet, spesifisitet, positive / negative prediktiv verdi

2.4.1 Reliabilitet

Test-retest reliabilitet av PRAP har tidligere blitt undersøkt i forhold til hvert enkelt spørsmål (Roach et al. 1997a). Reliabiliteten viste seg å være moderat til høy målt med kappa (K) for både korsryggsmerters respons til posisjon og aktivitet ($K = 0.46 - 0.89$) (tabell 2) og hoftel/bensmerters respons til posisjon og aktivitet ($K = 0.69 - 0.85$) (tabell 3) for de ulike spørsmålene. Totalt sett hadde spørsmål i forhold til bensmerter en tendens til å ha noe høyere test-retest reliabilitet.

Tabell 3; Test-retest reliabilitet for PRAP (Roach et al. 1997a)

Aktivitet	Korsryggsmerters (kappa)	Bensmerter (kappa)
Stå	0.70	0.83
Gå	0.67	0.82
Sitte	0.46	0.70
Bøye seg fremover	0.65	0.83
Kjøre bil	0.61	0.78
Ryggliggende	0.67	0.70
Sideliggende	0.69	0.73
Mageliggende	0.78	0.85
Hoste	0.75	0.75
Om morgenen	0.73	0.78
Om kvelden	0.72	0.84
Fra sittende til stående	0.64	0.78
Presse	0.77	0.69
Løfte	0.89	0.82
Bære	0.84	0.82

Test-retest reliabilitet er også undersøkt for algoritmen brukt for å klassifisere pasientene til de ulike diagnosene; skivesykdom, spinalstenose og skivesykdom med spinal stenose og godartet/uspesifikke korsryggsmerter. For de ulike kategoriene var test-retest reliabiliteten; $K = 0.89$ for skivesykdom, $K = 0.58$ for stenose, $K = 0.91$ skivesykdom med spinal stenose og $K = 0.57$ for godartete korsryggsmerter (Roach et al. 1997b).

De Carvalho et al. (2019) har undersøkt intrarater og interrater reliabilitet for den brasilianske utgaven av PRAP. Kappa-verdiene varierte fra 0.50 til 1.00 for korsryggsmerter og fra 0.23 til 0.84 for bensmerter, mens for korsryggsmerter varierte kappa-verdiene fra 0.25 til 0.63 og for bensmerter fra 0.18 til 0.53. Kappa-verdiene for intrarater og interrater reliabilitet for klassifisering etter algoritmen var henholdsvis 0.52 og 0.47.

2.4.2 Validitet

Roach et al. (1997b) undersøkte også sensitivitet, spesifisitet, positiv og negativ prediktiv verdi for algoritmen (tabell 4 og tabell 5). Det var stor variasjon i både sensitivitet og spesifisitet (tabell 4). Algoritmen for skivesykdom med spinal stenose hadde høyest sensitivitet (0.81) og negativ prediktiv verdi (0.94), mens algoritmen for godartete/uspesifikke korsryggsmerter hadde høyest spesifisitet (0.71) og positiv prediktiv verdi (0.40) (tabell 4)(Roach et al. 1997b).

Tabell 4; Sensitivitet og spesifisitet for de ulike kategoriene av korsryggsmerter (Roach et al. 1997b).

Kategori av korsryggsmerter	Sensitivitet	Spesifisitet
Godartete/ uspesifikke korsryggsmerter	0.56	0.72
Skivesykdom	0.65	0.47
Spinal stenose	0.52	0.74
Skivesykdom med spinal stenose	0.81	0.54

Tabell 5; Positiv og negativ prediktiv verdi for de ulike kategoriene av korsryggsmerter (Roach et al. 1997b).

Kategori av korsryggsmerter	Positiv prediktiv verdi	Negativ prediktiv verdi
Godartete/ uspesifikke korsryggsmerter	0.40	0.82
Skivesykdom	0.35	0.77
Spinal stenose	0.36	0.85
Skivesykdom med spinal stenose	0.24	0.94

I forhold til kategoriene skivesykdom, spinal stenose og spinal stenose med skivesykdom er det undersøkt sensitivitet, spesifisitet, samt negativ og positiv prediktiv verdi for utstrålende smerter til ben og pseudocladication. Utstrålende smerter hadde god sensitivitet både for skivesykdom (0.81) og skivesykdom med spinal stenose (0.94), spesifisiteten var derimot lav for dette symptomet. Den positive prediktive verdien var henholdsvis 0.29 (skivesykdom) og 0.19 (skivesykdom med spinal stenose) (tabell 6). Pseudocladication hadde relativt god sensitivitet (0.63) og spesifisitet (0.71) for spinal stenose, samt en positiv prediktiv verdi på 0.40 og en negativ prediktiv verdi på 0.85 (tabell 6). Av resultatene ser vi at den positive prediktive verdien for utstrålende smerter i ben var svært lav for både skivesykdom og skivesykdom med spinal stenose, bakgrunnen for dette kan være at flere ulike årsaker kan gi slike smerter (Roach et al. 1997b).

Tabell 6: Sensibilitet, spesifisitet, positiv og negativ prediktiv verdi for utstrålende smerter i ben og smerter i ben ved gange (Pseudocladication)(Roach et al. 1997b).

Kategori av ryggsmerter	Test	Sensitivitet	Spesifisitet	Positiv prediktiv verdi	Negativ prediktiv verdi
Skivesykdom	Utstrålende smerter i ben	0.81	0.18	0.29	0.68
Spinalstenose	Smerter i ben ved gange	0.63	0.71	0.40	0.86
Skivesykdom med spinal stenose	Utstrålende smerter i ben	0.94	0.21	0.19	0.95
Skivesykdom med spinal stenose	Smerter i ben ved gange	0.47	0.64	0.18	0.88

3.0 Hensikt og problemstilling

3.1 Hensikt

Denne masteroppgaven er en del av BACE-N studien. Hovedformålet med BACE-N prosjektet er å kartlegge; klinisk forløp av smerter, funksjonsbegrensninger og helserelatert livskvalitet, hos pasienter over 55 år som tar kontakt med primærhelsetjenesten på grunn av rygg smerter med eller uten utstrålende symptomer, samt å kartlegge prognostiske faktorer for vedvarende smerter og funksjonsbegrensninger.

I tillegg til de spesifikke målene inkluderer BACE-N flere metodestudier, da flere av måleinstrumentene i den originale BACE protokollen ikke tidligere har blitt benyttet i en norsk kontekst, ett av disse er PRAP.

PRAP er blitt oversatt til norsk i forbindelse med BACE-N, og i den forbindelse er det behov for å undersøke dets metodiske egenskaper som intern konsistens, begrepsvaliditet og test-retest reliabilitet (Grotle, 2020).

Hensikten med denne masteroppgaven er å undersøke test-retest reliabilitet for den norske utgaven av PRAP.

3.2 Problemstilling

Er test-retest reliabiliteten til den norske utgaven av PRAP tilfredsstillende?

4.0 Metode

4.1 Valg av forskningsdesign

BACE-N er en prospektiv kohortstudie, og en del av det internasjonale BACE samarbeidet (Grotle, 2020; Scheele et al., 2011). BACE ble etablert i 2008, for å etablere standardisert metodikk for kohortstudier, og for å innhente og dele data om byrden av ryggsmarter hos eldre personer (Scheele et al., 2011).

BACE-studier har blitt etablert i flere land blant annet Australia, Brasil, Nederland og Norge (Grotle, 2020). Hovedmålene er; å fastslå klinisk forløp og byrden av ryggsmarter hos eldre, å identifisere prognostiske faktorer for vedvarende ryggsmarter og nedsatt funksjon, samt og å utforske standard/vanlig/typisk behandling som tilbys i primærhelsetjenesten (Scheele et al., 2011).

Den norske grenen av studien (BACE-N) hadde oppstart, med en pilot studie, 15.01.2017. Full rekruttering fortsatte fra januar 2017 til februar 2020 uten store endringer i studiedesign. Studien har som mål å innhente informasjon fra primærhelsetjenesten, om prognostiske faktorer for nedsatt funksjon og klinisk forløp av smerter over to år, relatert til rygg hos eldre personer. Inkludert er også en grundig beskrivelse av utnyttelse og kostnader av helsetjenester. I tillegg til hovedmålene er det i BACE-N inkludert flere metodestudier, da flere av måleinstrumentene fra den originale BACE protokollen har blitt oversatt til norsk, og derfor bør valideres for norske forhold (Grotle, 2020).

4.2 Oversetelse av Pain Respons to Activity and Position Questionnaire

Den engelske versjonen av PRAP har blitt oversatt og adaptert til norske forhold etter internasjonale retningslinjer (Beaton et al., 2000). To oversettere, en kliniker og en filolog, har oversatt PRAP til norsk. Begge har norsk som morsmål. De har gjennomført oversettelsen først uavhengig av hverandre, for deretter å sammenstille disse til en norsk utgave. To personer med engelsk som morsmål oversatte deretter den norske versjonen av PRAP tilbake til engelsk. Disse to personene var blindet for den originale engelske versjonen av PRAP. Oversettelsen ble først oversatt av hver enkelt uavhengig av hverandre, og deretter sammenfattet til en engelsk versjon. Den norske versjonen ble til slutt evaluert og godkjent av to forskere, Margreth Grotle og Rikke Munk, i forskningsgruppen.

4.3 Pain Response to Activity and Position questionnaire

PRAP er et selvrappporterende spørreskjema utviklet for å screene for fire ulike typer korsryggsmerter (skivesykdom, spinal stenose, spinal stenose med skivesykdom og godartet/uspesifikke korsryggsmerter) basert på pasientenes rapportering av smerter i ulike posisjoner og aktiviteter. PRAP bruker pasientens beskrivelser av smerter, dens lokalisasjon, samt hvordan smerten endrer seg med posisjoner og aktivitet for å diagnostisere årsaken til korsryggsmerter. Spørreskjemaet inneholder 31 spørsmål som beskriver ulike aktiviteter eller posisjoner som kan forverre eller lette smerter i korsrygg, hofta eller ben hos pasienter med plager i korsryggen. 15 spørsmål tar for seg endringer av smerte i korsryggen, ett spørsmål lokalisasjon av smerter i ben og 15 spørsmål tar for seg endringer av smerte i hofta og ben. Pasientene rapporterer om de ved de ulike aktivitetene og posisjonene opplevde 1) ingen smerte, 2) ble bedre, 3) smertene var den samme, 4) smertene ble verre eller 5) ikke aktuelt om aktiviteten ikke ble gjennomført (Roach et al. 1997a).

4.4 Utvalg

Pasienter over 55 år som oppsøkte lege, fysioterapeut eller kiropraktorer som arbeider i primærhelsetjenesten i Norge for en ny episode med korsryggsmerter ble rekruttert til studien. I hovedstudien skal det inkluderes totalt 450 deltakere. For undersøkelse av test-retest reliabilitet blir pasienter forespurt om å fylle ut en retest hjemme 2 dager etter baseline. Dette ble gjort inntil nok antall hadde fullført retest. Retest ble fylt ut på data hjemme, ved behov ved det mulighet å få tilsendt en papirutgave av PRAP i posten (Grotle, 2020).

For metodestudiene skal det inkluderes >50 deltakere, noe som er tilstrekkelig i forhold til COSMIN retningslinjer for studier av test-retest reliabilitet (Mokkink, 2019).

4.4.1 Inklusjonskriterier

- Alder over 55 år.
- Oppsøkt hjelp i primærhelsetjenesten (fastlege, fysioterapeut eller kiropraktor).
- Ny episode med ryggsmerter.
- Har ikke besøkt primærhelsetjenesten for ryggsmerter de siste 6 månedene.

4.4.2 Eksklusjonskriterier

- Nedsatt kognitiv funksjon.
- Problemer med å snakke eller skrive norsk.
- Store utfordringer i forhold til mobilitet (kan ikke møte opp fysisk til klinisk undersøkelse).

4.5 Variabler

For test-retest reliabilitet i denne masteroppgaven vil det bli benyttet kategoriske, nominelle variabler.

Test-retest reliabilitet vil bli undersøkt for de 30 spørsmålene i PRAP. Det er 15 ulike spørsmål som besvares først for smertene i korsryggen og så for smertene i ben(et). Hvert av spørsmålene kan besvares med; ingen smerte, bedre, likt, verre eller ikke aktuelt (vedlegg 1).

Variablene som vil bli benyttet er:

1. Ingen smerte
2. Bedre
3. Uforandret
4. Verre
5. Ikke aktuelt

Disse variablene vil bli benyttet i to ulike analyser.

Det er utviklet en algoritme basert på de ulike symptomene og de fire diagnosene skjemaet screener for (vedlegg 2). Hver deltaker blir screenet til 1 av de 4 diagnosene; skivesykdom, spinal stenose, spinal stenose med skivesykdom og uspesifikke/godartete korsryggsmerter, basert på hvordan de besvarer de ulike spørsmålene.

Test-retest reliabiliteten vil bli undersøkt for diagnostisk enighet, det vil si om pasientene på test og retest blir klassifisert med samme diagnose, eller om dette har endret seg fra test til retest.

Den dikotome utfallsvariablen som vil bli benyttet er:

1. Samme diagnose
2. Endret diagnose

4.6 Datainnsamling

PRAP administreres til alle deltakerne som en del av materialet brukt i BACE-studien. I tillegg til PRAP blir det registrert sykehistorie, demografi (alder, kjønn, utdanning, jobb status) og flere andre spørreskjema (Scheele et al., 2011).

PRAP er et selvadministrerende spørreskjema som ble fylt ut ved bruk av nettbrett, alene i et separat rom. Test-retest reliabilitet ble målt ved å teste PRAP på et mindre utvalg som utførte testen 2 ganger. PRAP ble fylt ut ved baseline, samtidig ble deltakerne spurt om å fylle ut retest på nytt hjemme, inntil en stort nok antall deltakere (>50) hadde fullført retest. Deltakerne ble bedt om å fylle ut retest 2 dager etter baseline testing (Grotle, 2020).

4.7 Analyse

Innhentete data vil bli registrert og analysert i IBM SPSS statistics versjon 28.0.1.

Analysene vil bli gjennomført i henhold til COSMIN retningslinjene, og det vil det bli benyttet uvektet kappa (Mokkink, 2019).

Det vil bli gjennomført tre ulike test-retest analyser;

- Test-retest for alle personene samlet for hvert spørsmål. Kappa vil bli beregnet for hvert enkelt spørsmål.

Basert på disse analysene vil det beregnes en gjennomsnitts kappa for alle personene samlet. Gjennomsnitts kappa vil bli oppgitt med spredning i form av standardavvik.

- Test-retest for alle spørsmålene samlet for hver enkel person. Kappa vil bli beregnet per person.

Basert på disse analysene vil det beregnes en gjennomsnitts kappa for alle personene samlet. Gjennomsnitts kappa vil bli oppgitt med spredning i form av standardavvik.

- Test-retest for diagnostisk enighet mellom test og retest på grunnlag av prediksjonsalgoritmen. Kappa vil bli beregnet for alle personene samlet.

4.8 Klassifisering etter algoritmen

En person har ved hjelp av svarene i PRAP og screeningalgoritmen registret hvilken av de 4 diagnosene deltakerne passer inn i eller om de ikke tilfredsstiller kriteriene for noen av diagnosene. Om pasientene tilfredsstiller kravene til flere av diagnosene har deltakeren blitt registrert i begge/alle diagnosene. Det er i denne prosessen benyttet anonyme data.

4.9 Tolkning av kappa

I denne masteroppgaven vil kappa tolkes fra; svak enighet (0.0 - 0.19), rimelig enighet (0.20 - 0.39), moderat enighet (0.4 - 0.59), betydelig enighet (0.60 - 0.79), nesten perfekt enighet (0.80 - 0.99) til perfekt enighet (1.0). For en tilfredsstillende kappa vil det kreves en moderat enighet, det vil si en kappa fra 0.4 og over (De Vet, 2011; Landis & Koch, 1977).

4.10 Etikk

BACE-N ble klassifisert som en kvalitetssikringsstudie hos Regional Etisk Komite (REK) vest, referansenummer 2014/1634/REK vest, og ble godkjent av Norsk senter for forskningsdata (NSD), referansenummer 42149, i 2015.

For å inkludere et brukerperspektiv har, foruten et stort nettverk av forskere og klinikere, også brukerrepresentanter deltatt i etablering av BACE-N. Brukerrepresentanter vil også delta i tolkning, kommunikasjon og implementeringen av funnene fra studien.

For å sikre høy kvalitet på de publiserte artiklene fra BACE-N er protokollen publisert og registrert i clinicaltrials.gov. Resultatene fra studien vil bli rapportert uavhengig av funn.

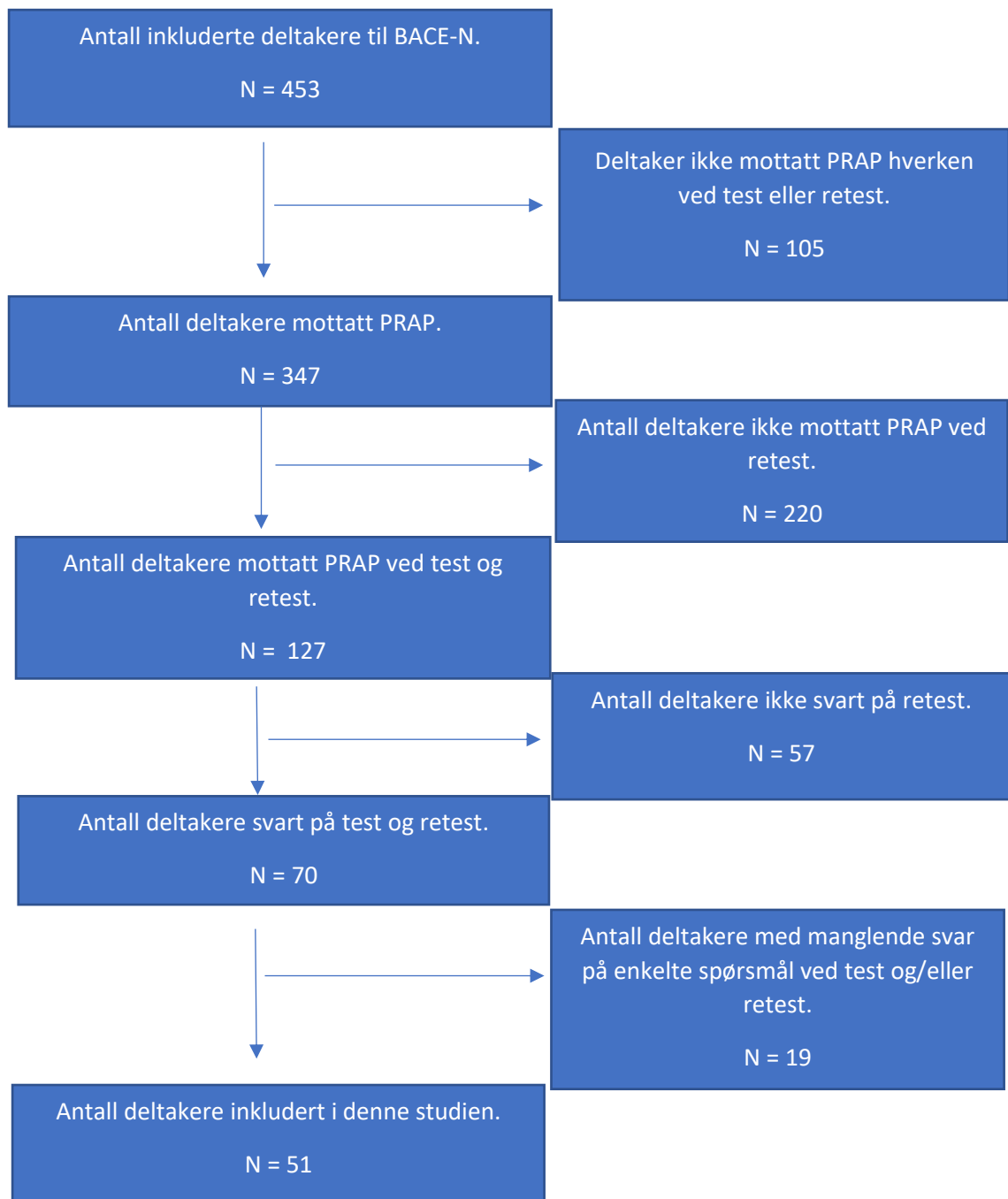
Deltakerne har signert samtykke til deltakelse, og fått informasjon om at de når som helst kan trekke seg fra studien (Grotle, 2020).

5.0 Resultat

BACE-N inkluderte 452 personer i alderen ≥ 55 år som oppsøkte primærhelsetjenesten på grunn av en ny episode med ryggmerter. Deltakerne ble rekruttert i perioden april 2015 til februar 2020 fra almenpraktiserende lege (n = 127), fysioterapeut (n = 130) og kiropraktor (n = 195). Median alder var 66 år (Interkvartil intervall 59 - 72 år). Pasientene hadde moderat til høy grad av funksjonsbegrensninger og de fleste hadde tilbakevendende rygg-episoder (67%).

Av 452 deltakere i BACE-N hadde kun 347 deltakere mottatt PRAP. Alle deltakerne fikk utdelt PRAP ved baseline (test), men kun 127 deltakere (36.6%) fikk utdelt PRAP ved retest.

Av de 127 deltakerne var det 57 (44.9%) som ikke returnerte retest, og ble ekskludert. Totalt 70 pasienter fylte ut både test og retest, av disse manglet 19 deltakere svar på enkelt spørsmål ved test og/eller retest og ble ekskludert. Totalt hadde 51 deltakere besvart alle spørsmålene ved både test og retest, og besvarelsene fra disse 51 er benyttet i analysene (figur 4).



Figur 4: Flytskjema.

Gjennomsnittlig antall dager mellom test og retest var på 3.5 (SD 2.95) (tabell 7). Dager mellom test og retest varierte fra 1 til 13 dager.

Tabell 7: Antall dager mellom test og retest.

	Antall (N)	Lavest	Høyest	Gjennomsnitt	Standardavvik
Antall dager mellom	40	1	13	3.5	2.95

5.1 Demografi

Av de 51 inkluderte deltakerne var 32 kvinner (63%). Deltakernes gjennomsnittsalder var 66 (SD 8.4) med en variasjon fra 53 til 89 år. 80% av deltakerne var 70 år eller yngre, mens 51% var 67 år eller yngre. 38 (75%) av deltakerne var gift og 48 (84%) hadde utdannelse over 12 år. 39 av 51 (76%) av deltakere hadde oppgitt arbeidsstatus og av disse var 22 (56%) i arbeid (tabell 8).

Alle de 51 deltakerne hadde alle (100%) hatt rygg smerter tidligere. Deltakernes gjennomsnittlige smerte siste uken var 5,6 (SD 1.9) målt ved NRS. 26 (51%) av deltakerne hadde hatt smerter i benet siste uke, og gjennomsnittlig antall tilfeller av korsrygg smerter den siste måneden var 3.3 (SD 1,0) (tabell 8).

Tabell 8: Deskriptive data

	Antall (N)	
Alder	51	66.3 (8,4)
Kjønn	51	Menn 19 (37%) Kvinner 32 (63%)
Sivilstatus	51	Gift 38 (74.5%) Enslig 3 (5.9%) Skilt 5 (9.8%) Enke/-mann 5 (9.8%)
Utdannelse over 12 år	50	48 (84%)
Arbeid	39	Ja 17 (43.6%) Nei 22 (56.4%)
Tidligere rygg smerter	51	Ja 51 (100%)
Frekvensen av smerter		
Korsrygg smerter siste mnd	51	3.3 (1.0)
Smerter i bein siste uke	51	Ja 26 (51%) Nei 25 (49%)
NRS siste uke	50	5,6 (1.9)
FABQ	51	9.88 (6.6)
RMDQ	51	9.22 (4.3)
PCS	51	11.4 (8.4)

NRS – Numeric Rating Scale, FABQ - Fear-Avoidance Belief Questionnaire, RMDQ - Roland-Morris Disability Questionnaire, PCS - Pain Catastrophizing Scale.

Gjennomsnittlig skår på frykt unngåelsesadferd målt ved FABQ var 9.88 (SD 6.6).

Funksjonsnedsettelse målt ved RMDQ var på 9.22 (SD 4.3), mens gjennomsnittlig skår på katastrofetanker målt med PCS var på 11.4 (SD 8.4) (Tabell 8).

5.2 Kappa for hvert spørsmål

For hvert enkelt spørsmål varierte kappa fra svak enighet (0-0.19) til betydelig enighet (0.6-0.79). Hovedvekten av svarene (70%) hadde en moderat enighet (0.4-0.59) (tabell 11) (tabell 9). 76.7% av spørsmålene oppnådde en akseptabel kappa, som vil si en moderat enighet eller høyere (tabell 11). Gjennomsnittlig kappa for alle spørsmålene samlet var på 0.449 (SD 0.105) (tabell 10), noe som gav en moderat enighet.

Det er en større variasjon i kappa for spørsmålene om smerter i korsrygg i forhold spørsmålene om smerter i bein. Kappa for spørsmålene for smerter i korsrygg varierte fra 0.144 til 0.626, med et gjennomsnitt på 0.433 (SD 0.124), mens for smerter i bein varierte kappa fra 0.322 til 0.660, med et gjennomsnitt på 0.465 (SD 0.083).

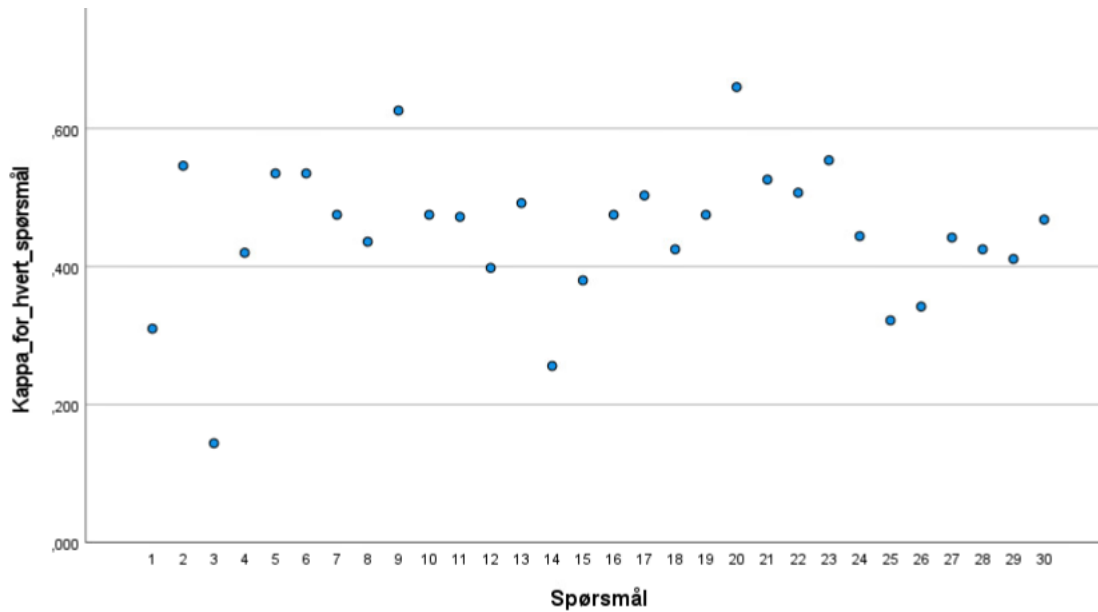
Tabell 9: Kappa for hvert enkelt spørsmål:

Spørsmål	Kappa - rygg Føler du smerte i ryggen når.....	Kappa - ben Føler du smerte i beinet når....
1. Når du står oppreist i mer enn 5 min	0.310	0.475
2. Når du går i et kvartal (ca. 100m) eller mer	0.546	0.503
3. Når du sitter i mer enn 5 minutter	0.144	0.425
4. Når du bøyer deg over vaskeservanten	0.420	0.475
5. Når du kjører bil	0.535	0.660
6. Når du ligger på ryggen	0.535	0.526
7. Når du ligger på siden	0.475	0.507
8. Når du ligger på magen	0.436	0.554
9. Når du hoster eller presser med magen	0.626	0.444
10. Når du står opp om morgenen	0.475	0.322
11. Mot slutten av dagen	0.472	0.342
12. Når du reiser deg fra sittende	0.398	0.442
13. Når du dytter noe tungt som en støvsuger eller gressklipper	0.492	0.425
14. Når du løfter noe tungt opp fra gulvet	0.256	0.411
15. Når du bærer noe tungt	0.380	0.468

Tabell 10: Gjennomsnittlig kappa per spørsmål.

	Antall (N)	Gjennomsnitt	Min	Maks	Standardavvik
Gjennomsnitt per spørsmål	30	0.449	0.144	06.60	0.105

Spørsmål nummer 3. «Når du sitter i mer enn 5 minutter» og spørsmål nummer 14. «Når du løfter noe tungt opp fra gulvet» skilte seg ut fra de resterende spørsmålene ved at det er en relativ større forskjell mellom kappa på svarene for ryggen ($K = 0.144/0.256$) og kappa på svarene for benet ($K = 0.425/0.411$), i forhold til forskjellen mellom kappa på svarene for korsrygg og bein på de resterende spørsmålene.



Figur 5: Scatterplot over kappa for hvert enkelt spørsmål

Tabell 11; Kategorisk kappa for hvert enkelt spørsmål

	Antall	Prosent
Ingen enighet	0	0.0
Svak enighet	1	3.5
Rimelig enighet	6	20.0
Moderat enighet	21	70.0
Betydelig enighet	2	6.7
Nesten perfekt enighet	0	0.0
Perfekt enighet	0	0.0
Total	30	100

5.3 Kappa for hver deltaker

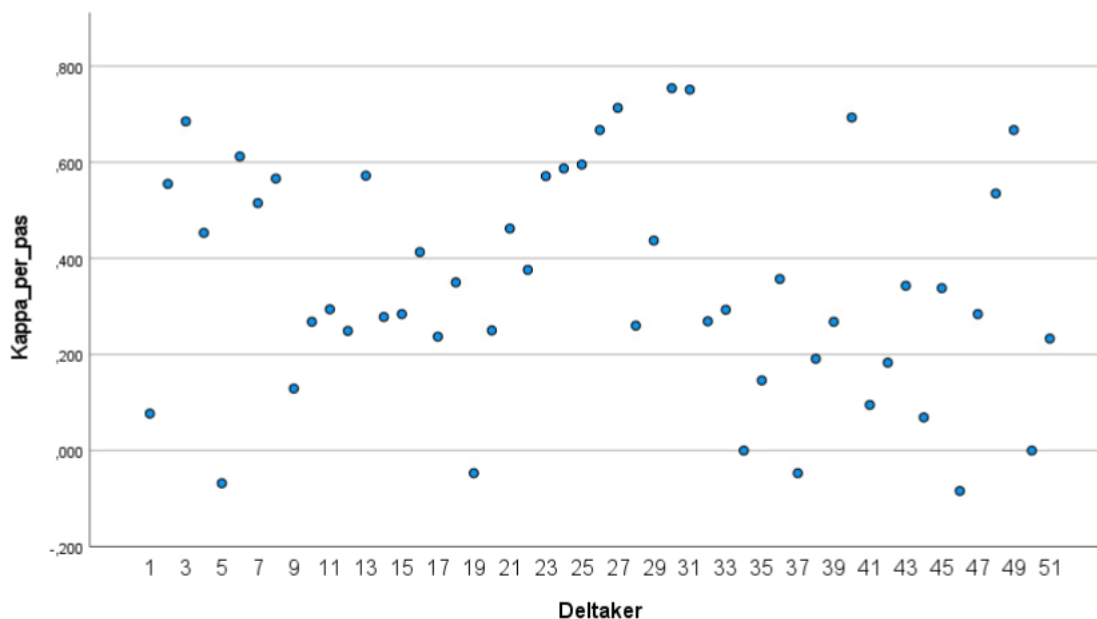
For hver enkelt pasient varierte kappa fra null enighet til betydelig enighet ($K = -0.084 - 0.745$) (tabell 12). Gjennomsnittlig kappa for hver enkelt pasient var 0.347 (SD 0.234) (tabell 13), som vil si rimelig enighet. Kun 40% av deltakerne oppnådde en akseptable kappa, som vil si moderat enighet eller høyere (tabell 14).

Tabell 12: Kappa per pasient:

Pasient	Kappa	Pasient	Kappa
1	0.077	27	0.713
2	0.555	28	0.260
3	0.685	29	0.437
4	0.453	30	0.754
5	-0.068	31	0.751
6	0.612	32	0.269
7	0.515	33	0.293
8	0.566	34	0.000
9	0.129	35	0.146
10	0.268	36	0.357
11	0.294	37	-0.047
12	0.249	38	0.191
13	0.572	39	0.268
14	0.278	40	0.693
15	0.284	41	0.095
16	0.413	42	0.183
17	0.237	43	0.343
18	0.350	44	0.069
19	-0.047	45	0.338
20	0.250	46	-0.084
21	0.462	47	0.284
22	0.376	48	0.535
23	0.571	49	0.667
24	0.587	50	0.000
25	0.595	51	0.233
26	0.667		

Tabell 13; Gjennomsnittlig kappa per deltaker

	Antall (N)	Gjennomsnitt	min	maks	Standardavvik
Gjennomsnitt per person	51	0.347	-0.084	0.754	0.234



Figur 6: Scatterplot over kappa for hver enkelt deltaker.

Tabell 14; Kategorisk kappa per person

	Antall	Prosent
Ingen enighet	4	7.8
Svak enighet	9	17.6
Rimelig enighet	18	35.3
Moderat enighet	12	23.5
Nesten perfekt enighet	8	15.7
Perfekt enighet	0	0.0
Total	51	100

5.4 Variasjon i utvalget

Ved inspeksjon av krysstabeller for både hvert spørsmål og hver deltaker kan det sees en homogenitet i utvalget. Det ser ut til at flertallet av svarene for hvert spørsmål er i 2/3 av de 5 svarkategoriene. Svarene for hver pasient viser en endra større homogenitet enn per spørsmål. Hos enkelte pasienter kan opp mot 100% av svarene være i samme kategori (Vedlegg 3 og 4).

5.5 Klassifisering etter algoritme

Ved klassifisering av pasientene ved både test og retest viste det deg at flertallet av pasientene ikke kunne klassifiseres/diagnostiseres eller at deltakerne passet kriteriene for 2 eller flere

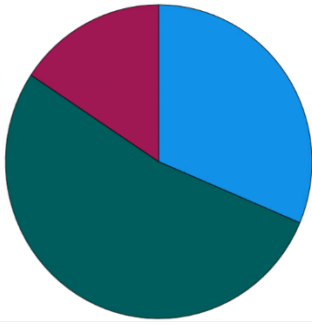
diagnoser (Vedlegg 5 og 6). Ved test tilfredsstilte kun 16 deltakere kriteriene (31.4%) til kun en diagnose, 27 (52.9%) deltakere tilfredsstilte kriteriene til 2 eller flere diagnoser, mens 8 (15.7%) av deltakerne tilfredsstilte ikke kriteriene til noen av de fire diagnosene spørreskjemaet screenet for (tabell 15) (figur 6a og 6b). Ved retest tilfredsstilte 18 (35.3%) av deltakerne kravene til kun en diagnose, 17 (33.3%) tilfredsstilte kravene til 2 eller flere diagnoser, mens 16 (31.4%) deltakere kunne ikke klassifiseres til noen av diagnosene (tabell 15).

Av de 16 som tilfredsstilte kun en diagnose ved test var det bare 6 deltakere som tilfredsstilte kriteriene for samme diagnose ved retest.

På bakgrunn av dette resultatet var det ikke mulig å gjennomføre en test-retest analyse og undersøke den diagnostiske enigheten for screening algoritmen.

Tabell 15; Klassifisering av deltakerne i henhold til diagnoser.

Klassifisering	Test		Retest	
	Antall (N)	Prosent (%)	Antall (N)	Prosent (%)
Kunne klassifiseres til en diagnose	16	31.4	18	35.3
Kunne klassifiseres til 2 eller flere diagnoser	27	52.9	17	33.3
Kunne ikke klassifiseres	8	15.7	16	31.4
Totalt antall	51	100	51	100



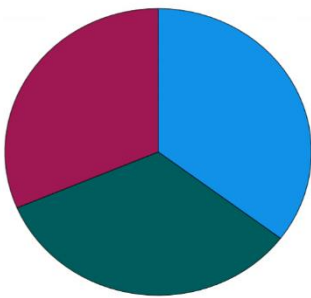
Diagnostisering etter algoritme, ved test:

Grønn: Kunne plasseres i 2 eller flere diagnoser 52.9%.

Blå: Kunne plasseres i en diagnose 31.4%.

Rød: Kunne ikke plasseres i noen av diagnosene 15.7%

Figur 7a: Fordeling av deltakere i forhold til antall diagnoser ved test.



Diagnostisering etter algoritme, ved retest:

Grønn: Kunne plasseres i 2 eller flere diagnoser 33.3%.

Blå: Kunne plasseres i en diagnose 35.3%.

Rød: Kunne ikke plasseres i noen av diagnosene 31.4%.

Figur 7b: Fordeling av deltakere i forhold til antall diagnoser ved retest.

6.0 Diskusjon

PRAP er et screeningskjema som har til hensikt å skille mellom fire ulike typer korsryggsmerter; spinal stenose, skivesykdom, spinal stenose med skivesykdom og uspesifikke/godartete korsryggsmerter (Roach et al. 1997b). Hensikten med en slik screening/diagnostisering er å guide beslutninger i forhold til behandlingen.

PRAP har nylig blitt oversatt til norsk i forbindelse med BACE-N studien. Formålet med denne studien er å undersøke test-retest reliabiliteten av den norske utgaven av PRAP, på et utvalg eldre enn 55 år. Test-retest reliabilitet for PRAP skulle bli undersøkt med tre ulike analyser; for hvert spørsmål, for hver deltaker, og for klassifiseringsalgoritmen. Det ble gjennomført test-retest analyser og oppgitt kappaverdier for hvert enkelt spørsmål og for hver enkelt deltaker, samt gjennomsnittlig kappa med standardavvik for disse to analysene. Det var ikke mulig å gjennomføre den tredje planlagte analysen, da klassifisering av deltakerne til de ulike diagnosene ikke lot seg gjennomføre, fordi flertallet av pasientene kunne klassifiseres til to eller flere diagnoser (52.9%), eller ingen av diagnosene (15.7%).

Gjennomsnittlig kappa for hvert enkelt spørsmål oppnådde moderat enighet, mens gjennomsnittlig kappa for hver enkelt deltaker oppnådde rimelig enighet.

6.1 Diskusjon av metode

En av styrkene ved studien er etterfølgelse av internasjonale retningslinjer for metodestudier (Mokkink, 2019). Det var i denne studien invitert 127 deltakere, det vil si deltakere som hadde mottatt både test og retest. 100 eller flere deltakere regnes som svært godt i henhold til COSMIN retningslinjene, i anbefalingene er det tatt hensyn til forventet frafall og manglende verdier (Mokkink, 2019). Til tross for at det ble invitert over 100 deltakere var det kun mulig å inkludere 51 av disse i analysene. Det var et frafall på 76 på grunn av manglende returnering av retest, samt noen av deltakerne manglet svar på enkeltspørsmål, enten på test eller retest. Til tross for et stor frafall er det fullt mulig å oppnå et rimelig sikkert estimat på reliabilitet med 51 analyserte deltakere (De Vet, 2011). Et antall mellom 50 og 99 deltakere regnes som adekvat i forhold til COSMIN retningslinjene (Mokkink, 2019). Det store frafallet gir derimot fare for skjevhet i utvalget. En analyse av deskriptive data for disse to gruppene kunne eventuelt ha avdekket slike skjevheter (Staff, 2015).

Det er benyttet «listwise delition» i behandlingen av manglende verdier, det vil si at om en invitert deltaker har manglet et enkelt svar har deltakeren blitt ekskludert fra analysene fra alle spørsmålene (Dong & Peng, 2013), noe som bidro til at det kun ble inkludert 51 deltakere i studien. Denne måten å håndtere manglende verdier på kan føre til «selection bias», som «sample bias» eller «lost to follow up bias» (Staff, 2015). I motsetning til en «listwise delition» kunne en «pairwise delition» ha styrket estimatene ytterligere. Ved «pairwise delition» ville deltakere som hadde svart på det aktuelle spørsmålet, men manglet svar på andre spørsmål ikke blitt ekskludert. Det ville uansett være stor fare for ulike bias ved denne måten å håndtere manglende verdier på (Dong & Peng, 2013). Bias i forskningen innebærer at resultatene i en studie er skjeve eller feilaktige ved at de avviker systematisk fra den virkelige populasjonen som undersøkes i studien (Grønmo, 2020). En sammenlikning av deskriptive data for de 127 inviterte deltakerne og de 51 som ble inkludert i analysene kunne ha avdekket mulige systematiske forskjeller mellom de to utvalgene.

En annen styrke ved studien er rekruttering av deltakere fra flere ulike helseaktører i primærhelsetjenesten, både fastleger, fysioterapeuter og kiropraktorer, som jobber med pasienter med plager i korsryggen. Det kan tenkes at det er ulike pasientpopulasjoner som oppsøker de ulike profesjonene slik at inkludering av pasienter fra flere ulike aktører kan styrke studiens eksterne validitet. Noe som også styrker studiets eksterne validitet ytterligere er etableringen av teststasjoner, og rekruttering av leger, fysioterapeuter og kiropraktorer i ulike geografiske rekrutteringsområder. Hensikten med BACE-N studien er å undersøke en populasjon over 55 år, og kun inklusjon av pasienter over 55 år begrenser den eksterne validiteten til denne aldersgruppen.

Det er ingen standard regel for tidsintervall mellom test og retest, men ved utfylling av test og retest er det viktig at avstanden mellom de to testene er lang nok til å unngå at deltakerne husker hva de svarte på første test, men samtidig kort nok til at situasjonen/fenomenet ikke har endret seg (Oladimeji, 2015). I denne studien ble det vurdert at to dager mellom test og retest var adekvat. Det viste seg at den gjennomsnittlige tiden mellom test og retest i studien ble 3.5 dager, og tiden mellom testene varierte fra 1 dag til 13 dager. Den lange tiden mellom test og retest hos enkelte av pasientene kan ha ført til at smertene i ryggen eller beinet har endret seg, og at deltakerne derfor svarte annerledes på retest. Deltakerne i BACE-N studien kunne ikke nektes å søke behandling mellom rekruttering, baseline testing og retesting. Om deltakere har hatt behandling mellom test og retest kan dette ha påvirket smertene og dermed

skåringen av PRAP ved retest. Endret smerteopplevelse kan dermed ha påvirket resultatet og reliabiliteten i negativ retning.

Klassifiseringen av deltakerne i henhold til algoritmen er kun gjennomført av en person (denne forfatteren). Skåring i forhold til algoritmen er standardisert og utformet på en måte som gir lite rom for tolkning, noe som reduserer sjansen for variasjon i klassifiseringen. Personen som har gjennomført klassifiseringen har heller ikke hatt kontakt med deltakerne og har håndtert anonyme data.

PRAP er et selvrapporterende skjema, og deltakerne i denne studien skulle fylle ut spørreskjemaet alene, både ved test og retest, for å redusere påvirkning fra andre personer. Det ble gitt informasjon på forhånd ved første test, og ved utfylling satt deltakerne i et rom alene og fylte ut skjemaet på et nettbrett. Retest hjemme ble fylt ut via data eller et skjema deltakerne fikk i posten. Vi vet at deltakerne ved test ikke har hatt noen påvirkning av andre personer under utfyllingen, noe som gjør det mer sikkert at svarene representerer det deltakerne opplever og ikke tolkninger fra andre personer. Det er derimot mer usikkert hvordan utfyllingen er foretatt hjemme, om dette er gjort alene, eller sammen med andre personer som kan påvirke utfyllingen av skjemaet, og dermed resultatet av studien. Deltakerne har verken under utfylling av test eller retest blitt påvirket av personer involvert i denne studien.

Manglende utfylling av enkeltspørsmål kan indikere at PRAP ikke er gunstig for bruk til utfylling på egenhånd uten oppfølging/veiledning av helsepersonell. Hos de Carvalho et al. (2019) ble det kun registret et frafall på 7 deltakere fra test til retest. Grunnen til dette kan være at PRAP skjemaet ble fylt ut gjennom et intervju av helsepersonell. Ved undersøkelse av test-retest reliabilitet av algoritmen hos Roach et al. (1997a) er det usikkerhet rundt eventuelt frafall. Det er ikke beskrevet frafall, men det kan tenkes at mulighet for veiledning fra helsepersonell underveis og at deltakerne ble kontakert ved retest kan ha begrenset frafall på grunn av manglende utfylling av retest eller manglende utfylling av enkelte spørsmål.

Det er ikke beskrevet hvilken informasjon deltakerne fikk på forhånd og om denne er standardisert slik at alle deltakerne fikk den samme informasjonen. Hvilken informasjon som gis, og om informasjonen er den samme til alle deltakerne, kan påvirke hvordan PRAP fylles ut (Mokkink, 2019). Om deltakerne har fått ulik informasjon før utfylling kan dette ha innvirkning på svarene deltakerne gir, for eksempel om uklarheter er oppklart før utfylling. Det kan tenkes at svarene vil ha større variasjon om det er uklart hvordan skjemaet skal fylles

ut. Informasjon ble bare gitt før test, ikke før retest, slik at hva den enkelte deltaker husker av informasjonen ved retest kan variere. Dette kan for eksempel føre til ulik tolkning av spørsmålene eller kategoriene fra test til retest.

Ved gjennomgang av svarene er det tydelig at noen har misforstått utfyllingen av PRAP skjemaet. Dette gjaldt spesielt avkryssninger på 'ikke aktuelt' som skulle velges dersom deltakeren ikke utførte den aktuelle aktiviteten eller inntok riktig posisjon. Det ser ut som enkelte deltakere har krysset av for 'ikke aktuelt' dersom de ikke hadde smerter i det hele tatt i rygg eller ben. Disse deltakerne har ikke vært mulig å klassifisere i henhold til de fire diagnosene, og dermed gjort det vanskeligere å gjennomføre en test-retest analyse av algoritmen. Om disse pasientene har tolket dette ulikt på test og retest kan også ha svekket estimatene på hvert enkelt spørsmål. Det er mulig at bedre informasjon eller noe veiledning under utfylling kunne forhindre slike misforståelser og muliggjort klassifisering og undersøkelse av test-retest reliabilitet av algoritmen, samt bedret estimatene på hvert enkelt spørsmål.

Det er velkjent av konteksten påvirker smerteopplevelsen (Rossettini et al., 2018), og for å unngå slik påvirkning anbefales det at utførelsen av test og retest er standardisert (Mokkink, 2019). Det er forskjell i konteksten er mellom test og retest i denne studien. Test ble utført på et testsenter sammen med andre skjemaer og en fysisk undersøkelse, mens retest ble fylt ut hjemme. Det er i BACE-N protokollen ikke beskrevet rekkefølgen på undersøkelsene. Det er flere forhold i konteksten som kan ha vært forskjellig fra test og retest, dette kan for eksempel ha vært reisen til/fra testsenter eller interaksjon med personer på testsenteret eller hjemme som kan ha påvirket utfyllingen av PRAP på ulike måter.

6.2 Diskusjon opp mot tidligere forskning

Etter søk i Pubmed ble det kun funnet en tidligere studie som hadde undersøkt reliabilitet hos PRAP, denne undersøkte inter- og intra-tester reliabilitet (de Carvalho et al. 2019). Ved gjennomgang av referanselisten til studien ble det funnet ytterligere to studier som undersøkte metodiske egenskaper hos PRAP. En studie undersøkte test-retest reliabilitet for hvert enkelt spørsmål (Roach et al. 1997a) og en undersøkte test-retest reliabilitet for algoritmen, i tillegg til sensitivitet og positiv/negativ prediktiv verdi (Roach et al. 1997b).

Jeg vil diskutere resultatene i disse tre studiene opp mot resultatene denne masteroppgaven. Da det kun er en studie som har undersøkt test-retest reliabilitet tidligere, har jeg valgt også å diskutere resultatene i denne oppgaven opp mot resultatene hos de Carvalho et al. (2019). Jeg har i diskusjonen benyttet resultatene fra interrater reliabilitetsanalysen.

Resultatene i denne studien skiller seg noe fra de tidligere studiene ved å ha noe lavere test-retest reliabilitet, og ved at det ikke var mulig å klassifisere pasientene og gjennomføre en reliabilitetstest av klassifiseringsalgoritmen (de Carvalho, 2019; Roach et al. 1997b; Roach et al. 1997a).

Sammen med de Carvalho et al. (2019) skiller denne oppgaven seg fra en tidligere test-retest studie på PRAP ved at det er undersøkt test-retest reliabilitet på en eldre populasjon (Roach et al. 1997a). Studiet til Roach et al (1997a) har undersøkt reliabiliteten til en generell ryggpopulasjon, og har derfor en større variasjon i alder enn denne studien. Til tross for dette er det ikke så stor forskjell mellom gjennomsnittlig alder mellom denne studien og Roach et al (1997a) (66.3/54.2). Gjennomsnittlig alder hos de Carvalho et al. (2019) var derimot noe høyere (71.1).

Det er svært vanlig med korsryggsmerter blant befolkningen generelt (Lærum, 2013) og alle deltakerne i denne studien oppgav å ha hatt ryggsmerter tidligere. I Roach et al. (1997a) var det derimot rundt 90% som hadde hatt ryggsmerter tidligere. Det var et betydelig større antall som oppga smerter i bein hos Roach et al. (1997a) (86%) i forhold til i denne studien (51%). Dette er noe som kan forklares med et noe yngre utvalg og større spredning (20-85 år) hos Roach et al. (1997a).

6.2.1 Hvert enkelt spørsmål

Resultatene viser varierende test-retest reliabilitet for hvert enkelt spørsmål, fra svak til betydelig enighet ($K = 0.144 - 0.626$), men hovedvekten (70%) av spørsmålene oppnådde en moderat enighet ($K = 0.40 - 0.59$). Totalt oppnådde 23 av 30 spørsmål (76.7%) en moderat enighet ($K = 0.40$) eller høyere. Den gjennomsnittlige kappaverdien for alle spørsmålene samlet oppnådde også en moderat enighet ($K = 0.449$).

To spørsmål oppnådde betydelig enighet ($K = 0.60 - 0.79$); «Føler du smerte i korsryggen når du hoster eller presser med magen» og «Føler du smerte i benet når du kjører bil», med en kappaverdi på henholdsvis $K = 0.626$ og $K = 0.660$. Fem spørsmål; «Føler du smerte i

korsryggen når du står oppreist mer enn 5 minutter», «Føler du smerte i korsryggen når du reiser deg fra sittende», «Føler du smerte i korsryggen når du bærer noe tungt», «Føler du smert i beinet når du står opp om morgenen» og «Føler du smerte i beinet mot slutten av dagen» tilfredsstilte ikke kravet til en moderat enighet. To spørsmål; «Føler du smerter i korsryggen når du sitter i mer enn 5 minuttet» og «Føler du smerte i korsryggen når du løfter noe tungt opp fra gulvet» oppnådde de laveste kappaverdiene med henholdsvis svak og rimelig enighet ($K = 0.144$ og $K = 0.256$). Disse to spørsmålene, «... når du sitter mer enn 5 minutter» og «... når du løfter noe opp fra gulvet» skilte seg også ut ved at det var betydelig høyere kappa verdi på bensmerter ($K = 0.425$ og $K = 0.411$) i forhold til korsryggmerter ($K = 0.114$ og $K = 0.256$), og oppnådde tilfredsstillende kappa for bensmerter, men ikke for korsryggmerter. Resultatene viste generelt noe bedre test-retest reliabilitet for spørsmålene i forhold til bensmerter enn i forhold til korsryggmerter, gjennomsnittlig kappa for både spørsmålene om smerter i bein og korsrygg oppnådde en tilfredsstillende kappa ($K=0.465$ og $K= 0.433$).

Roach et al. (1997a) viste mindre variasjon og høyere kappaverdier for hvert enkelt spørsmål i forhold til denne studien. Mens denne studien viste en variasjon i kappa fra 0.144 (svak enighet) til 0.660 (betydelig enighet), viste Roach et al (1997a) en variasjon fra 0.46 (moderat enighet) til 0.89 (nesten perfekt enighet) på de ulike spørsmålene. I likhet med Roach et al. (1997a) hadde spørsmålene om bensmerter, med en gjennomsnittlig kappa på 0.465, noe høyere test-retest reliabilitet i forhold til korsryggmerter, med en kappa på 0.433. Det ble ikke oppgitt gjennomsnittlig kappa i Roach et al. (1997a) sin studie, men tallene viste generelt høyere skår på hvert av de 30 spørsmålene. En analyse i SPSS av oppgitte tall viste en gjennomsnittlig kappa på 0.743 (SD 0.089) noe som er betydelig høyere enn i denne studien med en gjennomsnittlig kappa på 0.449 (SD 0.105).

De Carvalho et al. (2019) hadde derimot en større variasjon, fra rimelig enighet til perfekt enighet, i resultatene for intraraterreliabiliteten av den brasilianske utgaven av PRAP, noe som er mer på linje med det som er funnet i denne masteroppgaven, der kappaverdiene varierte fra 0.23 til 1.0 for de ulike spørsmålene. Det blir vanskelig å sammenlikne resultatene til de Carvalho et al. (2019) med denne studien da det hverken er oppgitt gjennomsnittlig kappa eller kappa for hvert enkelt spørsmål.

6.2.2 Hver enkelt pasient

Ingen av de andre studiene som har undersøkt test-retest reliabilitet på PRAP per deltaker (de Carvalho et al. 2019; Roach et al. 1997a).

Resultatene for test-retest reliabilitet per deltaker i denne studien viser betydelig større variasjon, fra ingen enighet ($K = 0$ og under) til nesten perfekt enighet ($K = 0.80 - 0.99$) sammenlignet med reliabiliteten for hvert enkelt spørsmål som varierte fra svak enighet ($K = 0 - 0.19$) til betydelig enighet ($K = 0.60 - 0.79$). Gjennomsnittlig kappa per deltaker viste seg å ha rimelig enighet ($K = 0.449$) som er lavere enn per spørsmål, som viste moderat enighet ($K = 0.347$).

Kun 20 deltakere (39.2%) oppnådde en moderat eller høyere enighet, med en kappa på 0.40 eller over, som er akseptabel. De resterende 31 deltakerne (60.8%) hadde lavere kappa enn 0.4, og ikke tilfredsstillende resultat. Hele 6 deltakere (6.8%) har lavere eller lik enighet enn ved tilfeldighet. Basert på disse resultatene ser det ut som det er større variasjon for hver deltaker enn for hvert enkelt spørsmål. Den store variasjonen i kappa per deltaker stiller spørsmålstegn ved hvilken nytte skjemaet vil ha i klinikken. Hele ¼ del av deltakerne oppnådde en svært lav kappaverdi. Ved bruk av PRAP i klinikken vil det derfor være en usikkerhet om pasientens skår er en «sann» skår eller ikke.

6.2.3 Variasjon i utvalget

Kappa-verdier alene er ikke så informativt, og det er derfor anbefalt at krysstabellene blir presentert sammen med kappa-verdien(e). Krysstabellene viser en eventuell heterogenitet eller homogenitet i utvalget som kan påvirke kappaverdiene. Ved hjelp krysstabellene kan en se fordelingen av variablene og vurdere om dette kan ha påvirket kappaverdiene (De Vet, 2011). Kappaverdiene for både hvert enkelt spørsmål og hver pasient viste en viss grad av homogenitet i besvarelsene, og betydelig større for hver pasient. Dette kan muligens forklare de noe lavere kappaverdiene per pasient i forhold til hvert enkelt spørsmål.

Det er mulig at forskjeller i kappaverdier kan forklares av forskjeller i homo-/heterogeniteten i det ulike utvalgene. Det er derimot vanskelig å vite om dette kan forklare forskjellene mellom resultatene i denne masteroppgaven og de to tidligere studiene til Roach et al. (1997a og b) og de Carvalho et al. (2019), da disse studiene ikke publiserte krysstabeller.

6.2.4 Algoritmen

Det ble valgt å ikke gjennomføre test-retest-analyse for prediktiv enighet i denne studien, da det viste seg at kun 31.4% og 35.3 % av deltakerne kunne kategoriseres til en diagnose ved henholdsvis test og retest og kun 6 av 51 deltakere hadde en og samme diagnose ved både test og retest. Roach et al. (1997b) kunne derimot kategorisere 99 av 106 deltakerne i 1 av 4 kategorier ved test (baseline). Det er derimot ikke oppgitt tall på klassifiseringen ved retest, men det ble gjennomført en test-retest analyse av algoritmen på bakgrunn av en klassifisering. Test-retest reliabilitet for algoritmen varierte fra moderat til nesten perfekt enighet med $K=0.57$ for uspesifikke/godartete korsrygg smerter, $K=0.89$ for skivesykdom, $K=0.58$ for spinal og $K=0.91$ for spinal stenose med skivesykdom (Roach et al. 1997b). En klassifisering var også mulig hos de Carvalho et al. (2019), og de fant moderat enighet ($K=0.52$) for intrarater reliabilitet for algoritmen.

PRAP er i utgangspunktet et selvrappporterende skjema som pasientene/deltakerne skal fylle ut på egenhånd. En viktig forskjell mellom denne studien og Roach et al. (1997b) og de Carvalho et al. (2019) er muligheten til å veilede deltakerne underveis i utfyllingen av skjemaet. I BACE-N studien fylte deltakerne ut spørreskjemaet på egenhånd/alene både ved test og retest. Dette for å redusere påvirkning fra andre personer. I studien til Roach et al. (1997b) var det derimot vært mulighet for påvirkning fra helsepersonell, gjennom veiledning under utfylling, mens hos de Carvelho et al. (2019) ble skjemaet fylt ut av helsepersonell gjennom et intervju, noe som også gir mulighet til påvirkning. Gjennom påvirkning/oppklaring av misforståelser og uklarheter under utfylling kan svarene ha vist en «tydeligere» profil, og dermed lettere å klassifisere.

Det kan se ut som om misforståelser i utfyllingen kan ha påvirket muligheten for å klassifisere pasientene til de fire ulike diagnosene. Det kan se ut som enkelte deltakelse har misforstått hvordan de skulle tolke svaralternativet «ikke aktuelt». Dette skulle benyttes dersom deltakeren ikke kunne gjennomføre den aktuelle aktiviteten eller innta den aktuelle posisjonen. Det kan se ut som enkelte deltakere har brukt dette svaralternativet på alle spørsmålene relatert til smerter i bein, når de ikke har hatt smerter i beina, mens riktig utfyllelse hadde vært ingen smerter. Dette vil ikke nødvendigvis påvirke test-retest reliabiliteten for hvert enkelt spørsmål, men kan gi vanskeligheter med å klassifisere pasientene til 1 av de 4 diagnosene, og dermed føre til problemer med å benytte algoritmen til

klassifisering. Slike og andre misforståelser kan muligens unngås/redueres ved veiledning under utfylling av skjemaet, som hos Roach et al. (1997b) og de Carvalho et al. (2019).

Det er ikke screenet for kognitiv funksjon, men deltakere med tydelig nedsatt kognitiv funksjon, som hindret dem å gjennomføre/fylle ut spørreskjemaene, ble ekskludert fra studien. De Carvalho et al. (2019) benyttet Mental State Mini-Exam (MMES) for å screene og utelukke deltakere med kognitiv svikt. Det er mulig at bruk av MMES eller liknende screeningskjema i denne studien ville ha plukket opp og ekskludert flere deltakere med kognitiv svikt som kan ha hatt utfordringer med å fylle ut PRAP.

6.4 Klinisk relevans

PRAP er et selv rapporterende spørreskjema som har som mål å; screene og klassifisere/diagnostisere pasienter til fire ulike typer korsryggsmerter for deretter å kunne styre behandlingen i en retning. En viktig del av PRAP er å kunne klassifisere hver enkelt pasient til de ulike diagnosene for korsryggsmerter og om dette ikke er mulig vil dens kliniske relevans reduseres drastisk. I denne studien viste det seg å være betydelig større variasjon i kappa mellom hver enkelt deltaker som vil kunne påvirke spørreskjemaets kliniske nytteverdi.

6.5 Videre forskning

Både denne studien og tidligere forskning har vist akseptabel test-retest reliabilitet for hvert enkelt spørsmål i PRAP. Det er derimot første gang test-retest reliabilitet er undersøkt for hver deltaker, og resultatene her viste svak test-retest reliabilitet og stor variasjon mellom deltakerne.

I motsetning til ved denne studien har det ved to tidligere studier vært mulig å klassifisere flertallet av deltakerne til de ulike diagnosene. Algoritmen har vist moderat intratester reliabilitet for hele algoritmen (de Carvalho et al. 2019), og fra moderat til høy test-retest reliabilitet for de enkelte diagnosene (Roach et al. 1997b). En forskjell mellom denne og de to andre studiene er veiledning underveis i utfyllingen, det kan se ut som dette er sentralt for å få svar som gjør det mulig å klassifisere pasientene i henhold til algoritmen.

For å kunne benytte PRAP i en klinisk setting, bør test-retest reliabilitet per deltaker og algoritmen undersøkes ytterligere.

7.0 Konklusjon

Både denne og tidligere studier (de Carvalho et al. 2019; Roach et al. 1997b) har vist varierende test-retest reliabilitet for de ulike spørsmålene i PRAP, med noe lavere verdier i denne studien. Enkelte spørsmål skilte seg ut med spesielt lave test-retest verdier, men 70% av spørsmålene oppnådde moderat enighet, som i denne studien regnes som et tilfredsstillende resultat for test-retest reliabiliteten. Det var en betydelig variasjon i test-retest reliabiliteten til hver enkelt deltaker fra ingen enighet til nesten perfekt enighet, og kun 39% av pasientene oppnådde akseptabel kappaverdi.

Da hovedmålet med PRAP er å klassifisere pasientene til fire ulike typer korsryggsmerte er det avgjørende at algoritmen og svarene fra PRAP kan benyttes til nettopp dette. Basert på denne og tidligere studier kan de se ut som det blir vanskelig å diagnostisere pasientene om de ikke får veiledning underveis ved utfylling (de Carvalho et al.2019; Roach et al. 1997b).

Referanseliste:

- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*, 25(24), 3186-3191. <https://doi.org/10.1097/00007632-200012150-00014>
- Benzakour, T., Igoumenou, V., Mavrogenis, A. F., & Benzakour, A. (2019). Current concepts for lumbar disc herniation. *Int Orthop*, 43(4), 841-851. <https://doi.org/10.1007/s00264-018-4247-6>
- Carter, R. a. L., J. (2015). *Rehabilitation Research; Principles and Applications* (Vol. 5). Saunders.
- COSMIN. (2022). *About the COSMIN initiative*. COSMIN. Retrieved 24.09 from
- de Carvalho, G. R., de Oliveira, E. A, Rocha, V. T. M., Pereira, D. S. and Pereira, S. M. (2019). Cross-cultural adaption and reliability of the pain response to activity and position questionnaire. *Advances in Rheumatology*, 4. <https://doi.org/doi.org/10.1186/s42358-019-0098-0>
- De Vet, H. C. W., Terwee, C.B., Mokkink, L.B & Knol,D.L. (2011). *Measurement in medicine*. Cambige Universety Press.
- Deer, T., Sayed, D., Michels, J., Josephson, Y., Li, S., & Calodney, A. K. (2019). A Review of Lumbar Spinal Stenosis with Intermittent Neurogenic Claudication: Disease and Diagnosis. *Pain Med*, 20(Suppl 2), S32-S44. <https://doi.org/10.1093/pm/pnz161>
- Dionne, C. E., Dunn, K. M., & Croft, P. R. (2006). Does back pain prevalence really decrease with increasing age? A systematic review. *Age Ageing*, 35(3), 229-234. <https://doi.org/10.1093/ageing/afj055>
- Docking, R. E., Fleming, J., Brayne, C., Zhao, J., Macfarlane, G. J., Jones, G. T., & Cambridge City over-75s Cohort Study, c. (2011). Epidemiology of back pain in older adults: prevalence and risk factors for back pain onset. *Rheumatology (Oxford)*, 50(9), 1645-1653. <https://doi.org/10.1093/rheumatology/ker175>
- Dong, Y., & Peng, C. Y. (2013). Principled missing data methods for researchers. *Springerplus*, 2(1), 222. <https://doi.org/10.1186/2193-1801-2-222>
- Elfving, B. o. L., D. (2015). *Reliabilitet - Reproducerbarhet och Vulkan.se*.
- Fairbank, J. C., & Pynsent, P. B. (2000). The Oswestry Disability Index. *Spine (Phila Pa 1976)*, 25(22), 2940-2952; discussion 2952. <https://doi.org/10.1097/00007632-200011150-00017>
- Ferreira, M. L., & de Luca, K. (2017). Spinal pain and its impact on older people. *Best Pract Res Clin Rheumatol*, 31(2), 192-202. <https://doi.org/10.1016/j.berh.2017.08.006>
- Grotle, M. (2020, 29. mars 2021). *BACK pain in Elders in Norway (BACE-N)*.
- Grotle, M., Brox, J. I., Veierod, M. B., Glomsrod, B., Lonn, J. H., & Vollestad, N. K. (2005). Clinical course and prognostic factors in acute low back pain: patients consulting primary care for the first time. *Spine (Phila Pa 1976)*, 30(8), 976-982. <https://doi.org/10.1097/01.brs.0000158972.34102.6f>
- Grønmo, S. (2020). *Bias i forskning*. Store Norske Leksikon. Retrieved 24.09.22 from
- Hartvigsen, J., Hancock, M. J., Kongsted, A., Louw, Q., Ferreira, M. L., Genevay, S., Hoy, D., Karppinen, J., Pransky, G., Sieper, J., Smeets, R. J., Underwood, M., & Lancet Low Back Pain Series Working, G. (2018). What low back pain is and why we need to pay attention. *Lancet*, 391(10137), 2356-2367. [https://doi.org/10.1016/S0140-6736\(18\)30480-X](https://doi.org/10.1016/S0140-6736(18)30480-X)
- Hoy, D., Brooks, P., Blyth, F., & Buchbinder, R. (2010). The Epidemiology of low back pain. *Best Pract Res Clin Rheumatol*, 24(6), 769-781. <https://doi.org/10.1016/j.berh.2010.10.002>
- Jensen, R. K., Harhangi, B. S., Huygen, F., & Koes, B. (2021). Lumbar spinal stenosis. *BMJ*, 373, n1581. <https://doi.org/10.1136/bmj.n1581>
- Kent, P. M., & Keating, J. L. (2008). Can we predict poor recovery from recent-onset nonspecific low back pain? A systematic review. *Man Ther*, 13(1), 12-28. <https://doi.org/10.1016/j.math.2007.05.009>

- Krismer, M. a. v. T., M. (2007). Low back pain (non-specific). *Best Practice & Research Clinical Rheumatology*, vol 21, 15. <https://doi.org/10.1016/j.berh.2006.08.004>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. <https://www.ncbi.nlm.nih.gov/pubmed/843571>
- Linton, S. J., & Hallden, K. (1998). Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain*, 14(3), 209-215. <https://doi.org/10.1097/00002508-199809000-00007>
- Lærum, E., Brox, J.I., Storheim, K., Espeland, A., Haldorsen, E., Munch-Ellingsen, J., Nielsen, L-L., Rossvoll, I., Skouen, J.S., Stig, L-C. og Werner, E. L. (2007). *Nasjonale kliniske retningslinjer for korsryggsmerter - med og uten nerverotsaffeksjon*.
- Lærum, E., Søren, B., Ihlebæk, C., Johnsen, K., Natvig, B og Aas, E. (2013). Et muskel- og skjelettregnskap - Forekomst og kostnader knyttet til skader, sykdommer og plager i muskel- og skejlett systemet. In F. f. m.-o. s. (FORMI) (Ed.), (2. utgave ed.). Muskel og Skjelett Tiåret (MST).
- Maher, C., Underwood, M., & Buchbinder, R. (2017). Non-specific low back pain. *Lancet*, 389(10070), 736-747. [https://doi.org/10.1016/S0140-6736\(16\)30970-9](https://doi.org/10.1016/S0140-6736(16)30970-9)
- Mokkink, L. B., Prinsen, C. A. C., Patrick, D., Alonso, J., Bouter, L. M., de Vet, H. C. W. and Terwee, C. B. (2019). COSMIN Study Design checklist for Patient-reported outcome measurement instruments.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*, 19(4), 539-549. <https://doi.org/10.1007/s11136-010-9606-8>
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010b). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*, 63(7), 737-745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
- Oladimeji, A. B. (2015). Principles and Methods og Validity and Reliability Testing og queationnaires used in Social and Health Science Reserches. *Nigerian Postgraduate Medical Journal*, 22(4), 195-201.
- Ostelo, R. W., & de Vet, H. C. (2005). Clinically important outcomes in low back pain. *Best Pract Res Clin Rheumatol*, 19(4), 593-607. <https://doi.org/10.1016/j.berh.2005.03.003>
- Paeck, T., Ferreira, M. L., Sun, C., Lin, C. W., Tiedemann, A., & Maher, C. G. (2014). Are older adults missing from low back pain clinical trials? A systematic review and meta-analysis. *Arthritis Care Res (Hoboken)*, 66(8), 1220-1226. <https://doi.org/10.1002/acr.22261>
- Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*, 56(1), 45-50. <https://doi.org/10.4103/0301-4738.37595>
- Roach, K. E., Brown, M. D., Albin, R. D., Delaney, K. G., Lipprandi, H. M., & Rangelli, D. (1997). The sensitivity and specificity of pain response to activity and position in categorizing patients with low back pain. *Phys Ther*, 77(7), 730-738. <https://doi.org/10.1093/ptj/77.7.730>
- Roach, K. E., Brown, M. D., Dunigan, K. M., Kusek, C. L., & Walas, M. (1997). Test-retest reliability of patient reports of low back pain. *J Orthop Sports Phys Ther*, 26(5), 253-259. <https://doi.org/10.2519/jospt.1997.26.5.253>
- Rossettini, G., Carlino, E., & Testa, M. (2018). Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. *BMC Musculoskelet Disord*, 19(1), 27. <https://doi.org/10.1186/s12891-018-1943-8>
- Scheele, J., Luijsterburg, P. A., Ferreira, M. L., Maher, C. G., Pereira, L., Peul, W. C., van Tulder, M. W., Bohnen, A. M., Berger, M. Y., Bierma-Zeinstra, S. M., & Koes, B. W. (2011). Back complaints in the elders (BACE); design of cohort studies in primary care: an international consortium. *BMC Musculoskelet Disord*, 12, 193. <https://doi.org/10.1186/1471-2474-12-193>
- Staff, A. (2015). *Bias*. De Nasjonale forskningsetiske komiteene. Retrieved 4.09 from

- Werner, E. L., & Indahl, A. (2005). [Knowledge, practice and attitudes to back pain among doctors, physiotherapists and chiropractors]. *Tidsskr Nor Laegeforen*, 125(13), 1794-1797.
<https://www.ncbi.nlm.nih.gov/pubmed/16012543> (Kunnskap, praksis og holdninger til rygglidelser hos leger, fysioterapeuter og kiropraktorer.)
- Wu, A., March, L., Zheng, X., Huang, J., Wang, X., Zhao, J., Blyth, F. M., Smith, E., Buchbinder, R., & Hoy, D. (2020). Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. *Annals of translational medicine*, 8(6).

Vedlegg:

Vedlegg 1: PRAP

Smertereaksjon ved aktivitet og kroppstilling

Føler du smerte i ryggen...	Ingen smerte	Bedre	Det samme	Verre	Ikke aktuelt
1. Når du står oppreist i mer enn 5 minutter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Når du går ett kvartal (ca. 100m) eller mer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Når du sitter i mer enn 5 minutter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Når du bøyer deg over vaskeservanten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Når du kjører bil	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Når du ligger på ryggen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Når du ligger på siden	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Når du ligger på magen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Når du hoster eller presser med magen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Når du står opp om morgenen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Mot slutten av dagen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Når du reiser deg fra sittende	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Når du dytter noe tungt som en støvsuger eller gressklipper	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Når du løfter noe tungt opp fra gulvet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Når du bærer noe tungt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Føler du smerte i beinet...	Ingen smerte	Bedre	Det samme	Verre	Ikke aktuelt
1. Når du står oppreist i mer enn 5 minutter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Når du går ett kvartal (ca. 100m) eller mer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Når du sitter i mer enn 5 minutter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Når du bøyer deg over vaskeservanten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Når du kjører bil	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Når du ligger på ryggen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Når du ligger på siden	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Når du ligger på magen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Når du hoster eller presser med magen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Når du står opp om morgenen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Mot slutten av dagen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Når du reiser deg fra sittende	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Når du dytter noe tungt som en støvsuger eller gressklipper	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Når du løfter noe tungt opp fra gulvet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Når du bærer noe tungt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Pain Response to Activity and Position Questionnaire (PRAP) etter Roach KE et al. [Phys Ther.](#) 1997 Jul;77(7):730-8: The sensitivity and specificity of pain response to activity and position in categorizing patients with low backpain. Oversatt av Grotle M og Munk R 2014, HiOA.

Smertereaksjon ved aktivitet og kroppsstilling

Algoritme for klassifisering av korsryggsmerter

Kriterier for klassifisering av spinal stenose

Hvis minst en av disse fem symptomene er til stede

1. Ryggsmertene forverres når du står oppreist i mer enn 5 minutter
2. Ryggsmertene forverres når du går ett kvartal (ca. 100m) eller mer
3. Smertene i benet forverres når du bærer tunge ting
4. Smertene i benet forverres når du dytter tunge ting
5. Smertene i benet forverres når du står oppreist i mer enn 5 minutter

Og hvis minst en av disse to symptomene er til stede

1. Ryggsmertene er fraværende, bedre eller uforandret når du sitter i mer enn 5 minutter
2. Smertene i benet er fraværende, bedre eller uforandret når du sitter mer enn 5 minutter

Og hvis minst en av disse to symptomene er til stede

1. Smertene i benet er fraværende, bedre eller uforandret når du står opp om morgenen
2. Smertene i benet er fraværende, bedre eller uforandret om kvelden

Da plasseres tilstanden i kategorien spinal stenose

Kriterier for klassifisering av smerter relatert til mellomvirvelskive

Hvis minst en av disse fire symptomene er til stede

1. Ryggsmertene forverres når du kjører bil
2. Ryggsmertene forverres når du sitter mer enn 5 minutter
3. Smertene i benet forverres når du kjører bil
4. Smertene i benet forverres når du sitter mer enn 5 minutter

Og hvis minst en av disse tre symptomene er til stede

1. Smertene i benet er fraværende, bedre eller uforandret når du dytter tunge ting
2. Smertene i benet er fraværende, bedre eller uforandret når du står oppreist i mer enn 5 minutter
3. Smertene i benet er fraværende, bedre eller uforandret når du ligger på magen

Da plasseres tilstanden i kategorien smerter relatert til mellomvirvelskive

Kriterier for klassifisering av smerter relatert til mellomvirvelskive og spinal stenose

Hvis minst en av disse tre symptomene er til stede

1. Ryggsmertene forverres når du dytter tunge ting
2. Smertene i benet forverres når du bærer tunge ting
3. Smertene i benet forverres når du står oppreist i mer enn 5 minutter

Og hvis minst en av disse to symptomene er til stede

1. Ryggsmertene forverres når du står opp om morgenen
2. Smertene i benet forverres når du kjører bil

Da plasseres tilstanden i kategorien smerter relatert til mellomvirvelskive og spinal stenose

Kriterier for klassifisering av godartede korsryggsmerter

Hvis minst en av disse to symptomene er til stede

1. Ryggsmertene forverres når du sitter i mer enn 5 minutter
2. Ryggsmertene forverres når du går ett kvartal (ca. 100m) eller mer

Og hvis minst en av disse tre symptomene er til stede

1. Ryggsmertene er fraværende, bedre eller uforandret når du endrer stilling fra sittende til stående
2. Smertene i benet er fraværende, bedre eller uforandret når du endrer stilling fra sittende til stående
3. Smertene i benet er fraværende, bedre eller uforandret når du går ett kvartal (ca. 100m) eller mer

Og hvis minst en av disse tre symptomene er til stede

1. Smertene i benet er fraværende, bedre eller uforandret når du sitter i mer enn 5 minutter
2. Smertene i benet er fraværende, bedre eller uforandret når du løfter tunge ting
3. Smertene i benet er fraværende, bedre eller uforandret når du bøyer deg framover

Da plasseres tilstanden i kategorien for godartede korsryggsmerter

Vedlegg 3: Krysstabeller per spørsmål

1. Når du står oppreist i mer enn 5 minutter * RE_PRAP_rygg_1. Når du står oppreist i mer enn 5 minutter Crosstabulation

Count

		RE_PRAP_rygg_1. Når du står oppreist i mer enn 5 minutter					Total
		1	2	3	4	5	
1. Når du står oppreist i mer enn 5 minutter	Ingen smerte	4	3	1	0	0	8
	Bedre	2	1	5	1	0	9
	Det samme	1	3	11	1	0	16
	Verre	1	2	4	9	2	18
Total		8	9	21	11	2	51

2. Når du går ett kvartal (ca. 100m) eller mer * RE_PRAP_rygg_2. Når du går ett kvartal (ca. 100m) eller mer Crosstabulation

Count

		RE_PRAP_rygg_2. Når du går ett kvartal (ca. 100m) eller mer					Total
		1	2	3	4	5	
2. Når du går ett kvartal (ca. 100m) eller mer	Ingen smerte	10	1	1	0	0	12
	Bedre	2	8	3	0	1	14
	Det samme	2	3	12	1	0	18
	Verre	0	1	1	4	0	6
	Ikke aktuelt	0	0	0	1	0	1
Total		14	13	17	6	1	51

3. Når du sitter i mer enn 5 minutter * RE_PRAP_rygg_3. Når du sitter i mer enn 5 minutter Crosstabulation

Count

		RE_PRAP_rygg_3. Når du sitter i mer enn 5 minutter					Total
		1	2	3	4	5	
3. Når du sitter i mer enn 5 minutter	Ingen smerte	6	2	2	0	0	10
	Bedre	3	2	2	0	0	7
	Det samme	1	4	5	3	1	14
	Verre	1	1	12	5	0	19
	Ikke aktuelt	0	0	1	0	0	1
Total		11	9	22	8	1	51

4. Når du bøyer deg over vaskeservanten * RE_PRAP_rygg_4. Når du bøyer deg over vaskeservanten Crosstabulation

Count

		RE_PRAP_rygg_4. Når du bøyer deg over vaskeservanten					Total
		1	2	3	4	5	
4. Når du bøyer deg over vaskeservanten	Ingen smerte	5	2	2	1	0	10
	Bedre	0	1	0	0	0	1
	Det samme	3	2	8	6	0	19
	Verre	0	0	4	14	1	19
	Ikke aktuelt	0	0	0	0	2	2
Total		8	5	14	21	3	51

5. Når du kjører bil * RE_PRAP_rygg_5. Når du kjører bil Crosstabulation

Count

		RE_PRAP_rygg_5. Når du kjører bil					Total
		1	2	3	4	5	
5. Når du kjører bil	Ingen smerte	13	1	2	1	0	17
	Bedre	1	2	2	0	0	5
	Det samme	2	2	13	4	0	21
	Verre	1	0	1	2	0	4
	Ikke aktuelt	0	0	0	0	4	4
Total		17	5	18	7	4	51

6. Når du ligger på ryggen * RE_PRAP_rygg_6. Når du ligger på ryggen Crosstabulation

Count

		RE_PRAP_rygg_6. Når du ligger på ryggen					Total
		1	2	3	4	5	
6. Når du ligger på ryggen	Ingen smerte	13	3	1	0	0	17
	Bedre	4	6	3	0	0	13
	Det samme	0	1	5	2	0	8
	Verre	1	1	2	6	0	10
	Ikke aktuelt	0	0	0	0	3	3
Total		18	11	11	8	3	51

7. Når du ligger på siden * RE_PRAP_rygg_7. Når du ligger på siden Crosstabulation

Count

		RE_PRAP_rygg_7. Når du ligger på siden					Total
		1	2	3	4	5	
7. Når du ligger på siden	Ingen smerte	13	1	0	1	0	15
	Bedre	4	6	4	1	0	15
	Det samme	1	4	6	1	0	12
	Verre	1	0	2	5	0	8
	Ikke aktuelt	0	0	0	0	1	1
Total		19	11	12	8	1	51

8. Når du ligger på magen * RE_PRAP_rygg_8. Når du ligger på magen Crosstabulation

Count

		RE_PRAP_rygg_8. Når du ligger på magen					Total
		1	2	3	4	5	
8. Når du ligger på magen	Ingen smerte	8	2	0	0	0	10
	Bedre	5	4	0	0	0	9
	Det samme	2	1	4	3	1	11
	Verre	0	0	7	3	1	11
	Ikke aktuelt	0	0	0	1	9	10
Total		15	7	11	7	11	51

9. Når du hoster eller presser med magen * RE_PRAP_rygg_9. Når du hoster eller presser med magen Crosstabulation

Count

		RE_PRAP_rygg_9. Når du hoster eller presser med magen				Total
		1	3	4	5	
9. Når du hoster eller presser med magen	Ingen smerte	23	2	0	0	25
	Det samme	4	6	3	0	13
	Verre	0	2	9	0	11
	Ikke aktuelt	1	0	0	1	2
Total		28	10	12	1	51

10. Når du står opp om morgenen * RE_PRAP_rygg_10. Når du står opp om morgenen Crosstabulation

Count

		RE_PRAP_rygg_10. Når du står opp om morgenen					Total
		1	2	3	4	5	
10. Når du står opp om morgenen	Ingen smerte	2	4	0	0	0	6
	Bedre	0	2	0	1	0	3
	Det samme	3	2	12	2	0	19
	Verre	2	2	3	15	0	22
	Ikke aktuelt	0	0	0	0	1	1
Total		7	10	15	18	1	51

11. Mot slutten av dagen * RE_PRAP_rygg_11. Mot slutten av dagen Crosstabulation

Count

		RE_PRAP_rygg_11. Mot slutten av dagen				Total
		1	2	3	4	
11. Mot slutten av dagen	Ingen smerte	4	0	0	1	5
	Bedre	0	3	2	1	6
	Det samme	3	1	9	5	18
	Verre	0	1	4	17	22
Total		7	5	15	24	51

12. Når du reiser deg fra sittende * RE_PRAP_rygg_12. Når du reiser deg fra sittende Crosstabulation

Count

		RE_PRAP_rygg_12. Når du reiser deg fra sittende					Total
		1	2	3	4	5	
12. Når du reiser deg fra sittende	Ingen smerte	4	1	1	1	0	7
	Bedre	0	0	2	0	0	2
	Det samme	1	0	11	3	0	15
	Verre	2	1	7	16	1	27
Total		7	2	21	20	1	51

13. Når du dytter noe tungt som en støvsuger eller gressklipper * RE_PRAP_rygg_13. Når du dytter noe tungt som en støvsuger eller gressklipper Crosstabulation

Count

		RE_PRAP_rygg_13. Når du dytter noe tungt som en støvsuger eller gressklipper					Total
		1	2	3	4	5	
13. Når du dytter noe tungt som en støvsuger eller gressklipper	Ingen smerte	4	0	0	0	0	4
	Det samme	1	1	10	8	0	20
	Verre	0	0	3	18	1	22
	Ikke aktuelt	1	1	0	1	2	5
Total		6	2	13	27	3	51

14. Når du løfter noe tungt opp fra gulvet * RE_PRAP_rygg_14. Når du løfter noe tungt opp fra gulvet Crosstabulation

Count

		RE_PRAP_rygg_14. Når du løfter noe tungt opp fra gulvet					Total
		1	2	3	4	5	
14. Når du løfter noe tungt opp fra gulvet	Ingen smerte	3	0	0	0	0	3
	Bedre	0	1	0	0	0	1
	Det samme	2	1	6	7	0	16
	Verre	2	1	5	14	3	25
	Ikke aktuelt	1	1	1	2	1	6
Total		8	4	12	23	4	51

15. Når du bærer noe tungt * RE_PRAP_rygg_15. Når du bærer noe tungt Crosstabulation

Count

		RE_PRAP_rygg_15. Når du bærer noe tungt					Total
		1	2	3	4	5	
15. Når du bærer noe tungt	Ingen smerte	3	0	0	1	0	4
	Bedre	0	1	0	0	0	1
	Det samme	2	0	4	4	0	10
	Verre	3	0	3	20	2	28
	Ikke aktuelt	0	1	0	4	3	8
Total		8	2	7	29	5	51

1. Når du står oppreist i mer enn 5 minutter * RE_PRAP_UE_1. Når du står oppreist i mer enn 5 minutter Crosstabulation

Count

		RE_PRAP_UE_1. Når du står oppreist i mer enn 5 minutter					Total
		1	2	3	4	5	
1. Når du står oppreist i mer enn 5 minutter	Ingen smerte	15	0	2	1	2	20
	Bedre	2	2	2	0	0	6
	Det samme	3	0	10	0	0	13
	Verre	2	0	4	3	0	9
	Ikke aktuelt	1	0	0	0	2	3
Total		23	2	18	4	4	51

2. Når du går ett kvartal (ca. 100m) eller mer * RE_PRAP_UE_2. Når du går ett kvartal (ca. 100m) eller mer Crosstabulation

Count

		RE_PRAP_UE_2. Når du går ett kvartal (ca. 100m) eller mer					Total
		1	2	3	4	5	
2. Når du går ett kvartal (ca. 100m) eller mer	Ingen smerte	14	0	1	1	1	17
	Bedre	3	5	2	0	0	10
	Det samme	2	0	8	1	0	11
	Verre	1	3	3	3	0	10
	Ikke aktuelt	1	0	0	0	2	3
Total		21	8	14	5	3	51

3. Når du sitter i mer enn 5 minutter * RE_PRAP_UE_3. Når du sitter i mer enn 5 minutter Crosstabulation

Count

		RE_PRAP_UE_3. Når du sitter i mer enn 5 minutter				Total
		1	3	4	5	
3. Når du sitter i mer enn 5 minutter	Ingen smerte	16	3	0	1	20
	Bedre	2	1	0	0	3
	Det samme	5	9	0	1	15
	Verre	0	6	5	0	11
	Ikke aktuelt	1	0	0	1	2
Total		24	19	5	3	51

4. Når du bøyer deg over vaskeservanten * RE_PRAP_UE_4. Når du bøyer deg over vaskeservanten Crosstabulation

Count

		RE_PRAP_UE_4. Når du bøyer deg over vaskeservanten					Total
		1	2	3	4	5	
4. Når du bøyer deg over vaskeservanten	Ingen smerte	17	1	4	0	1	23
	Bedre	1	0	1	0	0	2
	Det samme	3	0	10	1	0	14
	Verre	3	0	0	3	1	7
	Ikke aktuelt	1	0	1	0	3	5
Total		25	1	16	4	5	51

5. Når du kjører bil * RE_PRAP_UE_5. Når du kjører bil Crosstabulation

Count

		RE_PRAP_UE_5. Når du kjører bil					Total
		1	2	3	4	5	
5. Når du kjører bil	Ingen smerte	19	0	2	0	1	22
	Bedre	2	1	1	0	0	4
	Det samme	2	0	10	1	0	13
	Verre	0	0	1	5	0	6
	Ikke aktuelt	2	0	0	0	4	6
Total		25	1	14	6	5	51

6. Når du ligger på ryggen * RE_PRAP_UE_6. Når du ligger på ryggen Crosstabulation

Count

		RE_PRAP_UE_6. Når du ligger på ryggen					Total
		1	2	3	4	5	
6. Når du ligger på ryggen	Ingen smerte	19	0	1	0	1	21
	Bedre	6	3	3	0	0	12
	Det samme	2	0	5	0	0	7
	Verre	1	1	1	3	0	6
	Ikke aktuelt	1	0	0	0	4	5
Total		29	4	10	3	5	51

7. Når du ligger på siden * RE_PRAP_UE_7. Når du ligger på siden Crosstabulation

Count

		RE_PRAP_UE_7. Når du ligger på siden					Total
		1	2	3	4	5	
7. Når du ligger på siden	Ingen smerte	18	0	1	1	1	21
	Bedre	4	3	2	0	0	9
	Det samme	2	1	4	3	0	10
	Verre	0	0	2	6	0	8
	Ikke aktuelt	1	0	0	0	2	3
Total		25	4	9	10	3	51

8. Når du ligger på magen * RE_PRAP_UE_8. Når du ligger på magen Crosstabulation

Count

		RE_PRAP_UE_8. Når du ligger på magen					Total
		1	2	3	4	5	
8. Når du ligger på magen	Ingen smerte	15	0	1	0	1	17
	Bedre	5	2	3	0	0	10
	Det samme	2	1	5	0	0	8
	Verre	0	0	2	2	0	4
	Ikke aktuelt	1	0	1	0	10	12
Total		23	3	12	2	11	51

9. Når du hoster eller presser med magen * RE_PRAP_UE_9. Når du hoster eller presser med magen Crosstabulation

Count

		RE_PRAP_UE_9. Når du hoster eller presser med magen				Total
		1	3	4	5	
9. Når du hoster eller presser med magen	Ingen smerte	26	2	0	1	29
	Bedre	2	0	0	0	2
	Det samme	4	6	1	0	11
	Verre	1	3	1	0	5
	Ikke aktuelt	1	0	1	2	4
Total		34	11	3	3	51

10. Når du står opp om morgenen * RE_PRAP_UE_10. Når du står opp om morgenen Crosstabulation

Count

		RE_PRAP_UE_10. Når du står opp om morgenen					Total
		1	2	3	4	5	
10. Når du står opp om morgenen	Ingen smerte	12	2	6	1	1	22
	Bedre	1	0	2	0	0	3
	Det samme	4	1	5	1	0	11
	Verre	0	0	5	7	0	12
	Ikke aktuelt	1	0	0	0	2	3
Total		18	3	18	9	3	51

11. Mot slutten av dagen * RE_PRAP_UE_11. Mot slutten av dagen Crosstabulation

Count

		RE_PRAP_UE_11. Mot slutten av dagen					Total
		1	2	3	4	5	
11. Mot slutten av dagen	Ingen smerte	12	0	3	2	1	18
	Bedre	0	1	6	0	0	7
	Det samme	3	0	5	3	0	11
	Verre	1	0	5	6	0	12
	Ikke aktuelt	1	0	0	0	2	3
Total		17	1	19	11	3	51

12. Når du reiser deg fra sittende * RE_PRAP_UE_12. Når du reiser deg fra sittende Crosstabulation

Count

		RE_PRAP_UE_12. Når du reiser deg fra sittende					Total
		1	2	3	4	5	
12. Når du reiser deg fra sittende	Ingen smerte	12	0	3	2	1	18
	Bedre	0	2	3	0	0	5
	Det samme	4	1	6	1	0	12
	Verre	1	0	4	8	0	13
	Ikke aktuelt	1	0	0	0	2	3
Total		18	3	16	11	3	51

13. Når du dytter noe tungt som en støvsuger eller gressklipper * RE_PRAP_UE_13. Når du dytter noe tungt som en støvsuger eller gressklipper Crosstabulation

Count

		RE_PRAP_UE_13. Når du dytter noe tungt som en støvsuger eller gressklipper					Total
		1	2	3	4	5	
13. Når du dytter noe tungt som en støvsuger eller gressklipper	Ingen smerte	11	0	3	1	2	17
	Bedre	0	0	1	0	0	1
	Det samme	4	0	12	2	0	18
	Verre	0	0	2	5	1	8
	Ikke aktuelt	3	1	0	1	2	7
Total		18	1	18	9	5	51

14. Når du løfter noe tungt opp fra gulvet * RE_PRAP_UE_14. Når du løfter noe tungt opp fra gulvet Crosstabulation

Count

		RE_PRAP_UE_14. Når du løfter noe tungt opp fra gulvet					Total
		1	2	3	4	5	
14. Når du løfter noe tungt opp fra gulvet	Ingen smerte	13	0	1	1	2	17
	Bedre	0	0	1	0	0	1
	Det samme	5	0	7	2	0	14
	Verre	0	0	2	7	1	10
	Ikke aktuelt	3	1	2	1	2	9
Total		21	1	13	11	5	51

15. Når du bærer noe tungt * RE_PRAP_UE_15. Når du bærer noe tungt Crosstabulation

Count

		RE_PRAP_UE_15. Når du bærer noe tungt				Total
		1	3	4	5	
15. Når du bærer noe tungt	Ingen smerte	14	0	1	2	17
	Bedre	0	1	0	0	1
	Det samme	3	6	3	1	13
	Verre	0	2	8	1	11
	Ikke aktuelt	2	2	2	3	9
Total		19	11	14	7	51

Vedlegg 4: Krysstabeller per pasient

Pas_1_test * Pas_1_retest Crosstabulation

Count

		Pas_1_retest			Total
		1,00	3,00	4,00	
Pas_1_test	1,00	6	0	0	6
	2,00	7	0	0	7
	3,00	10	1	0	11
	4,00	5	0	1	6
Total		28	1	1	30

Pas_2_test * Pas_2_retest Crosstabulation

Count

		Pas_2_retest				Total
		1,00	2,00	3,00	4,00	
Pas_2_test	1,00	2	0	1	0	3
	2,00	0	2	2	0	4
	3,00	0	2	6	1	9
	4,00	0	1	2	11	14
Total		2	5	11	12	30

Pas_3_test * Pas_3_retest Crosstabulation

Count

		Pas_3_retest				Total
		1,00	3,00	4,00	5,00	
Pas_3_test	1,00	15	2	0	0	17
	3,00	0	1	2	0	3
	4,00	0	0	5	1	6
	5,00	0	0	1	3	4
Total		15	3	8	4	30

Pas_4_test * Pas_4_retest Crosstabulation

Count

		Pas_4_retest				Total
		1,00	2,00	3,00	4,00	
Pas_4_test	1,00	5	1	0	0	6
	2,00	6	5	1	0	12
	3,00	0	1	7	0	8
	4,00	0	0	3	1	4
Total		11	7	11	1	30

Pas_5_test * Pas_5_retest Crosstabulation

Count

		Pas_5_retest			Total
		1,00	2,00	3,00	
Pas_5_test	2,00	1	0	10	11
	3,00	1	1	7	9
	4,00	0	0	10	10
Total		2	1	27	30

Pas_6_test * Pas_6_retest Crosstabulation

Count

		Pas_6_retest			Total
		1,00	3,00	4,00	
Pas_6_test	1,00	19	1	0	20
	3,00	0	3	3	6
	4,00	0	2	2	4
Total		19	6	5	30

Pas_7_test * Pas_7_retest Crosstabulation

Count

		Pas_7_retest				Total
		1,00	2,00	3,00	4,00	
Pas_7_test	2,00	4	3	1	0	8
	3,00	0	1	6	1	8
	4,00	0	0	3	11	14
Total		4	4	10	12	30

Pas_8_test * Pas_8_retest Crosstabulation

Count

		Pas_8_retest				Total
		1,00	2,00	3,00	4,00	
Pas_8_test	1,00	5	6	3	0	14
	2,00	0	1	0	0	1
	3,00	0	0	6	0	6
	4,00	0	0	1	8	9
Total		5	7	10	8	30

Pas_9_test * Pas_9_retest Crosstabulation

Count

		Pas_9_retest				Total
		1,00	3,00	4,00	5,00	
Pas_9_test	1,00	4	4	1	0	9
	2,00	1	1	0	0	2
	3,00	4	2	0	1	7
	4,00	1	4	2	2	9
	5,00	1	0	0	2	3
Total		11	11	3	5	30

Pas_10_test * Pas_10_retest Crosstabulation

Count

		Pas_10_retest			Total
		2,00	3,00	4,00	
Pas_10_test	1,00	0	12	3	15
	2,00	2	0	0	2
	3,00	0	1	1	2
	4,00	1	0	10	11
Total		3	13	14	30

Pas_11_test * Pas_11_retest Crosstabulation

Count

		Pas_11_retest					Total
		1,00	2,00	3,00	4,00	5,00	
Pas_11_test	1,00	4	0	0	0	0	4
	2,00	6	0	0	0	0	6
	3,00	0	0	6	2	0	8
	4,00	0	0	3	1	2	6
	5,00	0	4	0	0	2	6
Total		10	4	9	3	4	30

Pas_12_test * Pas_12_retest Crosstabulation

Count

		Pas_12_retest			Total
		1,00	3,00	4,00	
Pas_12_test	2,00	0	1	0	1
	3,00	0	3	3	6
	4,00	0	0	7	7
	5,00	15	0	1	16
Total		15	4	11	30

Pas_13_test * Pas_13_retest Crosstabulation

Count

		Pas_13_retest					Total
		1,00	2,00	3,00	4,00	5,00	
Pas_13_test	1,00	16	0	2	0	0	18
	2,00	1	1	4	0	0	6
	3,00	0	0	0	1	0	1
	4,00	0	0	0	4	0	4
	5,00	0	0	0	0	1	1
Total		17	1	6	5	1	30

Pas_14_test * Pas_14_retest Crosstabulation

Count

		Pas_14_retest				Total
		1,00	2,00	3,00	4,00	
Pas_14_test	1,00	11	1	2	2	16
	3,00	0	0	1	0	1
	4,00	0	1	3	3	7
	5,00	2	0	0	4	6
Total		13	2	6	9	30

Pas_15_test * Pas_15_retest Crosstabulation

Count

		Pas_15_retest				Total
		1,00	2,00	3,00	4,00	
Pas_15_test	1,00	9	3	1	1	14
	2,00	2	1	0	0	3
	3,00	3	0	0	2	5
	4,00	2	0	0	6	8
Total		16	4	1	9	30

Pas_16_test * Pas_16_retest Crosstabulation

Count

		Pas_16_retest				Total
		1,00	2,00	3,00	4,00	
Pas_16_test	2,00	0	2	3	0	5
	3,00	2	0	5	4	11
	4,00	0	1	1	12	14
Total		2	3	9	16	30

Pas_17_test * Pas_17_retest Crosstabulation

Count

		Pas_17_retest			Total
		3,00	4,00	5,00	
Pas_17_test	2,00	5	0	0	5
	3,00	9	0	0	9
	4,00	11	3	0	14
	5,00	0	0	2	2
Total		25	3	2	30

Pas_18_test * Pas_18_retest Crosstabulation

Count

		Pas_18_retest			Total
		2,00	3,00	4,00	
Pas_18_test	2,00	4	3	1	8
	3,00	0	2	2	4
	4,00	0	6	12	18
Total		4	11	15	30

Pas_19_test * Pas_19_retest Crosstabulation

Count

		Pas_19_retest		Total
		1,00	2,00	
Pas_19_test	1,00	24	2	26
	4,00	4	0	4
Total		28	2	30

Pas_20_test * Pas_20_retest Crosstabulation

Count

		Pas_20_retest		Total
		1,00	4,00	
Pas_20_test	1,00	24	0	24
	4,00	3	1	4
	5,00	2	0	2
Total		29	1	30

Pas_21_test * Pas_21_retest Crosstabulation

Count

		Pas_21_retest				Total
		2,00	3,00	4,00	5,00	
Pas_21_test	3,00	1	10	3	0	14
	4,00	0	2	8	0	10
	5,00	0	3	1	2	6
Total		1	15	12	2	30

Pas_22_test * Pas_22_retest Crosstabulation

Count

		Pas_22_retest					Total
		1,00	2,00	3,00	4,00	5,00	
Pas_22_test	1,00	4	1	2	2	0	9
	2,00	0	0	1	0	0	1
	3,00	1	0	6	5	0	12
	4,00	0	0	1	3	0	4
	5,00	1	0	0	0	3	4
Total		6	1	10	10	3	30

Pas_23_test * Pas_23_retest Crosstabulation

Count

		Pas_23_retest				Total
		2,00	3,00	4,00	5,00	
Pas_23_test	2,00	1	3	1	0	5
	3,00	0	3	1	0	4
	4,00	0	2	11	0	13
	5,00	0	1	1	6	8
Total		1	9	14	6	30

Pas_24_test * Pas_24_retest Crosstabulation

Count

		Pas_24_retest				Total
		1,00	2,00	3,00	4,00	
Pas_24_test	1,00	13	0	1	1	15
	2,00	1	1	0	0	2
	3,00	2	0	3	3	8
	4,00	0	0	0	5	5
Total		16	1	4	9	30

Pas_25_test * Pas_25_retest Crosstabulation

Count

		Pas_25_retest				Total
		1,00	2,00	3,00	4,00	
Pas_25_test	1,00	5	0	0	0	5
	2,00	0	4	3	0	7
	3,00	2	0	3	4	9
	4,00	0	0	0	9	9
Total		7	4	6	13	30

Pas_26_test * Pas_26_retest Crosstabulation

Count

		Pas_26_retest				Total
		1,00	2,00	3,00	4,00	
Pas_26_test	1,00	1	0	0	0	1
	2,00	0	8	3	0	11
	3,00	1	0	7	1	9
	4,00	0	0	2	7	9
Total		2	8	12	8	30

Pas_27_test * Pas_27_retest Crosstabulation

Count

		Pas_27_retest			Total
		2,00	3,00	4,00	
Pas_27_test	2,00	3	0	0	3
	3,00	1	19	2	22
	4,00	0	1	4	5
Total		4	20	6	30

Pas_28_test * Pas_28_retest Crosstabulation

Count

		Pas_28_retest				Total
		1,00	3,00	4,00	5,00	
Pas_28_test	2,00	2	1	1	0	4
	3,00	14	1	0	0	15
	4,00	0	1	8	0	9
	5,00	0	0	0	2	2
Total		16	3	9	2	30

Pas_29_test * Pas_29_retest Crosstabulation

Count

		Pas_29_retest			Total
		1,00	3,00	4,00	
Pas_29_test	1,00	17	4	0	21
	2,00	3	0	0	3
	3,00	0	5	1	6
Total		20	9	1	30

Pas_30_test * Pas_30_retest Crosstabulation

Count

		Pas_30_retest					Total
		1,00	2,00	3,00	4,00	5,00	
Pas_30_test	2,00	1	2	0	0	0	3
	3,00	0	3	3	1	0	7
	4,00	0	0	0	5	0	5
	5,00	0	0	0	0	15	15
Total		1	5	3	6	15	30

Pas_31_test * Pas_31_retest Crosstabulation

Count

		Pas_31_retest			Total
		1,00	3,00	4,00	
Pas_31_test	1,00	17	2	0	19
	2,00	1	0	0	1
	3,00	0	1	0	1
	4,00	0	1	8	9
Total		18	4	8	30

Pas_32_test * Pas_32_retest Crosstabulation

Count

		Pas_32_retest					Total
		1,00	2,00	3,00	4,00	5,00	
Pas_32_test	1,00	8	0	0	0	0	8
	2,00	4	3	2	0	0	9
	3,00	4	1	0	2	0	7
	4,00	0	0	0	2	4	6
Total		16	4	2	4	4	30

Pas_33_test * Pas_33_retest Crosstabulation

Count

		Pas_33_retest				Total
		1,00	2,00	3,00	4,00	
Pas_33_test	1,00	2	0	0	0	2
	2,00	3	3	0	0	6
	3,00	5	1	4	1	11
	4,00	3	0	0	3	6
	5,00	3	0	0	2	5
Total		16	4	4	6	30

Pas_34_test * Pas_34_retest Crosstabulation

Count

		Pas_34_retest	
		3,00	Total
Pas_34_test	3,00	23	23
	4,00	6	6
	5,00	1	1
Total		30	30

Pas_35_test * Pas_35_retest Crosstabulation

Count

		Pas_35_retest			Total
		3,00	4,00	5,00	
Pas_35_test	2,00	1	0	0	1
	3,00	11	9	0	20
	4,00	4	3	0	7
	5,00	0	0	2	2
Total		16	12	2	30

Pas_36_test * Pas_36_retest Crosstabulation

Count

		Pas_36_retest		Total
		1,00	4,00	
Pas_36_test	1,00	22	0	22
	2,00	1	0	1
	3,00	5	0	5
	4,00	0	2	2
Total		28	2	30

Pas_37_test * Pas_37_retest Crosstabulation

Count

		Pas_37_retest		Total
		1,00	4,00	
Pas_37_test	1,00	27	2	29
	4,00	1	0	1
Total		28	2	30

Pas_38_test * Pas_38_retest Crosstabulation

Count

		Pas_38_retest		Total
		2,00	3,00	
Pas_38_test	2,00	1	1	2
	3,00	2	18	20
	4,00	3	5	8
Total		6	24	30

Pas_39_test * Pas_39_retest Crosstabulation

Count

		Pas_39_retest				Total
		1,00	2,00	3,00	4,00	
Pas_39_test	1,00	17	0	0	0	17
	2,00	3	0	0	0	3
	3,00	5	2	1	0	8
	4,00	0	0	1	1	2
Total		25	2	2	1	30

Pas_40_test * Pas_40_retest Crosstabulation

Count

		Pas_40_retest				Total
		2,00	3,00	4,00	5,00	
Pas_40_test	1,00	0	1	0	0	1
	2,00	3	1	0	0	4
	3,00	1	13	1	1	16
	4,00	0	0	7	0	7
	5,00	0	1	0	1	2
Total		4	16	8	2	30

Pas_41_test * Pas_41_retest Crosstabulation

Count

		Pas_41_retest			Total
		1,00	2,00	3,00	
Pas_41_test	1,00	2	1	1	4
	2,00	0	1	1	2
	3,00	0	6	6	12
	4,00	0	5	4	9
	5,00	0	1	2	3
Total		2	14	14	30

Pas_42_test * Pas_42_retest Crosstabulation

Count

		Pas_42_retest		Total
		4,00	5,00	
Pas_42_test	2,00	0	1	1
	3,00	0	2	2
	4,00	1	4	5
	5,00	0	22	22
Total		1	29	30

Pas_43_test * Pas_43_retest Crosstabulation

Count

		Pas_43_retest				Total
		1,00	2,00	3,00	4,00	
Pas_43_test	1,00	4	1	0	0	5
	2,00	1	0	0	0	1
	3,00	5	0	1	0	6
	4,00	3	0	3	12	18
Total		13	1	4	12	30

Pas_44_test * Pas_44_retest Crosstabulation

Count

		Pas_44_retest				Total
		2,00	3,00	4,00	5,00	
Pas_44_test	1,00	1	0	0	15	16
	2,00	1	3	0	0	4
	3,00	0	3	1	0	4
	4,00	1	5	0	0	6
Total		3	11	1	15	30

Pas_45_test * Pas_45_retest Crosstabulation

Count

		Pas_45_retest				Total
		1,00	2,00	4,00	5,00	
Pas_45_test	1,00	9	1	2	1	13
	2,00	1	1	1	0	3
	3,00	0	0	3	0	3
	4,00	0	1	1	1	3
	5,00	0	0	4	4	8
Total		10	3	11	6	30

Pas_46_test * Pas_46_retest Crosstabulation

Count

		Pas_46_retest			Total
		1,00	3,00	4,00	
Pas_46_test	3,00	9	8	0	17
	4,00	2	10	1	13
Total		11	18	1	30

Pas_47_test * Pas_47_retest Crosstabulation

Count

		Pas_47_retest			Total
		1,00	3,00	4,00	
Pas_47_test	1,00	2	4	0	6
	2,00	0	1	0	1
	3,00	0	6	9	15
	4,00	0	0	8	8
Total		2	11	17	30

Pas_48_test * Pas_48_retest Crosstabulation

Count

		Pas_48_retest			Total
		2,00	3,00	4,00	
Pas_48_test	2,00	3	0	0	3
	3,00	1	6	3	10
	4,00	0	4	13	17
Total		4	10	16	30

Pas_49_test * Pas_49_retest Crosstabulation

Count

		Pas_49_retest					Total
		1,00	2,00	3,00	4,00	5,00	
Pas_49_test	1,00	14	0	0	0	0	14
	3,00	1	5	3	0	0	9
	4,00	0	0	1	4	0	5
	5,00	0	0	0	0	2	2
Total		15	5	4	4	2	30

Pas_50_test * Pas_50_retest Crosstabulation

Count

		Pas_50_retest		Total
		1,00	2,00	
Pas_50_test	1,00	27	3	30
Total		27	3	30

Pas_51_test * Pas_51_retest Crosstabulation

Count

		Pas_51_retest					Total
		1,00	2,00	3,00	4,00	5,00	
Pas_51_test	1,00	13	2	2	3	3	23
	3,00	2	0	0	0	0	2
	4,00	0	0	0	3	1	4
	5,00	0	0	0	0	1	1
Total		15	2	2	6	5	30

Vedlegg 5: Klassifisering ved test.

	Spinal stenose	Smerter relatert til mellomvirvelskive	Smerter relatert til mellomvirvelskive og spinal stenose	Godartede korsryggsmerter	Kunne ikke klassifiseres
1	x	x			
2		x	x		
3	x				
4	x				
5		x	x	x	
6	x				
7	x		x	x	
8	x			x	
9		x		x	
10	x		x		
11		x			
12					x
13			x		
14	x		x		
15	x			x	
16	x	x	x	x	
17	x		x		
18			x		
19	x	x			
20			x		
21	x	x	x		
22	x				
23		x	x		
24	x				
25	x	x	x	x	
26		x		x	
27		x		x	
28		x		x	
29			x		
30					x
31		x	x	x	
32	x		x		
33					x
34					x
35		x			
36	x				
37					x
38	x				
39	x				
40		x	x		
41		x		x	
42					x
43	x	x		x	
44		x		x	

45	x				
46		x	x	x	
47		x			
48		x	x		
49					x
50					x
51		x		x	

Vedlegg 6: Klassifisering ved retest.

	Spinal stenose	Smerter relatert til mellomvirvelskive	Smerter relatert til mellomvirvelskive og spinal stenose	Godartede korsryggsmerter	Kunne ikke klassifiseres
1		x			
2		x	x	x	
3	x	x		x	
4					x
5					x
6	x			x	
7	x	x	x		
8	x			x	
9	x				
10	x	x	x	x	
11					x
12	x	x		x	
13			x		
14	x				
15	x			x	
16	x		x		
17					x
18	x		x		
19	x				
20					x
21		x			
22	x				
23	x	x	x		
24	x				
25		x	x		
26	x				
27					x
28	x	x		x	
29					x
30					x
31			x		
32			x		
33					x
34					x
35		x	x	x	
36	x				
37		x			
38					x
39	x				
40		x	x		
41					x
42					x
43	x				
44					x
45	x		x		
46					x
47		x	x		

48			x		
49					x
50					x
51		x			