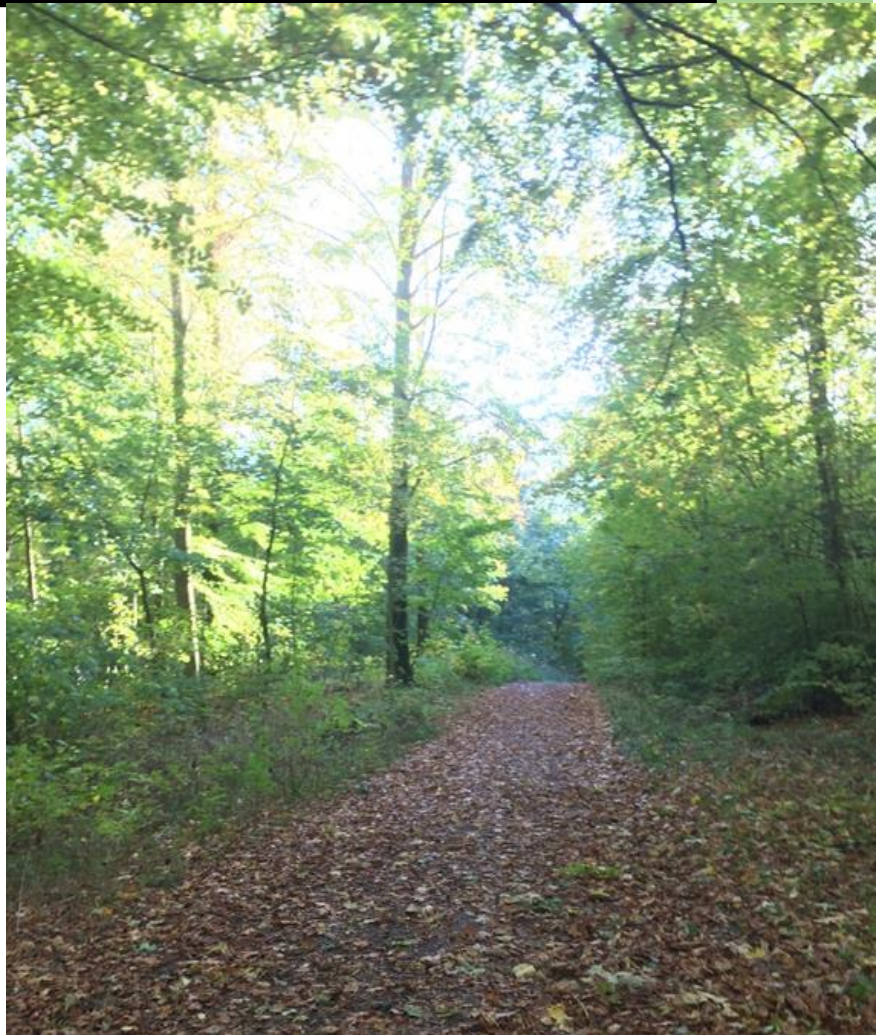


2019

Tampa Skala for Kinesiofobi: Test-retest studie af den danske version på kroniske smertepatienter



Master i smertevidenskab og
tværfaglig smertebehandling

Bettina Eiger, Mette Errebo og
Zerrin Sirek

Vejleder: Henrik Bjarke Vægter

15-05-2019

1 INDHOLD

2	Abstract	3
3	Introduktion	4
3.1	Fear-Avoidance overbevisninger	4
3.2	Tampa Scale of Kinesiophobia	5
3.3	COSMIN	5
3.4	Validitet af TSK	6
3.4.1	Content validity	6
3.4.2	Construct validity	6
3.5	Responsiveness af TSK	7
3.6	Reliabilitet for TSK	7
3.6.1	Test-retest reliabilitet	7
3.6.2	Internal Consistency	7
3.6.3	Measurement Error	8
4	Formål	9
5	Materiale og metoder	9
5.1	Litteratursøgning	9
5.2	Deltagere	9
5.3	Procedure	9
5.4	Måleinstrumenter	10
5.4.1	Demografiske data	10
5.4.2	Tampa Scale of Kinesiophobia	10
5.4.3	Numeric Rating Scale (NRS)	11
5.4.4	Pain Disability Index (PDI)	11
5.4.5	Hverdagsmotion	11
5.4.6	Fritidsaktivitet	11
5.5	Statistik	12

5.5.1	Demografiske og kliniske data	12
5.5.2	Test-retest reliabilitet	12
5.5.3	Measurement error	13
5.5.4	Cut-off scores	13
6	Resultatafsnit	14
6.1	Population og demografiske data	14
6.2	Test-retest reliabilitet	16
6.3	Measurement error	16
6.4	Cut-off score	17
6.4.1	TSK-17 og PDI _{total} og NRS _{avg}	19
6.4.2	TSK-17 og hverdagsmotion, fritidsaktivitet og siddende tid	19
7	Diskussion	21
7.1	Test-retest reliabilitet	21
7.2	Measurement Error	22
7.3	TSK cut-off scores	23
7.4	Metode diskussion	24
7.5	Kliniske implikationer	25
8	Konklusion	27
9	Perspektivering	27
10	Referencer	28

2 ABSTRACT

Introduction: Kinesiophobia is a known predictor for the degree of disability in chronic pain patients. The Tampa Scale of Kinesiophobia(TSK) has been extensively researched and translated into 14 languages. It has been found to be a valid and reliable tool in identifying kinesiophobic beliefs in patients. It has recently been translated into Danish, but to these authors knowledge, not yet tested for reliability. Although it has been implemented into research and in the clinic, there is no consensus on which cut-off score implies a clinically relevant degree of kinesiophobia. We have followed recommendations from COSMIN in planning and reporting this reliability study.

Aim and study design: We used a test-retest study design to establish the test-retest reliability of the TSK (Danish version) in a group of chronic pain patients referred to a regional pain center. The patients were stable between tests and received no interventions. Both tests were filled out via e-mail link. The test-retest reliability was calculated using ICC_{2,1}. We investigated measurement error in terms of SEM and SDC and explored existing cut-off scores for TSK-17 and TSK-13.

Results: Of 108 eligible patients, 77 were included in the analysis. The mean interval between tests was 8,4 days. We found ICC_{2,1}-values of 0,863 (CI95% 0,790 – 0,911), 0,882 (0,819 – 0,924) and 0,873 (0,807 – 0,918) for TSK-17, TSK-13 and TSK-11. We found SEM-values of 3,08, 2,42 and 2,10 and SDC95%-values of 8,53 (TSK-17), 6,71 (TSK-13) and 5,82 (TSK-11). With the cut-off scores of 40 on the TSK-17 and 31 on the TSK-13, both derived from the median of the population, the scale was found to be stable and reliable, with kappa values of 0,607 ($p<0,001$) and 0,688 ($p<0,001$).

Conclusion: This study showed that the Danish versions of TSK-17, TSK-13 and TSK-11 have good to excellent test-retest reliability, with small SEM-values and SDC of 8,53, 6,71 and 5,82. With a cut-off derived from the median in TSK-17 and TSK-13, the scale was found stable and there was a tendency towards the patients in TSK_{high} feeling more limited due to pain and being less active, lending evidence to the clinical relevance of these cut-off's. Further research is warranted to confirm these results.

3 INTRODUKTION

Kroniske smerter er en stor samfundsmæssig byrde og en stor byrde for den enkelte patient. Af den vestlige verdens befolkning lever omkring 20% med kroniske smerter¹, hvilket også er gældende i Danmark². Kroniske lænderyg- og nakkesmerter udgør den største andel³. Udviklingen fra akutte til kroniske smerter er ikke fuldt ud belyst og behandlingen af kroniske smerter er vanskelig. Tidligere tiders bio-mekanisk eller bio-medicinsk tankegang er erstattet med en bio-psyko-social forståelse^{4,5}, hvor ikke kun fysiske faktorer men også sociale og psykiske faktorer har betydning for udviklingen af kroniske smerter⁶.

3.1 FEAR-AVOIDANCE OVERBEVISNINGER

Forekomsten af "*fear-avoidance*" overbevisninger hos patienten er en af de faktorer, som er velunderbygget ift. at spille en rolle for overgangen til kroniske smerter^{7,8}. Forbindelsen mellem frygt og smerte blev første gang fremsat som teori i 1983 som: "*A fear-avoidance model of exaggerated pain perception*"⁹, og *Kinesiophobia*(kinesiofobi) blev i 1990 introduceret af Kori et al.¹⁰, og blev defineret som "an excessive, irrational, and debilitating fear of physical movement and activity resulting from a feeling of vulnerability to painful injury or reinjury" (overdreven, irrationel og invaliderende frygt for fysisk bevægelse og aktivitet, som stammer fra en følelse af sårbarhed overfor smertefuld eller gentagen skade).

Frygten for bevægelse blev sammenlignet med andre fobier¹⁰, hvor den ekstreme variant af frygt for bevægelse var kinesiofobi. At frygten for smerter kan være mere invaliderende end smerten selv påviste Crombez et al.¹¹, men smertens intensitet har også betydning for såvel akutte som kroniske smertepatienter pointerer Leeuw et al. i et review fra 2007⁷. Den teoretiske kognitive-adfærdsmodel bag fænomenet kendt som "*Fear-avoidance beliefs model*" fremsatte Vlaeyen et al. i 1995¹². Heri beskrives en negativ, selvforstærkende spiral for en mindre gruppe af patienter, der efter traume eller skade grundet katastrofering, undgåelsesadfærd med deraf følgende deconditionering, fastholdes i en kronisk smertetilstand. Begreberne *fear-avoidance* og kinesiofobi bruges ofte synonymt, på trods af at der er psykologiske gradforskelle^{11,13,14}. *Fear-avoidance* kan ses lidt bredere og dækker alle former for frygt relateret til smerte, mens kinesiofobi udtrykker en ekstrem bevægelsesfrygt.

I klinikken er det vigtigt at kunne identificere patienter med kinesiofobi, så interventionen målrettes en reduktion kinesiofobien. Det er påvist at *in vivo eksponering* reducerer kinesiofobi, at patienternes fysiske formåen derefter stiger og at risikoen for depression eller *pain related distress* aftager¹⁵. Først skal overbevisningerne hos patienterne dog identificeres.

3.2 TAMPA SCALE OF KINESIOPHOBIA

Tampa Scale of Kinesiophobia (TSK)^{12,16} var det første af flere spørgeskemaer, der identificerer *fear-avoidance* overbevisninger. TSK blev udviklet til patienter med kroniske lænderygmerter i 1991 af Miller, Kori og Todd, men blev først publiceret i 1995 sammen med den hollandske oversættelse¹². TSK har vist sig at være en robust prædikator for graden af funktionsnedsættelse og fysisk dysfunktion hos kroniske smertepatienter¹⁷. Der er fundet korrelation mellem scoren på TSK og graden af depression samt graden af smerterelateret angst^{11,18,19}. Desuden er der sammenhæng mellem scoren på TSK og den oplevede smerteintensitet⁸. Et netop publiceret, modificeret Delphi-studie anbefaler brug af TSK-11 og TSK-17 til brug i populationer med muskuloskeletale problematikker, når bevægeangst ønskes undersøgt²⁰. TSK udmærker sig i forhold til andre spørgeskemaer omhandlende bevægeangst, ved ikke at have arbejdsrelaterede spørgsmål. Dette opleves i klinikken som en fordel, da mange kroniske smertepatienter står udenfor arbejdsmarkedet.

TSK er anvendelig i forskellige patientgrupper udover kroniske lænderygpatienter, som den oprindeligt blev udviklet til, herunder til fibromyalgipatienter²¹, subakutte lænderygpatienter²², og post-operative patienter²³. Herudover findes TSK i flere modificerede, validerede udgaver blandt andet til patienter med cancer fatigue²⁴, 2 forskellige udgaver til kronisk træthedssyndrom^{25,26}, temporomandibulare problemer²⁷⁻²⁹, hjertesygdom³⁰⁻³², postoperative patienter^{23,33,34}, og til brug i den generelle befolkning³⁵. Der findes udover den oprindelige TSK-17 med 17 items, forkortede udgaver, hvoraf de mest anvendte er TSK-13^{13,21,36-39} og TSK-11⁴⁰⁻⁴⁴. TSK-13 er af Neblett et al.¹⁷ blevet inddelt i klinisk meningsfulde intervaller for at give klinikerne en mulighed for at tolke på TSK-scoren. De fandt at scoren på TSK på alle grupper korrelerede negativt med alle løfteopgaver, og korrelerede positivt med oplevet smerteintensitet, depression, smerterelateret angst og oplevet funktionsbegrænsning, hvilket kan fortælle noget om patienternes behov for intervention.

3.3 COSMIN

Når et måleredskab benyttes i klinikken såvel som i forskningssammenhænge, er det vigtigt at have kendskab til om måleredskabet er validt og reliabelt, og at kende til de psykometriske egenskaber⁴⁵. Der har igennem tiderne været brugt mange forskellige betegnelser for de psykometriske egenskaber, hvilket har skabt forvirring, og gjort det svært at sammenligne fund på tværs af studier. I 2010 opnåede et internationalt ekspertpanel konsensus om taksonomi, terminologi og definitioner vedr. de psykometriske egenskaber af patient rapporterede måleinstrumenter efter et Delphi studie ledet af Mokkink et al.⁴⁶. Dette har siden også udmøntet sig i en manual for systematiske reviews på området^{47,48}. Hele initiativet betegnes COSMIN

(*C*onsensus-based Standards for the selection of health Measurement *I*Nstruments). Vi valgte at benytte disse engelske terminologier og definitioner igennem hele masterprojektet (se Appendix 2, Tabel 5).

3.4 VALIDITET AF TSK

3.4.1 Content validity

Content validity angiver i hvilken grad måleinstrumentet repræsenterer det underliggende *construct* (eks. overbevisninger, adfærd) og *face validity* angiver om måleinstrumentet ser ud til at måle netop det, det er beregnet til at måle. *Content validity* og herunder *face validity* er undersøgt af Lundberg et al. på TSK-17. Ekspertgruppen, som bestod af specialiserede fysioterapeuter, fandt at TSK-17 både havde *face validity* og *content validity*, og at TSK-17 var et relevant instrument til at måle på *constructet* (fænomenet) ”frygt for bevægelse”. Antallet af *items* var generelt for mange og enkelte spørgsmål for ens, ifølge ekspertgruppen⁴⁹. Clark et al.³⁶ fandt ved faktoranalyser at TSK-13 var tilstrækkelig til at afdække bevægeangst, og at instrumentet fik en bedre *internal consistency* ved at fjerne de ”omvendte spørgsmål”. Woby et al. fandt i 2005 ved faktoranalyser at yderligere to spørgsmål var overflødige og kunne undlades⁴³.

Bunzli et al. har stillet spørgsmål til hvilket *construct* TSK egentlig måler. I et kvalitativt studie blev det undersøgt hvilke tanker, overbevisninger og forestillinger kroniske lænderyg patienter, som scorede højt på TSK-17, havde. Særligt to overbevisninger syntes at være fremherskende; En overbevisning om at smertefuld aktivitet vil resultere i skade og en overbevisning om at smertefuld aktivitet vil øge lidelsen og/eller funktionstab. Således konkluderede Bunzli et al., at TSK muligvis i højere grad måler på patienternes overbevisninger om, 1) at smertefuld aktivitet vil bidrage til skade, og 2) smertefuld aktivitet vil føre til øget lidelse og funktionstab⁵⁰.

3.4.2 Construct validity

Hver gang et spørgeskema oversættes repræsenterer det en ny udgave. Det er derfor nødvendigt at validere det op imod original udgaven, og undersøge dets reliabilitet^{47,51}. TSK er oversat til og valideret på mange sprog herunder hollandsk¹², svensk⁴⁹, norsk^{52,53}, brasiliansk portugisisk⁵⁴, spansk⁵⁵, italiensk²², persisk⁵⁶, Hong Kong kinesisk⁴⁴, portugisisk⁵⁷, tysk⁵⁸, japansk⁵⁹, ”simplified” kinesisk⁶⁰, Thai⁶¹, og dansk⁶².

Der er i litteraturen uenighed om, hvorvidt TSK er en uni- eller multidimensionel skala, og om den dækker en, to eller flere *constructs*. Vi identificerede 25 artikler, der ved at lave faktoranalyser undersøgte forskellige TSK-versioner (TSK 4, TSK-9, TSK-11, TSK-12, TSK-13 og TSK-17) og de underliggende subskalaer. (Se Appendix 1). De fandt alt fra 1 til 5 faktorer, men hyppigst forekommende var 2 faktorer, som blev identificeret i samlet 14 faktoranalyser. Af de 14 artikler

som fandt 2 faktorer eller subskalaer, fandt 8 studier at disse to subskalaer dækkede over henholdsvis *Activity Avoidance* (undgåelsesadfærd) og *Somatic Focus* (kropsligt fokus). Disse subskalaer blev fundet i forskellige populationer, nemlig kroniske smertepatienter^{41,44,63}, patienter med fibromyalgi³⁹, kroniske lænderygsmærter³⁹, lænderygsmærter⁵⁸, patienter fra specialiserede ortopædiske klinikker⁴⁴, postoperative patienter efter lumbal og cervical operation²³, patienter med slidgigt⁶⁴ samt hos personer med arbejdsrelaterede smerter i overekstremiteterne⁴⁰.

3.5 RESPONSIVENESS AF TSK

Responsiveness angiver måleinstrumentets evne til at registrere ændringer i *construct*et over tid (se Appendix 2, Tabel 5). TSK-17 og TSK-11 blev vurderet til at være et responsivt måleredskab, når ændringer hos kroniske lænderygpatienter skal måles⁴³. Ligeledes blev TSK-13 vurderet til at kunne måle ændringer over tid hos patienter, der gennemgik en kognitiv intervention efter lumbal deseoperation³⁴. Andre studier fandt en dårligere evne til at måle forandringer over tid hos patienter med kroniske lænderygsmærter og patienter med prolaps relateret iskias^{53,57}. Den metodiske kvalitet vurderes at være *inadequate* på disse studier^{53,57} jævnfør Appendix 2, Tabel 7.

3.6 RELIABILITET FOR TSK

3.6.1 Test-retest reliabilitet

Test-retest reliabilitet angiver det omfang, at scoren for patienter forbliver uændret ved gentagne målinger over tid, forudsat at patientens tilstand er stabil og er et udtryk for den relative reliabilitet (Appendix 2, Tabel 5)^{47,51}. Reliabiliteten for TSK-11, TSK-13 og TSK-17 er vel undersøgt i litteraturen. I alt 21 studier undersøgte test-retest reliabiliteten hos forskellige patientpopulationer som kroniske smertepatienter, rygpatienter (akutte, subakutte og kroniske), fibromyalgi patienter, whiplash patienter og skulderpatienter (se Appendix 2, Tabel 7). Der er fundet en god reliabilitet med ICC-værdier fra 0,72 - 0,99 i 19 studier, og et studie med en *weighted kappa* værdi på 0,73. Kun et enkelt studie fandt en ringe reliabilitet, vurderet ud fra COSMIN⁶⁵, med en *weighted kappa* på 0,63. Overordnet var der en tendens til at spørgeskemaet havde en bedre reliabilitet hos patienter med kroniske smerter end hos akutte patienter^{53,66-68}. Den metodiske kvalitet, vurderet ud fra COSMINs *Risk of bias checklist*, blev overordnet vurderet *inadequate*, idet en del studier ikke udfører begge test i samme *setting*, og ikke forholder sig til om patienterne er stabile i den mellemliggende periode.

3.6.2 Internal Consistency

Hos kroniske smertepatienter blev der fundet en fremragende *internal consistency* ved Cronbachs alpha⁶⁹ på TSK-11 på total (0,80-0,87), *Activity Avoidance (AA)* (0,71-0,74) og *Somatic*

Focus (SF) på 0,75-0,85^{41,63}. I øvrige populationer rangerede Cronbachs alpha på total (0,67-0,73), AA (0,56-0,67) og SF (0,60-0,76). På TSK-13 fandt Heuts et al. i en population med slidgigt en Cronbachs alpha på hhv. 0,743 (AA) og 0,64(SF) på subskalaerne⁶⁴, mens Roelofs et al. i en population af patienter med kroniske lænderygsmerter finder en god *internal consistency* ved Cronbachs alpha på total (0,81), AA(0,72) og SF(0,76)³⁹. Archer et al. fandt en samlet Cronbachs alpha på 0,76 i en population af postoperative patienter efter lumbal eller cervical operation²³. Det er problematisk at flere studier ikke forholder sig til om skalaen er unidimensionel eller multidimensionel ved udregning af Cronbachs alpha, idet Cronbachs alpha udelukkende kan bruges ved unidimensionelle skalaer og skal udregnes separat per subskala. Ved subskalaer bør Cronbachs alpha ikke opgives for den samlede skala⁷⁰.

3.6.3 Measurement Error

I både klinikken og i forskningsmæssige sammenhænge er det brugbart at kende til måleredskabets absolutte reliabilitet *measurement error*, f.eks. *standard error of measurement (SEM)* og *smallest detectable change (SDC)*^{47,71}. Herved kan kliniker og forsker bruge måleredskabet på individniveau, i forhold til at kunne udtrykke, med hvor stor sandsynlighed, en registreret ændring ved en intervention, er den "sande" ændring og ikke blot en tilfældig måleusikkerhed. Hvorvidt denne forandring er relevant for patienterne, udtrykker dette dog ikke. *Measurement error's* størrelse kan også have en betydning, når patienter stratificeres til forskellige grupper/interventioner. Jo større *measurement error*, jo mere "fejlbehæftet" vil det være hvor den enkelte patient stratificeres til. I litteraturen identificerede vi 5 artikler som omhandlede *measurement error* med varierende værdier og primært vurderet metodisk *inadequate* (Appendix 2, Tabel 7)^{43,66,72,73}.

Med baggrund i den ovenfor beskrevne litteratur, fandt vi det relevant at undersøge test-retest reliabiliteten på den danske version af TSK, da dette os bekendt ikke tidligere var gjort. For at lette overførbareheden til klinisk praksis er det væsentligt at kende til den absolutte reliabilitet udtrykt ved SEM og SDC. De studier vi har fundet er primært metodisk vurderet *inadequate*, og kun ét enkelt studie har undersøgt *measurement error* i en population af kroniske smertepatienter. Vi mener, derfor, at vi med denne undersøgelse bidrager med ny viden. Derudover undersøgte vi reliabiliteten for TSK opdelt ved *cut-off score*, da dette er anvendeligt i klinisk praksis ved stratificering af patienter til interventioner, og fordi det os bekendt ikke tidligere er gjort.

4 FORMÅL

Formålet med forskningsprojekt, er at undersøge:

- 1) de psykometriske egenskaber; test-retest reliabiliteten, *measurement error* i form af *standard error of measurement* (SEM) og *smallest detectable change* (SDC) for den danske version af spørgeskemaet TSK⁶² hos kroniske smertepatienter henvist til et tværfagligt smertecenter, Smertecenter Syd, Danmark.
- 2) reliabiliteten af eksisterende *cut-off scores* for TSK (TSK-17 og TSK-13) på denne population og vurdere om der er tendens til kliniske forskelle i de to grupperinger.

5 MATERIALE OG METODER

5.1 LITTERATURSØGNING

Der er foretaget systematisk litteratursøgning i PubMed, Embase og PsycINFO ud fra en fælles udarbejdet søgestreng, med afsæt i anbefalinger fra COSMINs manual for systematic review of PROMs (*Patient Reported Outcome Measures*). Første søgning blev foretaget i fællesskab, for at sikre ensartethed og for at opnå erfaring og konsensus om søgemetode. Efterfølgende søgninger i de to følgende databaser, blev foretaget simultant af to forfattere. Se Appendix 2, Figur 2 for flowchart og nærmere beskrivelse af søgning. Artiklerne blev inkluderet ud fra en gennemarbejdet udvælgelsesproces, bestående af 4 trin. Se Appendix 2, figur 1. Dernæst blev data udtrukket fra artiklerne. Alle artikler blev kvalitetsvurderet af minimum 2 forfattere ud fra COSMINs *Risk of Bias checklist*⁷⁴, og studiets resultat er vurderet ud fra *COSMINs Quality of results* (Appendix 2, Tabel 6).

5.2 DELTAGERE

En gruppe kroniske smertepatienter (smerter > 6 måneder) med overvejende muskuloskeletale problemstillinger henvist til Smertecenter Syd, Danmark blev rekrutteret til dette studie. Inklusionskriteriet var at patienterne kunne læse og forstå dansk, samt var tilmeldt e-boks. Patienterne var en del af et større igangværende studie.

5.3 PROCEDURE

Inden første besøg på smertecenteret blev en række spørgeskemaer, heriblandt TSK-17 og oplysningsskemaer om demografiske forhold, sendt til patienterne via link i e-boks. I forbindelse

med første besvarelse blev alle patienter inviteret til at deltage i dette test-retest studie. Såfremt patienten accepterede at deltage, blev der 7 dage efter første bevarelse automatisk tilsendt et link til TSK-17. Både første og anden besvarelse blev foretaget i eget hjem elektronisk via link. Indsamling af data fra patienter foregik i perioden d.3. november 2018 til 15. marts 2019. Det blev forudsat at patienterne var stabile i perioden mellem test og retest og patienterne modtog ingen intervention i den mellemliggende periode. Patienter, der behandles på tværfaglige smerteklinikker, har en langvarig smerteproblematik, og det er derfor overvejende sandsynligt at de er stabile i deres tilstand. Der er tidligere fundet sammenhæng mellem smerteintensitet og graden af kinesiofobi⁸. Ved at undersøge om patienternes score på oplevede gennemsnitssmerte, målt på *Numeric Rating Scale (NRS)*⁷⁵, ændredes fra test til retest, kunne vi få en indikation af om patienterne rent faktisk var stabile. Patienter med svar tider > 14 dage blev ekskluderet, og ved *missing items* ekskluderede vi besvarelserne i analyserne. Det er en metode brugt i andre studier^{49,63,72,76}.

Patienterne havde i anden sammenhæng givet tilladelse til, at deres data blev brugt i forskningsmæssig sammenhæng, og dette var allerede anmeldt og godkendt af datatilsynet, hvorfor der ikke krævedes særskilt godkendelse til denne undersøgelse. Alt data opbevares registreret i en sikret database Pain-Data kvalitets- og forskningsdatabase på Region Syddanmarks IT-service i anonymiseret form.

5.4 MÅLEINSTRUMENTER

5.4.1 Demografiske data

Følgende demografiske data blev indsamlet og registreret; Køn, alder, vægt, højde, ægteskabelig status, uddannelsesniveau og smertevarighed.

5.4.2 Tampa Scale of Kinesiophobia

TSK-17: *Tampa scale of Kinesiophobia* er et selvrapporteret spørgeskema, der er udviklet til at måle på frygten for bevægelse eller skade. Der er i dette studie benyttet den danske version af spørgeskemaet⁶², se Appendix 3. Det består af 17 *items* (spørgsmål) som besvares på en 4-punkts Likert-skala, med svarmulighed fra 1 "meget uenig" til 4 "meget enig". Minimumsscoren er 17, maksimum er 68.

TSK-13 blev udviklet af Clark et al.³⁶ efter at faktor analyser viste at de 4 "omvendt scorede spørgsmål", items 4, 8, 12 og 16, *loadede* dårligt, og blev herefter fjernet, hvilket efterlod items 1, 2, 3, 5, 6, 7, 9, 10, 11, 13, 14, 15 og 17. Den afkortede TSK-13 er siden fundet af flere forskere ved *confirmatory factor analysis*^{13,21,37-39}. Minimumsscore er 13, og maksimum er 52.

TSK-11 blev introduceret af Woby et al.⁴³ i 2005 efter at nye faktoranalyser viste en redundans på spørgsmålene 9 og 14, som efterlod 1, 2, 3, 5, 6, 7, 10, 11, 13, 15 og 17. Minimumsscoren er 11, og maksimum er 44. TSK-11 blev siden bekræftet ved *confirmatory factor analysis*⁴⁰⁻⁴².

Vi valgte at lave analyser på alle 3 udgaver, selvom patienterne alle besvarede den fulde version, TSK-17.

5.4.3 Numeric Rating Scale (NRS)

Den oplevede gennemsnitlige smerteintensitet (NRS_{avg}), højeste (NRS_{max}) og laveste (NRS_{min}) smerteintensitet, registreret gennem de sidste 24 timer, blev registreret ved første test og igen på retest dagen på *Numeric Rating Scale* (NRS)⁷⁵. NRS er et selvrapporeret spørgeskema som besvares på en Likert-skala fra 0-10, hvor 0 indikerer ingen smerte og 10 indikerer den værst tænkelige smerte. NRS er hos kroniske smertepatienter fundet valid og reliabel⁷⁷, og for kroniske smertepatienter er 2 point på scalaen den mindste klinisk relevante forandring(MCID)⁷⁸.

5.4.4 Pain Disability Index (PDI)

Patienternes fysiske begrænsninger blev registreret ved *Pain Disability Index* (PDI)⁷⁹. PDI er et spørgeskema, der er udviklet til at måle i hvilken grad patienterne oplever sig begrænset pga. smerter. Den består oprindeligt af 7 spørgsmål, men bruges ofte i denne forkortede version, da den har vist bedre *internal consistency*⁸⁰. Den forkortede version består af 5 spørgsmål, som hver især besvares ved en Likert-skala fra 0-10, hvor 0 indikerer ingen begrænsninger og 10 indikerer høj grad af begrænsninger pga. smerter indenfor områderne; familieliv, fritidsaktiviteter, sociale aktiviteter, beskæftigelse og seksualitet. Patienterne besvarede spørgeskemaet ved *baseline*.

5.4.5 Hverdagsmotion

Derudover indhentede vi information om patienternes hverdagsmotion med følgende spørgsmål: "Hvor meget tid bruger du en almindelig uge til hverdagsmotion, for eksempel gåture, cykling eller havearbejde? Medregn al den tid, du bruger (aktiviteten skal vare over 10 minutter ad gangen for at blive medregnet)." Hvor 0) 0 min/uge, 1) <30min/uge, 2) 30-60min/uge, 3) 60-90min/uge, 4) 90-150min/uge, 5) 150-300min/uge og 6) >300min/uge.

5.4.6 Fritidsaktivitet

Patienterne angav hvor aktive, de var i fritiden ud fra følgende muligheder:

- 1) Jeg sidder som regel og læser, ser TV, går i biografen - tilbringer fritiden med stillesiddende sysler.
- 2) Jeg går en tur, kører lidt på cykel, eller er i anden legemlig aktivitet mindst 4 timer om ugen (lettere fritidshobby, bordtennes, bowling og lign.)

- 3) Jeg er aktiv idrætsudøver, løber, svømmer, spiller tennis / badminton mindst 3 timer om ugen. Hvis jeg ikke dyrker sport, men ofte udfører tungt havearbejde, eller anden tungt fritidsarbejde, hører jeg også til i denne gruppe.
- 4) Jeg dyrker konkurrenceidræt, eller svømmer, cykler, løber lange distancer flere gange om ugen.

5.5 STATISTIK

Alle statistiske beregninger blev udført i statistikprogrammet IBM SPSS Statistics (version 25).

5.5.1 Demografiske og kliniske data

Data blev undersøgt for fejl og *outliers* ved at sortere data ascenderende i SPSS. For demografiske data blev der beregnet gennemsnitsværdier (*mean*) og standard deviation (SD) for normalfordelte data, og for ikke normalfordelte data blev median og minimum til maksimum værdier angivet. For nominale data blev frekvenser i form af procentsatser angivet. Vi undersøgte om data var normalfordelt ved Shapiro Wilks test⁸¹ og Kolmogorov-Smirnov. Vi undersøgte om deltagernes smerteintensitet var statistisk signifikant forskellige fra hinanden ved test og retest, ved at benytte Wilcoxon ranks test⁸². Signifikansniveauet(p) blev fastsat til $p < 0,05$ ⁴⁵.

5.5.2 Test-retest reliabilitet

Vi udarbejdede et *overlay scatterplot* for at få et visuelt overblik over fordelingen af data på TSK og eventuelle *outliers*. Vi undersøgte *Floor or Ceiling effect* (gulv eller loft effekt) på TSK ved at udregne procentdelen af patienter der enten havde maximum- eller minimumscore. En andel på mere end 15% indikerer, at der er *Floor or Ceiling effect*⁶⁵. Dette er en del af validiteten af spørgeskemaet, men hvis mange patienter har minimums- eller maximumscore(min.-max.), vil det påvirke reliabiliteten. Vi fandt *mean* (SD), median (min.-max.) af totalsummen af TSK-versionerne. Herefter brugte vi Wilcoxon ranks test for non-parametriske data til at beregne om forskellen på de to TSK-målinger (test og retest) var signifikant forskellige.

For at udtrykke test-retest reliabiliteten udregnede vi *Intraclass Correlation Coefficient* (ICC). Ved at benytte ICC_{2,1} (*two-way random effects model, total agreement*) får vi udover variansen ved måling af den enkelte patient, også inkluderet systematiske forskelle i gruppen mellem de to målinger. En ICC-værdi på 1 udtrykker en perfekt reliabilitet uden nogen målefejl, hvorimod en ICC-værdi på 0 udtrykker, at der ikke er nogen reliabilitet overhovedet^{70,83}. For at minimere risikoen for *recall bias* fastsatte vi retest til at foregå minimum 7 dage efter *baseline*. Da patienterne forudsættes at være stabile mellem *baseline* og retest, har vi fastsat et maksimalt retest interval til 14 dage. En fremragende reliabilitet fastsatte vi til $>0,90$, god 0,75 - 0,90, moderat 0,5 - 0,75, og

dårlig $<0,5^{84}$. Efter anbefaling fra COSMIN^{47,48,65} bestræbte vi os på at *settings* skulle være ens for begge besvarelser, så patienterne udfyldte begge gange spørgeskemaet via et elektronisk link.

5.5.3 Measurement error

Vi har beregnet *measurement error* ved at beregne *Standard Error of Measurement*(SEM)⁵¹ (Mokkink et al., 2018, p. 57). Vi har beregnet SEM_{agreement}, hvor både tilfældige og systematiske fejl inkluderes, da vi vil kunne skelne "sande" ændringer fra målefejlene⁶⁵ ved brug af formlen⁸⁵:

$$SEM = SD \cdot \sqrt{(1-ICC)}$$

hvor *Standard Deviationen* $SD = \sqrt{(SS_{total}/n-1)}$. SS_{total} er *Sum of Squares Total*, n = det samlede antal af både test + retest og ICC er den beregnede $ICC_{2,1}$.

Vi har konverteret SEM_{agreement} til SDC_{ind} ved hjælp af formlen^{65,86}:

$$SDC_{ind} = 1.96 \times \sqrt{2} \times SEM$$

hvor SDC_{ind} fortæller, at det er på individniveau for dermed at kunne udtrykke, hvornår en forandring på den enkelte patients score med 95% sandsynlighed kan siges, at skyldes en forandring i patientens score, der er udover testens målefejl⁶⁵. SDC_{ind} skrives nogle gange SDC_{95%}.

5.5.4 Cut-off scores

Inddeling af patienterne efter TSK-scoren i form af en klassificering i to grupper karakteriseret ved henholdsvis en høj (TSK_{høj}) og en lav (TSK_{lav}) score er ofte anvendelig i klinisk praksis, når patienterne skal stratificeres til interventioner¹². Der er beskrevet brug af *cut-off score* på TSK-17, hyppigst anvendt er den originale *cut-off score* på 37 fremsat af Vlaeyen et al.^{12,87,88}, som blev fundet ved medianen af den pågældende population af kroniske lænderygpatienter. Andre studier har ligeledes brugt populationens median til at inddele skalen dikotomisk, heriblandt Vlaeyen et al. 1999 ved *cut-off scores* på 40⁸⁹ samt Barke et al. 2012⁹⁰ ved 35,5. Vi valgte derfor dels at undersøge *cut-off score* 37, samt vores populations median som *cut-off score* på TSK-17.

På TSK-13 findes ikke egentlige *cut-off scores*, men Neblett et al. fandt frem til 4 klinisk meningsfulde grupperinger af TSK-13¹⁷. Ud fra disse valgte vi at opdele skalaen dikotomisk og opdele den i TSK_{lav} (*subclinical + mild, TSK*≤32) og TSK_{høj} (*moderate og severe, TSK*>32) med en *cut-off score* på 32. Ud fra populationens beregnede median fastsatte vi en *cut-off score* på TSK-13. Der er endnu ikke beskrevet *cut-off scores* ved brug af TSK-11, hvorfor vi ikke undersøgte dette.

Vi beregnede procentsatsen for hvor mange patienter, der var stabile, dvs. der blev i samme klassifikation mellem test og retest. Desuden udregnede vi kappaværdier⁹¹ for TSK-17 og TSK-13 med tilhørende signifikansværdier (p). At undersøge om patienterne stabilt tilhører enten TSK_{lav} eller TSK_{høj} klassifikation kan illustrere noget om måleredskabets styrke i forhold til at kunne stratificere patienterne. Jo færre patienter, der skifter fra den ene klassifikation til den anden, desto mere reliabelt er måleredskabet og dermed stiger dets anvendelighed i klinisk praksis. En kappaværdi på $<0,2$ betegnes dårlig, $0,21 - 0,40$ betegnes jævn, $0,41 - 0,60$ moderat, $0,61 - 0,80$ god og $0,81-1,00$ særdeles god^{45,70}.

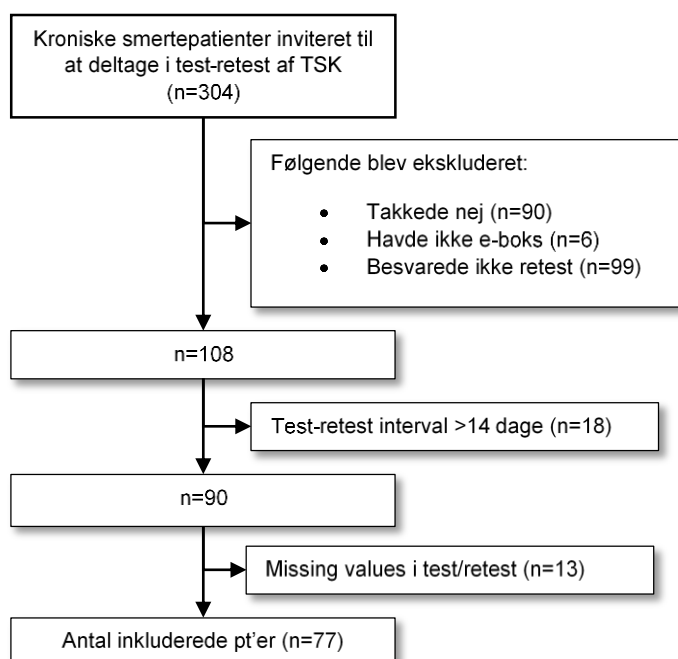
Vi undersøgte, om der var tendens til forskel mellem patienterne kategoriseret i hhv. TSK_{lav} og TSK_{høj}, på PDI_{total} og NRS_{avg}, ved brug af den non-parametriske test Mann Whitney U⁸², samt illustreret ved grafiske fremstillinger af siddende tid, hverdagsaktiviteter og fritidsaktiviteter.

6 RESULTATAFSNIT

6.1 POPULATION OG DEMOGRAFISKE DATA

304 patienter blev inviteret til at deltage i studiet, 213 ønskede at deltage. 105 besvarede ikke retesten, hvilket resulterede i 108 inkluderede patienter. Heraf blev 18 patienter ekskluderet da tidsintervallet var > 14 dage, og 13 patienter blev ekskluderet grundet mangelfuld udfyldelse af TSK-17 (se Figur 1), hvilket resulterede i en population på $n = 77$.

Figur 1. Flowchart over inklusion og eksklusion



Størstedelen var kvinder, 66,2 %, gennemsnitsalderen var 49,9 år (+/- 14,4), median på smertevarighed var 2.375 dage (221-14.056), og median på NRS_{avg} var 7(0-10). Hovedparten var gift eller samlevende (74%) og havde en erhvervsfaglig uddannelse (41,6%). I alt 49,4 % var hverdagsaktive ≤ 60 min dagligt. Ingen dyrkede konkurrenceidræt, eller svømmede, cyklede, løb lange distancer flere gange om ugen. Se Tabel 1.

Tabel 1. Populationens karakteristika (n=77)		
Karakteristika	Måleenhed	Mean (±SD) eller antal (%)
Køn	Kvinder	51 (66,2%)
Alder	År	49,9 (±14,4) fra 16-78
BMI(n=74)	(Kg/m ²)	29,54(14,69-56,75)
Ægteskabelig status	Gift/Samlevende	57 (74%)
	Enlig	20 (26%)
Uddannelsesniveau	Grundskole	11 (14,3%)
	Gymnasial uddannelse	4 (5,2%)
	Erhvervsfaglig uddannelse	32 (41,6%)
	Kort videregående uddannelse	4 (5,2%)
	Mellemlang videregående uddannelse	8 (10,4%)
	Bachelor	7 (9,1%)
	Lang videregående uddannelse	3 (3,9%)
Anden uddannelse		7 (9,1%)
Begrænsning pga. smerte	PDI	37 (0-50)
Siddende tid (n=75)	Næsten hele dagen	13,3 %
	7-15 timer	57,3 %
	1-6 timer	41,3 %
	Aldrig	1,3 %
Fritidsmotion (n=75)	1) Læser eller andre stillesiddende sysler.	50,6%
	2) Gå, cykelture el.lign. >4t./uge	44,2%
	3) Idrætsudøver, tungt have arbejde eller andet >3t./uge	2,6%
	4) Konkurrenceidræt, eller svømmer, cykler, løber lange distancer flere gange om ugen	0%
Hverdagsmotion	0 Min	7,8%
	Mindre end 30 min	22,1%
	30-60 min	19,5%
	60-90 min	6,5 %
	90-120 min	16,9%
	Mere end 120 min	14,3%
	6 timer	13 %
Gennemsnitssmerte	NRS _{avr} (test) (n=77)	7 (0-10)
	NRS _{avg} (re-test) (n=76)	6 (0-10)
Mindste smerte	NRS _{min} (test) (n=77)	5 (0-10)
	NRS _{min} (re-test) (n=76)	5 (0-10)
Maksimale smerte	NRS _{max} (test) (n=76)	8 (0-10)
	NRS _{max} (re-test) (n=76)	8 (0-10)
Smertevarighed	Dage: Median (min-max)	2375 (221-14277)

Normalfordelte data angives ved mean (±SD) og ikke normalfordelte data angives ved median (min - max). BMI, Body Mass Index, PDI, Pain Disability Index, NRS, Numeric Rating Scale, NRS_{avg}, gennemsnitssmerte seneste 24 timer, NRS_{min} mindste 24timer, NRS_{max}, maksimale smerte 24timer.

6.2 TEST-RETEST RELIABILITET

Medianen af TSK-scoren var ved retest højere end ved test i alle versioner, med en difference på hhv. -1,0 (TSK-17), -0,8 (TSK-13) og -0,6 (TSK-11). Udregnet ved Mann Whitney U er ingen af disse forskelle signifikante. Vi fandt at spørgeskemaet TSK har en god til fremragende reliabilitet i alle 3 versioner TSK-11, TSK-13 og TSK-17 med høje ICC-værdier på hhv. 0,873 (CI 95% 0,790 - 0,911), 0,882 (CI 95% 0,819 - 0,924) og 0,863 (CI 95% 0,807 – 0,918) (se nedenstående Tabel 2). Intervallet mellem test og retest var i gennemsnit 8,4 dage. Der blev ikke fundet *Floor or Ceiling effect* på nogen af spørgeskemaerne.

Tabel 2. Oversigt over *mean*, difference, median, ICC-værdier og *Floor or ceiling effect* på TSK-17, TSK-13 og TSK-11

	TSK-17		TSK-13		TSK-11	
Test	TSK-17 _{test}	TSK-17 _{retest}	TSK-13 _{test}	TSK-13 _{retest}	TSK-11 _{test}	TSK-11 _{retest}
<i>Mean</i> (SD)	40,1 (±8,5)	41,2(±8,1)	31,8(±7,3)	32,5(±6,8)	27,6(±6,2)	28,2(±5,6))
Median (min-max)	39 (20-66)	40 (21-66)	31 (14-52)	32 (15-51)	27 (12-44)	27 (13-44)
Difference (SD) (CI 95%) <i>p</i> -værdi	-1,0 (±4,3) (-2,0 - (-0,4)) <i>p</i> = 0,054		-7,9 (±3,4) (-1,6 - (-0,3)) <i>p</i> = 0,069		-0,6 (±2,9) (-1,3 - 0,1) <i>p</i> = 0,104	
ICC _{2,1} (CI 95%)	0,863 (0,790 - 0,911)		0,882 (0,819 - 0,924)		0,873 (0,807 – 0,918)	
<i>Floor or Ceiling effect</i> %	0,00		1,30		1,95	

TSK, *Tampa Scale of Kinesiophobia*, SD, Standard Deviation, ICC_{2,1}, *Intra Class Correlation – two-way random effects, total agreement*, CI 95%, *Confidence Interval* 95%, *p*, signifikansværdi

6.3 MEASUREMENT ERROR

SEM_{agreement} er udregnet til hhv. 3,08 (TSK-17), 2,42 (TSK-13) og 2,10 (TSK-11) (se Tabel 3), illustreret ved udregningerne for TSK-17 ved formlen⁸⁵:

$$SEM_{agreement} = SD \times \sqrt{1-ICC_{2,1}} = 8,31 \times \sqrt{(1-0,863)} = 3,08$$

hvor SD udregnes efter følgende formel⁸⁵:

$$SD = (\frac{SS_{total}}{(n-1)}) = (\frac{10.573,07}{(154-1)}) = 8,31$$

og efterfølgende er SDC_{ind} beregnet til hhv. 8,53 (TSK-17), 6,71 (TSK-13) og 5,82 (TSK-11) (se Tabel 3) på følgende måde^{65,86}, ligeledes illustreret ved udregninger for TSK-17:

$$SDC_{ind} = 1,96 \times \sqrt{2} \times SEM = 1,96 \times \sqrt{2} \times 3,08 = 8,53$$

SDC_{ind} svarer til den mindste forandring i en persons score som med 95% sandsynlighed vil kunne betegnes som en "sand" forandring, udover redskabets målefejl, på individniveau. For beregninger på TSK-13 og TSK-11 se Appendix 4.

Tabel 3. Oversigt over $SEM_{agreement}$ -værdier og SDC -værdier

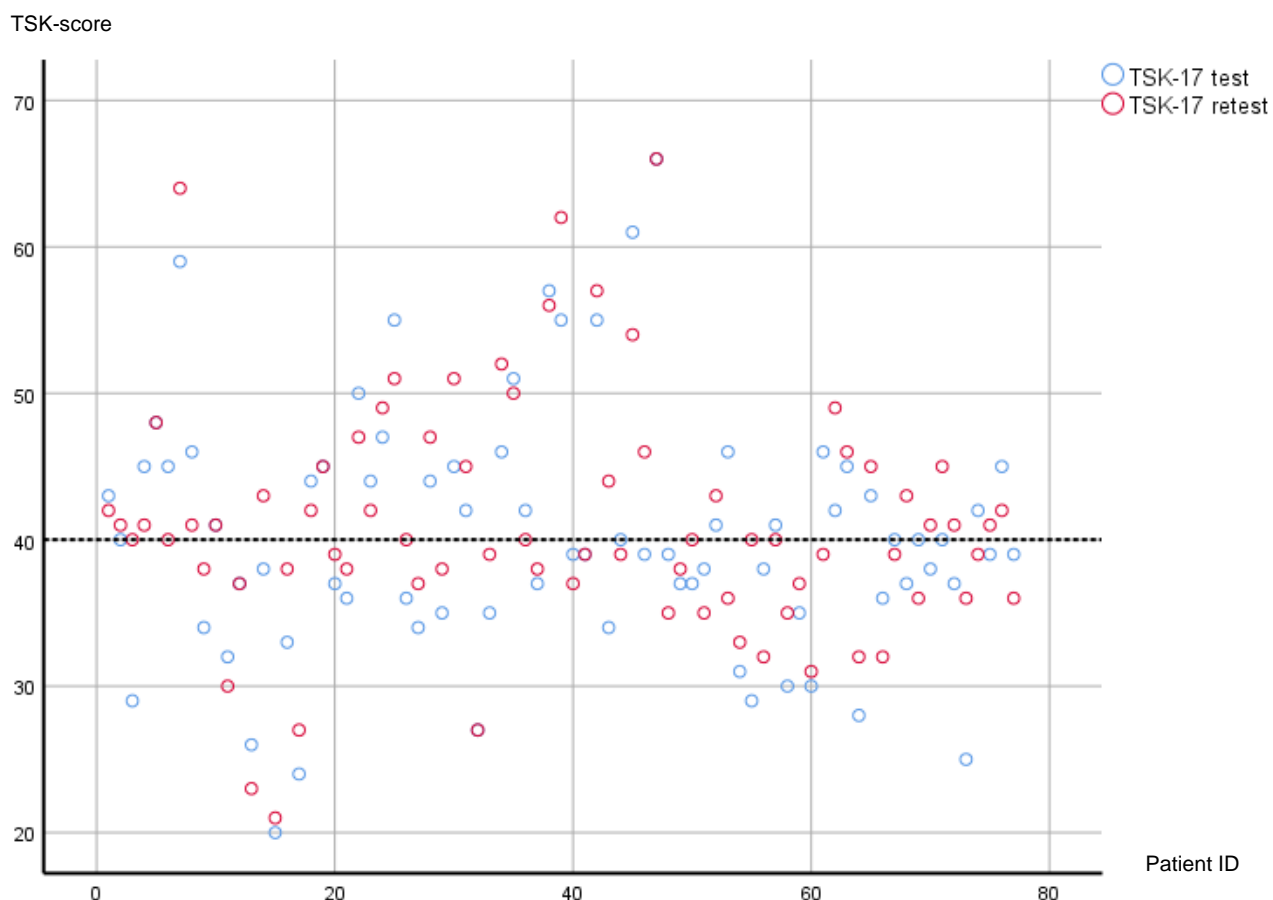
	TSK-17	TSK-13	TSK-11
$SEM_{agreement}$	3,08	2,42	2,10
SDC_{ind}	8,53	6,71	5,82

Tabel 3. $SEM_{agreement}$, *Standard Error of Measurement with total agreement*, SDC_{ind} , *Smallest Detectable Change = SDC95%*

6.4 CUT-OFF SCORE

TSK-17: Med en *cut-off score* på 37, som er kendt fra litteraturen¹², var der hhv. 54,54% stabile og 45,45 % som skiftede imellem klassifikationerne lav/høj med en kappaværdi på 0,219 ($p= 0,002$). Det betyder, at TSK-17 ikke er et stabilt måleredskab, når *cut-off scoren* er 37 i denne population af kroniske smertepatienter. Når datasættet blev opdelt ud fra den beregnede median på 40, var 80,51% stabile, og 19,5% skiftede gruppe imellem de to målinger. Den beregnede kappaværdi er god på 0,607 ($p < 0,001$). Se figur 2 med *scatterplot* over TSK-17 test-retest med *cut-off score* på 40. Grundet ovenstående beregninger, valgte vi i de nedenstående afsnit at benytte 40 som *cut-off score* ($n= 44$ i TSK_{lav} og $n=33$ i $TSK_{høj}$).

Figur 2. Scatterplot over TSK-17 test/retest med cut-off 40



TSK-13: Ved *cut-off score* på 32 på TSK-13, fastsat ud fra Neblett et al.', var der i alt 77 % som var stabile og blev i den samme klassifikation; TSK_{lav} eller TSK_{høj}. Det betyder, at 23% enten skifter fra høj til lav, eller fra lav til høj imellem de to målinger. Kappaværdien vurderes at være moderat på 0,55 ($p < 0,001$). Populationens median er 31 på TSK-13. Ved at fastsætte *cut-off score* til 31, fandt vi at 84,4% forblev stabile, og 16,6% skiftede klassifikation, med en god kappaværdi på 0,688 ($p < 0,001$).

Tabel 4. Oversigt over *cut-off scores*, procentvise stabile patienter, kappaværdier

Cut-off score	TSK-17		TSK-13	
	37	40	32	31
Stabile %	54,5 %	80,5 %	77,9 %	84,4 %
Kappa værdi (<i>p</i> -værdi)	0,219 (0,002)	0,607 ($< 0,001$)	0,551 ($< 0,001$)	0,688 ($< 0,001$)

TSK, Tampa scale of Kinesiophobia, *p*-værdi, signifikansværdi

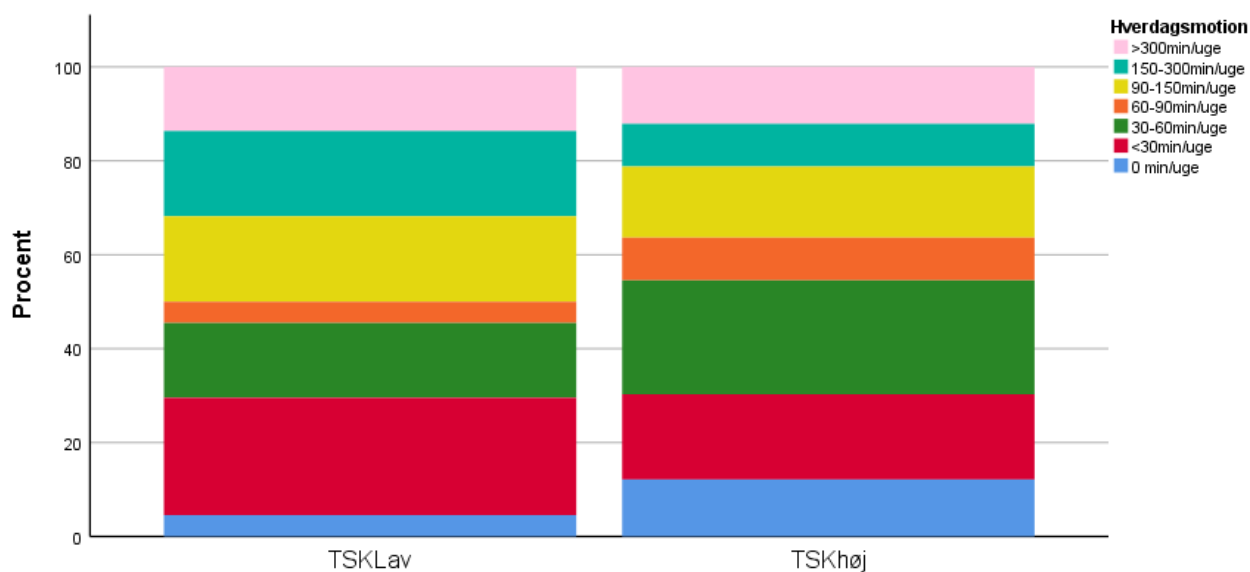
6.4.1 TSK-17 og PDI_{total} og NRS_{avg}

Udregnet med Mann Whitney U fandt vi en signifikant forskel på hhv. TSK_{lav} og TSK_{høj} grupperne på PDI_{total}, ($p=0,038$). PDI_{total} var højest i TSK_{høj} gruppen med en median på 41 og i TSK_{lav} gruppen var medianen 36. Således konkluderedes, at patienterne i TSK_{høj} gruppen havde en signifikant højere score på PDI_{total} sammenholdt med patienterne i TSK_{lav} gruppen på TSK-17. Derimod fandt vi ingen signifikant forskel ($p=0,514$) på patienternes NRS_{avg} i de to grupper, ved udregning med Mann Whitney U.

6.4.2 TSK-17 og hverdagsmotion, fritidsaktivitet og siddende tid

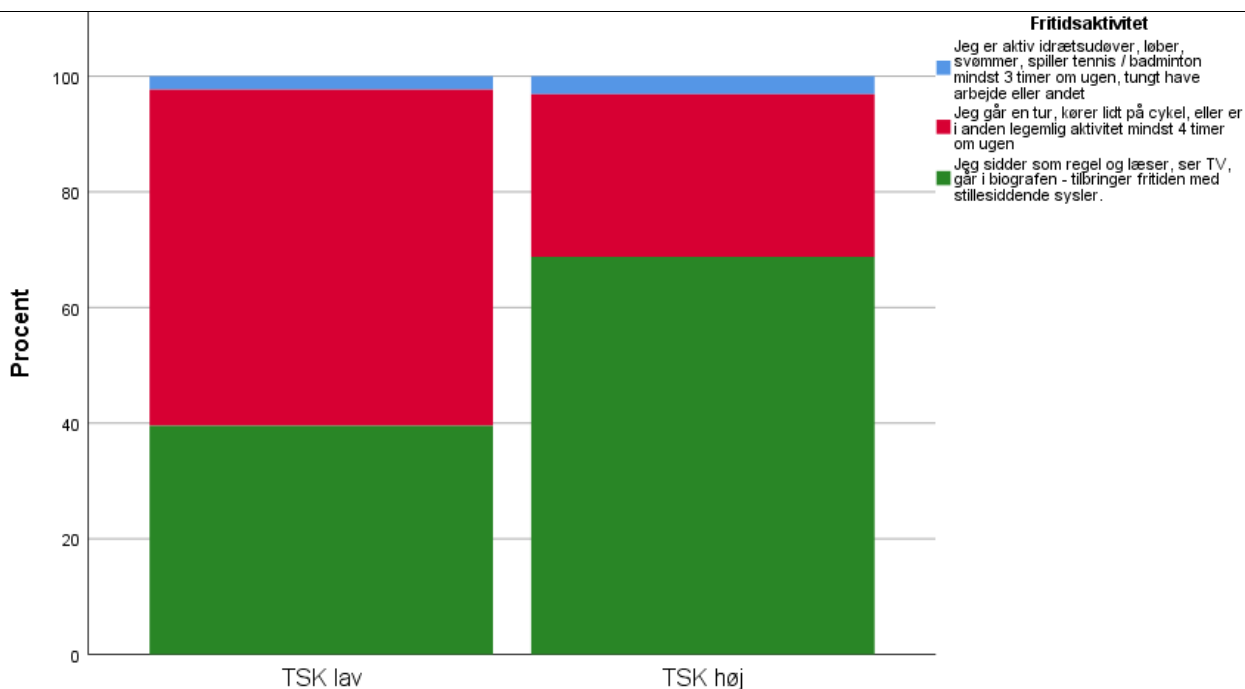
Med en *cut-off score* på 40 på TSK-17, ses der tendenser til at mængden af hverdagsmotion er af kortest varighed i TSK_{høj} gruppen, hvor 12,1% svarer at de dyrker 0 min pr uge, og i TSK_{lav} gruppen, er den markant lavere nemlig 4,5%. Samtidigt er færre aktive >150min i TSK_{høj} gruppen (21,2%) sammenlignet med TSK_{lav} gruppen (31,8%). Se Figur 3.

Figur 3. Hverdagsmotion for TSK-17 i hhv. TSK_{lav} og TSK_{høj} ved *cut-off score* 40 v/ *baseline*



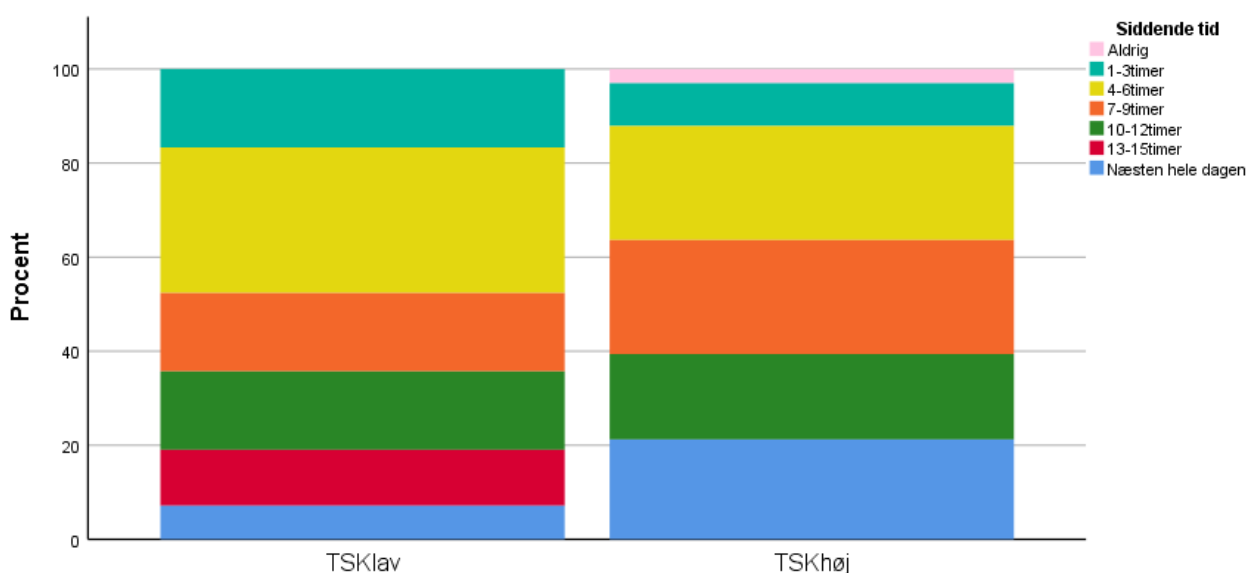
Der ses en tendens til, at en større andel i TSK_{høj} gruppen bruger deres fritid med stillesiddende sysler som tv, læsning og biografture end i TSK_{lav} gruppen med respektive 68,8% og 39,5%. Se Figur 4.

Figur 4. Fritidsmotion for TSK-17 i hhv. TSK_{lav} og TSK_{høj} ved cut-off score 40 v/ baseline



I alt 57,3% sidder 7 timer eller mere ned dagligt. I TSK_{høj} gruppen er der en større andel, der sidder ned "næsten hele dagen" (21,2%), end i TSK_{lav} gruppen (7,1%). Se Figur 5.

Figur 5. Siddende tid for TSK-17 i hhv. TSK_{lav} og TSK_{høj} ved cut-off score 40 v/ baseline



Således vurderes der at være tendens til at patienterne i TSK_{høj} gruppen dels er mere begrænset pga. smerter målt på PDI_{total}, generelt sidder ned en større andel af deres dag, samt at mængden af hverdagsmotion og fritidsmotion er mindre i TSK_{høj} sammenlignet med TSK_{lav}.

7 DISKUSSION

7.1 TEST-RETEST RELIABILITET

I alle tre danske versioner af TSK; TSK-17, TSK-13 og TSK-11 findes der i dette studie en god til fremragende test-retest reliabilitet med følgende ICC-værdier 0,863 (CI95% 0,790 – 0,911) og 0,882 (CI95% 0,819 - 0,924) og 0,873 (CI95% 0,807 – 0,918). På TSK-17 fandt vi 2 studier der undersøgte test-retest reliabilitet på kroniske smertepatienter. Disse fandt ICC_{2,1} værdier på hhv. 0,80 fra Askary-Ashtiany et al.⁹² og 0,9 fra Koho et al.⁷⁶, hvilket stemmer overens med vores fund, trods forskellen i tidsintervallet på hhv. 48 og 24 timer. Et andet studie fra Lamé et al. adskiller sig fra ovenstående ved at have et langt tidsinterval mellem testene på gennemsnitlig 52 dage, hvilket muligvis forklarer den moderate ICC-værdi på 0,72 på TSK-17, og på 0,74 på TSK-13⁹³. Yderligere forholder vi os kritisk til Lamé et al.'s⁹³ studie idet, at den valgte ICC model ikke er angivet, hvilket kan påvirke resultatet i begge retninger⁸⁵, og at de desuden valgte de at udregne ICC først for hele populationen, og dernæst fjernede de outliers og forbedrede dermed ICC-værdien. Vores resultat på TSK-11 stemmer overens med Hapidou et al.'s⁷², der ligeledes testede i en population af kroniske smertepatienter og med et test-retest interval på 5 dage. De beregnede en ICC_{2,1} værdi på 0,81 (0,58-0,93). I en gruppe af ældre kroniske smertepatienter fandt Larsson et al.⁶³ en god ICC værdi på 0,747, men har dog ikke angivet hvilken ICC model, der er brugt til udregning.

Vi har i dette studie været bevidste om valget af ICC model, og valgt *two-way random effects model (2,1)*, fordi denne model, udover variansen ved den enkelte patient, inkluderer systematiske forskelle og dette anbefales i litteraturen^{47,51,65,84}. ICC_{2,1} anvendes hyppigt i test-retest reliabilitetsstudier^{72,76,92}. Lamé et al.⁹³ og Larsson et al.⁶³ angav derimod ikke valg af model, hvilket kan gøre det sværere at sammenligne værdierne på tværs af studier. Det er vigtigt at forholde sig kritisk til studier, når valg af ICC-model ikke er angivet, samt tolke på resultatet med en vis forsigtighed⁸⁴.

Meget homogene besvarelser på skalaen kan bevirke at ICC-værdien påvirkes i en nedadgående retning⁷⁰, idet det er svært at diskriminere imellem patienterne. Ved udarbejdelse af *scatterplot* fandt vi en bred repræsentation af hele TSK-skalaen i vores population, altså en relativ heterogen gruppe, hvilket kan medvirke til den relative høje ICC-værdi.

Studierne, der undersøger test-retest reliabiliteten, er overordnet *inadequate* ved vurdering af *risk of bias* (se Appendix 2, Tabel 7), grundet manglende stillingtagen til patienternes stabilitet og skiftende *settings* ved de to besvarelser. Vores resultater af test-retest reliabiliteten adskiller sig ikke væsentligt fra tidligere fund, hvorfor det kan diskuteres hvor meget, det betyder at patienterne besvarer i samme *setting*, og det må antages, at patienterne i de andre studier har været stabile trods manglende beskrivelse af dette.

7.2 MEASUREMENT ERROR

I vores studie med kroniske smertepatienter finder vi på TSK-11 en SEM-værdi på 2,10. Ved litteratursøgning fandt vi kun et enkelt studie, som beregnede en SEM-værdi i en population af patienter med kroniske smerter. Hapidou et al.⁷² beregnede en SEM-værdi på 2,41 på TSK-11, hvilket stemmer overens med vores studie. Hapidou et al. opgav en SDC90% værdi på 5,62. Det anbefales at benytte et konfidensinterval (CI) på 95%, når der beregnes SDC til brug på individniveau⁶⁵. Derfor er det problematisk, når der opgives SDC-værdier med andre CI, da det vanskeliggør sammenligninger på tværs af studier, samt gør det vanskeligt for klinikerne at bruge informationen i klinisk praksis. Når Hapidou et al.'s SDC-værdi omregnes til SDC95%(=SDC_{ind}) bliver den 6,68, hvilket er større, end de 5,82 vi finder. Der indgår kun 18 patienter i dette test-retest studiet, hvorfor vi forholder os kritiske til styrken af dette fund. To andre studier forholder sig til SDC i TSK-11, nemlig George et al. samt Woby et al., begge i en population af patienter med kroniske lænderygsmerter^{43,73}. George et al. fandt en SEM-værdi på 1,7 og en tilhørende SDC på 4,8, som er mindre, end den vi finder. Det kan diskuteres om det korte tidsinterval mellem test og retest på 48 timer, kan medføre *recall bias*, hvilket kan give en "falsk" høj ICC- og dermed lavere SEM- og SDC-værdier. Woby et al. derimod opgiver en SDC på 3, som tilsyneladende blot er SEM-værdien rundet op. Hvis vi derimod ud fra den opgivne SEM på 2,54 beregner, får vi en SDC på 7,04, som er større end vores.

For TSK-13 har vi kun kendskab til ét andet studie som forholder sig til *measurement error*. Det er fra Haugen et al. som i en population med iskias smerter grundet lumbal prolaps⁵³. Her benyttede de ikke SEM til at udtrykke *measurement error*, og de to studier kan derfor vanskeligt sammenlignes.

For TSK-17 er der to studier der opgiver SDC værdier. I en population af patienter med akutte lænderygsmerter fandt Ostelo et al.⁶⁶ en SDC på 9,2 hvilket er lidt højere end vores fund på 8,53 i en population af kroniske smertepatienter, men dog sammenligneligt. Woby et al. opgiver for TSK-17 i en population af patienter med kroniske lænderygsmerter en SDC på 4, som igen ikke er udregnet, men blot SEM-værdien rundet op. Hvis vi udregner SDC med den angivne SEM, får vi SDC 8,76 hvilket stemmer overens med vores SDC.

Alle tidligere studier som omhandler *measurement error* er metodisk vurderet *inadequate* ud fra COSMINs Risk of Bias (se Appendix 2, Tabel 7), bortset fra Hapidou et al. som er vurderet *doubtful*. Med afsæt i vurderingen af den metodiske kvalitet af tidligere studier, forholder vi os kritiske til disse fund. I fremtidige studier kunne det være interessant at undersøge SEM og SDC hos flere populationer og i forskellige sektorer, idet det muligvis kan være afgørende hvilken population og i hvilken sektor dette undersøges.

Det er problematisk, at der i studier angives forskellige SDC-værdier (SDC90% og SDC95%) samt SDC-værdier som ikke er beregnet. Det besværliggør sammenligninger på tværs af studier, og er misvisende for klinikerne. Dette er en af begrundelserne for, hvorfor COSMIN-initiativet søger at opnå international enighed om taxonomi og metoder til rapportering af validitets- og reliabilitetsundersøgelser for PROMs, så sammenligninger på tværs af studier muliggøres.

7.3 TSK CUT-OFF SCORES

Ved fastsættelse af en *cut-off score* ud fra gruppens median på 40, finder vi en moderat til god kappaværdi på 0,607, $p < 0,000$ og en ringe kappaværdi 0,219, $p = 0,002$ ved den jævnligt anvendte *cut-off score* på 37, som blev fastsat ud fra den daværende populations median¹². Det tyder således på, at spørgeskemaet er mest stabilt når *cut-off scoren* fastsættes ud fra populationens median. Koho et al.⁹⁴ argumenterer for at det statistisk set kan give mening, at fastsætte *cut-off scores* ud fra populationens median, men at risikoen ved dette kan være, at der sker en overestimering af antallet af patienter med betydende kinesiofobi, når *cut-off score* fastsættes på denne måde.

På den anden side vil det ikke være brugbart i hverken klinikken eller forskningsmæssige sammenhænge, at benytte en *cut-off score*, hvor størstedelen af patienterne lige så vel kunne være stratificeret til den anden gruppe. Vi har ikke identificeret studier, som undersøger stabiliteten i TSK ved forskellige *cut-off scores*. Men det kan diskuteres om, det klinisk giver mening med en *cut-off score* på 40, i en population af kroniske smertepatienter, eller om for mange patienter klassificeres til at være i TSK_{høj} gruppen. I vores studie finder vi at PDI_{median} er signifikant forskellige fra hinanden i de to grupper med *cut-off scoren* 40, og at patienter i TSK_{høj} oplever sig mere begrænset i hverdagsfunktionerne grundet smerter, end de patienter som befinder sig i TSK_{lav} gruppen. Ligeledes finder vi en tendens til at mængden af hverdags- og fritidsmotion er af kortere varighed i TSK_{høj} sammenlignet med TSK_{lav}, at de sidder en større del af deres dag, og at en større andel i TSK_{høj} bruger deres fritid med stillesiddende sysler. Dette fund svarer til den omvendte korrelation mellem kinesiofobi og fritidsaktivitet som Koho et al. fandt i en gruppe af kroniske smertepatienter⁹⁴ og taler for brugen af *cut-off score* på 40 i denne population.

I vores studie finder vi ingen signifikant forskel i NRS_{avg} mellem $TSK_{høj}$ og TSK_{lav} grupperne, på TSK-17 ved *cut-off score* 40. Det kunne stemme overens med at Crombez et al. fandt, at kinesiofobi var vigtigere end smertens intensitet på patientens oplevede begrænsning¹¹. Det er i modsætning til Neblett et al. som fandt en positiv korrelation mellem smerteintensitet og score på TSK¹⁷. I vores studie er der kun få patienter i hver gruppe ($TSK_{høj}$ $n = 33$, TSK_{lav} $n = 44$), hvilket ikke giver meget *power*, som kan være en mulig forklaringen på forskellen i fund.

Fremtidige studier bør undersøge de kliniske forskelle mellem grupperne af patienter i hhv. TSK_{lav} og $TSK_{høj}$ for at konfirmere disse tendenser, og for at kunne undersøge om den stabile *cut-off score* på 40 for TSK-17 også er klinisk relevant.

7.4 METODE DISKUSSION

I dette studie har vi bevidst valgt at højne den metodiske kvalitet, ved at forholde os til patienternes stabilitet samt designe studiet, således at begge besvarelser er udført via e-mail link. Dette ville ved *risk of bias* vurdering ad modum COSMIN give vurderingen *very good*, hvilket styrker tilliden til egne fund.

Vi valgte at ekskludere de patienter, som ikke havde besvaret spørgeskemaet 100%, hvilket tidligere er gjort i andre studier^{49,63}. Vi ekskluderede i alt 14,4% grundet *missing items*, hvilket er på grænsen af de 15% som Devet et al. maksimalt anbefaler⁷⁰. Det er svarende til Larsson et al.⁶³(14,2%) og lavere end hvad Lundberg et al.⁴⁹(18,5%) og Koho et al.⁷⁶(29,9%) ekskluderer. De høje procentsatser for *missing items* på TSK-spørgeskemaet kan antyde, at spørgeskemaet er vanskeligt at besvare. Semple⁹⁵ fandt i et kvalitativt studie, at patienterne kæmper med at forstå eller besvare specifikke spørgsmål. Nogle spørgsmål var udformet, så de forekom irrelevante. Det kunne eks. være "Uheldet har gjort, at min krop vil være mere sårbar resten af livet", i tilfælde hvor patienten ikke har været ude for et uheld. Dette havde indflydelse på deres svar. Vi undersøgte ikke som anbefalet⁷⁰, om det er bestemte *items*, der figurerer som *missing items*. Dette kunne have givet en indikation af, om det er formuleringen af spørgsmålet, udtrætning hos patienten eller at spørgsmålet er irrelevant for målgruppen.

Problemet med *missing items* kan ligeledes skyldes, at det er vanskeligt at afdække overbevisninger som kinesiofobi, fordi det ikke er et veldefineret *construct* hvilket Lundberg et al.⁹⁶ påpeger i et *review* fra 2011. Antallet af faktoranalyser og de mange forskellige resultater på disse, med 1 -5 *subscales* vidner om at skalaen er problematisk. I forskellige populationer, findes forskellige *subscales*, men heller ikke disse er konsistente på tværs af studier. Derfor anbefales det, at der foretages faktoranalyse på nye versioner og i specifikke populationer^{47,51}. På trods af, at vi ikke med faktoranalyse har afdækket hvilken version og mulige *subscales*, der passede bedst på

vores data, valgte vi at udregne ICC-, SEM- og SDC-værdier på både TSK-17, TSK-13 og TSK-11. Det havde været optimalt først at lave *confirmatory factor analysis*, men vores population var for lille. I erkendelse af at vi ikke udførte faktoranalyse, og dermed ikke ved om TSK-17 i dansk version er unidimensionel, fravalgte vi at udregne Cronbachs alpha⁹⁷.

I vores studie har testdeltagerne besvaret TSK vha. link på computer eller tablet. TSK blev oprindeligt udviklet i en papirversion, og en undersøgelse fra 2014 viste at computerversionen havde højere ICC-værdier sammenlignet med papirversionen⁷⁶, og et Bland-Altman Plot⁹⁸ illustrerede at forskellen i *mean* testscore var 1,9. Dette er ikke i overensstemmelse med en metaanalyse af Gwaltney et al., som fandt at der ikke var signifikant forskel på papir og computer versioner af spørgeskemaer⁹⁹. I en digital verden vil det med fordel kunne inddrages at patienterne besvarer TSK elektronisk, dels for at lette journaliseringen og for at undgå fejl ved sammenregningen af TSK-scoren, eftersom der ikke lader til at være en forskel på hvor nemt eller svært det er at udfylde ift. besvarelsesmetode⁷⁶.

7.5 KLINISKE IMPLIKATIONER

I mødet med patienter med kinesiofobi, er det vigtigt for klinikerens at kunne identificere de patienter med høj grad af kinesiofobi med et reliabelt måleredskab, hvilket vi finder den danske version af TSK er. Når klinikerens skal differentiere mellem de patienter, der har en relevant høj grad af kinesiofobi, og de der ikke har, er det vigtigt at kende til den klinisk relevante *cut-off score*. I dette studie har vi forholdt os til statistisk relevante *cut-off scores*, og har forholdt os til om de to grupper adskilte sig fra hinanden ved *cut-off score* på hhv. 40 for TSK-17 og 31 for TSK-13, hvilket vi fandt. Patienter i TSK_{høj} gruppen tenderede til at være mere inaktive og oplevede sig mere begrænsede pga. smerter. Dette kan indikere at en reduktion af kinesiofobi med fordel kan inddrages i behandlingen patienter med score over 40 på TSK-17 og 31 på TSK-13.

Kognitiv adfærdsterapi har vist sig som en effektiv måde at reducere kinesiofobi hos kroniske lænderygpatienter¹⁰⁰. Ligeledes har *Graded exposure*, hvor patienten ved gradvis eksponering udsættes for de funktioner, som patienten frygter, vist at kunne reducere kinesiofobi hos patienter med kroniske lænderygsmerter¹⁵. Flere studier har også fundet at *in vivo* eksponering reducerer kinesiofobi hos patienter med kroniske smerter¹⁰¹. Reduktion af kinesiofobi forbedrer patienters fysiske formåen, reducerer depression og smerterelateret frygt¹⁷.

Når patienten har gennemgået et behandlingsforløb, er det væsentligt, at TSK kan identificere om kinesiofobien er blevet reduceret eller øget. SDC angiver om forandringen er ud over en evt. målefejl. Dette fortæller IKKE behandleren om en evt. reduktionen har haft betydning for patienten. *Minimal important change (MIC)* er et udtryk for hvor stor en ændring, der skal til på et givent måleredskab, for at patienterne oplever det som en væsentlig ændring^{47,51} Andre steder

kaldes det *Minimal clinically important difference* (MCID). Litteraturen på dette område er sparsom i forhold til TSK. For at kunne identificere MIC, er det som oftest nødvendigt at kende til en *golden standard* for det undersøgte område. Eftersom en *golden standard*, så vidt vi ved, ikke eksisterer for TSK, kan det være nødvendigt at opstille eksempelvis *Global rating scale* eller *Global perceived effect scale*, for på bedst mulig måde, at afdække hvorvidt patienterne oplever en relevant ændring. I litteraturgennemgangen identificerede vi et studie af Monticone et al.³⁴ som angav en MCID > 6 på TSK-13, hos patienter efter en lumbal deseoperation. Dette betyder at patienterne, skal have en reduktion på mindst 7 point på TSK-13, for at det vil opleves som en bedring. I en population med kroniske uspecifikke lændesmerter fandt samme forfatter en MCID > 5,5 på TSK-13, hvilket jo betyder at de kun skal forandre sig mindst 6 point på samme skala for at det vil opleves som en bedring¹⁰². Ovennævnte er relevant, for jævnfør COSMIN skal MIC være større end SDC for at måleredskabet kan benyttes^{47,65}. Ingen af ovenstående populationer er identisk med vores, men det kunne alligevel indikere at TSK-13 er et responsivt måleredskab, og at SDC < MIC.

En del studier anbefaler at vælge TSK-11 når graden af kinesiofobi skal vurderes hos kroniske smertepatienter, da den er fundet valid og reliabel^{41,43,63}. Den kortere udgave vil medføre en mindre belastning for patienten, hvilket rent etisk ville være at foretrække, og vil betyde et mindre tidsforbrug for såvel patient som behandler. Vi finder kun marginale forskelle på ICC-værdierne på de tre versioner, og kan derfor ikke herudfra fremhæve den ene version fremfor den anden. Men ved konfidensintervallerne for differencen mellem TSK_{test} og TSK_{retest}, adskiller TSK-11 sig fra TSK-17 og TSK-13 ved at passere 0. På TSK-17 og TSK-13 ligger $p > 0,05$ beregnet ved Mann Whitney U, men udregnet ved parret t-test er de signifikant forskellige fra hinanden (Data ikke præsenteret). Derimod er $p > 0,05$ både beregnet ved Mann Whitney U ($p=0,104$) samt ved den parrede t-test for TSK-11, og må vi antage at medianscoren er ens ved test og retest. Dette antyder, at TSK-11 er mere brugbar i en population af kroniske smertepatienter. Et nyligt publiceret modificeret Delphi studie anbefaler både TSK-11 og TSK-17 til brug ved kroniske muskuloskeletale smerter, og man kunne med fordel vælge den korte version af ovenstående grunde. Fremtidige studier kunne rette sig mod at identificere en eller flere klinisk relevante og statistisk stabile *cut-off score* for TSK-11 i populationen af kroniske smertepatienter, enten som en dikotomisk skala eller med flere inddelinger af skalaen som både Neblett et al. og Lüning Bergsten et al. argumenterer for^{17,103}.

8 KONKLUSION

Baseret på fundene i dette studie, indikerer det at den danske version af *Tampa Scale of Kinesiophobia* (TSK-17, TSK-13 og TSK-11) er et reliabelt spørgeskema til brug på danske smertecentre i populationer af kroniske smertepatienter med en god til fremragende ICC_{2,1}-værdi på hhv. 0,863 (CI95% 0,790 - 0,911) og 0,882 (CI95% 0,819 - 0,924) og 0,873 (CI95% 0,807 - 0,918). Der er lave SEM-værdier på hhv. 3,08 og 2,42 og 2,10 og SDC-værdier, der indikerer at forandring på hhv. >8,53, >6,71 og >5,82 med 95% sandsynlighed skyldes en forandring hos den enkelte patient, udover testens måleusikkerhed. Ved en *cut-off score* udledt fra populationens median, på hhv. 40 for TSK-17 og 31 for TSK-13, findes det stabilt og reliabelt med kappa-værdier på hhv. 0,607 ($p < 0,001$) og 0,688 ($p < 0,001$). Patienter i TSK_{høj} gruppen tenderede til at være mere inaktive og oplevede sig mere begrænsede pga. smerter. Dette kan indikere at behandlinger, rettet mod at reducere kinesiofobi, med fordel kan inddrages i behandlingsplanen for patienter med score over 40 på TSK-17 og 31 på TSK-13. Den oplevede smerteintensitet adskilte sig ikke fra hinanden i de to grupper TSK-17 ved *cut-off score* på 40. Der er ingen *Floor or Ceiling effect* i de tre udgaver. Det er usikkert om ovenstående fund kan overføres til andre populationer, eller til kroniske smertepatienter i andre sektorer.

9 PERSPEKTIVERING

I behandlingen af kroniske smertepatienter har det længe været implicit at den opdaterede tilgang til patienterne er ved integrationen af den bio-psyko-sociale model. Desværre har studier vist at undervisningen af sundhedsprofessionelle tager udgangspunkt i det biologiske domæne og ikke i den bio-psyko-sociale model¹⁰⁴ og at denne tilgang bidrager til behandlernes og dermed patienternes katastrofering¹⁰⁵. Katastrofering er kendt som en robust prædiktør for kronicitet¹⁰⁶ og er jf. "Fear-avoidance modellen" dét, der går forud for udviklingen af kinesiofobi¹⁰⁷. Selvom behandlere erkender at det psykiske domæne er væsentligt, føler de sig ikke kompetente til at behandle dette¹⁰⁸. Som fysioterapeut og ergoterapeut er bevægelse en naturlig tilgang til patienterne. I modsætning til PCS (Pain Catastrophizing Scale)¹⁰⁹ som identificerer katastrofetanker, identificerer TSK patienternes deraf følgende bevægeangst. Dette vil for mange terapeuter være en mere legitim tilgang til et psykologisk domæne, og det er vores håb, at vi med vores studie kan bidrage til at TSK-spørgeskemaet benyttes i klinisk praksis og dermed at flere terapeuter afprøver at arbejde i det psykologiske domæne, og oparbejder kompetencer der gør at de føler sig trygge, også i dette domæne.

10 REFERENCER

1. Vos T, Barber RM, Bell B, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2015;386(9995):743-800.
2. Flachs E, Eriksen L, Koch M, et al. *The Disease Burden in Denmark - Diseases [In Danish: Sygdomsbyrden i Danmark - Sygdomme]*. København; 2015.
3. Hoy D, March L, Brooks P, et al. The global burden of low back pain: Estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis*. 2014;73(6):968-974.
4. Pickering AP, Høgh M, Jensen N-H. *Smertebogen*. Munksgaard; 2015.
5. Engel GL. The Need for a New Medical Model: A Challenge for Biomedicine. *Science (80-)*. 1977;196(4286):129-136.
6. Andersen TE. Kognitiv adfærdsterapi for kroniske smerter. *Månedsskrift Prakt lægegering*. 2008;1104-1117.
7. Leeuw M, Goossens MEJB, Linton SJ, Crombez G, Boersma K, Vlaeyen JWS. The fear-avoidance model of musculoskeletal pain: Current state of scientific evidence. *J Behav Med*. 2007;30(1):77-94.
8. Vlaeyen JWS, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain*. 2000;85(3):317-332.
9. Lethem J, Slade PD, Troup JDG, Bentley G. Outline of a fear-avoidance model of exaggerated pain perception-I. *Behav Res Ther*. 1983;21(4):401-408.
10. Kori SH, Miller RP, Todd DD. Kinisophobia: A New View of Chronic Pain Behavior. *Pain Manag*. 1990;3:35-43.
11. Crombez G, Vlaeyen JWS, Heuts PHTG, Lysens R. Pain-related fear is more disabling than pain itself: Evidence on the role of pain-related fear in chronic back pain disability. *Pain*. 1999;80(1-2):329-339.
12. Vlaeyen JWS, Kole-Snijders AMJ, Boeren RGB, van Eek H. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. *Pain*. 1995;62(3):363-372.

13. Swinkels-Meewisse EJCM, Swinkels RAHM, Verbeek ALM, Vlaeyen JWS, Oostendorp RAB. Psychometric properties of the Tampa Scale for kinesiophobia and the fear-avoidance beliefs questionnaire in acute low back pain. *Man Ther.* 2003;8(1):29-36.
14. Lundberg M. Kinesiophobia - Various Aspects of Moving with Musculoskeletal Pain. 2006.
15. Vlaeyen JWS, de Jong J, Geilen M, Heuts P, van Breukelen G. The Treatment of Fear of Movement/(Re)injury in Chronic Low Back Pain: Further Evidence on the Effectiveness of Exposure In Vivo. *Clin J Pain.* 2002;18(4):251-261.
16. Miller RP, Todd DD, Kori SH. *Tampa Scale for Kinesiophobia.*; 1991.
17. Neblett R, Hartzell MM, Mayer TG, Bradford EM, Gatchel RJ. Establishing clinically meaningful severity levels for the Tampa Scale for Kinesiophobia (TSK-13). *Eur J Pain.* 2016;20(5):701-710.
18. de Moraes Vieira ÉB, de Góes Salvetti M, Damiani LP, de Mattos Pimenta CA. Self-Efficacy and Fear Avoidance Beliefs in Chronic Low Back Pain Patients: Coexistence and Associated Factors. *Pain Manag Nurs.* 2014;15(3):593-602.
19. Kreddig N, Rusu AC, Burkhardt K, Hasenbring MI. The German PASS-20 in Patients with Low Back Pain: New Aspects of Convergent, Divergent, and Criterion-Related Validity. *Int J Behav Med.* 2015;22(2):197-205.
20. Sleijser-Koehorst MLS, Bijker L, Cuijpers P, Scholten-Peeters GGM, Coppieters MW. Preferred self-administered questionnaires to assess fear of movement, coping, self-efficacy, and catastrophizing in patients with musculoskeletal pain-A modified Delphi study. *Pain.* 2019;160(3):600-606.
21. Goubert L, Crombez G, Van Damme S, Vlaeyen JWS, Bijttebier P, Roelofs J. Confirmatory Factor Analysis of the Tampa Scale for Kinesiophobia: Invariant Two-Factor Model Across Low Back Pain Patients and Fibromyalgia Patients. *Clin J Pain.* 2004;20(2):103-110.
22. Monticone M, Giorgi I, Baiardi P, Barbieri M, Rocca B, Bonezz C. Development of the Italian Version of the Tampa Scale of Kinesiophobia (TSK-I): Cross-Cultural Adaptation, Factor Analysis, Reliability, and Validity. *Spine (Phila Pa 1976).* 2010;35(12):1241-1246.
23. Archer KR, Phelps KD, Seebach CL, Song Y, Riley LH 3rd, Wegener ST. Comparative study of short forms of the Tampa Scale for Kinesiophobia: fear of movement in a surgical spine population. *Arch Phys Med Rehabil.* 2012;93(8):1460-1462.
24. Velthuis MJ, Van Den Bussche E, May AM, Gijsen BCM, Nijs S, Vlaeyen JWS. Fear of

movement in cancer survivors: Validation of the Modified Tampa Scale of Kinesiophobia - Fatigue. *Psychooncology*. 2012;21(7):762-770.

25. Nijs J, De Meirleir K, Duquet W. Kinesiophobia in chronic fatigue syndrome: Assessment and associations with disability. *Arch Phys Med Rehabil*. 2004;85(10):1586-1592.
26. Silver A, Haeney M, Vijayadurai P, Wilks D, Patrick M, Main CJ. The role of fear of physical movement and activity in chronic fatigue syndrome. *J Psychosom Res*. 2002;52(6):485-493.
27. Aguiar AS, Bataglion C, Visscher CM, Bevilaqua Grossi D, Chaves TC. Cross-cultural adaptation, reliability and construct validity of the Tampa scale for kinesiophobia for temporomandibular disorders (TSK/TMD-Br) into Brazilian Portuguese. *J Oral Rehabil*. 2017;44(7):500-510.
28. He S, Wang J, Ji P. Validation of the Tampa Scale for Kinesiophobia for Temporomandibular Disorders (TSK-TMD) in patients with painful TMD. *J Headache Pain*. 2016;17(1):109.
29. Visscher CM, Ohrbach R, van Wijk AJ, Wilkosz M, Naeije M. The Tampa Scale for Kinesiophobia for Temporomandibular Disorders (TSK-TMD). *Pain*. 2010;150(3):492-500.
30. Acar S, Savci S, Keskinoglu P, et al. Tampa scale of Kinesiophobia for Heart Turkish Version Study: cross-cultural adaptation, exploratory factor analysis, and reliability. *J Pain Res*. 2016;9:445-451.
31. Bäck M, Jansson B, Cider Å, Herlitz J, Lundberg M. Validation of a questionnaire to detect kinesiophobia (fear of movement) in patients with coronary artery disease. *J Rehabil Med*. 2012;44(4):363-369.
32. Moran RW, Rushworth WM, Mason J. Investigation of four self-report instruments (FABT, TSK-HC, Back-PAQ, HC-PAIRS) to measure healthcare practitioners' attitudes and beliefs toward low back pain: Reliability, convergent validity and survey of New Zealand osteopaths and manipulative physio. *Musculoskelet Sci Pract*. 2017;32:44-50.
33. George SZ, Lentz TA, Zeppieri G, Lee D, Chmielewski TL. Analysis of shortened versions of the tampa scale for kinesiophobia and pain catastrophizing scale for patients after anterior cruciate ligament reconstruction. *Clin J Pain*. 2012;28(1):73-80.
34. Monticone M, Ambrosini E, Rocca B, Foti C, Ferrante S. Responsiveness and minimal clinically important changes for the Tampa Scale of Kinesiophobia after lumbar fusion during cognitive behavioral rehabilitation. *Eur J Phys Rehabil Med*. 2017;53(3):351-358.

35. Houben RMA, Leeuw M, Vlaeyen JWS, Goubert L, Picavet HSJ. Fear of movement/injury in the general population: Factor structure and psychometric properties of an adapted version of the Tampa Scale for Kinesiophobia. *J Behav Med.* 2005;28(5):415-424.
36. Clark ME, Kori SH, Brockel J. Kinesiophobia and chronic pain: Psychometric characteristics and factor analysis of the Tampa Scale. In: *Proceedings of the American Pain Society Annual Meeting.* Vol 15. ; 1996:17.
37. Cook AJ, Brawer PA, Vowles KE. The fear-avoidance model of chronic pain: Validation and age analysis using structural equation modeling. *Pain.* 2006;121(3):195-206.
38. Geisser M, Haig A, Theisen M. Activity Avoidance and Function in Persons with Chronic Back Pain. *J Occup Rehabil.* 2000;10(3):215-227.
39. Roelofs J, Goubert L, Peters ML, Vlaeyen JWS, Crombez G. The Tampa Scale for Kinesiophobia: Further examination of psychometric properties in patients with chronic low back pain and fibromyalgia. *Eur J Pain.* 2004;8(5):495-502.
40. Roelofs J, Sluiter JK, Frings-Dresen MHW, et al. Fear of movement and (re)injury in chronic musculoskeletal pain: Evidence for an invariant two-factor model of the Tampa Scale for Kinesiophobia across pain diagnoses and Dutch, Swedish, and Canadian samples. *Pain.* 2007;131(1-2):181-190.
41. Tkachuk GA, Harris CA. Psychometric properties of the Tampa Scale for Kinesiophobia-11 (TSK-11). *J Pain.* 2012;13(10):970-977.
42. Walton D, Elliott JM. A Higher-Order Analysis Supports Use of the 11-Item Version of the Tampa Scale for Kinesiophobia in People With Neck Pain. *Phys Ther.* 2013;93(1):60-68.
43. Woby SR, Roach NK, Urmston M, Watson PJ. Psychometric properties of the TSK-11: A shortened version of the Tampa Scale for Kinesiophobia. *Pain.* 2005;117(1-2):137-144.
44. Wong WS, Kwok HY, Luk KDK, et al. Fear of movement/(re)injury in Chinese patients with chronic pain: Factorial validity of the Chinese version of the Tampa Scale for Kinesiophobia. *J Rehabil Med.* 2010;42(7):620-629.
45. Beyer N, Magnusson P, Thorborg K. *Målemetoder i Forebyggelse, Behandling Og Rehabilitering.* 2nd ed. Copenhagen: Munksgaard; 2012.
46. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63(7):737-745.

47. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27(5):1147-1157.
48. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res.* 2018;27(5):1171-1179.
49. Lundberg MKE, Styf J, Carlsson SG. A psychometric evaluation of the Tampa Scale for Kinesiophobia - From a physiotherapeutic perspective. *Physiother Theory Pract.* 2004;20(2):121-133.
50. Bunzli S, Smith A, Watkins R, Schutze R, O'Sullivan P. What Do People Who Score Highly on the Tampa Scale of Kinesiophobia Really Believe?: A Mixed Methods Investigation in People With Chronic Nonspecific Low Back Pain. *Clin J Pain.* 2015;31(7):621-632.
51. Mokkink LB, Prinsen CAC, Patrick DL, et al. *COSMIN Methodology for Systematic Reviews of Patient - Reported Outcome Measures (PROMs).*; 2018.
52. Damsgård E, Fors T, Anke A, Røe C. The tampa scale of kinesiophobia: A rasch analysis of its properties in subjects with low back and more wide spread pain. *J Rehabil Med.* 2007;39(9):672-678.
53. Haugen AJ, Grøvle L, Keller A, Grotle M. Cross-cultural adaptation and validation of the Norwegian version of the tampa scale for kinesiophobia. *Spine (Phila Pa 1976).* 2008;33(17):E595-E601.
54. De Souza FS, Marinho C da S, Siqueira FB, Maher CG, Costa LOP. Psychometric testing confirms that the Brazilian-Portuguese adaptations, the original versions of the fear-avoidance beliefs questionnaire, and the Tampa Scale of Kinesiophobia have similar measurement properties. *Spine (Phila Pa 1976).* 2008;33(9):1028-1033.
55. Gómez Pérez L, López Martínez AE, Ruíz Párraga GT. Reliability and validity of the spanish version of the Tampa Scale for Kinesiophobia (TSK) in chronic and acute pain patients. In: *European Journal of Pain.* Vol 13. L. Gómez Pérez, Departamento de Personalidad, Evaluación y Tratamiento Psicológico, Universidad de Málaga, Málaga, Spain; 2009:S220.
56. Jafari H, Ebrahimi I, Salavati M, Kamali M, Fata L. Psychometric Properties of iranian Version of Tampa Scale for Kinesiophobia in Low Back Pain Patients. *Arch Rehabil.* 2010;11(1):0-0.
57. Cordeiro N, Pezarat-Correia P, Gil J, Cabri J. Portuguese language version of the Tampa Scale for Kinesiophobia [13 items]. *J Musculoskelet Pain.* 2013;21(1):58-63.

58. Rusu AC, Kreddig N, Hallner D, Hülsebusch J, Hasenbring MI. Fear of movement/(re)injury in low back pain: Confirmatory validation of a German version of the Tampa Scale for Kinesiophobia. *BMC Musculoskelet Disord.* 2014;15(1):280.
59. Kikuchi N, Matsudaira K, Sawada T, Oka H. Psychometric properties of the Japanese version of the Tampa Scale for Kinesiophobia (TSK-J) in patients with whiplash neck injury pain and/or low back pain. *J Orthop Sci.* 2015;20(6):985-992.
60. Wei X, Xu X, Zhao Y, Hu W, Bai Y, Li M. The Chinese version of the Tampa Scale for Kinesiophobia was cross-culturally adapted and validated in patients with low back pain. *J Clin Epidemiol.* 2015;68(10):1205-1212.
61. Areeudomwong P, Butttagat V. Reliability and validity of the cross-culturally adapted thai version of the tampa scale for kinesiophobia in knee osteoarthritis patients. *Malaysian J Med Sci.* 2017;24(2):61-67.
62. Pedersen M, Fink P, Oernboel E, Carstensen T, Dantoft T, Frostholm L. *Translation and Psychometric Examination of the Danish Version of The Tampa Scale for Kinesiophobia.*; 2019.
63. Larsson C, Hansson EE, Sundquist K, Jakobsson U. Psychometric properties of the Tampa Scale of Kinesiophobia (TSK-11) among older people. *Physiother Theory Pract.* 2014;30(6):421-428.
64. Heuts PHTG, Vlaeyen JWS, Roelofs J, et al. Pain-related fear and daily functioning in patients with osteoarthritis. *Pain.* 2004;110(1-2):228-235.
65. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34-42.
66. Ostelo RWJG, Swinkels-Meewisse IJCM, Knol DL, Vlaeyen JWS, de Vet HCW. Assessing pain and pain-related fear in acute low back pain: what is the smallest detectable change? *Int J Behav Med.* 2007;14(4):242-248.
67. Gómez-Pérez L, López-Martínez AE, Ruiz-Párraga GT. Psychometric properties of the spanish version of the Tampa Scale for Kinesiophobia (TSK). *J Pain.* 2011;12(4):425-435.
68. Bunketorp L, Carlsson J, Kowalski J, Stener-Victorin E. Evaluating the reliability of multi-item scales: A non-parametric approach to the ordered categorical structure of data collected with the Swedish version of the Tampa Scale for Kinesiophobia and the Self-Efficacy Scale. *J Rehabil Med.* 2005;37(5):330-334.

69. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297-334.
70. Devet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine*. 8th ed. United Kingdom: Cambridge university press; 2017.
71. Mokkink LB, Terwee CB, Patrick DL, et al. 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*. 2010;19(4):539-549.
72. Hapidou EG, O'Brien MA, Pierrynowski MR, de Las Heras E, Patel M, Patla T. Fear and Avoidance of Movement in People with Chronic Pain: Psychometric Properties of the 11-Item Tampa Scale for Kinesiophobia (TSK-11). *Physiother Can*. 2012;64(3):235-241.
73. George SZSZ, Valencia C, Beneciuk JMJM. A Psychometric investigation of fear-avoidance model measures in patients with chronic low back pain. *J Orthop Sports Phys Ther*. 2010;40(4):197-205.
74. Mokkink LB. *COSMIN Risk of Bias Checklist*; 2018.
75. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. *J Clin Nurs*. 2005;14(7):798-804.
76. Koho P, Aho S, Kautiainen H, Pohjolainen T, Hurri H. Test-retest reliability and comparability of paper and computer questionnaires for the Finnish version of the Tampa Scale of Kinesiophobia. *Physiother (United Kingdom)*. 2014;100(4):356-362.
77. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2005;113(1):9-19.
78. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain*. 2004;8(4):283-291.
79. Pollard CA. Preliminary Validity Study of the Pain Disability Index. *Percept Mot Skills*. 2011;59(3):974-974.
80. Soer R, Köke AJA, Vroomen PCAJ, et al. Extensive validation of the pain disability index in 3 groups of patients with musculoskeletal pain. *Spine (Phila Pa 1976)*. 2013;38(9).
81. Shapiro SS, Wilk MB. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*. 1965;52(3/4):591.

82. Habicht A. *Medicinsk Statistik*. København: Munksgaard; 2017.
83. Streiner D, Norman G. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 3rd editio. New York: Oxford University Press Inc; 2003.
84. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163.
85. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240.
86. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59(10):1033-1039.
87. Wertli MM, Rasmussen-barr E, Weiser S, Bachmann LM, Brunner F. Re: Wertli MM, Rasmussen-Barr E, Weiser S, et al. The role of fear avoidance beliefs as a prognostic factor for outcome in patients with nonspecific low back pain: a systematic review. *Spine J* 2014;14:816–36.e4. *Spine J*. 2014;14(8):A18.
88. Bränström H, Fahlström M. Kinesiophobia in patients with chronic musculoskeletal pain: Differences between men and women. *J Rehabil Med*. 2008;40(5):375-380.
89. Vlaeyen JWS, Seelen HAM, Peters M, et al. Fear of movement/(re)injury and muscular reactivity in chronic low back pain patients: An experimental investigation. *Pain*. 1999;82(3):297-304.
90. Barke A, Baudewig J, Schmidt-Samoa C, Dechent P, Kröner-Herwig B. Neural correlates of fear of movement in high and low fear-avoidant chronic low back pain patients: An event-related fMRI study. *Pain*. 2012;153(3):540-552.
91. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;XX(1):37-46.
92. Askary-Ashtiani A, Ebrahimi-Takamejani I, Torkaman G, Amiri M, Mousavi SJ. Reliability and validity of the persian versions of the fear avoidance beliefs questionnaire and tampa scale of kinesiophobia in patients with neck pain. *Spine (Phila Pa 1976)*. 2014;39(18):1095-1102.
93. Lamé IE, Peters MLML, Kessels AGAGAG, et al. Test-retest stability of the pain catastrophizing scale and the Tampa Scale for Kinesiophobia in chronic pain patients over a longer period of time. *J Health Psychol*. 2008;13(6):820-826.

94. Koho P, Borodulin K, Kautiainen H, Kujala U, Pohjolainen T, Hurri H. Finnish version of the Tampa Scale of Kinesiophobia: Reference values in the Finnish general population and associations with leisure-time physical activity. *J Rehabil Med.* 2015;47(3):249-255.
95. Semple L. What do people with persistent pain think about when they complete the tampa scale of kinesiophobia (TSK)? a qualitative study. In: *British Journal of Pain.* Vol 12. L. Semple, Glasgow Pain Management Programme NHS Greater Glasgow and Clyde, United Kingdom; 2018:7-8.
96. Lundberg M, Grimby-Ekman A, Verbunt J, Simmonds MJ. Pain-related fear: A critical review of the related measures. *Pain Res Treat.* 2011;2011.
97. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ.* 2011;2:53-55.
98. Bland M. J, Altman G D. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8(99):135-160.
99. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of Electronic and Paper-and-Pencil Administration of Patient-Reported Outcome Measures : A Meta-Analytic Review. *Pharmacoeconomics.* 2008;11(2):322–333.
100. Woby SR, Watson PJ, Roach NK, Urmston M. Are changes in fear-avoidance beliefs, catastrophizing, and appraisals of control, predictive of changes in chronic low back pain and disability? *Eur J Pain.* 2004;8(3):201-210.
101. Boersma K, Linton S, Overmeer T, Jansson M, Vlaeyen J, De Jong J. Lowering fear-avoidance and enhancing function through exposure in vivo A multiple baseline study across six patients with back pain. *Pain.* 2004;(108):8-16.
102. Monticone M, Ambrosini E, Rocca B, Foti C, Ferrante S. Responsiveness of the Tampa Scale of Kinesiophobia in Italian subjects with chronic low back pain undergoing motor and cognitive rehabilitation. *Eur Spine J.* 2016;25(9):2882-2888.
103. Lüning Bergsten C, Lundberg M, Lindberg P, Elfving B. Change in kinesiophobia and its relation to activity limitation after multidisciplinary rehabilitation in patients with chronic back pain. *Disabil Rehabil.* 2012;34(10):852-858.
104. Loeser JD, Schatman ME. Chronic pain management in medical education: a disastrous omission. *Postgrad Med.* 2017;129(3):332-335.
105. Colleary G, O'Sullivan K, Griffin D, Ryan CG, Martin DJ. Effect of pain neurophysiology education on physiotherapy students' understanding of chronic pain, clinical

recommendations and attitudes towards people with chronic pain: a randomised controlled trial. *Physiother (United Kingdom)*. 2017;103(4):423-429.

106. Keefe FJ, Rumble ME, Scipio CD, Giordano LA, Perri LM. Psychological Aspects of Persistent Pain: Current State of the Science. *J Pain*. 2004;5(4):195-211.
107. Vlaeyen JWS, Kole-Snijders AMJ, Rotteveel AM, Ruesink R, Heuts PHTG. The role of fear of movement/(re)injury in pain disability. *J Occup Rehabil*. 1995;5(4):235-252.
108. Alexanders J, Anderson A, Henderson S. Musculoskeletal physiotherapists' use of psychological interventions: a systematic review of therapists' perceptions and practice. *Physiotherapy*. 2015;101(2):95-102.
109. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and Validation. *Am Psychol Assoc*. 1995;7(4):524-532.