



HØGSKOLEN  
I BERGEN

---

BERGEN UNIVERSITY COLLEGE

Kartlegging av pustebesvær hos palliative pasienter  
-en systematisk kunnskapsoversikt over tilgjengelige  
kartleggingsverktøy og deres måleegenskaper

Measurement of breathlessness in palliative care  
-a systematic review of available measurement  
instruments and their psychometric properties

**Kjersti Solvåg**

**Mastergrad i kunnskapsbasert praksis i helsefag**

**Senter for kunnskapsbasert praksis**

**Avdeling for helse- og sosialfag**

**Innleveringsdato 15.05.15**

# **Mastergradsprosjekt**

Kartlegging av pustebesvær hos palliative pasienter  
-en systematisk kunnskapsoversikt over tilgjengelige  
kartleggingsverktøy og deres måleegenskaper

Measurement of breathlessness in palliative care  
-a systematic review of available measurement  
instruments and their psychometric properties

**Student: Kjersti Solvåg**

**Veileder: Kjetil Gundro Brurberg**

**Innleveringsdato 15.05.15**

**Antall ord: 12 899**

## Forord

Dame Cicely Saunders introduserte banebrytende tanker om forståelsen av symptomer hos alvorlig syke med begrepet «total smerte» som innbefatter både fysisk, psykisk, sosial og åndelig smerte. Denne måten å forstå og tilnærme seg pasientene og deres plager er grunnleggende i lindrende behandling og omfatter ikke bare smerte, men også andre symptomer som pustebesvær. Cicely Saunders var også tidlig opptatt av at faget måtte bygge på like deler klinikk, undervisning og forskning.

Dette var til inspirasjon når jeg som sykepleier i spesialisthelsetjenesten for lindrende behandling ønsket å ta en masterutdanning i kunnskapsbasert praksis og når jeg valgte tema for mitt mastergradsprosjekt. Kunnskapsbasert praksis står med tredelingen av kunnskapskildene i forskningsbasert-, erfaringsbasert- og brukerkunnskap på mange måter for den samme tilnærmingen til faget som Cicely Saunders gjorde og bidrar til at vi bygger våre helsetjenester på beste tilgjengelige kunnskap og på pasientens ønsker og behov. Sammen med kolleger har jeg de siste årene hatt et spesielt fokus på pasienter med pustebesvær. For å evaluere effekt av tiltak og endringer i helsetilstand er det viktig å forstå pasientens egenopplevelse. Til det trenger vi gode kartleggingsverktøy. Mitt mastergradsprosjekt har derfor vært å oppsummere tilgjengelig forskning om kartleggingsverktøy som fanger både den fysiske dimensjonen og angsten som er relatert til pustebesværet.

Mange fortjener en takk for at det har vært mulig for meg å gjennomføre dette prosjektet. Først og fremst min aller nærmeste familie og mine beste venner. Min mann Kenneth, våre barn Amanda og Mathias, min mor, Astrid og Erik, har gitt meg gode pusterom og tålmodig holdt ut med at det meste av min tid og oppmerksomhet har handlet om studiet. Uten min fantastiske leder Fanny Henriksen og mine gode kolleger hadde det ikke vært mulig å kombinere en krevende jobb med deltidsstudier.

Tusen takk også til alle mine medstudenter og i særdeleshet til Kristine, Hilde og Jill! Sist, men ikke minst fortjener Kjetil Gundro Brurberg en stor takk! Gjennom sin omfattende fagkunnskap, trygge veiledning og optimistiske tilbakemeldinger har han alltid gitt meg troen på at det har vært mulig å komme i havn med prosjektet.

## Sammendrag

**Bakgrunn:** Pustebeviser er et vanlig og belastende symptom hos mange pasienter med alvorlig langtkommen sykdom. På grunn av symptomets subjektive natur trenger vi verktøy for kartlegging av pasientens egenopplevelse, pasientrapporterte resultatmål (PROMs).

**Hensikt:** å identifisere tilgjengelige PROMs for kartlegging av pustebeviser hos palliative pasienter og oppsummere deres måleegenskaper. Kartleggingsverktøyene måtte inneholde dimensjoner for både pustebeviser og angst for å bli vurdert.

**Metode:** Et systematisk litteratursøk ble gjennomført i mars 2014. To personer screenet alle identifiserte referanser for relevans og kritisk vurderte inkluderte studier uavhengig av hverandre. COSMIN-sjekklisten ble benyttet til vurdering av studienes metodiske kvalitet. Relevante data ble hentet ut av en og dataekstraksjonen ble kvalitetssikret av en annen. Måleegenskapene til de inkluderte kartleggingsverktøyene ble oppsummert i en kvalitativ syntese.

**Hovedresultat:** Vi screenet 1458 referanser for relevans og inkluderte 16 publikasjoner som evaluerte måleegenskaper ved fire ulike PROMs: CDS, SRI, DMQ og en respiratorisk symptomsjekkliste. Bare SRI er validert i norsk versjon. Ingen av de inkluderte PROMs er direkte validert for bruk i en palliativ setting. De tilgjengelige valideringsstudier antyder at de aktuelle kartleggingsverktøyene har positive måleegenskaper, men siden viktige måleegenskaper ikke er vurdert og siden mange studier har metodiske svakheter, er det vanskelig å trekke sikre konklusjoner.

**Konklusjon:** Metodiske svakheter i studiene og manglende evaluering av viktige måleegenskaper medfører at den samlede kvaliteten til PROMs for pustebeviser og angst er uavklart. Videre forskning er nødvendig for å oversette og validere kartleggingsverktøyene for bruk i lindrende behandling i Norge.

**Omfang:** Masteroppgaven består av en systematisk oversikt og en innledningsdel. I innledningdelen utdypes metodiske aspekt, i tillegg til en mer omfattende rapportering av resultater samt en utvidet diskusjon med metodekritikk.

**Nøkkelord:** Palliasjon, dyspné, oversikt, spørreskjema, psykometri

## **Abstract**

**Background:** Breathlessness is a common and distressing symptom for many patients with advanced disease. Due to the subjective nature of breathlessness, patient-reported outcome measures (PROMs) are required to measure the patient's own experience.

**Aim:** to identify PROMs used to measure breathlessness in palliative care and to synthesise their measurement properties. Instruments had to include dimensions for breathlessness and anxiety to be considered.

**Method:** We performed a systematic search for literature in March 2014. Two reviewers independently screened all references for relevance and critically appraised included studies with the COSMIN-checklist. One reviewer extracted relevant data, and a second reviewer evaluated the extractions. We performed a best evidence synthesis to summarize the measurement properties of each included PROMs.

**Results:** We screened 1458 references for relevance, and included 16 publications evaluating the measurement properties of four different PROMs: CDS, DMQ, SRI and a respiratory symptom checklist. Only SRI was available in a validated Norwegian version, and none of the included PROMs were directly validated for use in palliative care. The available PROMs have promising measurement properties, but it is difficult to draw firm conclusions because important measurement properties are not evaluated yet, and many available studies are associated with methodological shortcomings.

**Conclusion:** The overall quality of PROMs for breathlessness and anxiety remains unclear because important measurement properties are not yet evaluated and many validation studies are associated with methodological shortcomings. Further research is needed to translate and validate the PROMs for use in palliative care in Norway.

**Scope:** This master's thesis consists of a systematic review and an introduction section. The introduction section present an elaboration on methodology and includes a more comprehensive reporting of results as well as an extended discussion with methodology critique.

**Keyword:** Palliative care, dyspnea, review, questionnaires, psychometrics

# Innholdsfortegnelse

<b>1</b>	<b>Bakgrunn</b> .....	11
1.1	Lindrende behandling, populasjon og setting.....	11
1.2	Pustebesvær .....	12
1.3	Pasientrapportert resultatmål (PROM) .....	13
1.4	Måleegenskaper .....	14
1.5	Oversikt over tidligere forskning.....	17
<b>2</b>	<b>Hensikt og forskningsspørsmål</b> .....	18
<b>3</b>	<b>Metode</b> .....	18
3.1	COSMIN.....	19
3.2	Inklusjons- og eksklusjonskriterier.....	20
3.3	Søk etter litteratur .....	21
3.4	Utvelgelse av studier .....	22
3.5	Kvalitetsvurdering av inkluderte studier .....	23
3.6	Dataekstraksjon .....	24
3.7	Sammenstilling og gradering .....	24
<b>4</b>	<b>Resultater</b> .....	26
4.1	Populasjon og setting.....	29
4.2	Inkluderte studiers metodiske kvalitet.....	30
4.3	Resultater for inkluderte kartleggingsverktøy .....	35
4.4	Sammenstilling og gradering av kartleggingsverktøyenes måleegenskaper .....	45
<b>5</b>	<b>Diskusjon</b> .....	47
5.1	Konsekvenser av metodiske valg, oversiktens styrker og svakheter.....	47
5.2	Metodiske utfordringer ved de inkluderte studiene.....	50
5.3	Kartleggingsverktøyenes måleegenskaper .....	52
5.4	Betydning for praksis og videre forskning .....	53
<b>6</b>	<b>Konklusjon</b> .....	54
<b>7</b>	<b>Referanseliste</b> .....	55

Artikkel      Kartleggingsverktøy for pustebesvær hos palliative pasienter: en systematisk oversikt

Vedlegg I      Søkestrategier

Vedlegg II     COSMIN-sjekklisten med firepunkts-skala

Vedlegg III    Kategorisering av måleegenskaper

Vedlegg IV    Ekskluderte publikasjoner med begrunnelse

- Vedlegg V Dataekstraksjonsskjema for inkluderte publikasjoner
- Vedlegg VI Oversikt over metodisk kvalitet og måleegenskaper fra hver studie
- Vedlegg VII Oversikt over tilgjengelige samvariasjoner for hvert kartleggingsverktøy
- Vedlegg VIII Beskrivelse av kartleggingsverktøy brukt til sammenligninger
- Vedlegg IX Oversikt over samvariasjoner mellom SRI og SF-36
- Vedlegg X Oversikt over samvariasjoner mellom SRI og CRQ, HADS, GARS og MRC

## Oversikt over figurer og tabeller

Forkortelser og begrepsforklaringer

Figur 1 COSMIN-taksonomi over sammenhenger mellom måleegenskaper

Figur 2 Flytskjema over identifisert litteratur med begrunnelse for eksklusjon

Tabell 1 Gradering av tillit til kartleggingsverktøys måleegenskaper

Tabell 2 Karakteristika ved inkluderte studier

Tabell 3 Karakteristika ved inkluderte PROMs

Tabell 4 CDS: inkluderte studiers metodiske kvalitet per måleegenskap

Tabell 5 SRI: inkluderte studiers metodisk kvalitet per måleegenskap

Tabell 6 DMQ: inkluderte studiers metodiske kvalitet per måleegenskap

Tabell 7 Utvalgte samvariasjoner for CDS

Tabell 8 Utvalgte samvariasjoner mellom SRI og SF-36

Tabell 9 Utvalgte samvariasjoner mellom SRI og CRQ, HADS og MRC

Tabell 10 Intern konsistens SRI

Tabell 11 Samvariasjoner for DMQ-30

Tabell 12 Samvariasjoner for DMQ-CAT

Tabell 13 Kvalitativ syntese



## Forkortelser og begrepsforklaringer

Begrep	Forklaring
CAT	Computer Adaptive Test
CDS	Cancer Dyspnea Scale
CFA	Bekreftende faktoranalyse (Confirmatory Factor Analysis)
COSMIN	Consensus-based Standards for the selection of health Measurement Instruments
Dimensjon	Aspekt ved et fenomen. Dimensjon, domene og faktor benyttes til dels om hverandre i internasjonal litteratur, i denne oppgaven benyttes i hovedsak dimensjon. For PROMs med mange spørsmål gjennomføres oftest faktoranalyse for å bestemme hvor mange ulike dimensjoner ved fenomenet det måler og hvilke spørsmål som måler hvilken dimensjon.
DMQ	Dyspnea Management Questionnaire
ECOG	Eastern Cooperative Oncology Group
EFA	Utforskende faktoranalyse (Exploratory Factor Analysis)
ESAS-r	Edmonton Symptom Assessment System-revised
GOLD	Global Initiative for Chronic Obstructive Lung Disease
ICC	Intraklasse korrelasjonskoeffisient (Intraclass Correlation Coefficient)
Kartleggingsverktøy	Benyttes i denne oppgaven for alle typer måleinstrumenter, tester og spørreskjema som kartlegger pasientrapporterte resultatmål.
Kols	Kronisk obstruktiv lungesykdom (COPD: Chronic Obstructive Lung Disease)
LoA	Limits of agreement
LTMV	Langtids mekanisk ventilering (Long term mechanical ventilation). En ventilator er et mekanisk hjelpemiddel som helt eller delvis kan kompensere for sviktende ventilasjon og bidra til at den opprettholdes på et tilfredsstillende nivå med normalisering av blodgassene. Hjemmerespiratorbehandling benyttes også som begrep for denne typen behandling.
LTOT	Langtids oksygenbehandling (long-term oxygen therapy)
NRS	Numerisk skala
Palliasjon	Lindrende behandling, pleie og omsorg.
PROM	Pasientrapporterte resultatmål (Patient Reported Outcome Measure). Data som måles med PROMs som er kartleggingsverktøy der informasjonen kommer direkte fra pasienten selv og ikke er påvirket eller tolket av andre.
Pustebesvær	En subjektivopplevelse av ubehag ved pusting. Det er et multidimensjonalt symptom som kan variere i intensitet og består av kvalitativt ulike opplevelser. Tungpust, åndenød og dyspné benyttes ofte synonymt, i denne oppgaven benyttes i hovedsak pustebesvær.

<b>RSC</b>	Respiratorisk symptomsjekkliste (Respiratory Symptom Checklist)
<b>SEM</b>	Standardfeilen til målingen (Standard Error of Measurement)
<b>SRI</b>	Severe Respiratory Insufficiency Questionnaire
<b>SDC</b>	Minste målbare forskjell (Smallest Detectable Change)
<b>VAS</b>	Visuell analog skala (Visual Analog Scale)
<b>WHO</b>	Verdens helseorganisasjon (The World Health Organization)

# 1 Bakgrunn

Pustebeviser er et av de vanligste symptomene hos pasienter med langtkommet kreftsykdom og hovedsymptomet ved alvorlige stadier av lungekreft og kronisk obstruktiv lungesykdom (kols) (Helsedirektoratet, 2012, s. 115 ; 2013, s. 64). På verdensbasis er det årlig mer enn 75 millioner mennesker med pustebeviser (Higginson et al., 2014), og som følge av at bedre medisinsk behandling bedrer prognosen for de underliggende sykdommene forventer vi en økning i dette antallet (Gysels, Bausewein & Higginson, 2007).

I mitt kliniske arbeid som spesialsykepleier i et senter for lindrende behandling møter jeg daglig pasienter som plages med pustebeviser. Jeg opplever at dette symptomet medfører mye angst som igjen forverrer pusten. Ved senteret er jeg en del av et tverrfagligteam som arbeider for å bedre tilbudet til denne pasientgruppen. Vi mangler et kartleggingsverktøy på norsk som kan benyttes når vi har behov for en utvidet kartlegging av pustebeviser og måle effekt av tiltak. Det var derfor naturlig å velge et tema innenfor denne problemstillingen når jeg skulle skrive min masteroppgave i kunnskapsbasert praksis.

Formålet med masteroppgaven har vært å identifisere validerte kartleggingsverktøy for pustebeviser hos palliative pasienter som er egnet for bruk i klinisk praksis.

Kartleggingsverktøyet må minst inneholde dimensjoner for pustebeviser og angst. Oppgaven består av en systematisk oversiktsartikkel som planlegges publisert i Sykepleien Forskning og en innledningsdel. I innledningsdelen er det en utvidet beskrivelse av metodiske aspekt, en mer omfattende rapportering av resultater og en diskusjon med metodekritikk. I tillegg drøftes resultatene og relevans for praksis og fremtidig forskning. Det vil være noe overlapp i rapporteringen mellom innledningen og artikkelen for å få nødvendig sammenheng.

## 1.1 Lindrende behandling, populasjon og setting

Lindrende behandling eller palliasjon er aktiv behandling, pleie og omsorg for pasienter som har en uhelbredelig sykdom og begrenset levetid. Målet med lindrende behandling er å bidra til at pasienten og de pårørende har best mulig livskvalitet. Å forebygge og lindre symptomer er sentralt, sammen med tiltak for å forebygge og avhjelpe psykiske, sosiale og åndelige/eksistensielle problemer (European Association for Palliative Care, 2012 ; World Health Organization, 2015). Palliasjon praktiseres på alle områder og nivåer i helsetjenesten

(Helsedirektoratet, 2013, s. 10) og er uavhengig av om pasienten er hjemme eller på sykehus (European Association for Palliative Care, 2012).

Det mangler internasjonal enighet om når en pasient er i palliativ fase. I noen anbefalinger avgrenses populasjonen til pasienter med antatt forventet levetid mindre enn ni til tolv måneder. En bedre definisjon er antakelig at den palliative fasen starter når det erkjennes at sykdommen er uhelbredelig (Helsedirektoratet, 2013, s. 11). Verdens helseorganisasjon (2015) understreker at prinsippene for lindrende behandling er anvendbare også tidlig i sykdomsforløpet parallelt med livsforlengende behandling.

I følge Nasjonalt handlingsprogram for palliasjon i kreftomsorgen (2013, s. 11) har omtrent 95 % av pasientene på palliative enheter på sykehus kreft, og de er også den største gruppen i primærhelsetjenesten. Men det er en økende erkjennelse av behovet for å tilby lindrende behandling også til pasienter med avanserte stadier av alvorlig ikke-kreftrelatert sykdom (Boland et al., 2013 ; Johnson & Fallon, 2013).

I denne systematiske oversikten har jeg definert den palliative pasientpopulasjonen til å være uavhengig av diagnose og den palliative fasen fra det erkjennes at pasienten har en uhelbredelig sykdom.

## 1.2 Pustebesvær

Dyspné er en subjektiv opplevelse av pustebesvær ved sykdom som påvirker respirasjonssystemet (Helsedirektoratet, 2013, s. 63). American Thoracic Society (1999, s. 322) definerer dyspné som en subjektiv opplevelse av ubehag ved pusting som består av kvalitativt ulike opplevelser og som varierer i intensitet. Begreper som tungpust, pustebesvær, dyspné og åndenød benyttes ofte synonymt. I denne oppgaven har jeg i hovedsak benyttet pustebesvær.

Pustebesvær er et komplekst og sammensatt fenomen (Gysels, Bausewein & Higginson, 2007, s. 288). Det er som smerter multidimensjonalt i sin natur og består i tillegg til fysiske elementer også av følelsesmessige aspekt (Dudgeon, 2010, s. 303). Noen pasienter med langtkommet sykdom lider av kontinuerlig pustebesvær, men mange beskriver episodisk pustebesvær eller en kombinasjon av begge (Reddy et al., 2009 ; Simon et al., 2013b). Episoder med pustebesvær kan utløse angst, panikk og til og med frykt for å dø (Gysels, 2007,

referert i Simon et al., 2013a). Den følelsesmessige responsen på episodisk pustebesvær kan føre til en ond sirkel som i seg selv utløser pustebesvær (Bailey, 2004).

Diagnostisering og behandling av underliggende årsaker er å foretrekke når det er mulig (Parshall et al., 2012, s. 435; Helsedirektoratet, 2013, s. 65), men for mange pasienter er årsaksforholdene uklare eller pustebesværet vedvarende på tross av optimal behandling (Parshall et al., 2012, s. 435). Med bakgrunn i symptomets subjektive natur og manglende sammenheng mellom objektive funn og pasientens opplevelse, trenger vi i tillegg til tradisjonell diagnostikk å kartlegge pasientens egenopplevelse (Bausewein, Booth & Higginson, 2008 ; Parshall et al., 2012, s. 436-7 ; Helsedirektoratet, 2013, s. 12).

### 1.3 Pasientrapportert resultatmål (PROM)

For kartlegging av symptomer er kartleggingsverktøy for pasientrapporterte resultatmål (PROM) viktige verktøy (Bausewein et al., 2011, s. 3). PROMs defineres som kartleggingsverktøy der informasjon om helsetilstand og/eller livskvalitet kommer direkte fra pasienten og ikke er påvirket eller tolket av andre (U.S. Department of Health and Human Services Food and Drug Administration, 2009, s. 2).

PROMs kan være selvadministrerte, intervjubaserte eller en kombinasjon, og data kan samles inn papirbasert, elektronisk eller per telefon (U.S. Department of Health and Human Services Food and Drug Administration, 2009, s. 13-4). Kartleggingsverktøyene kan benyttes i klinisk arbeid, til kartlegging av praksis og til forskning (Bausewein et al., 2011, s. 4).

Vi deler PROMs i to hovedgrupper: generiske og spesifikke. Generiske kartleggingsverktøy er multidimensjonale og inkluderer fysiske, psykiske og sosiale helseaspekt. De kan benyttes for en lang rekke helse- og livskvalitetskonsept og ved ulike helsetilstander, populasjoner og tiltak. Spesifikke kartleggingsverktøy er utviklet for spesielle helsetilstander, symptomer, populasjoner eller lignende og har gjerne større evne enn generiske kartleggingsverktøy til å fange opp klinisk meningsfulle endringer (Bausewein et al., 2011, s. 6-7). Spesifikke kartleggingsverktøy er derfor med sitt målrettede fokus mer relevante for pasienter og klinikere enn generiske kartleggingsverktøy (Garratt & Bjertnæs, 2014).

PROMs kan bestå av enkelt spørsmål som for eksempel «hvor tung er du i pusten». Det kan også bestå av mange spørsmål, innenfor en dimensjon (endimensjonalt) eller innenfor flere

dimensjoner av et fenomen (multidimensjonalt) for eksempel fysiske og psykiske dimensjoner ved pustebesvær (Bausewein et al., 2011, s. 5-9).

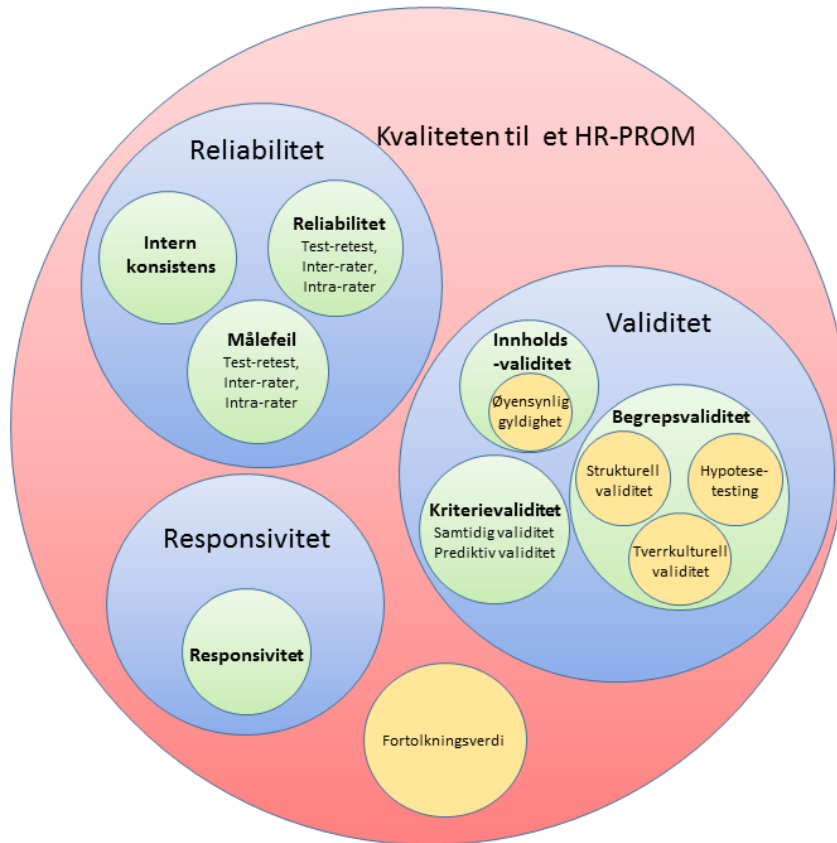
I lindrende behandling benytter vi validerte PROMs systematisk for å identifisere symptomer, innhente informasjon om sykdomsutvikling og vurdere effekt av behandling (Antunes, Harding & Higginson, 2014, s. 2) og vi registrerer regelmessig symptomer hos pasientene med Edmonton Symptom Assessment System-revised (ESAS-r) (Helsedirektoratet, 2013, s. 12). ESAS-r er et screeningverktøy som kartlegger tilstedeværelse og intensitet av de vanligste symptomene hos palliative pasienter med kreft i tillegg til velvære (Watanabe et al., 2011) og er også validert for andre pasientgrupper (Davison, Jhangri & Johnson, 2006). Det mangler dokumentasjon for hvilken terskelverdier som bør utløse en mer omfattende symptomkartlegging, men ofte benyttes en symptomskår på fire eller mer som tegn på at det er behov for en utvidet kartlegging (Oldenmenger et al., 2013).

For kartlegging av pustebesvær trenger vi et symptomspesifikt PROM som måler flere dimensjoner ved pustebesværet, for eksempel den fysiske opplevelsen av pusten og angst relatert til pustebesværet. Kvaliteten til et PROM er avhengig av målegenskapene. Vi trenger derfor kunnskap fra forskning om kartleggingsverktøyets målegenskaper for å vite om det er egnet for vårt formål, måler fenomenet det skal måle og om vi kan stole på resultatet (Mokkink et al., 2010b ; Bausewein et al., 2011, s. 12).

#### 1.4 Måleegenskaper

Ulike forfattere for eksempel Polit and Beck (2012) og nettstedet PsykTest (Nasjonalt kunnskapssenter for helsetjenesten, 2012) benytter forskjellige begrep og definisjoner på måleegenskaper. Ulik bruk av terminolog og forskjeller i definisjoner kan skape uklarhet om hvilke måleegenskaper det er snakk om, hvilke konsept måleegenskapen representerer og hvordan den bør evalueres (Mokkink et al., 2010c). I denne systematiske oversikten har vi valgt å benytte taksonomi, terminologi og definisjoner som det ble oppnådd internasjonal konsensus om i «The Cosmin study» (Mokkink et al., 2010c). «COnsensus-based Standards for the selection of health Measurement INstruments» (COSMIN) er også valgt fordi det er et metodisk rammeverk for utarbeidelse av systematiske oversikter over måleegenskaper. Dette er nærmere beskrevet i Kapittel 3.

Mokkink et al. (2010c) fordeler måleegenskapene på de tre hovedkategoriene reliabilitet, validitet og responsivitet, som hver har tilhørende måleegenskaper og undergrupper av disse (Figur 1).



**Figur 1** COSMIN-taksonomi over sammenhenger mellom måleegenskaper  
 Reproduisert fra Mokkink et al. (2010c) med tillatelse fra Elsevier.  
 HR-PROM (helserelaterte pasientrapporterte resultatmål)

**Reliabilitet** handler om konsistensen og påliteligheten til et kartleggingsverktøy og defineres av Mokkink et al. (2010c) som i hvilken grad en målingen er uten målefeil. Det kan være ulike kilder til variasjon eller målefeil ved gjentatte målinger, som for eksempel selve kartleggingsverktøyet, personen(e) som gjennomfører målingene, pasienten som er gjenstand for målingen og omstendighetene målingen gjennomføres under. Pålitelighet til et PROM er derfor avhengig av om pasienter som ikke har endret seg skårer det samme ved ulike sett med spørsmål fra samme kartleggingsverktøy (intern konsistens), over tid (test-retest), ulike personer ved samme anledning (inter-rater), eller samme person ved forskjellige anledninger (intra-rater). Reliabilitet består av måleegenskapene intern konsistens, reliabilitet og målefeil (Mokkink et al., 2010c).

- *Intern konsistens* er et mål på den indre sammenhengen mellom spørsmålene innenfor en dimensjon av et PROM og om spørsmålene måler samme konsept (Mokkink et al., 2010c ; de Vet et al., 2011, s. 80-1). Det mest kjente målet for intern konsistens er Cronbachs alfa (de Vet et al., 2011, s. 82).
- *Reliabilitet* er andelen av den totale variansen i målingene som skyldes faktiske forskjeller mellom pasienter og inkluderer test-retest, inter-rater og intra-rater (Mokkink et al., 2010c). Cohens kappas (vektet og uvektet kappas) og intraklasse korrelasjonskoeffisient (ICC) er mål for å vurdere reliabilitet (de Vet et al., 2011, s. 145).
- *Målefeil* er tilfeldige og systematiske feil i pasientens skår som ikke skyldes faktiske endringer i det som måles (Mokkink et al., 2010c). Standardfeilen til målingen (SEM) er ifølge Mokkink et al. (2010b) det foretrukne målet for vurdering av målefeil. Limits of agreement (LoA) og minste målbare forskjell (SDC) er også aktuelle mål og er begge direkte relatert til SEM (Mokkink et al., 2010b).

**Validitet** eller gyldighet defineres av Mokkink et al. (2010c) som i hvilken grad et PROM måler begrepet(ene) eller fenomenet(ene) det har til hensikt å måle, og består av måleegenskapene innholdsvaliditet, kriterievaliditet og begrepsvaliditet.

- *Innholdsvaliditet* handler om i hvilken grad kartleggingsverktøyets innhold reflekterer fenomenet som skal måles (Mokkink et al., 2010c). Viktige sider ved innholdsvaliditet er om alle spørsmålene er relevant for fenomenet, formålet og pasientene og at vesentlige spørsmål ikke mangler (Mokkink et al., 2010a).
- *Kriterievaliditet* er et mål på om skårene til et kartleggingsverktøy er en tilfredsstillende refleksjon av en «gullstandard» (Mokkink et al., 2010c). For PROMs er kriterievaliditet bare aktuelt når en kortversjon kan sammenlignes mot den originale langversjonen av kartleggingsverktøyet (Mokkink et al., 2010a).
- *Begrepsvaliditet* forteller om et kartleggingsverktøys skår samsvarer med hypoteser, basert på antagelsen om at kartleggingsverktøyet måler fenomenet(ene) som skal måles på gyldig måte. Begrepsvaliditet består av tre undergrupper: strukturell validitet, hypotesetesting og tverrkulturell validitet. *Strukturell validitet* beskriver om skårene til et PROM reflekterer fenomenets dimensjoner tilstrekkelig (Mokkink et al., 2010c). Faktoranalyse bør gjennomføres for å kartlegge eller bekrefte om kartleggingsverktøyet består av en eller flere dimensjoner. Utforskende faktoranalyse



(EFA) benyttes når vi ikke vet noe om dimensjonaliteten, mens bekreftende faktoranalyse (CFA) er mest hensiktsmessig når vi har informasjon om faktorstrukturen på forhånd (de Vet et al., 2011, s.71 ; 170-1). *Hypotesetesting* er et mål på om retningen og størrelsen på en samvariasjon eller forskjell er som forventet basert på fenomenet som måles (Mokkink et al., 2010a). Dette aspektet ved begrepsvaliditet evalueres gjerne ved å undersøke samvariasjonen mellom det aktuelle PROM med andre relevante kartleggingsverktøy (Mokkink et al., 2010a ; Mokkink et al., 2010c). *Tverrkulturell validitet* forteller om spørsmålene i en oversatt eller kulturelt tilpasset versjon av et PROM fungerer på samme måte som den opprinnelige versjonen (Mokkink et al., 2010c). Oversettelsesprosessen er en del av den tverrkulturelle validiteten og bør gjennomføres i flere trinn etter internasjonale standarder (de Vet et al., 2011, s. 182-4).

**Responsivitet** er et mål på kartleggingsverktøyets evne til å fange opp endring over tid i fenomenet som måles (Mokkink et al., 2010c).

## 1.5 Oversikt over tidligere forskning

Før jeg startet arbeidet med denne oversikten gjennomførte jeg et orienterende søk i Cochrane Library, MEDLINE, Prospero og på nettsiden til COSMIN (<http://www.cosmin.nl/>). Jeg identifiserte to systematiske oversikter som omhandler kartleggingsverktøy for pustebesvær hos palliative pasienter (Bausewein et al., 2007 ; Dorman, Byrne & Edwards, 2007). De gir en omfattende oversikt over tilgjengelige kartleggingsverktøy, men forfatterne påpeker at ingen er tilstrekkelig validert for bruk i lindrende behandling. Bausewein et al. (2007) gjennomførte litteratursøk i MEDLINE, mens Dorman, Byrne og Edwards (2007) søkte i flere relevante databaser. Ingen av oversiktene kunne konkludere med en «gullstandard» og anbefaler derfor en kombinasjon av ulike kartleggingsverktøy. I tillegg påpeker de at det er behov for videre evaluering av kartleggingsverktøyene for bruk i lindrende behandling (Bausewein et al., 2007 ; Dorman, Byrne & Edwards, 2007).

Flere valideringsstudier er gjennomført etter at de to oversiktene ble utarbeidet. Blant annet er Cancer Dyspnea Scale (CDS) som Dorman, Byrne og Edwards (2007) påpekte som lovende, validert i engelsk (Uronis et al., 2012) og svensk (Hench, Bergman & Gaston-Johansson, 2006) versjon.

Med bakgrunn i at det manglet oppsummert kunnskap om hvilke(n) PROM som er best egnet for kartlegging av pustebesvær hos palliative pasienter, at litteratursøkene i de to identifiserte oversiktene er fra 2005, og at det finnes nyere studier som ikke er oppsummert, mente jeg det var hensiktsmessig å gjennomføre en ny kunnskapsoppsummering.

## 2 Hensikt og forskningsspørsmål

Formålet med mastergradsprosjektet har vært å gi et godt beslutningsgrunnlag for valg av PROM for kartlegging av pustebesvær hos palliative pasienter i klinisk praksis, alternativt å peke på behov for ny forskning. Relevante kartleggingsverktøy må som minst inneholde dimensjoner for pustebesvær og angst. Den systematiske oversikten gir en oversikt over hvilke validerte PROM som er tilgjengelig for formålet samt kunnskap fra forskning om de ulike kartleggingsverktøys pålitelighet.

Oversiktens forskningsspørsmål er: Hvilke kartleggingsverktøy for pasientrapporterte resultatmål er egnet for kartlegging av pustebesvær og angst hos palliative pasienter med pustebesvær?

## 3 Metode

Masteroppgaven er en systematisk oversikt, som jeg vurderte å være en velegnet metode for å besvare forskningsspørsmålet. En systematisk oversikt er en oversikt over forskningsbasert kunnskap som passer forhåndsdefinerte kriterier for å svare på et klart definert forskningsspørsmål. Systematiske og eksplisitt metoder benyttes for formulering av spørsmål, identifisering, utvelgelse og kritisk vurdering av relevant forskning, og for å samle inn og analysere data fra de inkluderte studiene (Higgins & Green, 2011, s. 6 ; Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 4).

Det finnes ulike metodiske rammeverk som kan veilede arbeidet med planlegging og utarbeidelse av systematiske oversikter. Eksempler på dette er *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins & Green, 2011) som har hovedfokus på oversikter om effekt av tiltak. *Slik oppsummerer vi forskning: Håndbok for Nasjonalt kunnskapssenter for helsetjenesten* (Nasjonalt kunnskapssenter for helsetjenesten, 2013) er en mer generell håndbok som beskriver metoder for oppsummering av forskningsbasert kunnskap. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Deeks, Bossuyt & Gatsonis, 2013) omhandler utarbeidelse av oversikter over diagnostiske testers

nøyaktighet, mens *CRD's guidance for undertaking systematic reviews in health care* (Centre for Reviews and Dissemination, 2009) har et eget kapittel om systematiske oversikter over kliniske tester.

Diagnostiske tester skal bidra til å identifisere og utelukke diagnoser og må ha egenskaper som gjør at de kan skille mellom for eksempel syk/frisk og risiko/ikke risiko (Deeks, Bossuyt & Gatsonis, 2013). Dette er vesens forskjellig fra PROMs som skal kartlegge pasientens egen opplevelse av helsetilstand eller livskvalitet, og som dermed må ha måleegenskaper som sikrer at det pålitelig måler fenomenet det skal kartlegge og fanger opp viktige forskjeller over tid. Det stilles ulike krav til studier som skal vurdere diagnostiske testers nøyaktighet og måleegenskaper ved PROMs, og dermed også ulike metodiske krav til utarbeidelse av systematiske oversikter over disse. Til utarbeidelsen av denne systematiske oversikten trengte vi et metodisk rammeverk som er egnet for å oppsummere studier som har validert PROMs.

### 3.1 COSMIN

COSMIN er spesielt utviklet for å bidra til en bedre utvalgelse av helserelaterte PROM og har sitt utspring i manglende internasjonal enighet om terminologi og definisjoner av måleegenskaper for PROMs. Som en del av dette initiativet ble det gjennomført en Delphi studie hvor de oppnådde konsensus om taksonomi, terminologi og definisjoner, i tillegg til en sjekkliste med standarder for vurdering av den metodiske kvaliteten til studier av måleegenskaper. En Delphi studie er en konsensusprosess som går over flere runder, og har deltagelse fra eksperter fra relevante fagfelt. I COSMIN studien deltok 43 av 91 inviterte internasjonale eksperter innenfor fagfeltene psykologi, epidemiologi, statistikk og medisin i en firerunders konsensusprosess (Mokkink et al., 2010b). Nasjonalt kunnskapssenter for helsetjenesten var delaktig og har oversatt terminologi og definisjoner til norsk (Garratt & Bjertnæs, 2014).

*Measurement in Medicine: A Practical Guide* (de Vet et al., 2011, s. 275-314) beskriver de ulike trinnene i utarbeidelsen av en systematisk oversikt med særlig vekt på viktige metodiske aspekt ved oversikter over måleegenskaper til PROMs. Jeg identifiserte ingen reelle alternativer til COSMIN ved gjennomgang av metodelitteraturen og valgte derfor å benytte denne metodeboken sammen med ressurser og verktøy tilgjengelig fra nettsiden <http://www.cosmin.nl/> som hovedrammeverk for planlegging og utarbeidelse av oversikten. En utfordring har vært at det metodiske verktøyet fortsatt er under utvikling (de Vet et al.,

2011, s. 275 ; Terwee et al., 2012). Jeg har derfor også benyttet andre håndbøker, i hovedsak *Slik oppsummerer vi forskning: Håndbok for Nasjonalt kunnskapssenter for helsetjenesten* (Nasjonalt kunnskapssenter for helsetjenesten, 2013).

### 3.2 Inklusjons- og eksklusjonskriterier

Med bakgrunn i forskningsspørsmålet og anbefalinger fra de Vet et al. (2011 s. 282) utarbeidet jeg spesifikke inklusjons- og eksklusjonskriterier. For oversikt i tabell form henvises til artikkelens Tabell 1.

#### ***Inklusjonskriterier***

Populasjonen var palliative pasienter med pustebeviser som var 18 år eller eldre. Pasientene ble definert til å være i palliativ fase når de hadde en uheldredelig sykdom og vi inkluderte studier som omfattet palliative pasienter uavhengig av diagnose. Relevante kartleggingsverktøy var PROMs, som innbefatter alle typer instrumenter, spørreskjema, kartleggingsverktøy, tester eller tilsvarende der informasjonen kommer direkte fra pasienten og ikke er påvirket av andre. De kan være selv-administrerte, intervjuaserte, eller en kombinasjon. Kartleggingsverktøyet måtte minimum inneholde dimensjoner for pustebeviser og angst og være validert for bruk til pasienter som var inneliggende i sykehus, i dagbehandling, polikliniske eller fikk hjemmebaserte tjenester. Vi inkluderte studier som har utviklet og/eller evaluert måleegenskaper ved et relevant PROM, og som minst har evaluert én måleegenskap innenfor en av følgende kategorier: reliabilitet, validitet eller responsivitet.

#### ***Eksklusjonskriterier***

Pasienter med sykdom med mulighet for kurasjon eller med et akutt oppstått pustebeviser ble ekskludert. Vi ekskluderte generiske kartleggingsverktøy og kartleggingsverktøy som er utviklet for bruk i forbindelse med trening, i akuttsituasjoner eller ved innleggelse i akuttavdelinger. Det samme gjelder kartleggingsverktøy som baserer seg på observasjoner eller der pasientens svar er påvirket av andre. Studiedesign som ikke beskriver utvikling av og/eller evaluerer måleegenskaper ved et relevant PROM og systematiske oversikter over relevante PROM ble ekskludert. Aktuelle systematiske oversikter ble allikevel gjennomgått for å se etter mulig relevante studier som ikke var fanget opp av litteratursøket.

### 3.3 Søk etter litteratur

Et systematisk litteratursøk er viktig for å identifisere all relevant forskning som har vært utført for den aktuelle problemstillingen og forskningsspørsmålet, og styrker oversiktens gyldighet. Fremgangsmåten og grundigheten i litteratursøket er en viktig faktor som skiller en systematisk oversikt fra mer tradisjonelle oversikter (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 25-6).

Litteratursøk krever særskilt kompetanse og det anbefales å involvere bibliotekar tidlig i prosessen (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 26). Jeg har rådført meg med bibliotekar underveis i arbeidet og vi har hatt veiledningsseminar der bibliotekar, medstudent og veiledere har gitt tilbakemeldinger på søkestrategiene til mastergradsprosjektet. Endelige søkestrategier ble fagfellevurdert av bibliotekar med spesialkompetanse i søk etter medisinsk og helsefaglig litteratur (Sampson et al., 2009).

Valg av litteraturkilder er avhengig av blant annet hvilke type spørsmål, tema og studiedesign vi søker informasjon om (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 26-7). Med utgangspunkt i kildene som Bausewein et al. (2007) og Dorman, Byrne og Edwards (2007) benyttet, anbefalinger fra Kunnskapssenteret (Nasjonalt kunnskapssenter for helsetjenesten, 2013) og de Vet et al. (2011, s. 279) valgte jeg å søke i de store internasjonale medisin- og helsefagdatabasene og databaser som særlig indekserer publikasjoner om lindrende behandling og psykologi. I prosjektbeskrivelsen var det også planlagt å gjennomføre søk i databasen HaPI (Health and Psychosocial Instruments) som er en database for måleinstrumenter. Ettersom verken Høgskolen i Bergen, Helsebiblioteket eller min arbeidsplass abonnerer på HaPI, hadde jeg ikke mulighet til å søke i den.

Studier om måleegenskaper kan være særlig vanskelig å finne fordi det er stor variasjon i begrepene som benyttes, det kan være manglende eller ufullstendig indeksering av primærstudier og mangelfulle sammendrag (de Vet et al., 2011, s. 278). Søkestrategier for søk etter måleegenskaper ved PROMs bør bestå av søkeord for fenomenet vi er interessert i, målgruppen og måleegenskaper. Søkeord for type måleinstrument som for eksempel spørreskjema eller intervju benyttes ikke fordi dette medfører høy risiko for å miste relevante publikasjoner (de Vet et al., 2011, s. 280). Med bakgrunn i disse utfordringene ønsket jeg å gjennomføre et litteratursøk som var så sensitivt som mulig for å redusere faren for å miste relevante publikasjoner.

Opprinnelig plan var derfor å gjennomføre litteratursøket med emneord og tekstord for kategoriene pustebesvær og måleegenskaper for deretter å gjennomføre inklusjon/eksklusjon knyttet til pasientpopulasjon manuelt. Et prøvesøk resulterte i omtrent 8500 referanser og medførte behov for å gjøre søket mer spesifikt. Basert på oversiktens inklusjons- og eksklusjonskriterier bygget jeg derfor opp en søkestrategi med emneord og tekstord tilpasset hver database for kategoriene: 1) pustebesvær, 2) måleegenskaper og 3) palliative pasienter, kreft, kols (de Vet et al., 2011, s. 279-281). Relevante søkeord ble funnet ved hjelp av SveMed+, de enkelte databasenes emnetre og «scope notes». Jeg kombinerte emneord og tekstord innenfor hver kategori med OR og deretter kombinerte jeg de tre kategoriene med AND.

Det endelige litteratursøket gjennomførte jeg 20.-23. mars 2014 i databasene MEDLINE (Ovid) (1946-), Embase (Ovid) (1974-), PsycINFO (Ovid) (1806-), AMED (Ovid) (1985-), CINAHL (EBSCO) (1981-), Cochrane Library og SveMed+. I tillegg gjorde jeg gjorde siteringssøk på oversiktene til Bausewein et al. (2007) og Dorman, Byrne og Edwards (2007) i Science Citation Index og gjennomgikk referanser til relevante systematiske oversikter og inkluderte publikasjoner for å lete etter relevante studier som ikke var fanget opp i litteratursøket. Fullstendige søkestrategier er vedlagt (Vedlegg I). Jeg har ikke gjort avgrensninger på språk og tidsperiode i søket (de Vet et al., 2011, s. 281). I Embase og CINAHL er avgrensingen «exclude medline journals» benyttet. Alle referanser ble importert i referansehåndteringsverktøyet EndNote og dubletter fjernet.

For ikke å miste relevante studier er det også anbefalt å gjennomføre supplerende søk på navnene til aktuelle PROMs (de Vet et al., 2011, s. 281) og etter resultater fra studier og oversikter som ikke er tilgjengeliggjort i databaser, såkalt «grå litteratur» (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 28). På grunn av det initiale søkets størrelse har jeg ikke hatt kapasitet til å gjennomføre dette.

### 3.4 Utvelgelse av studier

Utvelgelse av studier bør gjennomføres i flere trinn av to personer uavhengig av hverandre (de Vet et al., 2011, s. 283-4 ; Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 35) for å sikre at alle relevante studier inkluderes.

Vi vurderte først titler og sammendrag for inklusjon. Deretter ble mulig relevante artikler innhentet i fulltekst og vurdert for inklusjon. Det var i noen grad også nødvendig å innhente

tilleggsinformasjon om aktuelle PROMs dimensjoner for å vurdere relevans. Alle delene av utvelgelsesprosessen ble gjort av to personer uavhengig av hverandre opp mot kriteriene for inklusjon og eksklusjon og uenighet ble løst ved diskusjon. De første 100 referansene fungerte som en pilot på om inklusjons- og eksklusjonskriteriene ble vurdert likt eller hadde behov for å justeres.

### 3.5 Kvalitetsvurdering av inkluderte studier

Kvalitetsvurdering av inkluderte studier er viktig for å avgjøre i hvilken grad vi kan stole på resultatene (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 38). Hvis en studie tilfredsstillende standarder for god metodisk kvalitet er risikoen for systematiske skjevheter begrenset (de Vet et al., 2011, s. 284). Det er allikevel slik at en studie kan ha betydelige systematiske skjevheter selv om den er utført i samsvar med standarder for god kvalitet, og omvendt kan det være metodiske svakheter som ikke nødvendigvis fører til risiko for systematiske skjevheter (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 38-9).

Det er utarbeidet mange ulike sjekklister som kan være til hjelp ved kritisk vurdering av ulike studiedesign. Cochrane Collaboration har for eksempel utarbeidet verktøy for å vurdere risiko for skjevheter i effektstudier og diagnostiske studier. I studier der man skal validere PROMs gjelder andre sett med kvalitetskriterier. I denne oversikten valgte vi derfor å benytte COSMIN-sjekklisten med firepunkts-skala (COSMIN, 2011) (Vedlegg II) og medfølgende manual (Mokkink et al., 2012) slik de Vet et al. (2011, s. 284-91) anbefaler.

COSMIN-sjekklisten er spesielt utviklet for vurdering av studier om måleegenskaper ved PROMs og består av tolv bokser, hvorav ni innehar spørsmål knyttet til metodologiske standarder for hvordan hver måleegenskap bør evalueres. Hvert spørsmål skåres som *utmerket*, *god*, *rimelig* eller *svak*. En samlet skår fastsettes deretter for hver boks (måleegenskap) etter prinsippet om at dårligste skår teller (Mokkink et al., 2010b ; Terwee et al., 2012). Sjekklisten tar hensyn til at det stilles ulike metodiske krav til evaluering av ulike måleegenskaper. En studie kan dermed skåre *utmerket* for metodisk kvalitet for en måleegenskap og *svak* for en annen.

Vi var to personer som gjennomførte kritisk vurdering uavhengig av hverandre. Uenighet eller usikkerhet ble løst ved hjelp av en tredje person. Kvalitetsvurderingen av de første studiene benyttet vi til å pilotere sjekklisten for å sikre at vi hadde en felles forståelse for kriteriene. Vi kvalitetssikret vår vurdering med veileder som har høy kompetanse på området. I noen

tilfeller har vi avviket fra prinsippet om at dårligste skår teller. Det har eksempelvis vært dersom spørsmålet ikke har gitt valgmulighet for alle fire skåralternativene og vi har vurdert kvaliteten til det alternativet som manglet, og i situasjoner der vi mener at den metodiske svakheten ikke har ført til betydelig risiko for systematiske skjevheter.

### 3.6 Dataekstraksjon

Dataekstraksjon innebærer uthenting av relevant informasjon fra de inkluderte studiene, gjerne ved hjelp av et dataekstraksjonsskjema, og bør gjennomføres av to personer uavhengig av hverandre (de Vet et al., 2011, s. 291 ; Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 42-3).

For å sikre at relevante og konsistente data ble hentet ut fra studiene utarbeidet jeg et dataekstraksjonsskjema etter kriteriene anbefalt av Vet et al. (2011, s. 291-4) og med hensyn til den informasjonen vi hadde behov for til sammenstilling og analyse av resultat i den systematiske oversikten. Dette inkluderte generell informasjon om publikasjonen, kjennetegn ved studien, det aktuelle kartleggingsverktøyet, studieutvalget, hvilke måleegenskaper som var vurdert, metodisk kvalitet og resultater for disse. Skjemaet ble deretter pilottestet og justert.

Siden denne oversikten er et mastergradsprosjekt var det ikke kapasitet til at to uavhengig av hverandre hentet ut data som anbefalt. Jeg hentet ut data fra alle de inkluderte studiene og veileder kontrollerte dataekstraksjonene.

### 3.7 Sammenstilling og gradering

I systematiske oversikter ønsker vi å oppsummere resultater og trekke konklusjoner på tvers av alle tilgjengelige enkeltstudier. Noen ganger er det mest hensiktsmessig å basere en slik oppsummering på en kvalitativ beskrivelse av relevante studier, andre ganger kan det være aktuelt å bruke tilgjengelige studier som utgangspunkt for nye statistiske analyser, såkalte meta-analyser (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 44). Det finnes ingen fasit som kan fortelle oss når det er hensiktsmessig å kombinere studier i en meta-analyse, det er en avgjørelse man må ta basert på om tilgjengelige studier er tilstrekkelig like med henblikk på utvalg, setting, versjon av kartleggingsverktøyet og administrasjonsform (de Vet et al., 2011, s. 296).



I dette tilfellet vurderte vi at de inkluderte studiene var for ulike til å forsvare en kvantitativ meta-analyse, isteden valgt vi å oppsummere resultater på tvers av enkeltstudier i en kvalitativ syntese der vi tok hensyn til tilgjengelige studiers metodisk kvalitet og dokumenterte måleegenskaper.

Den metodiske kvaliteten til enkeltstudiene vurderte vi ved hjelp av COSMIN-sjekklisten som beskrevet i Kapittel 3.5. Videre benyttet vi et sett med forhåndsdefinerte kriterier til å klassifisere måleegenskapene til ulike PROM som *positiv*, *uavklart* eller *negativ* (Vedlegg III). For eksempel vurderes intern konsistens for en endimensjonal skala å være *positiv* når Cronbachs alfa er større enn 0,70.

Som et siste trinn i den kvalitative oppsummeringen ønsket vi å trekke konklusjoner på tvers av tilgjengelig dokumentasjon (de Vet et al., 2011, s. 307-9). The Grading og Recommendations Assessment, Development and Evaluation (GRADE) er den mest kjente metoden for å vurdere kvaliteten på dokumentasjonen og styrken på anbefalinger (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 49 ; Vist et al., 2013). GRADE benyttes i retningslinjer og systematiske oversikter der man vurderer effekten av og gir anbefalinger om ulike behandlingsalternativer og skiller klart mellom kvaliteten på dokumentasjonen og styrken på anbefalingene (Vist et al., 2013). GRADE-kriteriene for vurdering av kvaliteten på dokumentasjonen er ikke direkte overførbare til studier som validerer måleegenskaper ved PROMs. Vi har derfor benyttet et sett kriterier som graderer den samlede dokumentasjonen om måleegenskapene til ulike PROM til *høy*, *moderat*, *lav*, *motstridende* eller *usikker* utfra på studien(e)s metodiske kvalitet, måleegenskapens resultat og grad av konsistens (Tabell 1).

**Tabell 1** Gradering av tillit til kartleggingsverktøys måleegenskaper tilpasset fra Schellingerhout et al. (2012)

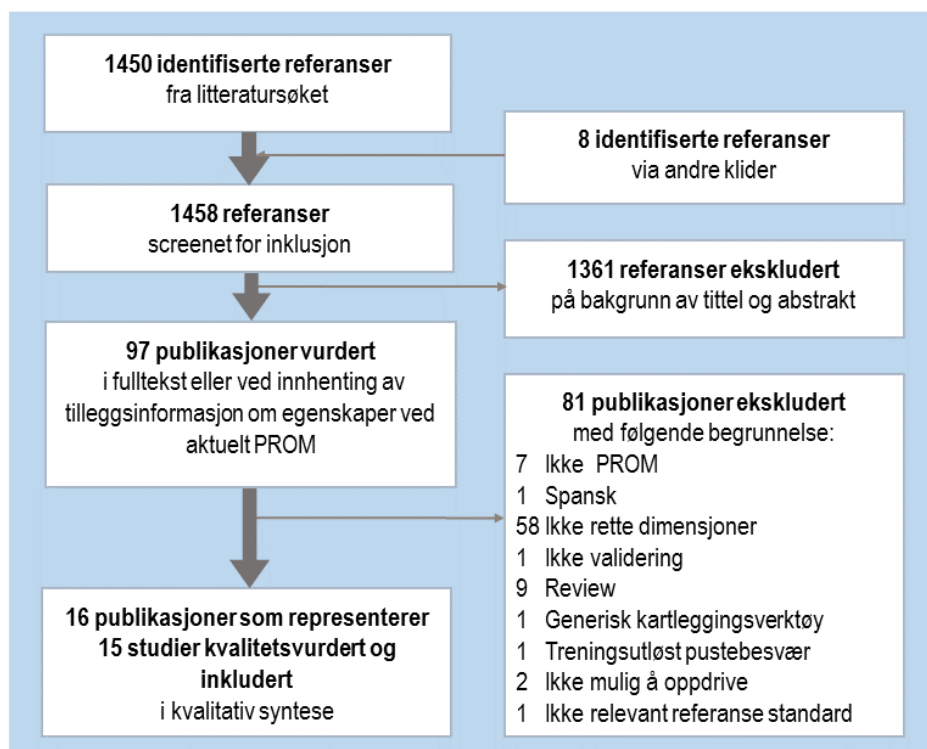
Nivå	Vurdering	Kriterier
Høy	+++ ELLER---	Konsistente funn i flere studier av <i>god</i> metodisk kvalitet ELLER i en studie av <i>utmerket</i> metodisk kvalitet.
Moderat	++ ELLER --	Konsistente funn i flere studier av <i>rimelig</i> metodisk kvalitet ELLER i en studie av <i>god</i> metodisk kvalitet.
Lav	+ ELLER -	En studie av <i>rimelig</i> metodisk kvalitet.
Motstridende	+/-	Motstridende funn.
Usikker	?	Bare studier av <i>svak</i> metodisk kvalitet.
	Ukjent	Ukjent resultat.

+ (positivt resultat); – (negativt resultat)

## 4 Resultater

Litteratursøket identifiserte totalt 1988 referanser (MEDLINE (n=1059), Embase (n=162), PsycINFO (n=88), AMED (n=90), CINAHL (n=101), Cochrane Library (n=152), SveMed+ (n=240), Science Citation Index (n=96)). I tillegg fant vi åtte referanser ved gjennomgang av referanselister til relevante publikasjoner og ved søk etter tilleggsinformasjon om inkluderte PROMs. Resultatet av litteratursøket etter fjerning av dubletter var 1458 titler som vi vurderte for inklusjon etter kriteriene i Kapittel 3.2. Figur 2 viser oversikt over artikkelflyt. I Vedlegg IV gir vi en oversikt over ekskluderte publikasjoner med begrunnelse.

Vi kritisk vurderte 16 publikasjoner som beskriver 15 studier og inkludert disse i den kvalitative syntesen. Tabell 2 gir en oversikt over inkluderte studiers karakteristika og dataekstraksjonsskjema for studiene er vedlagt (Vedlegg V). Studiene beskriver utvikling og validering av de fire kartleggingsverktøyene som tilfredsstilte inklusjonskriteriene våre: Cancer Dyspnea Scale (CDS), Severe Respiratory Insufficiency (SRI) Questionnaire, Dyspnea Management Questionnaire (DMQ) og en respiratorisk symptom sjekklister (RSC). DMQ har vi inkludert i tre ulike versjoner: DMQ-30, DMQ-56 og DMQ-CAT (Computer Adaptive Test). Karakteristika ved kartleggingsverktøyene er oppsummert i Tabell 3.



Figur 2 Flytskjema over identifisert litteratur med begrunnelse for eksklusjon

**Tabell 2 Karakteristika ved inkluderte studier**

PROM	Studie	Sykdomskarakteristika ved utvalget	Antall deltagere (n)	Land	Språkversjon	Setting (n)
<b>CDS</b>	Tanaka et al., 2000	Avansert eller residiverende lungekreft. 94 % var i ECOG funksjonsklasse 0-1.	166	Japan	Japansk	Sykehus (31) Poliklinisk (135)
	Tanaka et al., 2002	Avansert eller residiverende lungekreft. 95 % var i ECOG funksjonsklasse 0-1.	171	Japan	Japansk	Poliklinisk
	Henoch et al., 2006	Avansert lungekreft. 77 % var i ECOG funksjonsklasse 0-1.	99	Sverige	Svensk	Poliklinisk
	Uronis et al., 2012	Avansert lungekreft. 63 % var i ECOG funksjonsklasse 0-1.	109 (USA n=78, UK n=12, AUS n=22)	USA, UK, AUS	Engelsk	Sykehus*
<b>SRI</b>	Windisch et al., 2003	Alvorlig hyperkapnisk respirasjonssvikt. Behandlet med LTMV. Stabil fase. Innlagt til rutinekontroll for LTMV.	226 (kols n=78, nevromuskulære sykdommer n=49)	Tyskland	Tysk	Sykehus
	Duierveman et al., 2008	Kronisk hyperkapnisk respirasjonssvikt, kols GOLD III eller IV. Stabil fase. Ingen av pasientene var eller hadde vært behandlet med LTMV.	72	Nederland	Nederlandsk	Ikke oppgitt
	Lopez-Campos et al., 2008	Pasienter behandlet med LTMV av ulike årsaker. Stabil fase.	115 (kols n=15, nevromuskulære sykdommer n=18)	Spania	Spansk	Poliklinisk
	Windisch et al., 2008	Kronisk hyperkapnisk respirasjonssvikt forårsaket av kols. Langsiktig behandling med LTMV og LTOT. Stabil fase. Innlagt til rutinekontroll for LTMV.	162 (overlapp med kols-utvalget i Windisch et al., 2003)	Tyskland	Tysk	Sykehus
	Ghosh et al., 2012	Kronisk hyperkapnisk respirasjonssvikt. Behandlet med LTMV. Stabil fase.	152 (kols n=26, nevromuskulære sykdommer n=32)	England	Engelsk	Poliklinisk
	Struik et al., 2013	Kronisk hyperkapnisk respirasjonssvikt forårsaket av kols. Fortsatt hyperkapnisk 48 timer etter avsluttet ventilasjonsbehandling for en akutt forverring.	170	Nederland	Nederlandsk	Sykehus
	Markussen et al., 2014	Pasienter med alvorlig kronisk hyperkapnisk respirasjonssvikt. Behandlet med LTMV.	127 (kols n=26, nevromuskulære sykdommer n=54)	Norge	Norsk	Hjemmeboende
<b>DMQ-30</b>	Norweg et al., 2006	Kols og astma	85 (kols n=73, astma n=12)	USA	Ikke oppgitt	Ikke oppgitt
<b>DMQ-56</b>	Norweg et al., 2011a	Kols	107 (overlapp med kols-utvalget i Norweg et al., 2006)	USA	Engelsk	Ikke oppgitt
<b>DMQ-CAT</b>	Norweg et al., 2011b	Kols. GOLD stadium I 5,5 %, II 43,5 %, III 39,4 %, IV 12 %.	292	USA	Engelsk	Poliklinisk
<b>RSC</b>	Han et al., 2011	Pustebesvær som følge av ulike hjerte- og lungesykdommer.	396 (kols n=58, hjertesvikt n=62)	Kina	Ikke oppgitt	Sykehus*

PROM (pasientrapporterte resultatmål); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); ECOG (Eastern Cooperative Oncology Group); LTMV (langtids mekanisk ventilasjon); LTOT (langtids oksygenbehandling); kols (kronisk obstruktiv lungesykdom); GOLD (Global Initiative for Chronic Obstructive Lung Disease); USA (United States of America); UK (United Kingdom); AUS (Australia);\* (uklart om inneliggende, poliklinisk eller begge deler).

**Tabell 3 Karakteristika ved inkluderte PROMs**

PROM	Målgruppe	Antall spørsmål	Dimensjoner	Måleskala (per spørsmål)	Scoringssystem	Språk	Forventet utfyllingstid
<b>CDS</b>	Kreft	12	Følelse av anstrengelse (5 spm) Følelse av angst (4 spm) Følelse av ubehag (3 spm) Totalskår	5-punkts Likert	Beregner totalskår per dimensjon: anstrengelse (0-20); angst (0-16); ubehag (0-12). Høyere skår indikerer mer alvorlig pustebesvær.	Japansk Engelsk Svensk	2 minutter
<b>SRI</b>	Kronisk respirasjons-svikt med LTMV behandling	49	Åndedrettsplager (8 spm) Fysisk funksjon (6 spm) Tilleggssymptomer og søvn (7 spm) Sosiale relasjoner (6 spm) Angst (5 spm) Psykisk velvære (9 spm) Sosial funksjon (8 spm) Totalskår	5-punkts Likert	Beregner skår per dimensjon som prosent av maks (0-100). Høyere skår indikerer bedre livskvalitet	Tysk Nederlandsk Spansk Engelsk Norsk	20 minutter
<b>DMQ-30</b>	Kols (rehabilitering)	30	Dyspnéintensitet (7 spm) Dyspnérelatert angst (7 spm) Angstrelatert aktivitetsunnngåelse (7 spm) Tiltro til aktivitetsrelatert mestringsevne (5 spm) Tilfredshet med mestringsstrategier (4 spm)	7-punkts Likert	Beregner gjennomsnittsskår per dimensjon (0-6). Totalskår kalkuleres ved å summere dimensjonene: intensitet, angst og aktivitetsunnngåelse. Høyere skår indikerer bedre funksjon.	Engelsk	Ikke oppgitt
<b>DMQ-56</b>	Kols (rehabilitering)	56	Dyspnéintensitet (12 spm) Dyspnérelatert angst (16 spm) Angstrelatert aktivitetsunnngåelse (15 spm) Tiltro til aktivitetsrelatert mestringsevne (6 spm) Tilfredshet med mestringsstrategier (7 spm)	7-punkts Likert	Beregner gjennomsnittsskår per dimensjon (0-6). Høyere skår indikerer bedre funksjon.	Engelsk	Ikke oppgitt
<b>DMQ-CAT</b>	Kols (rehabilitering)	20	Dyspnéintensitet Dyspnérelatert angst Angstrelatert aktivitetsunnngåelse Tiltro til aktivitetsrelatert mestringsevne	6-punkts Likert	Ikke oppgitt	Engelsk	Ikke oppgitt
<b>RSC</b>	Pustebesvær som følge av hjerte-/lungesykdom	46	Sensoriske kvaliteter (23 spm) Affektive aspekt (12 spm) Atferdsmessig påvirkning (4 spm) Andre symptomer (7 spm)	5-punkts Likert	Ikke oppgitt	Ikke oppgitt	10-15 minutter

PROM (pasientrapporterte resultatmål); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); LTMV (langtids mekanisk ventilasjon); kols (kronisk obstruktiv lungesykdom).

## 4.1 Populasjon og setting

De fire studiene om **Cancer Dyspnea Scale** (CDS) omfattet tilsammen 545 deltagere fra Japan (Tanaka et al., 2000 ; 2002), Sverige (Hench, Bergman & Gaston-Johansson, 2006), USA, Storbritannia og Australia (Uronis et al., 2012), alle med avansert eller residiverende lungekreft. I hovedsak var det polikliniske pasienter, men også noen inneliggende. Av de inkluderte pasientene var 85 % i Eastern Cooperative Oncology Group (ECOG) funksjonsklasse 0-1, som innebærer at de var i normal aktivitet (0), eller hadde begrensninger i fysisk aktivitet, men var oppegående og kunne utføre lett arbeid (1).

Syv studier om **Severe Respiratory Insufficiency (SRI) Questionnaire** inkluderte til sammen 946 deltagere fra Tyskland (Windisch et al., 2003 ; Windisch et al., 2008), Nederland (Duiverman et al., 2008 ; Struik et al., 2013), Spania (Lopez-Campos et al., 2008), England (Ghosh et al., 2012) og Norge (Markussen et al., 2014). Studien omhandlet pasienter med ulike diagnoser hvorav omtrent 50 % med kols og 16 % med nevrologiske sykdommer. 704 av deltagerne var etablert med langtids mekanisk ventilasjon (LTMV) (Windisch et al., 2003 ; Lopez-Campos et al., 2008 ; Windisch et al., 2008 ; Ghosh et al., 2012 ; Markussen et al., 2014), 72 deltagere hadde ikke på aktuelt eller tidligere tidspunkt hatt behandling med LTMV (Duiverman et al., 2008) og 170 hadde avsluttet ventilasjonsstøtte for en akutt forverring (Struik et al., 2013). Størstedelen av de inkluderte pasientene var i stabil fase, innlagt i sykehus eller poliklinisk til rutine kontroll, mens studien til Struik et al. (2013) inkluderte pasienter med kols (n=170) som var innlagt med en akutt forverring. Ingen av studiene oppgav Global Initiative for Chronic Obstructive Lung Disease (GOLD) stadium for de inkluderte pasientene, men i to av studiene (Duiverman et al., 2008 ; Struik et al., 2013) var GOLD stadium III eller IV et inklusjonskriterium. De to studiene til Windisch et al. (2003 ; 2008) har delvis overlappende utvalg av pasienter med kols. Data fra Windisch et al. (2008) er derfor benyttet for intern konsistens og strukturell validitet.

Tre studier evaluerte tre ulike versjoner av **Dyspnea Management Questionnaire** (DMQ): DMQ-30 (Norweg et al., 2006), DMQ-56 (Norweg et al., 2011a) og DMQ-CAT (Norweg et al., 2011b), og inkluderte til sammen 411 deltagere. Av disse var det 399 pasienter med kols og tolv med astma. Studien om DMQ-CAT omhandlet polikliniske pasienter (Norweg et al., 2011b). De andre studiene oppgav ikke setting, men vi har vurdert det som sannsynlig at disse også var polikliniske (Norweg et al., 2011a ; Norweg et al., 2011b). To av studiene oppgav ikke GOLD stadium, mens deltagerne i den største studien (n=292) fordelte seg med omtrent

5,5 % i GOLD stadium I, 43,5 % i GOLD stadium II, 39 % i GOLD stadium III og 12 % i GOLD stadium IV (Norweg et al., 2011b).

**Respiratoriske symptomsjekkliste (RSC)** beskrives i en studie som inkluderte 396 pasienter med pustebesvær hvorav omtrent 14,5 % med kols og 15,5 % med hjertesvikt. Pasientene var på sykehus, sannsynligvis både inneliggende og polikliniske (Han et al., 2011).

## 4.2 Inkluderte studiers metodiske kvalitet

Vi kritisk vurderte de inkluderte studiene hjelp av COSMIN-sjekklisten og i de følgende avsnitt oppsummerer vi vår vurdering av den metodiske kvaliteten samt gir en kort beskrivelse av hovedsvakhete(n) ved studien når vi har gitt lavere skår enn utmerket.

Oversettelse er ikke en egen måleegenskap, men en del av tverrkulturell validitet. Mange av de inkluderte studiene har bare beskrevet oversettelsesprosessen. Vi har valgt å beskrive kvaliteten på oversettelsen utfra spørsmålene (4-11) som vurderer denne prosessen i sjekklistens boks om tverrkulturell validitet.

En oversikt for hvert kartleggingsverktøy over hvilke studier som er inkludert, hvilke måleegenskaper som er vurdert og resultatet av kvalitetsvurderingen er fremstilt i Tabellene 4-6. For en samlet oversikt henviser vi til Vedlegg VI.

### ***Cancer Dyspnea Scale (CDS)***

**Innholdsvaliditeten** er beskrevet i en studie av *utmerket* metodisk kvalitet (Tanaka et al., 2000).

**Strukturell validitet** er evaluert i to studier (Tanaka et al., 2000 ; Henoeh, Bergman & Gaston-Johansson, 2006) av *utmerket* og en studie (Uronis et al., 2012) av *god* metodisk kvalitet. Uronis et al. (2012) benyttet utforskende faktoranalyse (EFA), mens bekreftende faktoranalyse (CFA) ville vært mest hensiktsmessig når faktorene var predefiner fra studien til Tanaka et al. (2000).

**Hypotesetesting** er rapportert i fem publikasjoner (Tanaka et al., 2000 ; 2002 ; Henoeh, Bergman & Gaston-Johansson, 2006 ; Henoeh et al., 2008 ; Uronis et al., 2012). De to publikasjonene til Henoeh et al. (2006 ; 2008) omhandler samme studie, vi har derfor rapportert resultater fra publikasjonen fra 2006. De fire studiene (Tanaka et al., 2000 ; 2002 ;

Henoch, Bergman & Gaston-Johansson, 2006 ; Uronis et al., 2012) var av *god* metodisk kvalitet for alle sammenligningene med unntak av sammenligningen av CDS med Hospital Anxiety and Depression Scale (HADS), Functional Assessment of Cancer Therapy (FACT-G, FACT-L) og perifer oksygenmetning (SpO<sub>2</sub>) i studien til Uronis et al. (2012). Den var av *rimelig* kvalitet. Studiene hadde mangelfulle beskrivelser av forhåndsdefinerte hypoteser og i noen av studiene var det også mangelfulle beskrivelser av de kartleggingsverktøyene det ble sammenlignet med. For nevnte sammenligninger i studien til Uronis et al. (2012) har vi i tillegg vurdert at det er risiko for systematiske feil på grunn av utvalgets størrelse og manglete svar på en andel av spørreskjemaene.

Ingen av studiene har evaluert tverrkulturell validitet. **Oversettelsen** til engelsk versjon (Tanaka et al., 2000) var av *svak* metodisk kvalitet. Det er ikke beskrevet om det var brukerinvolvering eller om det ble gjennomført pilottesting. Deler av oversettelsesprosessen til svensk versjon (Henoch, Bergman & Gaston-Johansson, 2006) er mangelfullt beskrevet og oversettelsen er vurdert til å være av *god* metodisk kvalitet.

To (Tanaka et al., 2000 ; Uronis et al., 2012) av tre studier som har evaluert **intern konsistens** var av *utmerket* metodisk kvalitet, mens en studie (Henoch, Bergman & Gaston-Johansson, 2006) var av *god* kvalitet. I sistnevnte studie takket 30 % de forespurte nei til deltagelse og omtrent syv % av spørreskjemaene ble ekskludert fordi de ikke var korrekt utfylt.

**Test-retest-reliabilitet** er evaluert i en studie (Tanaka et al., 2000) av *rimelig* kvalitet.

Utvalget var moderat (n=37) og intervallet mellom test og retest på bare er en uke medfører risiko for hukommelsesskjevhet.

**Tabell 4 CDS: inkluderte studiers metodiske kvalitet per måleegenskap**

Studie	Intern konsistens	Reliabilitet	Innholdsvaliditet	Strukturell validitet	Hypotese-testing
Tanaka et al., 2000	Utmerket	Rimelig	Utmerket	Utmerket	God
Tanaka et al., 2002					God
Henoch et al., 2006	God			Utmerket	God
Uronis et al., 2012	Utmerket			God	God*/Rimelig**

CDS (Cancer Dyspnea Scale);\*sammenligningene med VAS (Visual Analog Scale), Borg skala, ECOG (Eastern Cooperative Oncology Group); \*\*sammenligningene med HADS (Hospital Anxiety and Depression Scale), FACT (Functional Assessment of Cancer Therapy), SpO<sub>2</sub> (perifer oksygenmetning).

## **Severe Respiratory Insufficiency (SRI) Questionnaire**

**Innholdsvaliditet** er beskrevet i en studie (Windisch et al., 2003) av *utmerket* og en studie (Markussen et al., 2014) av *god* metodisk kvalitet. Utvalget som ble benyttet til å vurdere relevans i studien til Markussen et al. (2014) var moderat (n=6).

**Strukturell validitet** er evaluert i tre studier (Lopez-Campos et al., 2008 ; Windisch et al., 2008 ; Ghosh et al., 2012) av *svak* metodiske kvalitet. Studiene hadde et betydelig mindre utvalg enn anbefalt i forhold til antall spørsmål i kartleggingsverktøyet.

**Hypotesetesting** er gjennomført i fem studier (Windisch et al., 2003 ; Duiverman et al., 2008 ; Ghosh et al., 2012 ; Struik et al., 2013 ; Markussen et al., 2014) av *god* og en studie (Lopez-Campos et al., 2008) av *rimelig* metodisk kvalitet. Struik et al. (2013) og Duiverman et al. (2008) har ikke beskrevet hvordan manglende data ble håndtert. I studiene som sammenligner SRI med SF-36 var det mangelfull informasjon om SF-36 (Windisch et al., 2003 ; Lopez-Campos et al., 2008 ; Ghosh et al., 2012 ; Markussen et al., 2014). Men fordi SF-36 er et internasjonalt anerkjent kartleggingsverktøy som er mye brukt og validert for en rekke sykdomsgrupper valgte vi å ikke la dårligste skår telle.

**Tverrkulturell validitet** er evaluert for norsk versjon av SRI i en studie (Markussen et al., 2014) av *svak* metodisk kvalitet. Utvalget i studien var mindre enn anbefalt i forhold til antall spørsmål i kartleggingsverktøyet og det var ikke gjennomført faktoranalyse. Det var heller ikke klargjort om utvalget i studien var sammenlignbart med det originale utvalget.

**Oversettelsesprosessen** (Markussen et al., 2014) var av *god* metodisk kvalitet. Oversettelsen til engelsk versjon (Ghosh et al., 2012) var av *rimelig* metodisk kvalitet. Prosessen var ikke formalisert og oversetterne jobbet ikke uavhengig av hverandre.

**Intern konsistens** er evaluert i seks studier (Duiverman et al., 2008 ; Lopez-Campos et al., 2008 ; Windisch et al., 2008 ; Ghosh et al., 2012 ; Struik et al., 2013 ; Markussen et al., 2014) av *svak* metodisk kvalitet. Utvalgene i studiene var for små i forhold til antall spørsmål i SRI. I flere av studiene var det heller ikke gjennomført faktoranalyse eller henvist til annen studie der dette er gjennomført (Duiverman et al., 2008 ; Struik et al., 2013 ; Markussen et al., 2014).

**Test-retest-reliabilitet** er evaluert i en studie (Duiverman et al., 2008) av *svak* og en studie (Lopez-Campos et al., 2008) av *rimelig* metodisk kvalitet. I studien til Duiverman et al.



(2008) var utvalget ved test-retest lite (n=23) og det var ikke klart beskrevet hvordan de håndterte manglende data. Tidsintervallet mellom test og retest var på hele tolv uker, noe som øker faren for at utvalget ikke var stabilt. I studien til Lopez-Campos et al. (2008) var det ikke gjort rede for om alle deltok ved test-retest, om det var frafall mellom test og retest og om deltagerne faktisk var stabile.

**Tabell 5 SRI: inkluderte studiers metodisk kvalitet per måleegenskap**

Studie	Intern konsistens	Reliabilitet	Innholdsvaliditet	Strukturell validitet	Hypotesetesting	Tverrkulturell validitet
Windisch et al., 2003			Utmerket		God	
Duiverman et al., 2008	Svak	Svak			God	
Lopez-Campos et al., 2008	Svak	Rimelig		Svak	Rimelig	
Windisch et al., 2008	Svak			Svak		
Ghosh et al., 2012	Svak			Svak	God	
Struik et al., 2013	Svak				God	
Markussen et al., 2014	Svak		God		God	Svak

SRI (Severe Respiratory Insufficiency Questionnaire)

### ***Dyspnea Management Questionnaire (DMQ)***

**Innholdsvaliditeten** til DMQ-30 (Norweg et al., 2006) og DMQ-CAT (Norweg et al., 2011b) er beskrevet i studier av *utmerket* metodisk kvalitet.

For DMQ-56 er **strukturell validitet** evaluert i en studie (Norweg et al., 2011a) av *svak* metodisk kvalitet. Størrelsen på utvalget var mindre enn anbefalt i forhold til antall spørsmål som ble vurdert. Strukturell validitet for DMQ-CAT er evaluert i en studie (Norweg et al., 2011b) av *rimelig* metodisk kvalitet. Utvalget var i underkant av det som er anbefalt. Vi mener allikevel at utvalget var stort (n=292) og nær kravet til rimelig, og har derfor valgt å ikke la dårligste skår telle.

**Hypotesetesting** DMQ-30 er gjennomført i en studie (Norweg et al., 2006). Studien var av *god* metodisk kvalitet for sammenligningene med SOLQ og SF-12, og *svak* metodisk kvalitet for sammenligning med HADS-angst. Hovedbegrunnelsen for denne forskjellen i skår er størrelsen på utvalgene som var inkludert i de ulike analysene. For DMQ-CAT er hypotesetesting gjennomført i en studie (Norweg et al., 2011b) av *rimelig* metodisk kvalitet.

Studien har mangelfulle beskrivelser av hypoteser om forventede samvariasjoner eller gjennomsnittsforskjeller og mangelfull rapportering av egenskaper ved kartleggingsverktøyene det er sammenlignet med.

**Kriterievaliditet** er evaluerte for DMQ-CAT i en studie (Norweg et al., 2011b) av *god* metodisk kvalitet. DMQ-CAT ble sammenlignet med den totale samlingen av spørsmål ved hjelp av en simuleringsstrategi. Metoden innebærer at svar fra den totale samlingen ble benyttet til å besvare DMQ-CAT. Vi mener dette kan ha medført risiko for systematiske skjevheter, men har vurdert risikoen som begrenset og derfor ikke valgt å la dårligste skår telle.

**Intern konsistens** er evaluert i studier (Norweg et al., 2006 ; Norweg et al., 2011a ; Norweg et al., 2011b) av *god* metodisk kvalitet for alle tre versjonene av DMQ. Utvalgene i studiene om DMQ-30 og DMQ-56 (Norweg et al., 2006 ; Norweg et al., 2011a) var mindre enn anbefalt, mens det i studien om DMQ-CAT (Norweg et al., 2011b) ikke var beskrevet om det var manglende data og hvordan dette eventuelt var håndtert.

**Test-retest-reliabilitet** for DMQ-30 er evaluert i en studie (Norweg et al., 2006) av *svak* og for DMQ-56 i en studie (Norweg et al., 2011a) av *rimelig* metodisk kvalitet. Studien om DMQ-30 hadde et lite utvalg (n= 21-26), mens utvalget i studien om DMQ-56 var moderat (n=31-33).

**Tabell 6 DMQ: inkluderte studiers metodiske kvalitet per måleegenskap**

PROM	Studie	Intern konsistens	Reliabilitet	Innholdsvaliditet	Strukturell validitet	Hypotese-testing	Kriterievaliditet
DMQ-30	Norweg et al., 2006	God	Svak	Utmerket		God*/Svak**	
DMQ-56	Norweg et al., 2011a	God	Rimelig		Svak		
DMQ-CAT	Norweg et al., 2011b	God		Utmerket	Rimelig	Rimelig	God

DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test). \*sammenligningene med SOLQ (Seattle Obstructive Lung Disease Questionnaire) og SF-12 (Short Form Health Survey); \*\* sammenligning med HADS (Hospital Anxiety Depression Scale)-angst.

### **Respiratorisk symptomsjekkliste (RSC)**

En respiratorisk sjekkliste (RSC) er beskrevet i en studie (Han et al., 2011) og **strukturell validitet** var evaluert. Studien var av *utmerket* metodisk kvalitet.

### 4.3 Resultater for inkluderte kartleggingsverktøy

I det følgende beskriver jeg de fire kartleggingsverktøyene vi har inkludert og resultatet for hver måleegenskap fra de inkluderte studiene. En oversikt over resultatene finnes i Vedlegg VI. Resultatet kan være positivt selv om vi har vurdert at den metodisk kvaliteten til studien er svak, og omvendt. Vi har kategorisert resultatene i *positiv*, *uavklart* eller *negativ* utfra forhåndsbestemte kriterier (Vedlegg III). Jeg har oversatt de ulike kartleggingsverktøyenes dimensjoner til norsk.

For hypotesetesting har vi valgt å vektlegge de sammenligninger vi har vurdert som mest relevant, og for CDS og SRI sammenligninger som er testet i minst to studier. I Vedlegg VII viser vi en oversikt over alle tilgjengelige samvariasjoner for hvert verktøy.

Kartleggingsverktøyene som er benyttet til de sammenligningene vi har vurdert er kort beskrevet i Vedlegg VIII.

#### ***Cancer Dyspnea Scale (CDS)***

Fem publikasjoner (Tanaka et al., 2000 ; 2002 ; Hensch, Bergman & Gaston-Johansson, 2006 ; Hensch et al., 2008 ; Uronis et al., 2012) har beskrevet utvikling og validering av CDS. Kartleggingsverktøyet ble opprinnelig utviklet i Japan for pasienter med kreft og måler aspekter ved pustebesvær de siste dagene.

CDS består av tolv spørsmål fordelt på tre dimensjoner: «følelse av anstrengelse», «følelse av angst» og «følelse av ubehag». Hvert spørsmål skåres på en fem-punkts Likert-skala og det kalkuleres en skår for hver dimensjon som kan legges sammen til en totalskår. Totalskåren er et mål for «totalt pustebesvær» (Tanaka et al., 2000). CDS er validert på japansk (Tanaka et al., 2000 ; 2002), svensk (Hensch, Bergman & Gaston-Johansson, 2006) og engelsk (Uronis et al., 2012) (Tabell 3).

CDS ble utviklet med involvering av representanter for pasientgruppen og eksperter på fagfeltet, og en preliminært versjon ble pilottestet blant 117 pasienter med kreft (Tanaka et al., 2000). Vi har vurdert at CDS har *positive* egenskaper med hensyn på **innholdsvaliditet**.

Tre studier (Tanaka et al., 2000 ; Hensch, Bergman & Gaston-Johansson, 2006 ; Uronis et al., 2012) evaluerte faktorstrukturen til CDS, og bekreftet tre-faktor strukturen og en samlefaktor for total dyspné. Noen av spørsmålene i svensk og engelsk versjon (Hensch, Bergman &

Gaston-Johansson, 2006 ; Uronis et al., 2012) lastet annerledes enn i originalversjonen (Tanaka et al., 2000). «Følelse av anstrengelse» forklarte 27 %, «følelse av angst» 21 % og «følelse av ubehag» 14 % av den totale variansen (Tanaka et al., 2000). Vi mangler informasjon om hvilken andel av variansen totalskåren forklarte og resultatet for **strukturell validitet** er derfor *uavklart*.

**Hypotesetesting** er gjennomført i alle de inkluderte studiene (Tanaka et al., 2000 ; 2002 ; Henoch, Bergman & Gaston-Johansson, 2006 ; Uronis et al., 2012) og CDS er sammenlignet med en rekke ulike kartleggingsverktøy (Vedlegg VII). Vi har vektlagt sammenligningene mellom CDS og HADS, Borg skala og visuell analog skala (VAS-dyspné) (Tabell 7). Dimensjonene «følelse av anstrengelse» og «følelse av angst» samvarierte sterkest med Borg skala (range 0,72-0,86; 0,41-0,64) og VAS-dyspné (range 0,61-0,80; 0,40-0,59). CDS totalskår samvarierte sterkest med Borg skala (range 0,67-0,87) og VAS-dyspné (range 0,63-0,82), men også med HADS total (range 0,35-0,57). Vi har vurdert at resultatet for hypotesetesting var *negativt* i to (Tanaka et al., 2002 ; Uronis et al., 2012) og *positivt* i to studier (Tanaka et al., 2000 ; Henoch, Bergman & Gaston-Johansson, 2006).

**Tabell 7** Utvalgte samvariasjoner for CDS

CDS	HADS-angst	HADS-depresjon	HADS-total	Borg skala	VAS-dyspné
Følelse av anstrengelse	0,30**	0,32**	0,34**	0,72*	0,77*
	0,36'	0,24'	0,53''	0,86''	0,61'
	0,50''	0,38''			0,80''
Følelse av angst	0,22**	0,21**	0,24**	0,41*	0,53*
	0,41'	0,38'	0,45''	0,64''	0,40'
	0,32''	0,44''			0,59''
Følelse av ubehag	0,21**	0,30**	0,28**	0,44*	0,40*
	0,24'	0,20'	0,50''	0,64''	0,55'
	0,40''	0,43''			0,59''
Total skår	0,30**	0,34**	0,35**	0,67*	0,72*
	0,36'	0,27'	0,57''	0,87''	0,63'
	0,49''	0,45''			0,82''

CDS (Cancer Dyspnea Scale); HADS (Hospital Anxiety and Depression Scale); VAS (Visual Analog Scale); \*Tanaka et al. (2000); \*\*Tanaka et al. (2002); 'Henoch et al. (2006); ''Uronis et al. (2012).

Tverrkulturell validitet er ikke evaluert, men **oversettelsesprosessen** til engelsk (Tanaka et al., 2000) og til svensk versjon (Henoch, Bergman & Gaston-Johansson, 2006) er beskrevet.

**Intern konsistens** er evaluert i tre studier og resultatet var *positivt* med Cronbachs alfa større enn 0,70 (range 0,81-0,94) for de tre dimensjonene (Tanaka et al., 2000 ; Henoeh, Bergman & Gaston-Johansson, 2006 ; Uronis et al., 2012), større enn 0,70 (range 0,71-0,90) for totalskår i to av studiene (Henoeh, Bergman & Gaston-Johansson, 2006 ; Uronis et al., 2012) og 0,69 i studien til Tanaka et al. (2000).

**Test-retest-reliabilitet** er evaluert i en studie og resultatet var *negativt* med Pearsons korrelasjonskoeffisient mindre enn 0,80 (0,58-0,71) (Tanaka et al., 2000).

### ***Severe Respiratory Insufficiency (SRI) Questionnaire***

Syv publikasjoner (Windisch et al., 2003 ; Duiverman et al., 2008 ; Lopez-Campos et al., 2008 ; Windisch et al., 2008 ; Ghosh et al., 2012 ; Struik et al., 2013 ; Markussen et al., 2014) har beskrevet utvikling og validering av SRI. SRI ble utviklet i Tyskland for å måle helse relatert livskvalitet siste uke, hos pasienter som lider av kronisk respirasjonssvikt og som får behandling med LTMV. Respirasjonssvikten skyldes ulike underliggende sykdommer (Windisch et al., 2003).

Spørreskjemaet består av 49 spørsmål fordelt på syv dimensjoner: åndedrettsplager (SRI-RC), fysisk funksjon (SRI-PF), tilleggssymptomer og søvn (SRI-AS), sosiale relasjoner (SRI-SR), angst (SRI-AX), psykisk velvære (SRI-WB), sosial funksjon (SRI-SF) og en samlefaktor for totalskår (SRI-SS) (Windisch et al., 2003). Hvert spørsmål skåres på en fem-punkts Likert-skala. Etter omkodning og summering av spørsmål kan hver dimensjon anta verdier mellom null og 100 (prosent av maks), der høyere skår indikerer bedre livskvalitet (Windisch et al., 2003 ; Duiverman et al., 2008 ; Windisch et al., 2008). SRI er validert på tysk (Windisch et al., 2003 ; Windisch et al., 2008), nederlandsk (Duiverman et al., 2008 ; Struik et al., 2013), spansk (Lopez-Campos et al., 2008), engelsk (Ghosh et al., 2012) og norsk (Markussen et al., 2014) (Tabell 3). I følge Ghosh et al. (2012) er SRI i tillegg også nylig profesjonelt oversatt til dansk, fransk, gresk, hebraisk, japansk, polsk og svensk.

Ved utarbeidelsen av spørreskjemaet ble spørsmålene utviklet basert på sosiale, psykologiske og fysiske helsedomener. Både pasienter og et ekspertpanel var involvert (Windisch et al., 2003). Markussen et al. (2014) piloterte norsk versjon av SRI blant seks LTMV-brukere. De ble også bedt om å vurdere kartleggingsverktøyets relevans, fullstendighet og gjennomførbarhet. I tillegg involverte de en ekspertgruppe, forfatteren av original versjonen

og en konsensusgruppe (Markussen et al., 2014). Vi har vurdert at SRI har *positive* egenskaper med hensyn på **innholdsvaliditet**.

**Strukturell validitet** er evaluert i tre studier (Lopez-Campos et al., 2008 ; Windisch et al., 2008 ; Ghosh et al., 2012) som alle fant at SRI består av flere faktorer enn de opprinnelige syv dimensjonene. Flere ulike analyser av faktorstrukturen ble gjennomført og resulterte i et ulikt antall faktorer. Både Windisch et al. (2008) og Ghosh et al. (2012) viste at faktorene innenfor en dimensjon samvarierte med hverandre noe som indikerer at de ikke er uavhengige faktorer. Begge studiene fant også at bare en faktor (SRI-SS) hadde en egenverdi større enn én, og bekreftet dermed en faktor for totalskår (SRI-SS). I studien til Windisch et al. (2008) forklarte SRI-SS 58,5 % av variansen for hele utvalget, mens den i studien til Ghosh et al. (2012) forklarte 70 %. Lopez-Campos et al. (2008) gjennomførte en utforskende faktoranalyse som forklarte 60 % av variansen, men som resulterte i 13 faktorer. Resultatet for strukturell validitet var *positivt*.

**Hypotesetesting** er gjennomført i seks av de inkluderte studiene (Windisch et al., 2003 ; Duiverman et al., 2008 ; Lopez-Campos et al., 2008 ; Ghosh et al., 2012 ; Struik et al., 2013 ; Markussen et al., 2014). Vedlegg VII viser en oversikt over tilgjengelige samvariasjoner. Vi har vektlagt sammenligningene mellom de dimensjonene vi har vurdert som mest relevant for SRI og Short Form Health Survey (SF-36) (Tabell 8), Chronic Respiratory Disease Questionnaire (CRQ), HADS og Medical Research Council Dyspnea Scale (MRC) (Tabell 9).

**Tabell 8 Utvalgte samvariasjoner mellom SRI og SF-36**

	SF-36 GH	SF-36 VT	SF-36 MH	SF-36 PHC	SF-36 MHC
<b>SRI-RC</b>	0,63 <sup>'''</sup>	0,52 <sup>'''</sup>	0,29 <sup>'''</sup>	0,53 <sup>'''</sup>	0,40 <sup>'''</sup>
	0,61 <sup>***</sup>	0,63 <sup>***</sup>	0,43 <sup>***</sup>	0,48 <sup>''</sup>	0,28 <sup>''</sup>
	0,47 <sup>''</sup>	0,54 <sup>''</sup>	0,24 <sup>''</sup>	0,42 <sup>*</sup>	0,45 <sup>*</sup>
	0,51 <sup>*</sup>	0,50 <sup>*</sup>	0,42 <sup>*</sup>		
<b>SRI-AX</b>	0,50 <sup>'''</sup>	0,44 <sup>'''</sup>	0,49 <sup>'''</sup>	0,42 <sup>'''</sup>	0,55 <sup>'''</sup>
	0,58 <sup>***</sup>	0,66 <sup>***</sup>	0,63 <sup>***</sup>	0,29 <sup>''</sup>	0,37 <sup>''</sup>
	0,48 <sup>''</sup>	0,37 <sup>''</sup>	0,37 <sup>''</sup>	0,28 <sup>*</sup>	0,52 <sup>*</sup>
	0,51 <sup>*</sup>	0,56 <sup>*</sup>	0,51 <sup>*</sup>		
<b>SRI-SS</b>	0,70 <sup>'''</sup>	0,65 <sup>'''</sup>	0,54 <sup>'''</sup>	0,62 <sup>'''</sup>	0,61 <sup>'''</sup>
	0,66 <sup>''</sup>	0,76 <sup>''</sup>	0,60 <sup>''</sup>	0,67 <sup>''</sup>	0,54 <sup>''</sup>
	0,63 <sup>*</sup>	0,73 <sup>*</sup>	0,70 <sup>*</sup>	0,46 <sup>*</sup>	0,67 <sup>*</sup>

SRI (Severe Respiratory Insufficiency Questionnaire); SF-36 (Short Form Health Survey); RC (åndedrettsplager); AX (angst); SS (totalskår); GH (generell helse); VT (vitalitet); MH (mental helse); PHC (fysisk helsekomponent); MHC (mental heleskomponent); <sup>'''</sup>Markussen et al. (2014); <sup>\*\*\*</sup>Gosh et al. (2012); <sup>''</sup>Lopez-Campos et al. (2008); <sup>\*</sup>Windisch et al. (2003).

SRI ble sammenlignet med det generiske verktøyet SF-36 i fire av studiene (Windisch et al., 2003 ; Lopez-Campos et al., 2008 ; Ghosh et al., 2012 ; Markussen et al., 2014). Vedlegg IX gir en oversikt over samvariasjonene mellom alle faktorene. Når vi vurderte studiene samlet for de to SRI dimensjonene som er spesifikke for respirasjon (SRI-RC) og angst (SRI-AX), var det sterkest samvariasjoner mellom SRI-RC og SF-36 dimensjonene vitalitet (VT) (range 0,50-0,63) og generell helse (GH) (range 0,47-0,63). For SRI-AX var det sterkest samvariasjoner med SF-36 dimensjonene mental helse (MH) (range 0,37-0,63) og GH (range 0,48-0,58). SRI-SS samvarierte sterkest med SF-36 VT (0,65-0,76) og SF-36 GH (0,63-0,70), men også med også med SF-36 MH (0,54-0,70) og de to oppsummerende faktorene til SF-36 fysisk helsekomponent (PHC) (range 0,46-0,67) og mental helsekomponent (MHC) (range 0,54-0,67) (Tabell 8). Vi har vurdert at resultatet var *positivt* i tre av studiene (Windisch et al., 2003 ; Ghosh et al., 2012 ; Markussen et al., 2014) og *negativt* i en studie (Lopez-Campos et al., 2008).

**Tabell 9 Utvalgte samvariasjoner mellom SRI og CRQ, HADS og MRC**

	CRQ- dyspné	CRQ- fatigue	CRQ- følelser	CRQ- mestring	CRQ- total	HADS- angst	HADS- depresjon	MRC
<b>SRI-RC</b>	0,57** 0,56'	0,59** 0,60'	0,40** 0,45'	0,34** 0,50'	0,54** 0,62'	0,50** 0,56'	0,41** 0,43'	0,43** 0,41'
<b>SRI-AX</b>	0,35** 0,49'	0,57** 0,52'	0,65** 0,67'	0,60** 0,70'	0,67** 0,74'	0,69** 0,74'	0,54** 0,59'	0,26** 0,33'
<b>SRI-SS</b>	0,53** 0,63'	0,74** 0,63'	0,73** 0,71'	0,62** 0,71'	0,81** 0,82'	0,75** 0,80'	0,68** 0,72'	0,41** 0,46'

SRI (Severe Respiratory Insufficiency Questionnaire); CRQ (Chronic Respiratory Disease Questionnaire); HADS (Hospital Anxiety and Depression Scale); MRC (Medical Research Council Dyspnea Scale); RC (åndedrettsplager); AX (angst); SS (totalskår); \*\*Duiverman et al. (2008); 'Struik et al. (2013).

SRI ble sammenlignet med CRQ, HADS, GARS og MRC i to studier (Duiverman et al., 2008 ; Struik et al., 2013). Vedlegg X gir en oversikt over samvariasjonene. GARS er et kartleggingsverktøy for aktiviteter i dagliglivet (ADL) og utdypes derfor ikke her. SRI-RC samvarierte sterkest med CRQ-fatigue (range 0,59-0,60) og CRQ-dyspné (range 0,56-0,57), mens SRI-AX samvarierte sterkest med HADS-angst (range 0,69-0,74), CRQ-følelser (range 0,65-0,67) og CRQ-mestring (range 0,60-0,70). SRI-SS samvarierte sterkest med CRQ-total (range 0,81-0,82), HADS- angst (range 0,75-0,80) og HADS-depresjon (0,68-0,72). Generelt var det sterkest samvariasjoner mellom SRI dimensjoner og dimensjoner for de andre

kartleggingsverktøyene som måler sammenlignbare aspekt ved livskvalitet, mens det var svakest samvariasjoner mellom dimensjoner som måler ulike aspekt (Tabell 9). Vi har vurdert at resultatet for hypotesetesting var *positivt* i begge studiene.

Markussen et al. (2014) har evaluert **tverrkulturell validitet** for den norske oversettelsen av SRI mens Ghosh et al. (2012) har beskrevet oversettelsesprosessen til britisk-engelsk versjon. I følge Lopez-Campos et al. (2008) er tverrkulturell validitet for spansk versjon av SRI evaluert i en studie fra 2006. Denne har vi bare identifisert i en spanskpråklig versjon og derfor ikke hatt mulighet til å vurdere. Resultatet for tverrkulturell validitet av norsk versjon av SRI (Markussen et al., 2014) var *negativt* fordi det ikke var gjennomført en sammenligning med originalversjonen for faktorstruktur eller differential item function (DIF).

**Intern konsistens** er evaluert i seks studier (Duiverman et al., 2008 ; Lopez-Campos et al., 2008 ; Windisch et al., 2008 ; Ghosh et al., 2012 ; Struik et al., 2013 ; Markussen et al., 2014). Resultatet var *positiv* for SRI-SS med Cronbachs alfa større enn 0,70 (range 0,88-0,94) (Lopez-Campos et al., 2008 ; Windisch et al., 2008 ; Ghosh et al., 2012 ; Markussen et al., 2014). Resultatene var også *positiv* for alle dimensjonene med Cronbachs alfa større eller lik 0,70 (range 0,70-0,89), bortsett fra SRI-AS (0,66 og 0,68) i to av studiene (Duiverman et al., 2008 ; Markussen et al., 2014) og SRI-SR (0,63 og 0,68) i to av studiene (Lopez-Campos et al., 2008 ; Struik et al., 2013) (Tabell 10). Resultatene indikerte at den interne konsistensen gjennomgående viste *positive* egenskaper.

**Tabell 10 Intern konsistens SRI**

	Duiverman et al., 2008	Lopez-Campos et al., 2008	Windisch et al., 2008	Ghosh et al., 2012	Struik et al., 2013	Markussen et al., 2014
<b>SRI-RC</b>	0,73	0,81	0,82	0,84	0,82	0,81
<b>SRI-PF</b>	0,73	0,82	0,82	0,77	0,80	0,76
<b>SRI-AS</b>	0,66	0,74	0,73	0,81	0,77	0,68
<b>SRI-SR</b>	0,78	0,63	0,74	0,81	0,68	0,82
<b>SRI-AX</b>	0,70	0,73	0,77	0,87	0,80	0,81
<b>SRI-WB</b>	0,84	0,85	0,82	0,89	0,85	0,88
<b>SRI-SF</b>	0,73	0,83	0,82	0,87	0,81	0,79
<b>SRI-SS</b>		0,93	0,88	0,93		0,94

SRI (Severe Respiratory Insufficiency Questionnaire); RC (åndedrettsplager); PF (fysisk funksjon); AS (tilleggssymptomer og søvn); SR (sosiale relasjoner); AX (angst); WB (psykisk velvære); SF (sosial funksjon); SS (totalskår).



**Test-retest- reliabilitet** er evaluert to studier (Duiverman et al., 2008 ; Lopez-Campos et al., 2008). Resultatet var *positivt* med ICC større enn 0,70 for SRI-SS (range 0,81-0,95) (Duiverman et al., 2008 ; Lopez-Campos et al., 2008) og for alle dimensjonene (range 0,84-0,86) (Lopez-Campos et al., 2008).

### ***Dyspnea Management Questionnaire (DMQ)***

Tre studier har beskrevet utvikling og validering av tre ulike versjoner av DMQ; DMQ- 30 (Norweg et al., 2006), DMQ-56 (Norweg et al., 2011a), og DMQ-CAT (Norweg et al., 2011b). Kartleggingsverktøyet ble utviklet i USA for å måle effekt av lungerehabilitering og endring over tid hos pasienter med kols, og for å tilrettelegge for mer målrettet behandling relevant for fagområder som for eksempel ergoterapi og psykologi (Norweg et al., 2006).

DMQ- 30 består av 30 spørsmål (Norweg et al., 2006). Som et ledd i utviklingen av en elektronisk versjon evaluerte Norweg et al. (2011a) en utvidet samling av spørsmål, og utviklet DMQ-56 som består av 56 spørsmål. Deretter utviklet og testet de en prototype på en elektronisk versjon, DMQ-CAT, som består av 20 spørsmål (Norweg et al., 2011b). Ett inngangsspørsmål stilles, deretter skreddersys DMQ-CAT til den enkelte pasient slik at bare de mest relevante og informative spørsmålene administreres. DMQ har fem konseptuelt utledete dimensjoner: «dyspnéintensitet», «dyspnérelatert angst», «angstrelatert aktivitetsunngåelse», «tiltro til aktivitetsrelatert mestringssevne» og «tilfredshet med mestringsstrategier». DMQ-30 og DMQ-56 består av alle de fem dimensjonene, mens DMQ-CAT ikke har med «tilfredshet med mestringsstrategier» (Norweg et al., 2006 ; Norweg et al., 2011a ; Norweg et al., 2011b). For DMQ-30 og DMQ-56 skåres spørsmålene på en syv-punkts Likert-skala. Høyere skår representerer bedre funksjon. Det kalkuleres en skår for hver dimensjon og for DMQ-30 kalkuleres det i tillegg en totalskår ved å summere «dyspnéintensitet», «dyspnérelatert angst» og «angstrelatert aktivitetsunngåelse» (Norweg et al., 2006 ; Norweg et al., 2011a). For DMQ-CAT skåres spørsmålene på en seks-punkts Likert-skala. Det er ikke beskrevet hvordan algoritmen for skår er (Norweg et al., 2011b). DMQ er validert på engelsk (Norweg et al., 2006 ; Norweg et al., 2011a ; Norweg et al., 2011b) (Tabell 3).

DMQ-30 ble utviklet ved hjelp av data fra kvalitative intervju, en litteraturgjennomgang på områdene pustebesvær, angst, unngåelsesadferd, funksjonsstatus, helserelatert livskvalitet, brukertilfredshet og lungerehabilitering, og en gjennomgang av andre kartleggingsverktøy. En

preliminær versjon ble pilottestet av tre voksne med kols. Både eksperter og pasienter var involvert i utviklingen (Norweg et al., 2006). Vi har vurdert at DMQ-30 har *positive* egenskaper med hensyn på **innholdsvaliditet**. Utviklingen av DMQ-CAT var basert på en utvidelse av samlingen av spørsmålene fra DMQ-56. Involvering av lungemedisinske spesialister, pasienter med kols og en omfattende litteraturgjennomgang dannet grunnlaget for utvidelsen av spørsmål (Norweg et al., 2011b). Vi har derfor vurdert at DMQ-CAT også har *positive* egenskaper for innholdsvaliditet.

**Strukturell validitet** er evaluert for DMQ-56 og DMQ-CAT (Norweg et al., 2011a ; Norweg et al., 2011b). Bekreftende faktoranalyse og testing av modelltilpasning støttet den opprinnelige fem-faktormodellen og viste at fem distinkte og meningsbærende dimensjoner var den beste løsningen. «Dyspnéintensitet» forklarer 71 %, «dyspnérelatert angst» 31 %, «angstrelatert aktivitetsunngåelse» 20 %, «tiltro til aktivitetsrelatert mestringssevne» 9 %, og «tilfredshet med mestringsstrategier» 31 % av variansen (Norweg et al., 2011a). Resultatet for strukturell validitet for DMQ-56 var *positivt*. Ulike faktoranalyser ble gjennomført og item respons theory (IRT) modeller testet for å finne faktorstrukturen til DMQ-CAT. Modellen som passet best var en flerdimensjonal ikke-hierarkisk modell med fire dimensjoner, som forklarte en samlet varians på 63 % (Norweg et al., 2011b). Resultatet for strukturell validitet for DMQ-CAT var *positivt*.

**Hypotesetesting** for DMQ-30 (Norweg et al., 2006) viste at faktoren «dyspnéintensitet» samvarierer sterkest med Seattle Obstructive Lung Disease Questionnaire (SOLQ)-fysisk funksjon (0,69), SOLQ-følelsesmessig funksjon (0,56) og SF-12 «fysisk komponent oppsummert» (0,54). «Dyspnérelatert angst» samvarierte sterkest med SOLQ-mestringsferdigheter (0,72), SOLQ-følelsesmessig funksjon (0,65) og HADS-angst (0,60) (Tabell 11). Generelt var det sterkest samvariasjoner mellom DMQ-30 dimensjoner og dimensjoner for de andre kartleggingsverktøyene som måler sammenlignbare aspekt, mens det var svakest samvariasjoner mellom dimensjoner som måler ulike aspekt. Vi har vurdert at resultatet for hypotesetesting var *positivt*.

**Tabell 11 Samvariasjoner for DMQ-30**

DMQ-30	SOLQ- fysisk funksjon	SOLQ- følelsesmessig funksjon	SOLQ- mestrings- ferdigheter	SOLQ- tilfredshet med omsorg	SF-12 PCS	SF-12 MCS	HADS-angst
Dyspnéintensitet	0,69	0,56	0,52	-0,04	0,54	0,31	0,39
Dyspnérelatert angst	0,44	0,65	0,72	-0,07	0,30	0,41	0,60
Angstrelatert aktivitetsunngåelse	0,67	0,65	0,61	0,01	0,40	0,56	0,59
Tiltro til aktivitetsrelatert mestringsevne	0,22	0,28	0,25	-0,01	0,23	0,04	0,05
Tilfredshet med mestringsstrategier	0,02	0,16	0,21	0,27	0,06	0,03	0,01
<b>DMQ composite*</b>	0,83	0,81	0,74	-0,05	0,57	0,46	0,65

DMQ (Dyspnea Management Questionnaire); SOLQ (Seattle Obstructive Lung Disease Questionnaire); SF-12 (Short Form Health Survey); HADS (Hospital Anxiety and Depression Scale); PCS (fysisk komponent oppsummert); MCS (mental komponent oppsummert); \* sumskår av dyspnéintensitet, dyspnérelatert angst og angstrelatert aktivitetsunngåelse.

Sammenligningene av DMQ-CAT med University of California, San Diego Shortness of Breath Questionnaire (UCSD SOBQ), CRQ, COPD Self-Efficacy Scale (CSES) og HADS viste at «dyspnéintensitet» samvarierte sterkest med UCSD SOBQ, CRQ-mestring, CSES og CRQ-dyspné (range - 0,80-0,47), mens DMQ-CAT «dyspnérelatert angst» samvarierte sterkest med CRQ-mestring, CSES, CRQ-følelser og HADS-angst (range - 0,61-0,75) (Norweg et al., 2011b) (Tabell 12). Generelt var det sterkest samvariasjoner mellom DMQ-CAT dimensjoner og dimensjoner for de andre kartleggingsverktøyene som måler sammenlignbare aspekt, mens det var svakest samvariasjoner mellom dimensjoner som måler ulike aspekt. Vi har vurdert at resultatet for hypotesetesting var *positivt*.

**Tabell 12 Samvariasjoner for DMQ-CAT**

DMQ-CAT	UCSD SOBQ	CRQ- dyspné	CRQ- følelser	CRQ- mestring	CSES	HADS- angst
Dyspnéintensitet	-0,80	0,47	0,41	0,53	0,51	- 0,31
Dyspnérelatert angst	-0,59	0,45	0,62	0,75	0,66	- 0,61
Angstrelatert aktivitetsunngåelse	-0,72	0,37	0,45	0,58	0,47	- 0,36
Tiltro til aktivitetsrelatert mestringsevne	-0,69	0,36	0,48	0,64	0,71	- 0,42

DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); UCSD SOBQ (University of California, San Diego Shortness of Breath Questionnaire); CRQ (Chronic Respiratory Disease Questionnaire); CSES (COPD Self-Efficacy Scale); HADS (Hospital Anxiety and Depression Scale).

**Kriterievaliditet** for DMQ-CAT ble evaluert ved en sammenligning med den totale samlingen på 71 spørsmål (Norweg et al., 2011b). Samvariasjonene var høye for alle dimensjonene (range 0,94-0,97). Resultatet for kriterievaliditet var *positivt*.

Resultatet for **intern konsistens** var *positivt* for alle de tre versjonene av DMQ med Cronbachs alpha større enn 70 for de fem faktorene til DMQ-30 (range 0,87-0,93) og DMQ-56 (range 0,85-0,96), og for de fire faktorene til DMQ-CAT (range 0,92-0,98) (Norweg et al., 2006 ; Norweg et al., 2011a ; Norweg et al., 2011b).

ICC var større enn 0,70 for alle fem faktorene i DMQ-30 (range 0,71-0,92), for fire av faktorene i DMQ-56 (range 0,83-0,92) og 0,69 for faktoren «tilfredshet med mestringsstrategier» (Norweg et al., 2006 ; Norweg et al., 2011a). Resultatet for **test-retest-reliabilitet** var *positivt* for DMQ-30 og DMQ-56.

### **Respiratorisk symptomsjekkliste (RSC)**

En respiratorisk symptomsjekkliste bestående av spontant rapporterte utsagn som beskriver sensoriske erfaringer med pustebesvær, følelsesmessige aspekt og atferdsmessige komponenter ble administrert til pasienter med ulike hjerte- og lungelidelser for å identifisere ulike dimensjoner av pustebesvær. Pasientene svarte hvor ofte de hadde opplevd utsagnet siste måned på en fem-punkts Likert-skala (Han et al., 2011) (Tabell 3).

Det er bare måleegenskapen **strukturell validitet** som var evaluert for RSC. Faktoranalyse viste tolv sammenhengende og ikke-korrelerte faktorer som forklarte 64 % av den totale variansen. Ytterligere kriterier ble benyttet og resulterte i syv faktorer som synes å måle tre dimensjoner av pustebesvær: sensoriske kvaliteter (fire faktorer), affektive aspekt (to faktorer) og atferdsmessig påvirkning (en faktor). De resterende to faktorene «prikking» og «hoste» sorterte under andre symptomer (Han et al., 2011). Resultatet for strukturell validitet var *positivt*.

#### 4.4 Sammenstilling og gradering av kartleggingsverktøyenes måleegenskaper

I dette kapittelet har vi oppsummert resultatene på tvers av enkeltstudiene i en kvalitativ syntese (Tabell 13), der vi har tatt hensyn til den metodiske kvaliteten til de tilgjengelige studier (Kapittel 4.2) og de dokumenterte måleegenskapene (Kapittel 4.3).

**Tabell 13 Kvalitativ syntese**

	Intern konsistens	Reliabilitet	Innholdsvaliditet	Strukturell validitet	Hypotese-testing	Kriterievaliditet	Tverrkulturell validitet
<b>CDS</b>	+++	-	+++	ukjent	+/-	na	na
<b>SRI</b>	?	+	+++	?	+++	na	?
<b>DMQ-30</b>	++	?	+++	na	++	na	na
<b>DMQ-56</b>	++	+	na	?	na	na	na
<b>DMQ-CAT</b>	++	na	+++	+	+	++	na
<b>RSC</b>	na	na	na	+++	na	na	na

CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); +++ eller ---, høy evidens positivt/negativt resultat; ++ eller --, moderat evidens positivt/negativt resultat; + eller -, lav evidens positivt/negativt resultat; ± motstridene evidens; ?, usikker på grunn av svak metodisk kvalitet; ukjent, på grunn av ukjent resultat; na, ingen informasjon tilgjengelig.

#### **Cancer Dyspnea Scale (CDS)**

Dokumentasjonen visert at vi kan ha høy tillit til at CDS har god innholdsvaliditet. På grunn av uklar rapportering er det mer usikkerhet knyttet til den strukturelle validiteten. Den samlede dokumentasjonen om hypotesetesting er motstridende.

Samlet dokumentasjon viser at vi kan ha høy tillit til at internkonsistensen til CDS er god. Dokumentasjonen viser at vi kan ha begrenset tillit at resultat for test-retest-reliabiliteten er negativt.

### ***Severe Respiratory Insufficiency (SRI) Questionnaire***

Dokumentasjonen visert at vi kan ha høy tillit til at SRI har god innholdsvaliditet. Det er mer usikkerhet knyttet til den strukturelle validiteten, noe som først og fremst skyldes metodiske begrensninger i studiene der dette er evaluert. På bakgrunn av den samlede dokumentasjonen kan vi være sikre på at SRI samvarierer med andre PROMs som måler tilsvarende egenskaper. Uklar rapportering medfører at det er mer usikkerhet knyttet til den tverrkulturell validitet av norsk versjon.

På grunn av studienes svake metodiske kvalitet er det usikkerhet knyttet til den interne konsistensen. Dokumentasjonen viser at vi kan ha begrenset tillit til at test-retest-reliabiliteten er god.

### ***Dyspnea Management Questionnaire (DMQ)***

Dokumentasjonen visert at vi kan ha høy tillit til at DMQ-30 og DMQ-CAT har god innholdsvaliditet. Vi kan ha begrenset tillit til at den strukturelle validiteten til DMQ-CAT er god, mens er det mer usikkerhet knyttet til den strukturelle validiteten til DMQ-56 som følge av metodisk begrensninger i studien der dette ble evaluert. Dokumentasjonen om hypotesetesting viser at vi kan ha moderat tillit til at DMQ-30 samvarierer og begrenset tillit til at DMQ-CAT samvarierer med andre PROMs som måler tilsvarende egenskaper. Vi kan ha moderat tillit til at kriterievaliditeten til DMQ-CAT er god.

Dokumentasjonen viser at vi kan ha moderat tillit til at den interne konsistensen til alle tre versjonene av DMQ er god. Vi kan ha begrenset tillit til at test-retest-reliabiliteten er god for DMQ-56, mens er det mer usikkerhet knyttet til test-retest-reliabiliteten for DMQ-30 noe som først og fremst skyldes metodiske begrensninger i studien der dette ble målt.

### ***Respiratorisk symptomsjekkliste (RSC)***

Dokumentasjonen viser at vi kan ha høy tillit til at den strukturelle validiteten til RSC er god, men for andre måleegenskaper mangler vi dokumentasjon.

## 5 Diskusjon

Vi identifiserte fire PROMs for kartlegging av pustebevis og som tilfredsstilte våre inklusjonskriterier. De måler alle sammensatte aspekt ved symptomet og innehar dimensjoner for angst og pustebevis. Ingen av de inkluderte PROMs er direkte validert for bruk i lindrende behandling til kartlegging av pustebevis hos palliative pasienter uavhengig av grunnsykdom. Bare et av kartleggingsverktøyene (SRI) er oversatt til norsk.

Ingen av studiene evaluerte målefeil og responsivitet. Fem av studiene evaluerte innholdsvaliditet. Ni studier evaluerte strukturell validitet, tolv studier hypotesetesting, en studie tverrkulturell validitet og en studie evaluerte kriterievaliditet. Intern konsistens er evaluert i tolv studier og test-retest-reliabilitet i fem studier. Resultatene for måleegenskapene er i all hovedsak vurdert som positive. Studiene har metodiske svakheter for de fleste måleegenskapene, noe som særlig skyldes for små utvalg og mangelfulle hypoteser. Dette begrenser tilliten til resultatene.

Den systematiske oversikten oppsummerer hvilke PROMs som er tilgjengelig for formålet og hvilken dokumentasjon som er tilgjengelig om deres måleegenskapene. 13 av de 16 studiene er publisert etter at litteratursøket til Bausewein et al. (2007) og Dorman, Byrne og Edwards (2007) var avsluttet.

Metodiske valg har jeg i hovedsak diskutert og begrunnet i Kapittel 3. I dette kapitlet drøfter jeg konsekvenser av de metodiske valgene og oversiktsartikkelens styrker og svakheter. Deretter diskuterer jeg metodiske utfordringer ved de inkluderte studiene før resultatene for de inkluderte PROMs drøftes. Avslutningsvis diskuteres relevans for praksis og fremtidig forskning.

### 5.1 Konsekvenser av metodiske valg, oversiktens styrker og svakheter

Jeg valgte å utarbeide en systematisk oversikt for å besvare masteroppgavenes forskningsspørsmål. En fordel ved dette studiedesignet er at det oppsummerer all tilgjengelig forskning om en aktuell problemstilling på en gjennomiktig og etterprøvbar måte. Leseren kan få svar på spørsmål fra oppsummert kunnskap ved å lese en artikkel og kan gjøre egne vurderinger på bakgrunn av oversiktens funn og tillit til kvaliteten av det arbeidet som er gjort. Sykepleien Forsknings forfatterveiledning har begrensninger som medfører at all dokumentasjon som er nødvendig for å sikre en fullt ut gjennomiktig prosess ikke kan

tilgjengeliggjøres i tidsskriftet. Dokumentasjonen vil være tilgjengelig i denne innledningen som skal publiseres på Bergen Open Research Archive (BORA) (<http://bora.hib.no/nb/>).

Bruk av COSMIN som metodisk rammeverk har sikret at jeg har ivaretatt viktige faktorer og hensyn spesielt for utarbeidelse av systematiske oversikter over måleegenskaper ved PROMs og benyttet taksonomi, terminologi og definisjoner av måleegenskaper som det er internasjonal konsensus for. Ved å benytte COSMIN har vi også anvendt internasjonalt anerkjente kvalitetskrav ved vurdering av studienes metodiske kvalitet. En utfordringen med COSMIN er at metodeverket fortsatt er under utvikling slik Terwee et al. (2012) og de Vet et al. (2011, s. 275) påpeker. For å klassifisere resultater og oppsummere disse har vi derfor benyttet kriterier fra nyere studier om måleegenskaper (Schellingerhout et al., 2011 ; Schellingerhout et al., 2012) som også har benyttet COSMIN som metodisk rammeverk. Dersom disse kriteriene blir endret i fremtiden og vil det kunne påvirke resultatet av den kvalitative syntesen vi har gjennomført.

### ***Identifisering av studier***

Oversikten bygger på et omfattende og systematisk litteratursøk i alle databasene vi har vurdert som relevant for problemstillingen og som jeg har hatt tilgang til. I tillegg har jeg gjennomgått alle inkluderte studier og relevante systematiske oversikter for å fange opp andre aktuelle studier. Jeg har gjort et stort arbeid for å sikre at jeg har funnet så mange relevante søkeord som mulig og søket ble gjennomført uten begrensninger for språk og tidsperiode. Det tette samarbeidet mellom meg som kliniker og forfatter av artikkelen, og bibliotekar har vært en styrke for litteratursøket.

En begrensning ved litteratursøket er at jeg på grunn av prøvesøkets størrelse måtte legge til søkeord for pasientpopulasjonen i søkestrategien. Den palliative pasientpopulasjonen omfatter alle alvorlig syke som ikke har mulighet for å bli frisk. Det er altså i utgangspunktet en populasjon der aktuell helsetilstand og behov er av større betydning enn diagnose. Med bakgrunn i oversiktene til Bausewein et al. (2007) og Dorman, Byrne og Edwards (2007), og med egen kjennskap til fagfeltet var det klart at søkeord for den palliative populasjonen ikke ville fange opp alle relevante PROMs. Søkeord for relevante diagnoser ble derfor også lagt til populasjonen. Litteratursøket og oppgavens størrelse medførte at jeg ikke hadde mulighet for å søke etter alle relevante diagnoser og valgte derfor å begrense til de største gruppene som er kreft og kols. Det medfører en risiko for at jeg ikke har fanget opp potensielt relevante



PROMs som bare er validert for eksempelvis hjertesvikt og nevrologisk sykdom. En annen svakhet ved litteratursøket er at jeg ikke inkluderte søkeord for respirasjonssvikt. Dette oppdaget jeg ved at flere studier om SRI ikke var fanget opp i søket, men ble funnet ved gjennomgang av relevante studier. Det er derfor en risiko for at jeg har mistet relevante studier som har benyttet dette eller andre søkeord for pustebesvær som ikke var inkludert i min søkestrategi.

Oversiktsartikkelen kan ha publikasjonsskjevheter da jeg ikke har hatt kapasitet til å gjennomføre supplerende søk på de inkluderte PROMs, søke etter «grå litteratur» eller oppdatere søket før ferdigstilling av oversikten.

### ***Utvelgelse, kvalitetsvurdering og dataekstraksjon***

Vi pilottestet inklusjons- og eksklusjonskriterier, COSMIN-sjekklisten og skjemaet for dataekstraksjon før vi startet med utvelgelse, kvalitetsvurdering og dataekstraksjon. Pilottesting er viktig for å sikre en felles oppfatning av kriteriene for det arbeidet som skal utføres, og for å kunne foreta nødvendige avklaringer, justeringer og presiseringer før selve gjennomføringen (Higgins & Green, 2011, s. 154). Pilottesting viste at vi i liten grad hadde behov for å gjøre tilpasninger av kriteriene for inklusjon og eksklusjon eller for dataekstraksjonsskjemaet. Pilottesting av COSMIN-sjekklisten viste imidlertid at vi hadde behov for å innhente mer kunnskap om måleegenskaper og evaluering av disse før vi kunne gjennomføre kritisk vurdering.

Utvelgelse og kvalitetsvurdering av studier ble utført av to uavhengig av hverandre. Det reduserer sjansen for at relevante publikasjoner ble forkastet og sikrer en bedre og mer konsistent kvalitetsvurdering. Dataekstraksjon bør ideelt sett også utføres av to uavhengig av hverandre, men som mastergradsprosjekt hadde vi ikke kapasitet til dette. Alle dataekstraksjoner ble uthentet av en og kontrollert av en annen. Uenigheter i prosessen med utvelgelse, kritisk vurdering og dataekstraksjon er løst ved diskusjon eller ved å involvere en tredje person.

Bruk av COSMIN-sjekklisten gjorde det mulig for oss å vurdere studienes metodiske kvalitet utfra de særskilte metodiske krav som stilles for hver enkelt måleegenskap. Vi har dermed fått en differensiert kvalitetsvurdering som bidrar i vurderingen av hvilken tillit vi kan ha til et PROMs samlede kvalitet. For å sikre at vi har vurdert studiene konsistent har vi avslutningsvis gjennomgått kvalitetsvurderingen av alle studiene i en sammenheng. En

utfordring med sjekklister er at mange av spørsmålene krever en subjektiv vurdering noe som medfører en mulighet for at andre ville vurdert annerledes enn oss. Det er også en begrensning at det ikke alltid har vært mulig å skille mellom mangelfull rapportering og dårlig kvalitet. Det er en svakhet ved kvalitetsvurderingen at jeg ikke har hatt kapasitet til å kontakte studienes forfattere for å avklare uklarhet omkring dette.

## 5.2 Metodiske utfordringer ved de inkluderte studiene

Metodiske utfordringer og uklar rapportering var en betydelig utfordring ved de inkluderte studiene. Av de fem studiene som evaluerte innholdsvaliditet ble en vurdert til å være av *god* og fire av *utmerket* metodisk kvalitet. Ni studier vurdere strukturell validitet av disse ble fire vurdert å være av *svak*, en av *rimelig*, en av *god* og tre av *utmerket* metodisk kvalitet. Av studiene som evaluerte hypotesetesting ble en sammenligning vurdert å være av *svak*, tre av *rimelig* og ti av *god* metodisk kvalitet. En studie av *god* kvalitet har vurdert kriterievaliditet, mens tverrkulturell validitet var evaluert i en studie av *svak* metodisk kvalitet. Av de tolv studiene som har vurdert intern konsistens er seks vurdert å være av *svak*, fire av *god* og to av *utmerket* metodisk kvalitet. Evaluering av test-retest reliabilitet skårer *svakt* for metodisk kvalitet i to og *rimelig* i tre studier.

Jeg har identifisert én systematisk oversikt (Terwee et al., 2011) som oppsummerer metodisk kvalitet ved studier om måleegenskaper og benytter denne i noen grad til sammenligning med våre funn. Den har sett på kartleggingsverktøy for nakkesmerter og funksjonsnedsettelse, men det er ifølge forfatterne sannsynlig at tilsvarende metodiske styrker og svakheter finnes på andre felt.

Det var særlig to hovedutfordringer ved den metodiske kvaliteten til studiene vi har inkludert. Den ene var størrelsen på utvalget i studiene. Dette samsvarer med funnene til Terwee et al. (2011) der utvalgsstørrelsen også var en betydelig utfordring for evaluering av flere av måleegenskapene. For å skåre *utmerket* må utvalget som inkluderes i analysene være minimum 100 for alle måleegenskapene med unntak av for innholdsvaliditet (Mokkink et al., 2012, s. 22 ; Terwee et al., 2012). Dette har spesielt vært en utfordring ved evaluering av test-retest-reliabilitet der ingen av studiene hadde tilstrekkelig antall respondenter. For evaluering av strukturell validitet og faktorenes endimensjonalitet anbefales det at utvalget som inkluderes i analysene skal være fem til syv ganger antall spørsmål og minimum 100 (Mokkink et al., 2012, s. 22 ; Terwee et al., 2012). Dette har særlig slått ut for de

kartleggingsverktøyene som består av mange spørsmål. SRI er et eksempel på dette der alle studiene fikk skåren *svak* for metodisk kvalitet for intern konsistens og strukturell validitet.

Studier om håndtering av pustebesvær er ofte preget av metodiske svakheter og lav styrke (Dorman et al., 2009). Manglende data og frafall er et kjent problem og kan skyldes at pasientene blir dårligere eller dør mens studien pågår. Dette er forventet utfall ved mange studier som involverer palliative pasienter (Preston et al., 2013). Ellers godt metodisk gjennomførte studier kan dermed komme dårlig ut i kvalitetsvurdering. Forskning i lindrende behandling bør fortrinnsvis gjennomføres som multisenterstudier for å sikre tilstrekkelig rekruttering (Dorman et al., 2009). For å bedre kvaliteten på studiene har MORECare (Methods Of Researching End of life Care) blant annet utviklet retningslinjer for hvordan manglende data og frafall i studier om lindrende behandling og omsorg ved livets slutt bør håndteres (Preston et al., 2013).

En annen hovedutfordringen ved studienes metodisk kvalitet var knyttet til hypotesetesting. Flere av studiene oppgav ikke forhåndsdefinerte hypoteser om samvariasjon med kartleggingsverktøyene de ble sammenlignet med. Det var også en svakhet i flere av studiene at verktøyene for sammenligning var mangelfullt beskrevet. Noen studier oppgav hypoteser, men disse var i all hovedsak vage og lite spesifikke, særlig om forventet retning og størrelse på samvariasjonen. Denne utfordringen kan eksemplifiseres med CDS der dokumentasjonen for hypotesetesting er motstridende. I tillegg til at flere av samvariasjonene var mindre enn 0,50 var samvariasjonen med HADS-angst i to av tre studier sterkere med faktoren «følelse av anstrengelse» enn med «følelse av angst», mens HADS-total samvarierte sterkere med faktorene «følelse av anstrengelse» og «følelse av ubehag» enn med «følelse av angst». Når studiene mangler tydelige forhåndsdefinerte hypoteser er det fristende å tenke ut alternative forklaringer på lave samvariasjoner istedenfor å konkluderer med at kartleggingsverktøyet kanskje ikke er valid (de Vet et al., 2011, s. 172-5). Kan resultatene tolkes som en bekreftelse på den nære sammenhengen mellom pustebesvær og angst, er det et resultat av utfordringen med å finne kartleggingsverktøy som måler sammenlignbare aspekt eller betyr resultatene at CDS ikke er valid? Terwee et al. (2011) fant i sin oversikt at i underkant av 50 % av studiene hadde forhåndsformulerte hypoteser og av disse igjen hadde bare 48 % oppgitt forventet retning og størrelse på samvariasjonen. Manglende hypoteser øker risikoen for systematiske skjevheter (de Vet et al., 2011, s. 211).

Studiens metodiske kvalitet er avgjørende for hvorvidt vi kan stole på kartleggingsverktøyets kvalitet. En mulig forklaring på de metodiske utfordringen kan være at studiene ble gjennomført før det var internasjonal konsensus om hvordan måleegenskaper bør evalueres og mangel på kjennskap til COSMIN-konseptets krav til rapportering av valideringsstudier. Terwee et al. (2011, s. 261) oppsummerer den metodiske kvalitets betydning slik: Hvis kvaliteten til en studie er tilfredsstillende er resultatene gyldig og kartleggingsverktøyet kan være nyttig i klinisk arbeid eller forskning. Hvis kvaliteten til en studie ikke er tilfredsstillende så kan vi ikke stole på resultatet og kvaliteten til kartleggingsverktøyet forblir uklart.

### 5.3 Kartleggingsverktøyenes måleegenskaper

Innholdsvaliditet er en viktig egenskap ved et kartleggingsverktøy og fokuserer på om kartleggingsverktøyets innhold samsvarer med fenomenet det har til hensikt å måle, både i forhold til relevans og fullstendighet (de Vet et al., 2011, s. 150 ; s. 154-8). Vi mangler informasjon om innholdsvaliditeten for DMQ-56 og RSC, mens den er vurdert å være god for CDS, SRI, DMQ-30 og DMQ-CAT.

Kartleggingsverktøyenes begrepsvaliditet er avhengig av den strukturelle validiteten og hypotesetesting. For oversatte PROM er i tillegg den tverrkulturelle validiteten en viktig måleegenskap. Vi mangler informasjon om den strukturelle validiteten til DMQ-30, mens den er uavklart for CDS, SRI og DMQ-56. For DMQ-CAT har vi begrenset tillit til en positiv strukturell validitet. Av de inkluderte PROM er det bare for RSC vi har høy tillit til den strukturelle validiteten. Vi mangler informasjon om hypotesetesting for DMQ-30 og RSC. For DMQ-CAT og DMQ-30 har vi begrenset til moderat tillit til samvariasjonen med andre kartleggingsverktøy som måler tilsvarende egenskaper. SRI er det eneste av de inkluderte PROM der vi har høy tillit til at det samvarierer med andre kartleggingsverktøy som måler tilsvarende egenskaper. Både CDS og SRI er oversatt til flere språk. Det er svakheter ved prosessen ved alle oversettelsene. Tverrkulturell validitet er bare vurdert for oversettelsen av SRI til norsk og måleegenskapen er uavklart fordi dokumentasjonen er av svak kvalitet. Dermed har vi ikke tilstrekkelig informasjon om begrepsvaliditeten til noen av de inkluderte PROM og kan ikke være sikker på at de måler som forventet utfra eksisterende kunnskap om fenomenet (de Vet et al., 2011, s. 150).

For at vi gyldig skal kunne summere spørsmålene innenfor en dimensjon trengs moderate til høye samvariasjoner mellom spørsmålene (Cronbachs alfa lik eller over 0,70). Vi mangler informasjon om intern konsistensen til RSC. For SRI er den interne konsistensen ukjent, mens vi har moderat tillit til at intern konsistensen til de tre versjonene av DMQ og høy tillit til at den interne konsistensen til CDS er god. Vi kan derfor regne med at det er forsvarlig å summere skårene innenfor hver dimensjon til en sumskår for de tre versjonene av DMQ og CDS og i tillegg en totalskår for CDS.

Reliabiliteten til et kartleggingsverktøy gir oss informasjon om hvor stor del av variasjonen mellom målinger som skyldes faktiske forskjeller og ikke er et resultat av målefeil (Mokkink et al., 2010c). Vi mangler informasjon om test-retest-reliabiliteten til RSC og DMQ-CAT. Test-retest-reliabiliteten til DMQ-30 er uavklart, mens vi har begrenset tillit til at den er negativ for CDS. For SRI og DMQ-56 har vi begrenset tillit til at test-retest-reliabiliteten er positiv. Det er dermed en viss usikkerhet om de endringene som måles med de inkluderte PROMs skyldes faktiske endringer hos pasientene eller kan være resultat av målefeil.

Vi har ikke informasjon om noen av kartleggingsverktøyenes responsivitet. Dette er en viktig måleegenskap for å fange opp endringer over tid i fenomenet som måles (Mokkink et al., 2010c). Vi mangler dermed informasjon om de inkluderte PROMs evne til å måle endringer ved for eksempel symptomutvikling eller effekt av tiltak.

#### 5.4 Betydning for praksis og videre forskning

Det er altså betydelig mangler i den samlede dokumentasjonen vi har identifisert om måleegenskaper for alle de inkluderte PROM. Det er derfor ingen som peker seg ut som mer fordelaktig enn andre med tanke på kvalitet. Begrenset tillit til dokumentasjonen betyr ikke at måleegenskapene er dårlige, men at de fortsatt er uavklart og at ytterligere validering er nødvendig.

Viktige barrierer for bruk i praksis kan være at kartleggingsverktøyet er for langt for pasienter å besvare eller krever mye tid til administrering som for eksempel kompliserte skåringssystemer, kostnader relatert til bruk av kartleggingsverktøyet, eller dårlig tilgjengelighet. Dette er særlig viktig i lindrende behandling der pasientenes helsetilstand kan være betydelig svekket samt at tid og krefter er begrenset. Kartleggingsverktøyene bør og være egnet for bruk i ulike settinger som i hjemmet, poliklinisk eller på institusjon (Bausewein et al., 2011, s. 13-4). Både RSC, SRI og DMQ-56 er PROMs som inneholder

mange spørsmål, og forventet utfyllingstid er ti til 15 minutter for RSC og 20 minutter for SRI. CDS er i så måte mer akseptabelt med tolv spørsmål og en forventet utfyllingstid på to minutter. DMQ-CAT synes også å være brukervennlig ved at det skreddersys til den enkelte pasients svar slik at bare de mest informative og relevante spørsmålene administrerers. DMQ-CAT vil kreve tilgjengelighet til datautstyr og kompetanse til å bruke dette.

Bortsett fra SRI må de inkluderte kartleggingsverktøyene gjennomgå en oversettelsesprosess etter internasjonale standarder før de er tilgjengelig i norsk versjon. Videre forskning er nødvendig for å validere kartleggingsverktøyene for bruk i lindrende behandling til en populasjon som ikke er diagnosespesifikk. Dette samsvarer med funn fra studiene til Bausewein et al. (2007) og Dorman, Byrne og Edwards (2007). Våre funn pekere også på et betydelig behov for valideringsstudier av god kvalitet som har store nok utvalg og tydelig definerer hypoteser om forventede samvariasjoner. Det er viktig at det gjennomføres studier som vurderer responsivitet. COSMIN kan med fordel benyttes som veiledning for planlegging og rapportering av nye valideringsstudier.

## **6 Konklusjon**

Vi identifiserte fire kartleggingsverktøy for pustebesvær som tilfredstilte våre inklusjonskriterier: CDS, SRI, DMQ og RSC. Den samlede kvaliteten til kartleggingsverktøyene er uavklart som følge av metodiske svakheter ved studiene og at det mangler evaluering av viktig egenskaper som blant annet responsivitet. Selv om noen av kartleggingsverktøyene er bedre studert enn andre er det foreløpig ikke mulig å si om noen av de inkluderte verktøyene har bedre måleegenskaper enn de andre. CDS og DMQ-CAT synes mest brukervennlige med kortest tid til utfylling og minst belastning for pasientene. SRI er det eneste av kartleggingsverktøyene som er validert i norsk versjon.

Videre forskning er nødvendig for å oversette og validere kartleggingsverktøyene for bruk i lindrende behandling i Norge. Nye studier bør sikre tilstrekkelig store utvalg og en tydelig formulering av hypoteser om forventet samvariasjon med andre kartleggingsverktøy.

## 7 Referanseliste

American Thoracic Society (1999) Dyspnea: mechanisms, assessment, and management. A consensus statement *American Journal of Respiratory and Critical Care Medicine*, 159 (1), s. 321-40.

Antunes, B., Harding, R. & Higginson, I. J. (2014) Implementing patient-reported outcome measures in palliative care clinical practice: A systematic review of facilitators and barriers. *Palliative Medicine*, 28 (2), s. 158-75.

Bailey, P. H. (2004) The dyspnea-anxiety-dyspnea cycle--COPD patients' stories of breathlessness: "It's scary /when you can't breathe". *Qualitative Health Research*, 14 (6), s. 760-78.

Bausewein, C., Booth, S. & Higginson, I. J. (2008) Measurement of dyspnoea in the clinical rather than the research setting. *Current Opinion in Supportive & Palliative Care*, 2 (2), s. 95-9.

Bausewein, C., Daveson, B., Benalia, H., Simon, S. T. & Higginson, I. J. (2011) *Outcome Measurement in Palliative Care: The Essentials*. London, Prisma.

Bausewein, C., Farquhar, M., Booth, S., Gysels, M. & Higginson, I. J. (2007) Measurement of breathlessness in advanced disease: a systematic review. *Respiratory Medicine*, 101 (3), s. 399-410.

Boland, J., Martin, J., Wells, A. U. & Ross, J. R. (2013) Palliative care for people with non-malignant lung disease: summary of current evidence and future direction. *Palliative Medicine*, 27 (9), s. 811-6.

Centre for Reviews and Dissemination (2009) *Systematic Reviews. CRD's guidance for undertaking reviews in health care*. York, CRD University of York.

COSMIN (2011) *COSMIN checklist with 4-point scale* [Internett], COSMIN. Tilgjengelig fra: [http://www.cosmin.nl/cosmin\\_checklist.html](http://www.cosmin.nl/cosmin_checklist.html) [Nedlastet 25.11. 2013].

Davison, S. N., Jhangri, G. S. & Johnson, J. A. (2006) Cross-sectional validity of a modified Edmonton symptom assessment system in dialysis patients: a simple assessment of symptom burden. *Kidney International*, 69 (9), s. 1621-5.

de Vet, H. C. W., Terwee, C. B., Mokkink, L. B. & Knol, D. L. (2011) *Measurement in medicine: a practical guide*. Cambridge, Cambridge University Press.

Deeks, J. J., Bossuyt, P. M. & Gatsonis, C. (2013) *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* [Internett]. Tilgjengelig fra: <<http://srdta.cochrane.org/>> [Nedlastet 15.04.15].

Dorman, S., Byrne, A. & Edwards, A. (2007) Which measurement scales should we use to measure breathlessness in palliative care? A systematic review. *Palliative Medicine*, 21 (3), s. 177-91.

Dorman, S., Jolley, C., Abernethy, A., Currow, D., Johnson, M., Farquhar, M., Griffiths, G., Peel, T., Moosavi, S., Byrne, A., Wilcock, A., Alloway, L., Bausewein, C., Higginson, I. & Booth, S. (2009) Researching breathlessness in palliative care: consensus statement of the National Cancer Research Institute Palliative Care Breathlessness Subgroup. *Palliative Medicine*, 23 (3), s. 213-227.

Dudgeon, D. (2010) Dyspnea, Death Rattle, and Cough. I: Ferrell, B. R. & Coyle, N. red. *Oxford Textbook of Palliative Nursing*. New York, Oxford University Press, s. 303-18.

Duiverman, M. L., Wempe, J. B., Bladder, G., Kerstjens, H. A. & Wijkstra, P. J. (2008) Health-related quality of life in COPD patients with chronic respiratory failure. *European Respiratory Journal*, 32 (2), s. 379-86.

European Association for Palliative Care (04.05 2012) *EAPC By-laws* [Internett]. Tilgjengelig fra: <<http://www.eapcnet.eu/Corporate/AbouttheEAPC/EAPCOfficialdocuments/EAPCBylaws.aspx>> [Nedlastet 02.03.15].

Garratt, A. & Bjertnæs, Ø. A. (2014) *Pasientrapporterte resultatmål (PROMS) som nasjonale kvalitetsindikatorer: kvalitetskriterier og standard utviklings prosess* [Internett], Oslo, Nasjonalt kunnskapssenter for helsetjenesten. Tilgjengelig fra: <<http://www.kunnskapssenteret.no/publikasjoner/pasientrapporterte-resultatmal-proms-som-nasjonale-kvalitetsindikatorer-kvalitetskriterier-og-standard-utviklingsprosess>> [Nedlastet 01.12.2014].

Ghosh, D., Rzehak, P., Elliott, M. W. & Windisch, W. (2012) Validation of the English Severe Respiratory Insufficiency Questionnaire. *European Respiratory Journal*, 40 (2), s. 408-15.

Gysels, M., Bausewein, C. & Higginson, I. J. (2007) Experiences of breathlessness: a systematic review of the qualitative literature. *Palliative & supportive care*, 5 (3), s. 281-302.

Han, J. N., Xiong, C. M., Yao, W., Fang, Q. H., Zhu, Y. J., Cheng, X. S. & Van de Woestijne, K. P. (2011) Multiple dimensions of cardiopulmonary dyspnea. *Chinese Medical Journal*, 124 (20), s. 3220-6.



Helsedirektoratet (2012) *Kols: Nasjonalfaglig retningslinje og veileder for forebygging, diagnostisering og oppfølging* [Internett], Oslo. Tilgjengelig fra: <https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/847/Nasjonalfaglig-retningslinje-og-veileder-for-forebygging-diagnostisering-og-oppf%C3%B8lging-IS-2029.pdf> [Nedlastet 01.03.15].

Helsedirektoratet (2013) *Nasjonalt handlingsprogram for palliasjon i kreftomsorgen* [Internett], Oslo. Tilgjengelig fra: <https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/387/nasjonalt-handlingsprogram-for-palliasjon-i-kreftomsorgen.pdf> [Nedlastet 26.11.13].

Henoch, I., Bergman, B. & Gaston-Johansson, F. (2006) Validation of a Swedish version of the Cancer Dyspnea Scale. *Journal of Pain and Symptom Management*, 31 (4), s. 353-61.

Henoch, I., Bergman, B., Gustafsson, M., Gaston-Johansson, F. & Danielson, E. (2008) Dyspnea experience in patients with lung cancer in palliative care. *European Journal of Oncology Nursing*, 12 (2), s. 86-96.

Higgins, J. P. T. & Green, S. (2011) *Cochrane Handbook for Systematic Reviews of Interventions*. West Sussex, England, WILEY-BLACKWELL, A John Wiley & Sons, LTD, Publication.

Higginson, I. J., Bausewein, C., Reilly, C. C., Gao, W., Gysels, M., Dzingina, M., McCrone, P., Booth, S., Jolley, C. J. & Moxham, J. (2014) An integrated palliative and respiratory care service for patients with advanced disease and refractory breathlessness: a randomised controlled trial. *Lancet Respiratory Medicine*, 2 (12), s. 979-87.

Johnson, M. & Fallon, M. (2013) Just good care? The palliative care of those with non-malignant disease. *Palliative Medicine*, 27 (9), s. 803-4.

Lopez-Campos, J. L., Failde, I., Masa, J. F., Benitez-Moya, J. M., Barrot, E., Ayerbe, R., Leon-Jimenez, A. & Windisch, W. (2008) Transculturally adapted Spanish SRI questionnaire for home mechanically ventilated patients was viable, valid, and reliable. *Journal of Clinical Epidemiology*, 61 (10), s. 1061-6.

Markussen, H., Lehmann, S., Nilsen, R. M. & Natvig, G. K. (2014) The Norwegian version of the Severe Respiratory Insufficiency Questionnaire. *International Journal of Nursing Practice*.

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., Bouter, L. M. & de Vet, H. C. (2010a) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Medical Research Methodology*, 10, s. 22.

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 checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*, 19 (4), s. 539-49.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M. & de Vet, H. C. (2010c) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63 (7), s. 737-45.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M. & de Vet, H. C. (2012) *COSMIN checklist manual* [Internett], COSMIN. Tilgjengelig fra: <[http://www.cosmin.nl/cosmin\\_checklist.html](http://www.cosmin.nl/cosmin_checklist.html)> [Nedlastet 25.11.2013].

Nasjonalt kunnskapssenter for helsetjenesten (08.10 2012) *PsykTest* [Internett]. Tilgjengelig fra: <<http://www.psyktest.no/om-m%C3%A5leegenskaper>> [Nedlastet 20.11.2014].

Nasjonalt kunnskapssenter for helsetjenesten (2013) *Slik oppsummerer vi forskning. Håndbok for Nasjonalt kunnskapssenter for helsetjenesten. 3.2. revidert utg.* Oslo, Nasjonalt kunnskapssenter for helsetjenesten.

Norweg, A., Jette, A. M., Ni, P., Whiteson, J. & Kim, M. (2011a) Outcome measurement for COPD: reliability and validity of the Dyspnea Management Questionnaire. *Respiratory Medicine*, 105 (3), s. 442-53.

Norweg, A., Ni, P., Garshick, E., O'Connor, G., Wilke, K. & Jette, A. M. (2011b) A multidimensional computer adaptive test approach to dyspnea assessment. *Archives of Physical Medicine and Rehabilitation*, 92 (10), s. 1561-9.

Norweg, A., Whiteson, J., Demetis, S. & Rey, M. (2006) A new functional status outcome measure of dyspnea and anxiety for adults with lung disease: the dyspnea management questionnaire. *Journal of Cardiopulmonary Rehabilitation*, 26 (6), s. 395-404.

Oldenmenger, W. H., de Raaf, P. J., de Klerk, C. & van der Rijt, C. C. (2013) Cut points on 0-10 numeric rating scales for symptoms included in the Edmonton Symptom Assessment Scale in cancer patients: a systematic review. *Journal of Pain and Symptom Management*, 45 (6), s. 1083-93.

Parshall, M. B., Schwartzstein, R. M., Adams, L., Banzett, R. B., Manning, H. L., Bourbeau, J., Calverley, P. M., Gift, A. G., Harver, A., Lareau, S. C., Mahler, D. A., Meek, P. M. & O'Donnell, D. E. (2012) An Official American Thoracic Society Statement: Update on the Mechanisms, Assessment, and Management of Dyspnea. *American Journal of Respiratory and Critical Care Medicine*, 185 (4), s. 435-452.

Polit, D. F. & Beck, C. T. (2012) *Nursing research: generating and assessing evidence for nursing practice*. Philadelphia, Pa., Wolters Kluwer Health.

Preston, N. J., Fayers, P., Walters, S. J., Pilling, M., Grande, G. E., Short, V., Owen-Jones, E., Evans, C. J., Benalia, H., Higginson, I. J. & Todd, C. J. (2013) Recommendations for managing missing data, attrition and response shift in palliative and end-of-life care research: part of the MORECare research method guidance on statistical issues. *Palliative Medicine*, 27 (10), s. 899-907.

Reddy, S. K., Parsons, H. A., Elsayem, A., Palmer, J. L. & Bruera, E. (2009) Characteristics and Correlates of Dyspnea in Patients with Advanced Cancer. *Journal of Palliative Medicine*, 12 (1), s. 29-36.

Sampson, M., McGowan, J., Cogo, E., Grimshaw, J., Moher, D. & Lefebvre, C. (2009) An evidence-based practice guideline for the peer review of electronic search strategies. *Journal of Clinical Epidemiology*, 62 (9), s. 944-52.

Schellingerhout, J. M., Heymans, M. W., Verhagen, A. P., de Vet, H. C., Koes, B. W. & Terwee, C. B. (2011) Measurement properties of translated versions of neck-specific questionnaires: a systematic review. *BMC Medical Research Methodology*, 11, s. 87.

Schellingerhout, J. M., Verhagen, A. P., Heymans, M. W., Koes, B. W., de Vet, H. C. & Terwee, C. B. (2012) Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Quality of Life Research*, 21 (4), s. 659-70.

Simon, S. T., Higginson, I. J., Benalia, H., Gysels, M., Murtagh, F. E., Spicer, J. & Bausewein, C. (2013a) Episodes of breathlessness: types and patterns - a qualitative study exploring experiences of patients with advanced diseases. *Palliative Medicine*, 27 (6), s. 524-32.

Simon, S. T., Higginson, I. J., Benalia, H., Gysels, M., Murtagh, F. E., Spicer, J. & Bausewein, C. (2013b) Episodic and continuous breathlessness: a new categorization of breathlessness. *Journal of Pain and Symptom Management*, 45 (6), s. 1019-29.

Struik, F. M., Kerstjens, H. A., Bladder, G., Sprooten, R., Zijnen, M., Asin, J., van der Molen, T. & Wijkstra, P. J. (2013) The Severe Respiratory Insufficiency Questionnaire scored best in the assessment of health-related quality of life in chronic obstructive pulmonary disease. *Journal of Clinical Epidemiology*, 66 (10), s. 1166-74.

Tanaka, K., Akechi, T., Okuyama, T., Nishiwaki, Y. & Uchitomi, Y. (2000) Development and validation of the Cancer Dyspnoea Scale: a multidimensional, brief, self-rating scale. *British Journal of Cancer*, 82 (4), s. 800-5.

Tanaka, K., Akechi, T., Okuyama, T., Nishiwaki, Y. & Uchitomi, Y. (2002) Factors correlated with dyspnea in advanced lung cancer patients: organic causes and what else? *Journal of Pain and Symptom Management*, 23 (6), s. 490-500.

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M. & de Vet, H. C. (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research*, 21 (4), s. 651-7.

Terwee, C. B., Schellingerhout, J. M., Verhagen, A. P., Koes, B. W. & de Vet, H. C. (2011) Methodological quality of studies on the measurement properties of neck pain and disability questionnaires: a systematic review. *Journal of Manipulative and Physiological Therapeutics*, 34 (4), s. 261-72.

U.S. Department of Health and Human Services Food and Drug Administration (2009) *Guidance for industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. Silver Spring, U.S. Department of Health and Human Services Food and Drug Administration.

Uronis, H. E., Shelby, R. A., Currow, D. C., Ahmedzai, S. H., Bosworth, H. B., Coan, A. & Abernethy, A. P. (2012) Assessment of the psychometric properties of an English version of the cancer dyspnea scale in people with advanced lung cancer. *Journal of Pain and Symptom Management*, 44 (5), s. 741-9.

Vist, G. E., Sæterdal, I., Vandvik, P. O. & Flottorp, S. A. (2013) Gradering av kvaliteten på dokumentasjonen. *Norsk Epidemiologi* 23 (2), s. 151-156.

Watanabe, S. M., Nekolaichuk, C., Beaumont, C., Johnson, L., Myers, J. & Strasser, F. (2011) A multicenter study comparing two numerical versions of the Edmonton Symptom Assessment System in palliative care patients. *Journal of Pain and Symptom Management*, 41 (2), s. 456-68.

Windisch, W., Budweiser, S., Heinemann, F., Pfeifer, M. & Rzehak, P. (2008) The Severe Respiratory Insufficiency Questionnaire was valid for COPD patients with severe chronic respiratory failure. *Journal of Clinical Epidemiology*, 61 (8), s. 848-853.

Windisch, W., Freidel, K., Schucher, B., Baumann, H., Wiebel, M., Matthys, H. & Petermann, F. (2003) The Severe Respiratory Insufficiency (SRI) Questionnaire A specific measure of health-related quality of life in patients receiving home mechanical ventilation. *Journal of Clinical Epidemiology*, 56 (8), s. 752-759.

World Health Organization (2015) *WHO Definition of Palliative Care* [Internett].  
Tilgjengelig fra: <<http://www.who.int/cancer/palliative/definition/en/>> [Nedlastet 02.03.15].

# **Kartleggingsverktøy for pustebesvær hos palliative pasienter**

## **– en systematisk oversikt**

**Kandidat:** Kjersti Solvåg

**Mastergradsprogram:** Mastergrad i kunnskapsbasert praksis i helsefag

**Avdeling:** Avdeling for helse- og sosialfag

**Utdanningsinstitusjon:** Høgskolen i Bergen

**Planlagt publisert i:** Sykepleien Forskning

# **Kartleggingsverktøy for pustebeviser hos palliative pasienter**

## **– en systematisk oversikt**

### **Forfatter:**

Kjersti Solvåg, spesialsykepleier/fagutviklingsykepleier ved Sunniva senter for lindrende behandling, Medisinsk klinikk, Haraldsplass Diakonale Sykehus.

### **Kontaktperson:**

Kjersti Solvåg  
Tømmerbakken 6  
5101 Eidsvågneset

E-post: [kjerstisolvaag@hotmail.com](mailto:kjerstisolvaag@hotmail.com)

Tlf: 92 88 34 58

Antall tegn med mellomrom: 21 270

Antall ord: 2999

Antall figurer: 1

Antall tabeller: 5

Antall vedlegg: 5

## Sammendrag

**Bakgrunn:** Pustebeviser er et vanlig og belastende symptom hos mange pasienter med alvorlig langtkommen sykdom. På grunn av symptomets subjektive natur trenger vi verktøy for kartlegging av pasientens egenopplevelse, pasientrapporterte resultatmål (PROMs).

**Hensikt:** å identifisere tilgjengelige PROMs for kartlegging av pustebeviser hos palliative pasienter og oppsummere deres måleegenskaper. Kartleggingsverktøyene måtte inneholde dimensjoner for både pustebeviser og angst for å bli vurdert.

**Metode:** Et systematisk litteratursøk ble gjennomført i mars 2014. To personer screenet alle identifiserte referanser for relevans og kritisk vurderte inkluderte studier uavhengig av hverandre. COSMIN-sjekklisten ble benyttet til vurdering av studienes metodiske kvalitet. Relevante data ble hentet ut av en og dataekstraksjonen ble kvalitetssikret av en annen. Måleegenskapene til de inkluderte kartleggingsverktøyene ble oppsummert i en kvalitativ syntese.

**Hovedresultat:** Vi screenet 1458 referanser for relevans og inkluderte 16 publikasjoner som evaluerte måleegenskaper ved fire ulike PROMs: CDS, SRI, DMQ og en respiratorisk symptomsjekkliste. Bare SRI er validert i norsk versjon, og ingen av de inkluderte PROMs er direkte validert for bruk i en palliativ setting. De tilgjengelige valideringsstudier antyder at de aktuelle kartleggingsverktøyene har positive måleegenskaper, men siden viktige måleegenskaper ikke er vurdert og siden mange studier har metodiske svakheter, er det vanskelig å trekke sikre konklusjoner.

**Konklusjon:** Den samlede kvaliteten til PROMs for pustebeviser og angst er fortsatt uavklart fordi viktige måleegenskaper ennå ikke er evaluert og mange valideringsstudier er forbundet med metodiske svakheter. Videre forskning er nødvendig for å oversette og validere kartleggingsverktøyene for bruk i lindrende behandling i Norge.

**Nøkkelord:** Palliasjon, dyspné, oversikt, spørreskjema, psykometri

# Measurement instruments for breathlessness in palliative care - a systematic review

## Abstract

**Background:** Breathlessness is a common and distressing symptom for many patients with advanced disease. Due to the subjective nature of breathlessness, patient-reported outcome measures (PROMs) are required to measure the patient's own experience.

**Aim:** to identify PROMs used to measure breathlessness in palliative care and to synthesise their measurement properties. Instruments had to include dimensions for breathlessness and anxiety to be considered.

**Method:** We performed a systematic search for literature in March 2014. Two reviewers independently screened all references for relevance and critically appraised included studies with the COSMIN-checklist. One reviewer extracted relevant data, and a second reviewer evaluated the extractions. We performed a best evidence synthesis to summarize the measurement properties of each included PROMs.

**Results:** We screened 1458 references for relevance, and included 16 publications evaluating the measurement properties of four different PROMs: CDS, DMQ, SRI and a respiratory symptom checklist. Only SRI was available in a validated Norwegian version, and none of the included PROMs were directly validated for use in palliative care. The available PROMs have promising measurement properties, but it is difficult to draw firm conclusions because important measurement properties are not evaluated yet, and many available studies are associated with methodological shortcomings.

**Conclusion:** The overall quality of PROMs for breathlessness and anxiety remains unclear because important measurement properties are not yet evaluated and many validation studies are associated with methodological shortcomings. Further research is needed to translate and validate the PROMs for use in palliative care in Norway.

**Keyword:** Palliative care, dyspnea, review, questionnaires, psychometrics



## Introduksjon

Dyspné eller pustebesvær er et komplekst og sammensatt symptom hos mange pasienter med langtkommen alvorlig sykdom (1, 2). Studier har vist at 94 % av pasienter med kronisk lungesykdom og 78 % med lungekreft lider av pustebesvær siste leveår (3), og at pustebesvær kan assosieres med dårlig livskvalitet, angst, nedsatt funksjonsevne og reduserte leveutsikter (1, 2).

Pasientens opplevelse av pustebesvær samsvarer ofte ikke med sykdommens alvorlighetsgrad og objektive mål (4, 5). I lindrende behandling er systematisk bruk av verktøy som kartlegger pasientenes egenopplevelse derfor viktig for å identifisere symptomer, innhente informasjon om sykdomsutvikling og vurdere effekt av tiltak (6). Data som er basert på slike kartleggingsverktøy omtales gjerne som pasientrapporterte resultatmål (PROMs).

Før vi startet arbeidet med denne oversikten gjennomførte vi et orienterende litteratursøk for å se om det allerede fantes oversikter som besvarte våre spørsmål. Søket resulterte i to relevante treff (7, 8). Begge oversiktene er basert på ti år gamle litteratursøk, og begge viste at det var nødvendig med videre evaluering av hvilke PROMs som egner seg for kartlegging av pustebesvær i en palliativ setting (7, 8). Hensikten med denne systematiske oversikten har vært å oppsummere hvilke PROMs som er tilgjengelig for kartlegging av pustebesvær hos palliative pasienter og hvilke måleegenskaper disse kartleggingsverktøyene har. Kartleggingsverktøy måtte inneholde dimensjoner for både pustebesvær og angst for å bli vurdert for inklusjon.

## Metode

Vi har benyttet et metodisk rammeverk kalt «COnsensus-based Standards for the selection of health Measurement INSTRUMENTS» (COSMIN). Rammeverk som ble utarbeidet i en internasjonal konsensusprosess, og gir spesifikke anbefalinger om terminologi, taksonomi og metodebruk i studier som omhandler PROMs og deres måleegenskaper (9-11).

## Litteratursøk

Vi søkte systematisk etter litteratur i mars 2014 i databasene MEDLINE (1946-), Embase (1974-), PsycINFO (1806-), AMED (1985-), CINAHL (1982-), Cochrane Library og

SveMed+. Søkestrategiene bestod av emneord og tekstord tilpasset hver database for kategoriene 1) pustebesvær, 2) måleegenskaper og 3) palliative pasienter, kreft eller kronisk obstruktiv lungesykdom (kols). Vi gjennomførte siteringssøk for oversiktene til Bausewein og medarbeidere (7) og Dorman og medarbeidere (8) i Science Citation Index, og gjennomgikk referanselister til aktuelle publikasjoner. Søkene ble gjennomført uten begrensninger i språk og tid, og ble fagfellevurdert av spesialbibliotekar. Søkestrategi for MEDLINE er vedlagt (Vedlegg I), fullstendig søkestrategier kan fås fra forfatter.

## Utvelgelse

Titler og sammendrag ble vurdert for inklusjon på bakgrunn av forhåndsbestemte kriterier (Tabell 1). Antatt relevante artikler ble innhentet i fulltekst og vurdert for inklusjon. Alle trinnene i seleksjonsprosessen ble gjennomført av to personer uavhengig av hverandre, og uenighet ble løst ved diskusjon.

## Måleegenskaper

Før man tar i bruk nye kartleggingsverktøy er det viktig å forsikre seg om at verktøyene har gode måleegenskaper, det vil si om de måler det de er tenkt å måle og om vi kan stole på resultatene. I COSMIN-taksonomien deles måleegenskapene inn i tre hovedkategorier: reliabilitet, validitet og responsivitet. **Reliabilitet** forteller om feilkildene er tilstrekkelig små og om resultatene er tilstrekkelige stabile til at vi kan stole på målingene. I COSMIN-taksonomien består reliabilitet av tre underkategorier som kalles *intern konsistens*, *reliabilitet* (her test-retest-reliabilitet) og *målefeil*. **Validitet** beskriver om et kartleggingsverktøy måler de egenskapene det har til hensikt å måle, for eksempel om personer som skårer ulikt på en depresjonsskala også skårer forskjellig på andre tilsvarende skalaer. COSMIN-taksonomien definerer tre underordnede validitetsegenskaper (*innholdsvaliditet*, *kriterievaliditet*, og *begrepsvaliditet*), der begrepsvaliditeten deles inn i ytterligere tre kategorier kalt *strukturell validitet*, *hypotesetesting* og *tverrkulturell validitet*. **Responsivitet** er et mål på om kartleggingsverktøyet evner å fange opp viktige endring over tid (10).

## Vurdering av metodisk kvalitet

Vi vurderte metodisk kvalitet og risiko for systematiske feil i tråd med sjekklisten fra COSMIN (12, 13). Sjekklisten består av ulike skjema som fylles ut for å vurdere om en kan stole på måleegenskapene som rapporteres i en valideringsstudie. Hver valideringsstudie

måler gjerne et begrenset antall måleegenskaper, og man fyller bare ut skjema med relevans opp mot måleegenskapene som studien søker å evaluere.

Alle kvalitetsvurderinger ble gjennomført av to personer, og uavhengig av hverandre. Uenighet ble løst ved diskusjon eller ved å involvere en tredje person.

### **Dataekstraksjon og kategorisering av måleegenskaper**

Én person gikk gjennom alle inkluderte studier og registrerte bakgrunnsinformasjon om studiet, deltakere, kartleggingsverktøy, måleegenskaper, metodisk kvalitet og resultater ved hjelp av et dataekstraksjonsskjema. Dataekstraksjonen ble deretter kvalitetssikret av en annen. For hver studie vurderte vi måleegenskapene til det aktuelle kartleggingsverktøyet som *positive*, *uavklarte* eller *negative* opp mot et sett med forhåndsdefinerte kriterier, Vedlegg II.

### **Samlet dokumentasjon om kartleggingsverktøyenes egenskaper**

Vi oppsummerte den samlede dokumentasjonen om måleegenskapene til de ulike kartleggingsverktøy i en kvalitativ syntese som tar hensyn til metodisk kvalitet, måleegenskaper og grad av konsistens. Kvaliteten på dokumentasjonen vurderes som *høy*, *moderat*, *lav*, *motstridende* eller *usikker*, etter kriteriene i Vedlegg III.

## **Resultater**

Vi identifiserte 1458 referanser, og vi endte opp med å inkludere 16 publikasjoner (14-29) basert på 15 ulike studier (Figur 1). Tabell 2 gir en oversikt over inkluderte studier, mens oversikt over ekskluderte studier kan fås fra forfatter.

Vi identifiserte fire relevante kartleggingsverktøy: Cancer Dyspnea Scale (CDS), Severe Respiratory Insufficiency (SRI) Questionnaire, Dyspnea Management Questionnaire (DMQ) og Respiratory Symptom Checklist (RSC). Vi har oppsummert karakteristika ved kartleggingsverktøyene i Tabell 3. Tabell 4 oppsummerer vår vurdering av metodisk kvalitet og risiko for skjevheter, samt måleegenskapene i de inkluderte studiene. Verdier for intern konsistens, test-retest-reliabilitet og strukturell validitet er vist i Vedlegg IV og verdier for utvalgte samvariasjoner i Vedlegg V. Oversikt over alle tilgjengelige samvariasjoner kan fås fra forfatter. Den kvalitative syntesen som oppsummerer resultater på tvers av alle inkluderte studier er presentert i Tabell 5. I det følgende går vi kort gjennom de fire ulike

kartleggingsverktøyene og oppsummerer tilgjengelig dokumentasjonen. For hypotesetesting har vi valgt å vektlegge sammenligninger som vi har vurdert som mest relevant.

### **Cancer Dyspnea Scale (CDS)**

CDS ble utviklet i Japan for pasienter med kreft og måler aspekt ved pustebesvær (14).

Utvikling og validering er beskrevet i fem publikasjoner (14-18), hvorav to publikasjoner (16, 17) er basert på det samme datamaterialet.

Både representanter for pasientgruppen og fagekspertene var involvert i utviklingen av CDS, og kartleggingsverktøyet ble pilottestet i målgruppen (14). Dokumentasjonen viser at vi kan ha høy tillit til at CDS har god *innholdsvaliditet*.

Tre studier (14, 16, 18) bekrefter at CDS har en tredelt faktorstruktur, og en samlefaktor for totalt pustebesvær, men vi har ikke tilstrekkelige data til å vurdere kartleggingsverktøyets *strukturelle validitet*.

For *hypotesetesting* har vi vektlagt sammenligningene mellom CDS og Hospital Anxiety and Depression Scale (HADS), Borg skala og visuell analog skala (VAS-dyspné). Resultatet for hypotesetesting var *negativt* i to (15, 18) og *positivt* i to studier (14, 16). Den samlede dokumentasjonen om hypotesetesting er dermed motstridende.

Den *interne konsistensen* til CDS var *positiv* (14, 16, 18) og samlet dokumentasjon viser at vi kan ha høy tillit til at internkonsistensen til CDS er god. Resultatet for *test-retest-reliabilitet* var *negativt* (14), men vi har begrenset tillit til dokumentasjonen som foreligger.

### **Severe Respiratory Insufficiency (SRI) Questionnaire**

SRI ble utviklet i Tyskland for å måle helse relatert livskvalitet hos pasienter som behandles med langtids mekanisk ventilasjon (LTMV) på grunn av kronisk respirasjonssvikt.

Respirasjonssvikten skyldes ulike underliggende sykdommer (19). Syv publikasjoner (19-25) har beskrevet utvikling og validering av kartleggingsverktøyet. To publikasjoner (19, 22) har delvis overlappende utvalg, og data for strukturell validitet og intern konsistens er derfor hentet fra studien til Windisch og medarbeidere (22).

SRI er utviklet basert på sosiale, psykologiske og fysiske helsedomener, og både pasienter og fagekspertene var involvert (19). Norsk versjon av SRI ble pilottestet blant brukere av LTMV

(25). **Innholdsvaliditeten** vurderes som *positiv*, og dokumentasjonen viser at vi kan ha høy tillit til at SRI har god innholdsvaliditet.

Faktorstrukturen er evaluert i tre studier (21-23) som alle fant flere faktorer enn de opprinnelige syv dimensjonene. To studier (22, 23) viste at faktorene innenfor en dimensjon samvarierte og valgte derfor å beholde den opprinnelige syv-faktorstrukturen. En studie (21) fant 13 faktorer. Alle tre studiene hadde *positivt* resultat for **strukturell validitet** med total varians over 50 %. Det er allikevel usikkerhet knyttet til den strukturelle validiteten, noe som først og fremst skyldes metodiske begrensninger i studiene der dette er evaluert.

For **hypotesetesting** har vi vektlagt sammenligningene mellom SRI og Short Form Health Survey (SF-36), Chronic Respiratory Disease Questionnaire (CRQ), HADS og Medical Research Council Dyspnea Scale (MRC) (19-21, 23-25). Generelt var det sterkest samvariasjon mellom SRI dimensjoner og dimensjoner for de andre kartleggingsverktøyene som måler relaterte aspekt, mens det var svakest samvariasjon mellom dimensjoner som måler ulike aspekt. Samlet dokumentasjonen om hypotesetesting viser at vi kan være sikre på at SRI samvarierer med andre PROMs som måler tilsvarende egenskaper.

For norsk versjon av SRI (25) var resultatet *negativt* for **tverrkulturell validitet** fordi det ikke er gjennomført sammenligning av faktorstrukturen mot originalversjonen. Uklar rapportering medfører at det er usikkerhet knyttet til tverrkulturell validitet av norsk versjon.

**Intern konsistensen** til SRI (20-25) var gjennomgående *positiv*, men på grunn av metodiske begrensninger er det likevel usikkerhet knyttet til den interne konsistensen. **Test-retest-reliabiliteten** (20, 21) var *positiv*, men dokumentasjonen viser at vi kan ha begrenset tillit til at test-retest-reliabiliteten er god, først og fremst på grunn av metodiske svakheter i studiene der dette er evaluert.

### **Dyspnea Management Questionnaire (DMQ)**

DMQ ble utviklet i USA for å måle effekt av lungerehabilitering og endring over tid hos pasienter med kols, og for å tilrettelegge for mer målrettet behandling relevant for fagområder som for eksempel ergoterapi og psykologi (26). Tre studier beskriver utvikling og validering av tre ulike versjoner; DMQ-30 (26), DMQ-56 (27) og DMQ-CAT (28).

Data fra kvalitative intervju, en litteraturgjennomgang på områdene pustebesvær, angst, unngåelsesadferd, funksjonsstatus, helserelatert livskvalitet, brukertilfredshet og

lungerehabilitering samt en gjennomgang av andre kartleggingsverktøy dannet grunnlaget for utviklingen av DMQ-30. Både fagekspertene og pasientene var involvert og en preliminær versjon ble pilottestet blant voksne med kols (26). DMQ-CAT er en elektronisk versjon og ble utviklet basert på en utvidet samling av spørsmålene fra DMQ-56. Involvering av fagekspertene og pasientene, og en omfattende litteraturgjennomgang dannet grunnlaget for utvidelsen (28). **Innholdsvaliditeten** til DMQ-30 og DMQ-CAT er vurdert som *positiv*. Dokumentasjonen viser at vi kan ha høy tillit til at DMQ-30 og DMQ-CAT har god innholdsvaliditet.

Den opprinnelige fem-faktormodellen for DMQ-56 ble bekreftet (27). Evaluering av faktorstrukturen til DMQ-CAT (28) viste at en fire-faktormodell passet best. Faktorene forklarte mer enn 50 % av variansen for begge versjonene og resultatet for **strukturell validitet** var derfor *positivt* for både DMQ-56 og DMQ-CAT (27, 28). Dokumentasjonen viser at vi kan ha begrenset tillit til at den strukturelle validiteten til DMQ-CAT er god, mens er det mer usikkerhet knyttet til den strukturelle validiteten til DMQ-56, som følge av metodisk begrensninger i studien der dette ble evaluert.

For **hypotesetesting** er DMQ-30 sammenlignet med Seattle Obstructive Lung Disease Questionnaire (SOLQ), Short Form Health Survey (SF-12) og HADS (26). DMQ-CAT ble sammenlignet med University of California, San Diego Shortness of Breath Questionnaire (UCSD SOBQ), CRQ, COPD Self-Efficacy Scale (CSES) og HADS (28). For begge versjonene var det sterkest samvariasjon mellom DMQ dimensjoner og dimensjoner for de andre kartleggingsverktøyene som måler relaterte aspekter, mens det var svakest samvariasjon mellom dimensjoner som måler ulike aspekter. Resultatet for hypotesetesting av DMQ-30 og DMQ-CAT var *positivt*. Dokumentasjonen viser at vi kan ha moderat tillit til at DMQ-30 samvarierer med andre PROMs som måler tilsvarende egenskaper og begrenset tillit til at DMQ-CAT samvarierer med andre PROMs som måler tilsvarende egenskaper.

DMQ-CAT ble også sammenlignet med den totale samlingen på 71 spørsmål og resultatet for **kriterievaliditet** var *positivt* for alle dimensjonene med Pearsons korrelasjonskoeffisient 0,94-0,97 (28). Dokumentasjonen viser at vi kan ha moderat tillit til at kriterievaliditeten til DMQ-CAT er god.

Den **interne konsistensen** var *positivt* for alle de tre versjonene av DMQ (26-28). Dokumentasjonen viser at vi kan ha moderat tillit til at den interne konsistensen til DMQ-30, DMQ-56 og DMQ-CAT er god.

*Test-retest-reliabilitet* var *positivt* for DMQ-30 (26) og DMQ-56 (27). Dokumentasjonen viser at vi kan ha begrenset tillit til at test-retest-reliabiliteten er god for DMQ-56, mens det er mer usikkerhet knyttet til test-retest-reliabiliteten for DMQ-30, noe som først og fremst skyldes metodiske begrensninger i studien der dette ble målt.

### **Respiratorisk symptomsjekkliste (RSC)**

RSC ble utviklet i Kina for å måle multidimensjonale aspekt ved pustebevis hos pasienter med hjerte- og lungesykdom (29).

Evaluerings av faktorstrukturen viste tolv sammenhengende og ikke-korrelerte faktorer, og ved å legge til ytterligere kriterier fant Han og medarbeidere (29) syv faktorer som synes å måle tre dimensjoner av pustebevis. To resterende faktorer sorterte under andre symptomer. Faktorene forklarte 64 % av den totale variansen (29) og resultatet for *strukturell validitet* var *positivt*. Dokumentasjonen viser at vi kan ha høy tillit til at den strukturelle validiteten til RSC er god.

## **Diskusjon**

Vi indentifiserte fire PROMs for kartlegging av pustebevis som tilfredsstilte våre inklusjonskriterier. De måler alle sammensatte aspekt og innehar dimensjoner for angst og pustebevis. Ingen av de inkluderte PROMs var direkte validert for bruk i en palliativ setting. Måleegenskapene til de fire kartleggingsverktøyene var evaluert i 15 studier og beskrevet i 16 publikasjoner. De tilgjengelige kartleggingsverktøyene ble i all hovedsak vurdert å ha positive måleegenskaper med unntak av CDS som hadde negativ test-retest egenskaper, uavklart strukturell validitet og uavklart samvariasjon med andre kartleggingsverktøy. Responsivitet og målefeil var ikke evaluert i noen av studiene. Alle studiene hadde metodiske svakheter, særlig små utvalg og mangelfulle hypoteser, som gjør det vanskelig å trekke sikre konklusjoner.

### **Styrker og svakheter ved studien**

Utarbeidelsen av denne systematiske oversikten ble gjort i tråd med anbefalinger fra COSMIN-initiativet som tar hensyn til de spesielle aspektene ved validering av PROMs. Oversikten bygger på et bredt og systematisk litteratursøk og alle trinnene i utarbeidelsen ble gjennomført av to personer uavhengig av hverandre, eller utført av en og kvalitetssikret av en

annen person. En utfordring for litteratursøk til oversikter over PROMs er mangler ved sammendrag og indeksering av studier om måleegenskaper (9). En svakhet ved vårt litteratursøk er at søkeord for respirasjonssvikt ikke var inkludert. Flere av studiene om SRI ble derfor først identifisert ved gjennomgang av relevante studier. Dette medfører en risiko for at det finnes andre relevante PROMs som vi ikke har identifisert. Vi har ikke hatt kapasitet til å søke etter «grå litteratur», gjennomføre supplerende søk på inkluderte PROMs, eller kontakte fagpersoner. Samlet sett innebærer dette en risiko for at det finnes relevante publikasjoner som vi ikke har identifisert.

Vi kvalitetsvurderte de inkluderte studiene med COSMIN-sjekklisten etter først å ha gjennomført en grundig pilotering. Sjekklisten gjorde det mulig å gi en separat vurdering av de ulike måleegenskapene som var evaluert i en og samme studie. En utfordring har vært at vi ikke alltid har hatt mulighet for å skille mellom mangelfull rapportering og dårlig kvalitet i studien. Vi har ikke hatt kapasitet til å kontakte forfattere for avklaringer omkring dette.

### **Studiens metodiske kvalitet**

To hovedutfordringer ved de inkluderte studienes metodiske kvalitet er størrelsen på utvalgene og mangelfull rapportering om hypoteser. Størrelsen på utvalgene var ofte for små, og hver for seg hadde ingen av studiene nok respondenter til å evaluere test-retest-reliabilitet. Ved evaluering av intern konsistens og strukturell validitet er anbefalt utvalgsstørrelse fem til syv ganger antall spørsmål og minimum 100 (13, 30). Dette medfører at kartleggingsverktøy med mange spørsmål som for eksempel SRI skårer svakt for internkonsistens og strukturell validitet for alle de inkluderte studiene. Terwee og medarbeider (31) som har oppsummert metodisk kvalitet ved studier om måleegenskaper i en systematisk oversikt, fant også at utvalgsstørrelsen var en betydelig utfordring for evaluering av mange måleegenskaper.

Flere av studien hadde ikke forhåndsformulerte hypoteser eller de var mangelfullt formulert med hensyn til forventet størrelse eller retning på samvariasjonen. Det var ikke alltid klart om det skyldes mangler ved studien eller ved rapporteringen. Terwee og medarbeidere (31) fant i sin studie at mindre enn 50 % av studiene hadde forhåndsdefinerte hypoteser og bare 48 % av disse hadde beskrevet forventet retning og størrelse på samvariasjonen. Ved manglende hypoteser er det fare for at en finner alternative forklaringer på resultatene og risikoen for systematiske skjevheter øker (9).



## **Betydning for praksis og videre forskning**

Oppsummering av resultatene og den kvalitative syntesen viste at vi har begrenset tillit til dokumentasjonen og mangler kunnskap om viktige måleegenskaper for alle de inkluderte kartleggingsverktøyene. At vi har begrenset tillit til dokumentasjonen betyr ikke at måleegenskapene er dårlige, men at de fortsatt er uavklart (31). Selv om noen av de identifiserte kartleggingsverktøyene er bedre studert enn andre, så er det foreløpig vanskelig å gi et klart svar på om noen av de tilgjengelige verktøyene har bedre måleegenskaper enn de andre.

For PROMs som skal benyttes til kartlegging av palliative pasienter i klinisk praksis er det viktig at kartleggingsverktøyene ikke er for omfattende eller krever for mye ressurser for utfylling og til administrering (32). Her peker CDS seg ut med tolv spørsmål og en forventet utfyllingstid på to minutter. DMQ-CAT skreddersyr spørsmålene til respondenten utfra ett inngangsspørsmål (28). Belastningen for pasienten blir dermed redusert ved at bare de mest informative og relevante spørsmålene må besvares. For å benytte DMQ-CAT kreves det tilgang til og kompetanse til å bruke elektronisk verktøy.

Ingen av kartleggingsverktøyene vi identifiserte er validert for bruk til palliative pasienter uavhengig av diagnose og bare SRI er oversatt til norsk. I tillegg til de metodiske svakhetene som er påpekt og måleegenskaper som ennå ikke er evaluert, innebærer det at de relevante PROMs må videre valideres i studier av god kvalitet. Studiene bør særlig ha fokus på store nok utvalg, forhåndsformulering av hypoteser, god og tilstrekkelig rapportering av studiens gjennomføring. Det er også viktig at kartleggingsverktøyenes responsivitet evalueres slik at vi får kunnskap om det evner å fange opp viktige forskjeller som for eksempel effekt av behandling eller forverring av helsetilstand. COSMIN kan med fordel benyttes som veiledning for planlegging og rapportering av nye valideringsstudier.

## **Konklusjon**

Vi identifiserte fire kartleggingsverktøy for pustebesvær som tilfredstilte våre inklusjonskriterier. Noen kartleggingsverktøy er bedre studert enn andre, men foreløpig er det ikke mulig å gi et klart svar på om noen av de tilgjengelige verktøyene har bedre måleegenskaper enn de andre. CDS og DMQ-CAT synes mest brukervennlige med hensyn til tid til utfylling og belastning for pasientene, mens SRI er det eneste av verktøyene som er

tilgjengelig i norsk versjon. Videre forskning er nødvendig for å validere og gjøre kartleggingsverktøyene tilgjengelig for bruk i en palliativ setting i Norge.

## **Takksigelser**

Takk til Kjetil Gundro Brurberg for faglig og metodisk veiledning. Takk også til min medstudent Kristine Berg Titlestad for hjelp til kritisk vurdering av inkluderte studier og medstudent og bibliotekar Hilde Strømme som har fagfellevurdert litteratursøkene.

## Referanser

1. **Bausewein C, Jolley C, Reilly C, Lobo P, Kelly J, Bellas H, et al.** Development, effectiveness and cost-effectiveness of a new out-patient Breathlessness Support Service: study protocol of a phase III fast-track randomised controlled trial. *BMC Pulm Med.* 2012;12:58.
2. **Helsedirektoratet** [Internett]. Nasjonalt handlingsprogram for palliasjon i kreftomsorgen. Oslo; 2013 [hentet 26.11.13]. Tilgjengelig fra: <https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/387/nasjonalt-handlingsprogram-for-palliasjon-i-kreftomsorgen.pdf>
3. **Edmonds P, Karlsen S, Khan S, Addington-Hall J.** A comparison of the palliative care needs of patients dying from chronic respiratory diseases and lung cancer. *Palliat Med.* 2001;15(4):287-95.
4. **Bausewein C, Booth S, Higginson IJ.** Measurement of dyspnoea in the clinical rather than the research setting. *Curr Opin Support Palliat Care.* 2008;2(2):95-9.
5. **American Thoracic Society.** Dyspnea: mechanisms, assessment, and management. A consensus statement. *Am J Respir Crit Care Med.* 1999;159(1):321-40.
6. **Antunes B, Harding R, Higginson IJ.** Implementing patient-reported outcome measures in palliative care clinical practice: A systematic review of facilitators and barriers. *Palliat Med.* 2014;28(2):158-75.
7. **Bausewein C, Farquhar M, Booth S, Gysels M, Higginson IJ.** Measurement of breathlessness in advanced disease: a systematic review. *Respir Med.* 2007;101(3):399-410.
8. **Dorman S, Byrne A, Edwards A.** Which measurement scales should we use to measure breathlessness in palliative care? A systematic review. *Palliat Med.* 2007;21(3):177-91.
9. **de Vet HCW, Terwee CB, Mokkink LB, Knol DL.** Measurement in medicine: a practical guide. Cambridge: Cambridge University Press; 2011. 338 s.
10. **Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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):737-45.
11. **Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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):539-49.
12. **COSMIN.** COSMIN checklist with 4-point scale [Internett]. COSMIN; 2011 [hentet 25.11. 2013]. Tilgjengelig fra: <http://www.cosmin.nl/>.

13. **Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al.** COSMIN checklist manual [Internet]. COSMIN; 2012 [hentet 25.11.2013]. Tilgjengelig fra:  
<http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf>.
14. **Tanaka K, Akechi T, Okuyama T, Nishiwaki Y, Uchitomi Y.** Development and validation of the Cancer Dyspnoea Scale: a multidimensional, brief, self-rating scale. *Br J Cancer*. 2000;82(4):800-5.
15. **Tanaka K, Akechi T, Okuyama T, Nishiwaki Y, Uchitomi Y.** Factors correlated with dyspnea in advanced lung cancer patients: organic causes and what else? *J Pain Symptom Manage*. 2002;23(6):490-500.
16. **Henoch I, Bergman B, Gaston-Johansson F.** Validation of a Swedish version of the Cancer Dyspnea Scale. *J Pain Symptom Manage*. 2006;31(4):353-61.
17. **Henoch I, Bergman B, Gustafsson M, Gaston-Johansson F, Danielson E.** Dyspnea experience in patients with lung cancer in palliative care. *Eur J Oncol Nurs*. 2008;12(2):86-96.
18. **Uronis HE, Shelby RA, Currow DC, Ahmedzai SH, Bosworth HB, Coan A, et al.** Assessment of the psychometric properties of an English version of the cancer dyspnea scale in people with advanced lung cancer. *J Pain Symptom Manage*. 2012;44(5):741-9.
19. **Windisch W, Freidel K, Schucher B, Baumann H, Wiebel M, Matthys H, et al.** The Severe Respiratory Insufficiency (SRI) Questionnaire A specific measure of health-related quality of life in patients receiving home mechanical ventilation. *J Clin Epidemiol*. 2003;56(8):752-9.
20. **Duiverman ML, Wempe JB, Bladder G, Kerstjens HA, Wijkstra PJ.** Health-related quality of life in COPD patients with chronic respiratory failure. *Eur Respir J*. 2008;32(2):379-86.
21. **Lopez-Campos JL, Failde I, Masa JF, Benitez-Moya JM, Barrot E, Ayerbe R, et al.** Transculturally adapted Spanish SRI questionnaire for home mechanically ventilated patients was viable, valid, and reliable. *J Clin Epidemiol*. 2008;61(10):1061-6.
22. **Windisch W, Budweiser S, Heinemann F, Pfeifer M, Rzehak P.** The Severe Respiratory Insufficiency Questionnaire was valid for COPD patients with severe chronic respiratory failure. *J Clin Epidemiol*. 2008;61(8):848-53.
23. **Ghosh D, Rzehak P, Elliott MW, Windisch W.** Validation of the English Severe Respiratory Insufficiency Questionnaire. *Eur Respir J*. 2012;40(2):408-15.
24. **Struik FM, Kerstjens HA, Bladder G, Sprooten R, Zijnen M, Asin J, et al.** The Severe Respiratory Insufficiency Questionnaire scored best in the assessment of

- health-related quality of life in chronic obstructive pulmonary disease. *J Clin Epidemiol.* 2013;66(10):1166-74.
25. **Markussen H, Lehmann S, Nilsen RM, Natvig GK.** The Norwegian version of the Severe Respiratory Insufficiency Questionnaire. *Int J Nurs Pract.* 2014.
  26. **Norweg A, Whiteson J, Demetis S, Rey M.** A new functional status outcome measure of dyspnea and anxiety for adults with lung disease: the dyspnea management questionnaire. *J Cardiopulm Rehabil.* 2006;26(6):395-404.
  27. **Norweg A, Jette AM, Ni P, Whiteson J, Kim M.** Outcome measurement for COPD: reliability and validity of the Dyspnea Management Questionnaire. *Respir Med.* 2011;105(3):442-53.
  28. **Norweg A, Ni P, Garshick E, O'Connor G, Wilke K, Jette AM.** A multidimensional computer adaptive test approach to dyspnea assessment. *Arch Phys Med Rehabil.* 2011;92(10):1561-9.
  29. **Han JN, Xiong CM, Yao W, Fang QH, Zhu YJ, Cheng XS, et al.** Multiple dimensions of cardiopulmonary dyspnea. *Chin Med J.* 2011;124(20):3220-6.
  30. **Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC.** Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21(4):651-7.
  31. **Terwee CB, Schellingerhout JM, Verhagen AP, Koes BW, de Vet HC.** Methodological quality of studies on the measurement properties of neck pain and disability questionnaires: a systematic review. *J Manipulative Physiol Ther.* 2011;34(4):261-72.
  32. **Bausewein C, Daveson B, Benalia H, Simon ST, Higginson IJ.** Outcome Measurement in Palliative Care: The Essentials. London: PRISMA, 2011. 40 s.
  33. **Schellingerhout JM, Heymans MW, Verhagen AP, de Vet HC, Koes BW, Terwee CB.** Measurement properties of translated versions of neck-specific questionnaires: a systematic review. *BMC Med Res Methodol.* 2011;11:87.
  34. **Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB.** Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res.* 2012;21(4):659-70.
  35. **Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al.** Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34-42.

**Tabell 1      Inklusjons- og eksklusjonskriterier**

	<b>Inklusjonskriterier</b>	<b>Eksklusjonskriterier</b>
<b>Populasjon</b>	Voksne ( $\geq 18$ år) palliative pasienter med pustebesvær	Pasienter med akutt oppstått pustebesvær, sykdommer med mulighet for kurasjon
<b>Setting</b>	Inneliggende i sykehus, dagbehandling, poliklinisk eller hjemmebaserte tjenester	
<b>Kartleggingsverktøy</b>	PROM for kartlegging av pustebesvær og som minimum inkluderer dimensjoner for pustebesvær og angst	Generiske PROMs. Kartleggingsverktøy utviklet for bruk i forbindelse med trening eller i akutsituasjoner
<b>Måleegenskaper</b>	Innen kategoriene: reliabilitet, validitet, responsivitet	
<b>Studiedesign</b>	Utvikling og/eller valideringsstudie	Andre studiedesign

PROM (pasientrapporterte resultatmål); kols (kronisk obstruktiv lungesykdom).

**Tabell 2      Karakteristika ved inkluderte studier**

<b>PROM</b>	<b>Studie</b>	<b>Utvalg</b>	<b>N</b>	<b>Land</b>	<b>Setting</b>
<b>CDS</b>	Tanaka et al. (14)	Lungekreft	166	Japan	Poliklinisk Sykehus
	Tanaka et al. (15)	Lungekreft	171	Japan	Poliklinisk
	Henoch et al. (16)	Lungekreft	99	Sverige	Poliklinisk
	Uronis et al. (18)	Lungekreft	109	USA, UK, AUS	Sykehus*
<b>SRI</b>	Windisch et al. (19)	Respirasjonssvikt. LTMV. Ulike diagnoser	226	Tyskland	Sykehus
	Duierveman et al. (20)	Respirasjonssvikt. Kols	72	Nederland	Ikke oppgitt
	Lopez-Campos et al. (21)	LTMV. Ulike diagnoser	115	Spania	Poliklinisk
	Windisch et al. (22)	Respirasjonssvikt. LTMV. Kols	162	Tyskland	Sykehus
	Ghosh et al. (23)	Respirasjonssvikt. LTMV. Ulike diagnoser	152	England	Poliklinisk
	Struik et al. (24)	Respirasjonssvikt. Kols	170	Nederland	Sykehus
	Markussen et al. (25)	Respirasjonssvikt. LTMV. Ulike diagnoser	127	Norge	Hjemmeboende
<b>DMQ-30</b>	Norweg et al. (26)	Kols og astma	85	USA	Ikke oppgitt
<b>DMQ-56</b>	Norweg et al. (27)	Kols	107	USA	Ikke oppgitt
<b>DMQ-CAT</b>	Norweg et al. (28)	Kols	292	USA	Poliklinisk
<b>RSC</b>	Han et al. (29)	Hjerte-/lungesykdommer	396	Kina	Sykehus*

PROM (pasientrapporterte resultatmål); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); LTMV (langtids mekanisk ventilasjon); kols (kronisk obstruktiv lungesykdom); USA (United States of America); UK (United Kingdom); AUS (Australia); \* (uklart om inneliggende, poliklinisk eller begge deler).

**Tabell 3      Karaktetistika ved inkluderte PROMs**

PROM	Målgruppe	Antall spørsmål	Dimensjoner	Måleskala (per spørsmål)	Scoringssystem	Språk	Forventet utfyllingstid
CDS	Kreft	12	Følelse av anstrengelse (5 spm) Følelse av angst (4 spm) Følelse av ubehag (3 spm) Totalskår	5-punkts Likert	Beregner totalskår per dimensjon: anstrengelse (0-20); angst (0-16); ubehag (0-12). Høyere skår indikerer mer alvorlig pustebesvær.	Japansk Engelsk Svensk	2 minutter
SRI	Kronisk respirasjons-svikt med LTMV behandling	49	Åndedrettsplager (8 spm) Fysisk funksjon (6 spm) Tilleggssymptomer og søvn (7 spm) Sosiale relasjoner (6 spm) Angst (5 spm) Psykisk velvære (9 spm) Sosial funksjon (8 spm) Totalskår	5-punkts Likert	Beregner skår per dimensjon som prosent av maks (0-100). Høyere skår indikerer bedre livskvalitet	Tysk Nederlandsk Spansk Engelsk Norsk	20 minutter
DMQ-30	Kols (rehabilitering)	30	Dyspnéintensitet (7 spm) Dyspnérelatert angst (7 spm) Angstrelatert aktivitetssunngåelse (7 spm) Tiltro til aktivitetsrelatert mestringsevne (5 spm) Tilfredshet med mestringsstrategier (4 spm)	7-punkts Likert	Beregner gjennomsnittsskår per dimensjon (0-6). Totalskår kalkuleres ved å summere dimensjonene: intensitet, angst og aktivitetssunngåelse. Høyere skår indikerer bedre funksjon.	Engelsk	Ikke oppgitt
DMQ-56	Kols (rehabilitering)	56	Dyspnéintensitet (12 spm) Dyspnérelatert angst (16 spm) Angstrelatert aktivitetssunngåelse (15 spm) Tiltro til aktivitetsrelatert mestringsevne (6 spm) Tilfredshet med mestringsstrategier (7 spm)	7-punkts Likert	Beregner gjennomsnittsskår per dimensjon (0-6). Høyere skår indikerer bedre funksjon.	Engelsk	Ikke oppgitt
DMQ-CAT	Kols (rehabilitering)	20	Dyspnéintensitet Dyspnérelatert angst Angstrelatert aktivitetssunngåelse Tiltro til aktivitetsrelatert mestringsevne	6-punkts Likert	Ikke oppgitt	Engelsk	Ikke oppgitt
RSC	Pustebesvær som følge av hjerte-/lungesykdom	46	Sensoriske kvaliteter (23 spm) Affektive aspekt (12 spm) Atferdsmessig påvirkning (4 spm) Andre symptomer (7 spm)	5-punkts Likert	Ikke oppgitt	Ikke oppgitt	10-15 minutter

PROM (pasientrapporterte resultatmål); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); LTMV (langtids mekanisk ventilasjon); kols (kronisk obstruktiv lungesykdom).



**Tabell 4 Oversikt over metodisk kvalitet og måleegenskaper fra hver studie**

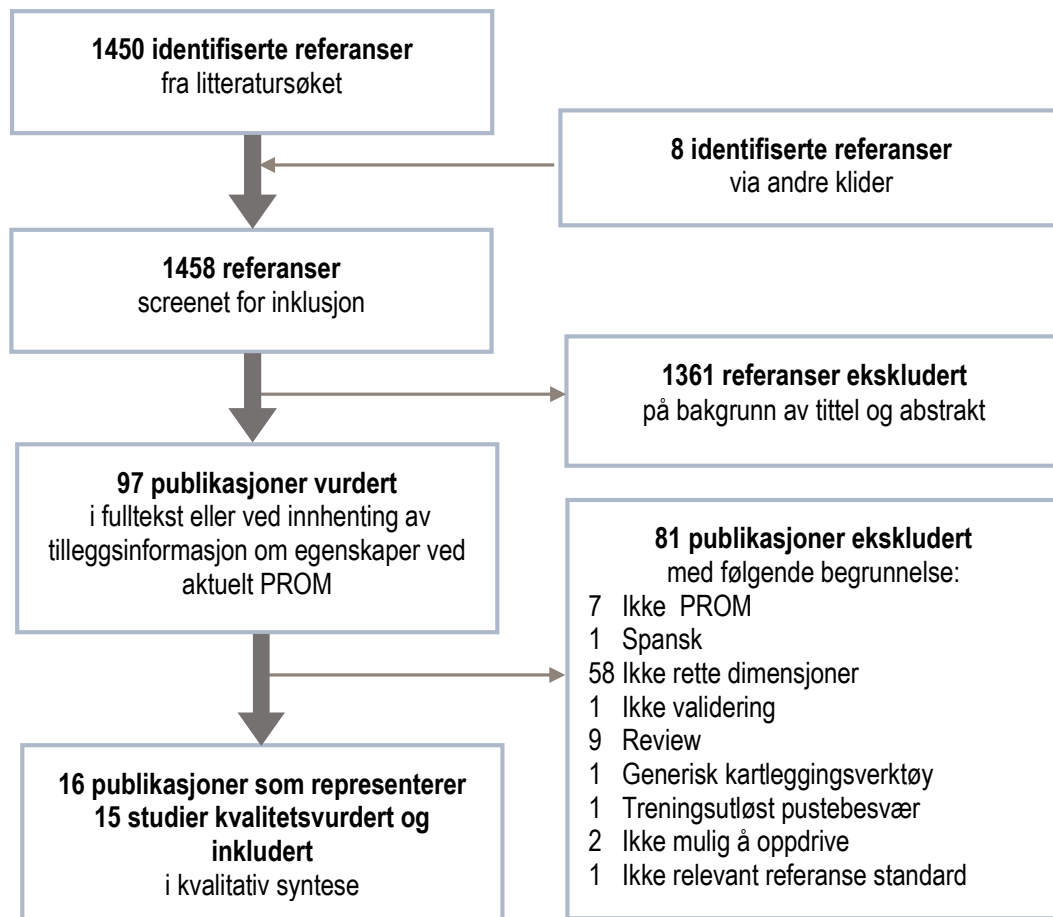
PROM	Studie	Intern konsistens		Reliabilitet		Innholdsvaliditet		Strukturell validitet		Hypotesetesting		Kriterievaliditet		Tverrkulturell Validitet	
		M	R	M	R	M	R	M	R	M	R	M	R	M	R
<b>CDS</b>	Tanaka et al. (14)	Utmerket	+	Rimelig	-	Utmerket	+	Utmerket	?	God	+				
	Tanaka et al. (15)									God	-				
	Henoch et al. (16)	God	+					Utmerket	?	God	+				
	Uronis et al. (18)	Utmerket	+					God	?	God/Rimelig'	-				
<b>SRI</b>	Windisch et al. (19)					Utmerket	+			God	+				
	Duierveman et al. (20)	Svak	+	Svak	+					God	+				
	Lopez-Campos et al. (21)	Svak	+	Rimelig	+			Svak	+	Rimelig	-				
	Windisch et al. (22)	Svak	+					Svak	+						
	Ghosh et al. (23)	Svak	+					Svak	+	God	+				
	Struik et al. (24)	Svak	+							God	+				
	Markussen et al. (25)	Svak	+			God	+			God	+			Svak	-
<b>DMQ-30</b>	Norweg et al. (26)	God	+	Svak	+	Utmerket	+			God/Svak''	+				
<b>DMQ-56</b>	Norweg et al. (27)	God	+	Rimelig	+			Svak	+						
<b>DMQ-CAT</b>	Norweg et al. (28)	God	+			Utmerket	+	Rimelig	+	Rimelig	+	God	+		
<b>RSC</b>	Han et al. (29)							Utmerket	+						

PROM (Patient-reported Outcome Measures); M (metodisk kvalitet); R (måleegenskapens resultat); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); ' (sammenligningene med VAS (Visual Analog Scale), BORG, ECOG (Eastern Cooperative Oncology Group) / HADS (Hospital Anxiety and Depression Scale) ,FACT (Functional Assessment of Cancer Therapy), SpO2 (perifer oksygenmetning); '' (sammenligningene med SOLQ (Seattle Obstructive Lung Disease Questionnaire), SF-12 (Short Form Health Survey) / HADS-angst.

**Tabell 5**      **Kvalitativ syntese**

	<b>Intern konsistens</b>	<b>Reliabilitet</b>	<b>Innholdsvaliditet</b>	<b>Strukturell validitet</b>	<b>Hypotese-testing</b>	<b>Kriterievaliditet</b>	<b>Tverrkulturell validitet</b>
<b>CDS</b>	+++	-	+++	ukjent	+/-	na	na
<b>SRI</b>	?	+	+++	?	+++	na	?
<b>DMQ-30</b>	++	?	+++	na	++	na	na
<b>DMQ-56</b>	++	+	na	?	na	na	na
<b>DMQ-CAT</b>	++	na	+++	+	+	++	na
<b>RSC</b>	na	na	na	+++	na	na	na

CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); +++ eller ---, høy evidens positivt/negativt resultat; ++ eller --, moderat evidens positivt/negativt resultat; + eller -, svak evidens positivt/negativt resultat; ±, motstridene evidens; ?, usikker på grunn av svak metodisk kvalitet; ukjent, på grunn av ukjent resultat; na, ingen informasjon tilgjengelig.



**Figur 1** Flytskjema over identifisert litteratur

## Vedlegg I    Søkestrategier

Database: Ovid **MEDLINE(R)** In-Process & Other Non-Indexed Citations and Ovid  
MEDLINE(R) <1946 to Present>

Dato for søk: 20.03.14

Antall treff: 1059

- 1    Dyspnea/ (14962)
- 2    dyspne\*.tw. (24457)
- 3    dyspnoe\*.tw. (7097)
- 4    breathless\*.tw. (3356)
- 5    (breath\* adj3 (short\* or difficult\*).tw. (6742)
- 6    or/1-5 (46450)
- 7    Validation Studies/ (65774)
- 8    exp "Reproducibility of Results"/ (267397)
- 9    Psychometrics/ (53417)
- 10    "Sensitivity and Specificity"/ (273130)
- 11    Cross-Cultural Comparison/ (21211)
- 12    Factor Analysis, Statistical/ (19819)
- 13    valid\*.tw. (389395)
- 14    reliabilit\*.tw. (98830)
- 15    reproducibilit\*.tw. (49623)
- 16    (internal adj2 consistenc\*).tw. (15916)
- 17    (test adj2 retest).tw. (15052)
- 18    ((inter or intra) adj2 rater).tw. (4882)
- 19    (measurement adj3 error).tw. (6045)
- 20    (hypothes\* adj test\*).tw. (4331)
- 21    (transcultural adj2 stud\*).tw. (133)

- 22 (comparison\* adj3 (cross adj1 cultural)).tw. (693)
- 23 psychometr\*.tw. (25748)
- 24 specificit\*.tw. (339236)
- 25 sensitivit\*.tw. (538209)
- 26 responsive\*.tw. (164312)
- 27 (factor adj2 analys\*).tw. (27571)
- 28 (minimal\* adj3 ((important or detectable) adj2 (change or difference))).tw. (1179)
- 29 or/7-28 (1639652)
- 30 6 and 29 (3293)
- 31 Palliative Care/ (39512)
- 32 Hospice Care/ (4534)
- 33 Terminal Care/ (20716)
- 34 Terminally Ill/ (5252)
- 35 ((palliative or hospice or terminal or supportive) adj2 care).tw. (26306)
- 36 ("end of life" adj2 care).tw. (5111)
- 37 exp Pulmonary Disease, Chronic Obstructive/ (35130)
- 38 copd.tw. (24932)
- 39 (chronic adj4 (obstruct\* or air\*) adj4 (pulmonar\* or lung or respiratory) adj4 diseas\*).tw. (30174)
- 40 exp Neoplasms/ (2516827)
- 41 (cancer\* or neoplasm\* or tumo?r\*).tw. (1868679)
- 42 or/31-41 (3044025)
- 43 30 and 42 (1059)

## Vedlegg II Kategorisering av måleegenskaper (1-3)

Måleegenskap	Vurdering	Kvalitetskriterier
<b>Reliabilitet</b>		
Intern konsistens	+	Endimensjonal (sub)skala OG Cronbachs alfa(s) $\geq 0,70$
	?	Ukjent dimensjonalitet ELLER ikke fastsatt Cronbachs alfa
	-	Ikke endimensjonal (sub)skala ELLER Cronbachs alfa (s) $< 0,70$
Reliabilitet	+	ICC/vektet kappa $\geq 0,70$ ELLER Pearsons $r \geq 0,80$
	?	Hverken ICC/vektet kappa eller Pearsons $r$ fastsatt
	-	ICC/vektet kappa $< 0,70$ ELLER Pearsons $r < 0,80$
<b>Validitet</b>		
Innholdsvaliditet	+	Målgruppen vurderer alle spørsmålene i spørreskjemaet som relevant
	?	Ingen målgruppe involvering
	-	Målgruppen vurderer testledd i spørreskjemaet til å være irrelevant ELLER vurderer spørreskjemaet til å være ufullstendig
Begrepsvaliditet		
Tverrkulturell validitet	+	Opprinnelig faktorstruktur bekreftet ELLER ingen betydningsfull DIF
	?	Bekreftelse av opprinnelig faktorstruktur OG DIF ikke omtalt
	-	Opprinnelig faktorstruktur ikke bekreftet ELLER betydningsfull DIF
Strukturell validitet	+	Faktorene bør forklare minst 50 % av variansen
	?	Forklart varians ikke nevnt
	-	Faktorene forklarer $< 50$ % av variansen
Hypotesetesting	+	Samvariasjon med en test som måler samme fenomen $\geq 0,50$ ELLER minst 75 % av resultatene er i samsvar med hypotesene OG samvariasjon med relaterte fenomen er høyere enn med ikke-relaterte fenomen
	?	Samvariasjon utelukkende fastsatt med ikke-korrelerte fenomen
	-	Samvariasjon $< 0,50$ med en test som måler samme fenomen ELLER lavere samvariasjon med relaterte fenomen enn med ikke-relaterte fenomen

ICC (intraklasse korrelasjonskoeffisient); DIF (differential item functioning)  
 + (positive måleegenskaper); ? (uavklarte måleegenskaper); - (negative måleegenskaper)

### Vedlegg III Gradering av tillit til kartleggingsverktøys måleegenskaper tilpasset fra Schellingerhout m.fl. (2)

---

Nivå	Vurdering	Kriterier
Høy	+++ ELLER---	Konsistente funn i flere studier av <i>god</i> metodisk kvalitet ELLER i en studie av <i>utmerket</i> metodisk kvalitet
Moderat	++ ELLER --	Konsistente funn i flere studier av <i>rimelig</i> metodisk kvalitet ELLER i en studie av <i>god</i> metodisk kvalitet
Lav	+ ELLER -	En studie av <i>rimelig</i> metodisk kvalitet
Motstridende	+/-	Motstridende funn
Usikker	?	Bare studier av <i>svak</i> metodisk kvalitet
	Ukjent	Ukjent resultat

---

+ (positivt resultat); – (negativt resultat)

## Vedlegg IV Resultater for intern konsistens, test-retest-reliabilitet og strukturell validitet

PROM	Studie	Intern konsistens (Cronbachs alfa)	Test-retest-reliabilitet (Pearsons korrelasjonskoeffisient/Intraklasse korrelasjonskoeffisient)	Strukturell validitet (% av total varians)
CDS	Tanaka et al. (14)	Følelse av anstrengelse 0,83; Følelse av angst 0,81; Følelse av ubehag 0,94; Total skår 0,64	Følelse av anstrengelse $r=0,71$ ; Følelse av angst $r=0,69$ ; Følelse av ubehag $r=0,58$ ; Total skår $r=0,69$	Følelse av anstrengelse 27; Følelse av angst 21; Følelse av ubehag 14
	Henoch et al. (16)	Følelse av anstrengelse 0,81; Følelse av angst 0,84; Følelse av ubehag 0,88; Total skår 0,90		
	Uronis et al. (18)	Følelse av anstrengelse 0,84; Følelse av angst 0,80; Følelse av ubehag 0,84; Total skår 0,71		
SRI	Duiverman et al. (20)	SRI-RC 0,73; SRI-PF 0,73; SRI-AS 0,66; SRI-SR 0,78; SRI-AX 0,70; SRI-WB 0,84; SRI-SF 0,73	ICC 0,81	
	López-Campos et al. (21)	SRI-RC 0,81; SRI-PF 0,82; SRI-AS 0,74; SRI-SR 0,63; SRI-AX 0,73; SRI-WB 0,85; SRI-SF 0,83; SRI-SS 0,93	ICC: SRI-RC 0,90; SRI-PF 0,96; SRI-AS 0,91; SRI-SR 0,89; SRI-AX 0,84; SRI-WB 0,91; SRI-SF 0,92; SRI-SS 0,95	60
	Windisch et al. (22)	SRI-RC 0,82; SRI-PF 0,82; SRI-AS 0,73; SRI-SR 0,74; SRI-AX 0,77; SRI-WB 0,82; SRI-SF 0,82; SRI-SS 0,88		SRI-SS 58,5
	Ghosh et al. (23)	SRI-RC 0,84; SRI-PF 0,77; SRI-AS 0,81; SRI-SR 0,81; SRI-AX 0,87; SRI-WB 0,89; SRI-SF 0,87; SRI-SS 0,93		SRI-SS 70
	Struik et al. (24)	SRI-RC 0,82; SRI-PF 0,80; SRI-AS 0,77; SRI-SR 0,68; SRI-AX 0,80; SRI-WB 0,85; SRI-SF 0,81		
	Markussen et al. (25)	SRI-RC 0,81; SRI-PF 0,76; SRI-AS 0,68; SRI-SR 0,82; SRI-AX 0,81; SRI-WB 0,88; SRI-SF 0,79; SRI-SS 0,94		
DMQ-30	Norweg et al. (26)	Dyspneintensitet 0,93; Dyspnerelatert angst 0,92; Angstrelatert aktivitetsunnngåelse 0,93; Tiltro til aktivitetsrelatert mestringsevne 0,87; Tilfredshet med mestringsstrategier 0,90; DMQ totalskår (21 testledd) 0,96	ICC: Dyspneintensitet 0,88; Dyspnerelatert angst 0,92; Angstrelatert aktivitetsunnngåelse 0,90; Tiltro til aktivitetsrelatert mestringsevne 0,71; Tilfredshet med mestringsstrategier 0,72; DMQ totalskår (21 testledd) 0,95	
DMQ-56	Norweg et al. (27)	Dyspneintensitet 0,95; Dyspnerelatert angst 0,95; Angstrelatert aktivitetsunnngåelse 0,96; Tiltro til aktivitetsrelatert mestringsevne 0,86; Tilfredshet med mestringsstrategier 0,85		Dyspneintensitet 71; Dyspnerelatert angst 31; Angstrelatert aktivitetsunnngåelse 20; Tiltro til aktivitetsrelatert mestringsevne 9; Tilfredshet med mestringsstrategier 31
DMQ-CAT	Norweg et al. (28)	0,92-0,98 for de fire faktorene		63
RSC	Han et al. (29)			64

PROM (pasientrapporterte resultatmål); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); LTMV (langtids mekanisk ventilasjon); kols (kronisk obstruktiv lungesykdom); SRI-RC (Åndedrettsplager); SRI-PF (Fysisk funksjon); SRI-AS (Tilleggssymptomer og søvn); SRI-SR (Sosiale relasjoner); SRI-AX (Angst); SRI-WB (Psykisk velvære); SR-SF (Sosial funksjon); SRI-SS (Totalskår:  $r$  (Pearsons korrelasjonskoeffisient); ICC (intraklasse korrelasjonskoeffisient).



## Vedlegg V Samvariasjoner

### Utvalgte samvariasjoner for CDS

CDS	HADS-angst	HADS-depresjon	HADS-total	Borg skala	VAS-dyspné
Følelse av anstrengelse	0,30** 0,36' 0,50''	0,32** 0,24' 0,38''	0,34**  0,53''	0,72*  0,86''	0,77*  0,61' 0,80''
Følelse av angst	0,22** 0,41' 0,32''	0,21** 0,38' 0,44''	0,24**  0,45''	0,41*  0,64''	0,53*  0,40' 0,59''
Følelse av ubehag	0,21** 0,24' 0,40''	0,30** 0,20' 0,43''	0,28**  0,50''	0,44*  0,64''	0,40*  0,55' 0,59''
Total skår	0,30** 0,36' 0,49''	0,34** 0,27' 0,45''	0,35**  0,57''	0,67*  0,87''	0,72*  0,63' 0,82''

CDS (Cancer Dyspnea Scale); HADS (Hospital Anxiety and Depression Scale); VAS (Visual Analog Scale); \*Tanaka et al. (14); \*\*Tanaka et al. (15); 'Hench et al.(16); ''Uronis et al. (18).

### Utvalgte samvariasjoner for SRI

	SF-36 GH	SF-36 VT	SF-36 MH	SF-36 PHC	SF-36 MHC	CRQ-dyspné	CRQ-fatigue	CRQ-følelser	CRQ-mestring	CRQ-Total	HADS-angst	HADS-depresjon	MRC
SRI-RC	0,63''' 0,61*** 0,47'' 0,51*	0,52''' 0,63*** 0,54'' 0,50*	0,29''' 0,43*** 0,24'' 0,42*	0,53'''  0,48'' 0,42*	0,40'''  0,28'' 0,45*	0,57** 0,56'	0,59** 0,60'	0,40** 0,45'	0,34** 0,50'	0,54** 0,62'	0,50** 0,56'	0,41** 0,43'	0,43** 0,41'
SRI-AX	0,50''' 0,58*** 0,48'' 0,51*	0,44''' 0,66*** 0,37'' 0,56*	0,49''' 0,63*** 0,37'' 0,51*	0,42'''  0,29'' 0,28*	0,55'''  0,37'' 0,52*	0,35** 0,49'	0,57** 0,52'	0,65** 0,67'	0,60** 0,70'	0,67** 0,74'	0,69** 0,74'	0,54** 0,59'	0,26** 0,33'
SRI-SS	0,70''' 0,66'' 0,63*	0,65''' 0,76'' 0,73*	0,54''' 0,60'' 0,70*	0,62''' 0,67'' 0,46*	0,61''' 0,54'' 0,67*	0,53** 0,63'	0,74** 0,63'	0,73** 0,71'	0,62** 0,71'	0,81** 0,82'	0,75** 0,80'	0,68** 0,72'	0,41** 0,46'

SRI (Severe Respiratory Insufficiency Questionnaire); SF-36 (Short Form Health Survey); CRQ (Chronic Respiratory Disease Questionnaire); HADS (Hospital Anxiety and Depression Scale); MRC (Medical Research Council Dyspnea Scale); RC (åndedrettsplager); AX (angst); SS (totalskår); GH (generell helse); VT (vitalitet); MH (mental helse); PHC (fysisk helsekomponent); MHC (mental helsekomponent); '''Markussen et al. (25); \*\*\*Gosh et al. (23); ''Lopez-Campos et al. (21); \*Windisch et al. (19); \*\*Duiverman et al. (20); 'Struik et al. (24).

## Samvariasjoner for DMQ-30

DMQ-30	SOLQ- fysisk funksjon	SOLQ- følelsesmessig funksjon	SOLQ- mestrings ferdigheter	SOLQ- tilfredshet med omsorg	SF-12 PCS	SF-12 MCS	HADS- angst
Dyspnéintensitet	0,69	0,56	0,52	-0,04	0,54	0,31	0,39
Dyspnérelatert angst	0,44	0,65	0,72	-0,07	0,30	0,41	0,60
Angstrelatert aktivitetsunngåelse	0,67	0,65	0,61	0,01	0,40	0,56	0,59
Tiltro til aktivitetsrelatert mestringsevne	0,22	0,28	0,25	-0,01	0,23	0,04	0,05
Tilfredshet med mestringsstrategier	0,02	0,16	0,21	0,27	0,06	0,03	0,01
DMQ composite*	0,83	0,81	0,74	-0,05	0,57	0,46	0,65

DMQ (Dyspnea Management Questionnaire); SOLQ (Seattle Obstructive Lung Disease Questionnaire); SF-12 (Short Form Health Survey); HADS (Hospital Anxiety and Depression Scale); PCS (Physical Component Summary); MCS (Mental Component Summary). \* sumskår av dyspnéintensitet, dyspnérelatert angst og angstrelatert aktivitetsunngåelse.

## Samvariasjoner for DMQ-CAT

DMQ-CAT	UCSD SOBQ	CRQ- dyspné	CRQ- følelser	CRQ- mestring	CSES	HADS- angst
Dyspnéintensitet	-0,80	0,47	0,41	0,53	0,51	- 0,31
Dyspnérelatert angst	-0,59	0,45	0,62	0,75	0,66	- 0,61
Angstrelatert aktivitetsunngåelse	-0,72	0,37	0,45	0,58	0,47	- 0,36
Tiltro til aktivitetsrelatert mestringsevne	-0,69	0,36	0,48	0,64	0,71	- 0,42

DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); UCSD SOBQ (University of California, San Diego Shortness of Breath Questionnaire); CRQ (Chronic Respiratory Disease Questionnaire); CSES (COPD Self-Efficacy Scale); HADS (Hospital Anxiety and Depression Scale).

## Vedlegg I Søkestrategier

Database: Ovid **MEDLINE**(R) In-Process & Other Non-Indexed Citations and Ovid

MEDLINE(R) <1946 to Present>

Dato for søk: 20.03.14

Antall treff: 1059

- 1 Dyspnea/ (14962)
- 2 dyspne\*.tw. (24457)
- 3 dyspnoe\*.tw. (7097)
- 4 breathless\*.tw. (3356)
- 5 (breath\* adj3 (short\* or difficult\*)).tw. (6742)
- 6 or/1-5 (46450)
- 7 Validation Studies/ (65774)
- 8 exp "Reproducibility of Results"/ (267397)
- 9 Psychometrics/ (53417)
- 10 "Sensitivity and Specificity"/ (273130)
- 11 Cross-Cultural Comparison/ (21211)
- 12 Factor Analysis, Statistical/ (19819)
- 13 valid\*.tw. (389395)
- 14 reliabilit\*.tw. (98830)
- 15 reproducibilit\*.tw. (49623)
- 16 (internal adj2 consistenc\*).tw. (15916)
- 17 (test adj2 retest).tw. (15052)
- 18 ((inter or intra) adj2 rater).tw. (4882)
- 19 (measurement adj3 error).tw. (6045)
- 20 (hypothes\* adj test\*).tw. (4331)
- 21 (transcultural adj2 stud\*).tw. (133)
- 22 (comparison\* adj3 (cross adj1 cultural)).tw. (693)
- 23 psychometr\*.tw. (25748)
- 24 specificit\*.tw. (339236)
- 25 sensitivit\*.tw. (538209)
- 26 responsive\*.tw. (164312)
- 27 (factor adj2 analys\*).tw. (27571)
- 28 (minimal\* adj3 ((important or detectable) adj2 (change or difference))).tw. (1179)
- 29 or/7-28 (1639652)

- 30 6 and 29 (3293)
- 31 Palliative Care/ (39512)
- 32 Hospice Care/ (4534)
- 33 Terminal Care/ (20716)
- 34 Terminally Ill/ (5252)
- 35 ((palliative or hospice or terminal or supportive) adj2 care).tw. (26306)
- 36 ("end of life" adj2 care).tw. (5111)
- 37 exp Pulmonary Disease, Chronic Obstructive/ (35130)
- 38 copd.tw. (24932)
- 39 (chronic adj4 (obstruct\* or air\*) adj4 (pulmonar\* or lung or respiratory) adj2 diseas\*).tw. (30174)
- 40 exp Neoplasms/ (2516827)
- 41 (cancer\* or neoplasm\* or tumo?r\*).tw. (1868679)
- 42 or/31-41 (3044025)
- 43 30 and 42 (1059)

Database: **Embase** <1974 to 2014 March 19>

Dato for søk: 20.03.14

Antall treff: 162

- 1 dyspnea/ (84082)
- 2 dyspne\*.tw. (35435)
- 3 dyspnoe\*.tw. (9917)
- 4 breathlessness.tw. (4255)
- 5 (breath\* adj3 (short\* or difficult\*).tw. (10793)
- 6 or/1-5 (104961)
- 7 validation study/ (40838)
- 8 exp reproducibility/ (137233)
- 9 psychometry/ (42090)
- 10 exp reliability/ (103639)
- 11 factorial analysis/ (25011)
- 12 "sensitivity and specificity"/ (194461)
- 13 cultural factor/ (48143)
- 14 factorial analysis/ (25011)
- 15 instrument validation/ (1751)
- 16 valid\*.tw. (499882)

- 17 reliabilit\*.tw. (118054)
- 18 reproducibilit\*.tw. (59486)
- 19 (internal adj2 consistenc\*).tw. (18577)
- 20 (test adj2 retest).tw. (17385)
- 21 ((inter or intra) adj2 rater).tw. (6620)
- 22 (measurement adj3 error).tw. (6541)
- 23 measurement error/ (3909)
- 24 (hypothes\* adj test\*).tw. (5007)
- 25 (transcultural adj2 stud\*).tw. (167)
- 26 (comparison\* adj3 (cross adj1 cultural)).tw. (774)
- 27 psychometr\*.tw. (31518)
- 28 specificit\*.tw. (395985)
- 29 sensitivit\*.tw. (638818)
- 30 responsive\*.tw. (186750)
- 31 (factor adj2 analys\*).tw. (32892)
- 32 (minimal\* adj3 ((important or detectable) adj2 (change or difference))).tw. (1543)
- 33 or/7-32 (1841064)
- 34 6 and 33 (6020)
- 35 palliative therapy/ (47963)
- 36 hospice care/ (5908)
- 37 terminal care/ (23149)
- 38 exp terminally ill patient/ (5965)
- 39 ((palliative or hospice or terminal or supportive) adj2 care).tw. (37458)
- 40 ("end of life" adj2 care).tw. (6582)
- 41 chronic obstructive lung disease/ (69985)
- 42 copd.tw. (36676)
- 43 (chronic adj4 (obstruct\* or air\*) adj4 (pulmonar\* or lung or respiratory) adj2 diseas\*).tw. (38625)
- 44 exp neoplasm/ (3312035)
- 45 (cancer\* or neoplasm\* or tumo?r).tw. (2135228)
- 46 or/35-45 (3803323)
- 47 34 and 46 (2234)
- 48 limit 47 to exclude medline journals (162)

Database: **PsycINFO** <1806 to March Week 3 2014>

Dato for søk: 20.03.14

Antall treff: 88

- 1 dyspnea/ (283)
- 2 dyspne\*.tw. (847)
- 3 dyspnoe\*.tw. (161)
- 4 breathless\*.tw. (275)
- 5 (breath adj3 (short\* or difficult\*).tw. (354)
- 6 or/1-5 (1559)
- 7 exp test validity/ (52630)
- 8 exp factor analysis/ (12073)
- 9 Psychometrics/ (36847)
- 10 Test Reliability/ (36915)
- 11 interrater reliability/ (2664)
- 12 "Error of Measurement"/ (1200)
- 13 valid\*.tw. (174441)
- 14 reliabilit\*.tw. (72191)
- 15 reproducibilit\*.tw. (1403)
- 16 (internal adj2 consistenc\*).tw. (17779)
- 17 (test adj2 retest).tw. (12751)
- 18 ((inter or intra) adj2 rater).tw. (2501)
- 19 (measurement adj3 error).tw. (2264)
- 20 (hypothes\* adj test\*).tw. (4047)
- 21 (transcultural adj2 stud\*).tw. (149)
- 22 (comparison\* adj3 (cross adj1 cultural)).tw. (2062)
- 23 psychometr\*.tw. (52778)
- 24 specificit\*.tw. (25870)
- 25 sensitivit\*.tw. (65549)
- 26 responsive\*.tw. (28650)
- 27 (factor adj2 analys\*).tw. (42674)
- 28 (minimal\* adj3 ((important or detectable) adj2 (change or difference))).tw. (158)
- 29 or/7-28 (355931)
- 30 6 and 29 (208)
- 31 palliative care/ (7106)
- 32 terminally ill patients/ (4005)
- 33 ((palliative or hospice or terminal or supportive) adj2 care).tw. (7587)

- 34 ("end of life" adj2 care).tw. (2824)
- 35 exp chronic obstructive pulmonary disease/ (783)
- 36 copd.tw. (920)
- 37 (chronic adj4 (obstruct\* or air\*) adj4 (pulmonar\* or lung or respiratory) adj diseases\*).tw. (1438)
- 38 exp neoplasms/ (33818)
- 39 (cancer\* or neoplasm\* or tumo?r\*).tw. (47759)
- 40 or/31-39 (60459)
- 41 30 and 40 (88)

Database: **AMED** (Allied and Complementary Medicine) <1985 to March 2014>

Dato for søk: 20.03.14

Antall treff: 90

- 1 dyspnea/ (296)
- 2 dyspne\*.tw. (674)
- 3 dyspnoe\*.tw. (244)
- 4 breathless\*.tw. (212)
- 5 (breath\* adj3 (short\* or difficult\*)).tw. (118)
- 6 or/1-5 (1058)
- 7 "consistency and reliability"/ (1898)
- 8 "reproducibility of results"/ (1965)
- 9 psychometrics/ (709)
- 10 cross cultural comparison/ (218)
- 11 valid\*.tw. (9392)
- 12 reliabilit\*.tw. (6824)
- 13 reproducibilit\*.tw. (2403)
- 14 (internal adj2 consistenc\*).tw. (1162)
- 15 (test adj2 retest).tw. (1607)
- 16 ((inter or intra) adj2 rater).tw. (587)
- 17 (measurement adj3 error).tw. (409)
- 18 (hypothes\* adj test\*).tw. (155)
- 19 (transcultural adj2 stud\*).tw. (6)
- 20 (comparison\* adj3 (cross adj1 cultural)).tw. (244)
- 21 psychometr\*.tw. (1655)
- 22 specificit\*.tw. (1170)

- 23 sensitivit\*.tw. (2657)  
 24 responsive\*.tw. (1490)  
 25 (factor adj2 analys\*).tw. (958)  
 26 (minimal\* adj3 ((important or detectable) adj2 (change or difference))).tw. (183)  
 27 or/7-26 (18274)  
 28 6 and 27 (116)  
 29 palliative care/ (5253)  
 30 exp terminal care/ (4613)  
 31 terminal illness/ (2750)  
 32 ((palliative or hospice or terminal or supportive) adj2 care).tw. (12240)  
 33 ("end of life" adj2 care).tw. (1971)  
 34 exp pulmonary disease chronic obstructive/ (2804)  
 35 copd.tw. (999)  
 36 (chronic adj4 (obstruct\* or air\*) adj4 (pulmonar\* or lung or respiratory) adj2 diseas\*).tw. (1014)  
 37 exp neoplasms/ (12237)  
 38 (cancer\* or neoplasm\* or tumo?r\*).tw. (15873)  
 39 or/29-38 (29048)  
 40 28 and 39 (90)

Database: **CINAHL**  
 Dato for søk: 20.03.14  
 Antall treff: 101

S1	(MH "Dyspnea")	Search modes - Boolean/Phrase	3,806
S2	TX dyspne*	Search modes - Boolean/Phrase	5,976
S3	TX dyspnoe*	Search modes - Boolean/Phrase	567
S4	TX breathless*	Search modes - Boolean/Phrase	752
S5	TX (breath* N3 (short* or difficult*))	Search modes - Boolean/Phrase	1,379
S6	S1 OR S2 OR S3 OR S4 OR S5	Search modes - Boolean/Phrase	7,669
S7	(MH "Validation Studies")	Search modes - Boolean/Phrase	15,725



S8	(MH "Instrument Validation")	Search modes - Boolean/Phrase	18,352
S9	(MH "Reliability and Validity+")	Search modes - Boolean/Phrase	112,746
S10	(MH "Reproducibility of Results")	Search modes - Boolean/Phrase	14,035
S11	(MH "Psychometrics")	Search modes - Boolean/Phrase	7,641
S12	(MH "Sensitivity and Specificity")	Search modes - Boolean/Phrase	27,207
S13	(MH "Factor Analysis")	Search modes - Boolean/Phrase	16,633
S14	(MH "Measurement Error+")	Search modes - Boolean/Phrase	387
S15	TX valid*	Search modes - Boolean/Phrase	83,808
S16	TX reliabilit*	Search modes - Boolean/Phrase	52,285
S17	TX reproducibilit*	Search modes - Boolean/Phrase	15,351
S18	TX (internal N2 consistenc*)	Search modes - Boolean/Phrase	14,338
S19	TX (test N2 retest)	Search modes - Boolean/Phrase	12,548
S20	TX ((inter or intra) N2 rater)	Search modes - Boolean/Phrase	1,823
S21	TX (measurement N3 error)	Search modes - Boolean/Phrase	1,309
S22	TX hypothes* N0 test*	Search modes - Boolean/Phrase	991
S23	TX (transcultural N2 stud*)	Search modes - Boolean/Phrase	98
S24	TX psychometr*	Search modes - Boolean/Phrase	12,094
S25	TX specificit*	Search modes - Boolean/Phrase	34,035
S26	TX sensitivit*	Search modes - Boolean/Phrase	58,937
S27	TX responsive*	Search modes - Boolean/Phrase	8,760
S28	TX (factor N2 analys*)	Search modes - Boolean/Phrase	18,909
S29	TX (comparison* N3 (cross N1 cultural))	Search modes - Boolean/Phrase	237
S30	TX (minimal* N3 ((important or detectable)	Search modes - Boolean/Phrase	438

	N2 (change or difference)))		
S31	S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30	Search modes - Boolean/Phrase	210,110
S32	(MH "Palliative Care")	Search modes - Boolean/Phrase	17,487
S33	(MH "Terminal Care")	Search modes - Boolean/Phrase	10,231
S34	(MH "Hospice Care")	Search modes - Boolean/Phrase	5,541
S35	(MH "Terminally Ill Patients+")	Search modes - Boolean/Phrase	7,235
S36	TX ((palliative or hospice or terminal or supportive) N2 care)	Search modes - Boolean/Phrase	43,745
S37	TX ("end of life" N0 care)	Search modes - Boolean/Phrase	4,573
S38	(MH "Lung Diseases, Obstructive+")	Search modes - Boolean/Phrase	29,164
S39	TX copd	Search modes - Boolean/Phrase	5,873
S40	TX (chronic N4 (obstruct* or air*) N4 (pulmonar* or lung or respiratory) N0 diseas*)	Search modes - Boolean/Phrase	8,852
S41	(MH "Neoplasms+")	Search modes - Boolean/Phrase	175,100
S42	TX (cancer* or neoplasm* or tumo#r*)	Search modes - Boolean/Phrase	218,142
S43	S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42	Search modes - Boolean/Phrase	296,018
S44	S6 AND S31	Search modes - Boolean/Phrase	1,051
S45	S43 AND S44	Search modes - Boolean/Phrase	518

S46	S43 AND S44	Limiters - Exclude MEDLINE records Search modes - Boolean/Phrase	101
-----	-------------	---	-----

<b>Database: The Cochrane Library</b>		
Dato for søk: 20.03.14		
Antall treff: 152		
#1	MeSH descriptor: [Dyspnea] this term only	736
#2	dyspne*:ti,ab,kw (Word variations have been searched)	2761
#3	dyspnoe*:ti,ab,kw (Word variations have been searched)	764
#4	breathless*:ti,ab,kw (Word variations have been searched)	583
#5	(breath* near/4 (short* or difficult*)):ti,ab,kw (Word variations have been searched)	523
#6	#1 or #2 or #3 or #4 or #5	4007
#7	MeSH descriptor: [Validation Studies] this term only	0
#8	MeSH descriptor: [Reproducibility of Results] explode all trees	8634
#9	MeSH descriptor: [Psychometrics] this term only	2223
#10	MeSH descriptor: [Sensitivity and Specificity] this term only	10003
#11	MeSH descriptor: [Factor Analysis, Statistical] this term only	499
#12	valid*:ti,ab,kw (Word variations have been searched)	15857
#13	reliabilit*:ti,ab,kw (Word variations have been searched)	4367
#14	reproducibilit*:ti,ab,kw (Word variations have been searched)	10052
#15	(internal near/3 consistenc*):ti,ab,kw (Word variations have been searched)	626
#16	(test near/3 retest):ti,ab,kw (Word variations have been searched)	881
#17	((inter or intra) near/3 rater):ti,ab,kw (Word variations have been searched)	310
#18	(measurement near/4 error):ti,ab,kw (Word variations have been searched)	395
#19	(hypothes* near/1 test*):ti,ab,kw (Word variations have been searched)	341
#20	(transcultural near/3 stud*):ti,ab,kw (Word variations have been searched)	5
#21	psychometr*:ti,ab,kw (Word variations have been searched)	3782
#22	specificit*:ti,ab,kw (Word variations have been searched)	15140
#23	sensitivit*:ti,ab,kw (Word variations have been searched)	29111
#24	responsive*:ti,ab,kw (Word variations have been searched)	5644
#25	factor near/3 analys*:ti,ab,kw (Word variations have been searched)	5487
#26	(comparison* near/4 (cross near/1 cultural)):ti,ab,kw (Word variations have been	211

	searched)	
#27	(minimal* near/2 ((important or detectable) near/3 (change or difference))):ti,ab,kw (Word variations have been searched)	251
#28	#7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27	64174
#29	#6 and #28	388
#30	MeSH descriptor: [Palliative Care] this term only	1330
#31	MeSH descriptor: [Hospice Care] this term only	76
#32	MeSH descriptor: [Terminal Care] this term only	227
#33	((palliative or hospice or terminal or supportive) near/2 care):ti,ab,kw (Word variations have been searched)	2567
#34	«end of life» near/2 care:ti,ab,kw (Word variations have been searched)	96
#35	MeSH descriptor: [Pulmonary Disease, Chronic Obstructive] explode all trees	2349
#36	copd:ti,ab,kw (Word variations have been searched)	6441
#37	(chronic near/4 (obstruct* or air*) near/4 (pulmonar* or lung or respiratory) near/1 diseas*):ti,ab,kw (Word variations have been searched)	5359
#38	MeSH descriptor: [Neoplasms] explode all trees	49382
#39	(cancer* or neoplasm* or tumo?r*):ti,ab,kw (Word variations have been searched)	71600
#40	#30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39	87413
#41	#29 and #40	152

Database: **SveMed+**

Dato for søk: 23.03.14

Antall treff: 240

1	noexp:"dyspnea"	85
2	dyspne*	88
3	dyspnoe*	8
4	breath*	160
5	#1 OR #2 OR #3 OR #4	240

Database: **Science Citation Index**

Dato for søk: 23.03.15

Antall treff: 96

Siteringssøk på:

Bausewein, C., Farquhar, M., Booth, S., Gysels, M. & Higginson, I. J. (2007) Measurement of breathlessness in advanced disease: a systematic review. *Respiratory Medicine*, 101 (3), s. 399-410.

Dorman, S., Byrne, A. & Edwards, A. (2007) Which measurement scales should we use to measure breathlessness in palliative care? A systematic review. *Palliative Medicine*, 21 (3), s. 177-91.



## COSMIN checklist with 4-point scale

### Contact

CB Terwee, PhD  
VU University Medical Center  
Department of Epidemiology and Biostatistics  
EMGO Institute for Health and Care Research  
1081 BT Amsterdam  
The Netherlands  
Website: [www.cosmin.nl](http://www.cosmin.nl), [www.emgo.nl](http://www.emgo.nl)  
E-mail: [cb.terwee@vumc.nl](mailto:cb.terwee@vumc.nl)

### Instructions

This version of the COSMIN checklist is recommended for use in systematic reviews of measurement properties. With this version it is possible to calculate overall methodological quality scores per study on a measurement property. A methodological quality score per box is obtained by taking the lowest rating of any item in a box ('worse score counts'). For example, if for a reliability study one item in the box 'Reliability' is scored poor, the methodological quality of that reliability study is rated as poor. The Interpretability box and the Generalizability box are mainly used as data extraction forms. We recommend to use the Interpretability box to extract all information on the interpretability issues described in this box (e.g. norm scores, floor-ceiling effects, minimal important change) of the instruments under study from the included articles. Similar, we recommend to use the Generalizability box to extract data on the characteristics of the study population and sampling procedure. Therefore no scoring system was developed for these boxes.

This scoring system is described in this paper:

Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research* 2011, July 6 [epub ahead of print].

**Step 1. Evaluated measurement properties in the article**

	Internal consistency	Box A
	Reliability	Box B
	Measurement error	Box C
	Content validity	Box D
	Structural validity	Box E
	Hypotheses testing	Box F
	Cross-cultural validity	Box G
	Criterion validity	Box H
	Responsiveness	Box I

## Step 2. Determining if the statistical method used in the article are based on CTT or IRT

Box General requirements for studies that applied Item Response Theory (IRT) models		excellent	good	fair	poor
1	Was the IRT model used adequately described? e.g. One Parameter Logistic Model (OPLM), Partial Credit Model (PCM), Graded Response Model (GRM)	IRT model adequately described	IRT model not adequately described		
2	Was the computer software package used adequately described? e.g. RUMM2020, WINSTEPS, OPLM, MULTILOG, PARSCALE, BILOG, NLMIXED	Software package adequately described	Software package not adequately described		
3	Was the method of estimation used adequately described? e.g. conditional maximum likelihood (CML), marginal maximum likelihood (MML)	Method of estimation adequately described	Method of estimation not adequately described		
4	Were the assumptions for estimating parameters of the IRT model checked? e.g. unidimensionality, local independence, and item fit (e.g. differential item functioning (DIF))	assumptions of the IRT model checked	assumptions of the IRT model partly checked	assumptions of the IRT model not checked or unknown	

To obtain a total score for the methodological quality of studies that use IRT methods, the 'worse score counts' algorithm should be applied to the IRT box in combination with the box of the measurement property that was evaluated in the IRT study. For example, if IRT methods are used to study internal consistency and item 4 in the IRT box is scored fair, while the items in the internal consistency box (box A) are all scored as good or excellent, the methodological quality score for internal consistency will be fair. However, if any of the items in box A is scored poor, the methodological quality score for internal consistency will be poor.



**Step 3. Determining if a study meets the standards for good methodological quality**

<b>Box A. Internal consistency</b>		<b>excellent</b>	<b>good</b>	<b>fair</b>	<b>poor</b>
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model? <i>Design requirements</i>	Percentage of missing items described	Percentage of missing items NOT described		
2	Was the percentage of missing items given?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was there a description of how missing items were handled?	Adequate sample size ( $\geq 100$ )	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size ( $< 30$ )
4	Was the sample size included in the internal consistency analysis adequate?	Factor analysis performed in the study population	Authors refer to another study in which factor analysis was performed in a similar study population	Authors refer to another study in which factor analysis was performed, but not in a similar study population	Factor analysis NOT performed and no reference to another study
5	Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	7* #items and $\geq 100$	5* #items and $\geq 100$ OR 6-7* #items but $< 100$	5* #items but $< 100$	$< 5^*$ #items
6	Was the sample size included in the unidimensionality analysis adequate?				

7	Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	Internal consistency statistic calculated for each subscale separately	Internal consistency statistic NOT calculated for each subscale separately
8	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>			
9	for Classical Test Theory (CTT), continuous scores: Was Cronbach's alpha calculated?	Cronbach's alpha calculated	No Cronbach's alpha and no item-total correlations calculated
10	for CTT, dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated
11	for IRT: Was a goodness of fit statistic at a global level calculated? E.g. $\chi^2$ , reliability coefficient of estimated latent trait value (index of (subject or item) separation)	Goodness of fit statistic at a global level calculated	Goodness of fit statistic at a global level NOT calculated

NB. Item 1 is used to determine whether internal consistency is relevant for the instrument under study. It is not used to rate the quality of the study.

**Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)**

		excellent	good	fair	poor
<i>Design requirements</i>					
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size ( $\geq 100$ )	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size ( $< 30$ )
4	Were at least two measurements available?	At least two measurements			Only one measurement
5	Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6	Was the time interval stated?	Time interval stated		Time interval NOT stated	
7	Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8	Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate

9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>					
11	for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
12	for dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
13	for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
14	for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

**Box C. Measurement error: absolute measures**

		excellent	good	fair	poor
<i>Design requirements</i>					
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size ( $\geq 100$ )	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size ( $< 30$ )
4	Were at least two measurements available?	At least two measurements			Only one measurement
5	Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6	Was the time interval stated?	Time interval stated		Time interval NOT stated	
7	Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8	Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate

<p>9 Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions</p> <p>10 Were there any important flaws in the design or methods of the study?</p> <p><i>Statistical methods</i></p> <p>11 for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?</p>	<p>Test conditions were similar (evidence provided)</p> <p>No other important methodological flaws in the design or execution of the study</p> <p>SEM, SDC, or LoA calculated</p>	<p>Assumable that test conditions were similar</p> <p>Other minor methodological flaws in the design or execution of the study</p> <p>Possible to calculate LoA from the data presented</p>	<p>Unclear if test conditions were similar</p> <p>Other important methodological flaws in the design or execution of the study</p> <p>SEM calculated based on Cronbach's alpha, or on SD from another population</p>	<p>Test conditions were NOT similar</p>
---	---	---	--	---

**Box D. Content validity (including face validity)**

	<p><b>excellent</b>                      <b>good</b>                      <b>fair</b>                      <b>poor</b></p>
<p><i>General requirements</i></p> <p>1 Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?</p>	<p>Assessed if all items refer to relevant aspects of the construct to be measured</p> <p>Aspects of the construct to be measured poorly described AND this was not taken into consideration</p> <p>NOT assessed if all items refer to relevant aspects of the construct to be measured</p>

<p>2 Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)</p>	<p>Assessed if all items are relevant for the study population in adequate sample size (<math>\geq 10</math>)</p> <p>Assessed if all items are relevant for the study population in moderate sample size (5-9)</p> <p>Assessed if all items are relevant for the study population in small sample size (&lt;5)</p> <p>NOT assessed if all items are relevant for the study population OR target population not involved</p>
<p>3 Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)</p>	<p>Assessed if all items are relevant for the purpose of the application</p> <p>Purpose of the instrument was not described but assumed</p> <p>NOT assessed if all items are relevant for the purpose of the application</p>
<p>4 Was there an assessment of whether all items together comprehensively reflect the construct to be measured?</p>	<p>Assessed if all items together comprehensively reflect the construct to be measured</p> <p>No theoretical foundation of the construct and this was not taken into consideration</p> <p>NOT assessed if all items together comprehensively reflect the construct to be measured</p>
<p>5 Were there any important flaws in the design or methods of the study?</p>	<p>No other important methodological flaws in the design or execution of the study</p> <p>Other minor methodological flaws in the design or execution of the study</p> <p>Other important methodological flaws in the design or execution of the study</p>

<b>Box E. Structural validity</b>		<b>excellent</b>	<b>good</b>	<b>fair</b>	<b>poor</b>
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model?				
	<i>Design requirements</i>				
2	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
4	Was the sample size included in the analysis adequate?	7* #items and $\geq 100$	5* #items and $\geq 100$ OR 5-7* #items but $< 100$	5* #items but $< 100$	$< 5$ * #items
5	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. rotation method not described)	Other important methodological flaws in the design or execution of the study (e.g. inappropriate rotation method)



<i>Statistical methods</i>	
6 for CTT: Was exploratory or confirmatory factor analysis performed?	<p>Exploratory or confirmatory factor analysis performed and type of factor analysis appropriate in view of existing information</p> <p>Exploratory factor analysis performed while confirmatory would have been more appropriate</p> <p>No exploratory or confirmatory factor analysis performed</p>
7 for IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?	<p>IRT test for determining (uni)dimensionality performed</p> <p>IRT test for determining (uni)dimensionality NOT performed</p>

<b>Box F. Hypotheses testing</b>				
<i>Design requirements</i>				
	<b>excellent</b>	<b>good</b>	<b>fair</b>	<b>Poor</b>
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size ( $\geq 100$ per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size ( $< 30$ per analysis)

<p>4 Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?</p> <p>5 Was the expected <i>direction</i> of correlations or mean differences included in the hypotheses?</p> <p>6 Was the expected absolute or relative <i>magnitude</i> of correlations or mean differences included in the hypotheses?</p> <p>7 for convergent validity: Was an adequate description provided of the comparator instrument(s)?</p> <p>8 for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?</p>	<p>Multiple hypotheses formulated a priori</p> <p>Expected direction of the correlations or differences stated</p> <p>Expected magnitude of the correlations or differences stated</p> <p>Adequate description of the constructs measured by the comparator instrument(s)</p> <p>Adequate measurement properties of the comparator instrument(s) in a population similar to the study population</p>	<p>Minimal number of hypotheses formulated a priori</p> <p>Expected direction of the correlations or differences NOT stated</p> <p>Expected magnitude of the correlations or differences NOT stated</p> <p>Adequate description of most of the constructs measured by the comparator instrument(s)</p> <p>Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population</p>	<p>Hypotheses vague or not formulated but possible to deduce what was expected</p> <p>Poor description of the constructs measured by the comparator instrument(s)</p> <p>Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population</p>	<p>Unclear what was expected</p> <p>NO description of the constructs measured by the comparator instrument(s)</p> <p>No information on the measurement properties of the comparator instrument(s)</p>
---	--	---	---	---

<p>9 Were there any important flaws in the design or methods of the study?</p> <p><i>Statistical methods</i></p> <p>10 Were design and statistical methods adequate for the hypotheses to be tested?</p>	<p>No other important methodological flaws in the design or execution of the study</p> <p>Statistical methods applied appropriate</p> <p>Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented</p> <p>Statistical methods applied NOT optimal</p> <p>Statistical methods applied appropriate</p> <p>Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)</p> <p>Other important methodological flaws in the design or execution of the study</p>
--	--

<b>Box G. Cross-cultural validity</b>					
<i>Design requirements</i>		<b>excellent</b>	<b>good</b>	<b>fair</b>	<b>poor</b>
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	

<p>3 Was the sample size included in the analysis adequate?</p> <p>4 Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?</p> <p>5 Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages</p> <p>6 Did the translators work independently from each other?</p> <p>7 Were items translated forward and backward?</p> <p>8 Was there an adequate description of how differences between the original and translated versions were resolved?</p>	<p>CTT: 7* #items and ≥100 IRT: ≥200 per group</p> <p>Both source language and target language described</p> <p>Expertise of the translators described with respect to disease, construct, and language</p> <p>Translators worked independently</p> <p>Multiple forward and multiple backward translations</p> <p>Adequate description of how differences between translators were resolved</p> <p>CTT: 5* #items and ≥100 OR 5-7* #items but &lt;100 IRT: ≥200 in 1 group and 100-199 in 1 group</p> <p>Expertise of the translators with respect to disease or construct poor or not described</p> <p>Assumable that the translators worked independently</p> <p>Multiple forward translations but one backward translation</p> <p>Poorly or NOT described how differences between translators were resolved</p> <p>CTT: 5* #items but &lt;100 IRT: 100-199 per group</p> <p>Expertise of the translators with respect to language not described</p> <p>Unclear whether translators worked independently</p> <p>One forward and one backward translation</p> <p>CTT: &lt;5* #items IRT: (&lt;100 in 1 or both groups</p> <p>Source language NOT known</p> <p>Only a forward translation</p>
---	---

9	Was the translation reviewed by a committee (e.g. original developers)?	Translation reviewed by a committee (involving other people than the translators, e.g. the original developers)	Translation NOT reviewed by (such) a committee
10	Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	Translated instrument pre-tested in the target population	Translated instrument pre-tested, but unclear if this was done in the target population
11	Was the sample used in the pre-test adequately described?	Sample used in the pre-test adequately described	Sample used in the pre-test NOT (adequately) described
12	Were the samples similar for all characteristics except language and/or cultural background?	Shown that samples were similar for all characteristics except language /culture	Stated (but not shown) that samples were similar for all characteristics except language /culture
13	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study

<i>Statistical methods</i>	
14 for CTT: Was confirmatory factor analysis performed?	Multiple-group confirmatory factor analysis performed
15 for IRT: Was differential item function (DIF) between language groups assessed?	DIF between language groups assessed
	Multiple-group confirmatory factor analysis NOT performed
	DIF between language groups NOT assessed

<b>Box H. Criterion validity</b>	
<i>Design requirements</i>	
	excellent      good      fair      poor
1 Was the percentage of missing items given?	Percentage of missing items described
2 Was there a description of how missing items were handled?	Described how missing items were handled
3 Was the sample size included in the analysis adequate?	Adequate sample size ( $\geq 100$ )
4 Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)
	Percentage of missing items NOT described
	Not described but it can be deduced how missing items were handled
	Good sample size (50-99)
	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'
	Not clear how missing items were handled
	Moderate sample size (30-49)
	Unclear whether the criterion used can be considered an adequate 'gold standard'
	Small sample size (<30)
	Criterion used can NOT be considered an adequate 'gold standard'

5	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>				
6	for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated		Correlations or AUC NOT calculated
7	for dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated		Sensitivity and specificity NOT calculated

<b>Box I. Responsiveness</b>		<b>excellent</b>	<b>good</b>	<b>fair</b>	<b>poor</b>
<i>Design requirements</i>					
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size ( $\geq 100$ )	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4	Was a longitudinal design with at least two measurement used?	Longitudinal design used			No longitudinal design used
5	Was the time interval stated?	Time interval adequately described			Time interval NOT described

<p>6 If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?</p>	<p>Anything that occurred during the interim period (e.g. treatment) adequately described</p> <p>Assumable what occurred during the interim period</p> <p>Unclear or NOT described what occurred during the interim period</p>
<p>7 Was a proportion of the patients changed (i.e. improvement or deterioration)?</p>	<p>Part of the patients were changed (evidence provided)</p> <p>NO evidence provided, but assumable that part of the patients were changed</p> <p>Unclear if part of the patients were changed</p> <p>Patients were NOT changed</p>
<p><b>Design requirements for hypotheses testing</b></p>	
<p>For constructs for which a gold standard was not available:</p>	
<p>8 Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?</p>	<p>Hypotheses formulated a priori</p> <p>Hypotheses vague or not formulated but possible to deduce what was expected</p> <p>Unclear what was expected</p>
<p>9 Was the expected <i>direction</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?</p>	<p>Expected direction of the correlations or differences stated</p> <p>Expected direction of the correlations or differences NOT stated</p>
<p>10 Were the expected absolute or relative <i>magnitude</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?</p>	<p>Expected magnitude of the correlations or differences stated</p> <p>Expected magnitude of the correlations or differences NOT stated</p>



<p>11 Was an adequate description provided of the comparator instrument(s)?</p> <p>12 Were the measurement properties of the comparator instrument(s) adequately described?</p> <p>13 Were there any important flaws in the design or methods of the study?</p> <p><i>Statistical methods</i></p> <p>14 Were design and statistical methods adequate for the hypotheses to be tested?</p>	<p>Adequate description of the constructs measured by the comparator instrument(s)</p> <p>Adequate measurement properties of the comparator instrument(s) in a population similar to the study population</p> <p>No other important methodological flaws in the design or execution of the study</p> <p>Statistical methods applied appropriate</p> <p>Poor description of the constructs measured by the comparator instrument(s)</p> <p>Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population</p> <p>Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population</p> <p>Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)</p> <p>Statistical methods applied NOT optimal</p> <p>Statistical methods applied appropriate</p> <p>NO description of the constructs measured by the comparator instrument(s)</p> <p>NO information on the measurement properties of the comparator instrument(s)</p> <p>Other important methodological flaws in the design or execution of the study</p> <p>Statistical methods applied NOT appropriate</p>
---	---

<b>Design requirement for comparison to a gold standard</b>	
For constructs for which a gold standard was available:	
15 Can the criterion for change be considered as a reasonable gold standard?	<p>Criterion used can be considered an adequate 'gold standard' (evidence provided)</p> <p>No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'</p> <p>Unclear whether the criterion used can be considered an adequate 'gold standard'</p> <p>Criterion used can NOT be considered an adequate 'gold standard'</p>
16 Were there any important flaws in the design or methods of the study?	<p>No other important methodological flaws in the design or execution of the study</p> <p>Other minor methodological flaws in the design or execution of the study</p> <p>Other important methodological flaws in the design or execution of the study</p>
<b>Statistical methods</b>	
17 for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	<p>Correlations or Area under the ROC Curve (AUC) calculated</p> <p>Correlations or AUC NOT calculated</p>
18 for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	<p>Sensitivity and specificity calculated</p> <p>Sensitivity and specificity NOT calculated</p>

### Interpretability

We recommend to use the Interpretability box to extract all information on the interpretability issues described in this box of the instruments under study from the included articles.

<b>Box Interpretability</b>	
Percentage of missing items	
Description of how missing items were handled	
Distribution of the (total) scores	
Percentage of the respondents who had the lowest possible (total) score	
Percentage of the respondents who had the highest possible (total) score	
Scores and change scores (i.e. means and SD) for relevant (sub) groups, e.g. for normative groups, subgroups of patients, or the general population	
Minimal Important Change (MIC) or Minimal Important Difference (MID)	

## Generalizability

We recommend to use the Generalizability box to extract data on the characteristics of the study populations and sampling procedures of the included studies.

<b>Box Generalisability</b>	
Median or mean age (with standard deviation or range)	
Distribution of sex	
Important disease characteristics (e.g. severity, status, duration) and description of treatment	
Setting(s) in which the study was conducted (e.g. general population, primary care or hospital/rehabilitation care)	
Countries in which the study was conducted	
Language in which the HR-PRO instrument was evaluated	
Method used to select patients (e.g. convenience, consecutive, or random)	
Percentage of missing responses (response rate)	

## Vedlegg III

## Kategorisering av måleegenskaper

(Schellingerhout et al., 2011 ; Schellingerhout et al., 2012 ; Terwee et al., 2007)

Måleegenskap	Vurdering	Kvalitetskriterier
<b>Reliabilitet</b>		
Intern konsistens	+	Endimensjonal (sub)skala OG Cronbachs alfa(s) $\geq 0,70$
	?	Ukjent dimensjonalitet ELLER ikke fastsatt Cronbachs alfa
	-	Ikke endimensjonal (sub)skala ELLER Cronbachs alfa (s) $< 0,70$
Reliabilitet	+	ICC/vektet kappa $\geq 0,70$ ELLER Pearsons $r \geq 0,80$
	?	Hverken ICC/vektet kappa eller Pearsons $r$ fastsatt
	-	ICC/vektet kappa $< 0,70$ ELLER Pearsons $r < 0,80$
<b>Validitet</b>		
Innholdsvaliditet	+	Målgruppen vurderer alle spørsmålene i spørreskjemaet som relevant
	?	Ingen målgruppe involvering
	-	Målgruppen vurderer testledd i spørreskjemaet til å være irrelevant ELLER vurderer spørreskjemaet til å være ufullstendig
Begrepsvaliditet		
Tverrkulturell validitet	+	Opprinnelig faktorstruktur bekreftet ELLER ingen betydningsfull DIF
	?	Bekreftelse av opprinnelig faktorstruktur OG DIF ikke omtalt
	-	Opprinnelig faktorstruktur ikke bekreftet ELLER betydningsfull DIF
Strukturell validitet	+	Faktorene bør forklare minst 50 % av variansen
	?	Forklart varians ikke nevnt
	-	Faktorene forklarer $< 50$ % av variansen
Hypotesetesting	+	Samvariasjon med en test som måler samme fenomen $\geq 0,50$ ELLER minst 75 % av resultatene er i samsvar med hypotesene OG samvariasjon med relaterte fenomen er høyere enn med ikke-relaterte fenomen
	?	Samvariasjon utelukkende fastsatt med ikke-korrelerte fenomen
	-	Samvariasjon $< 0,50$ med en test som måler samme fenomen ELLER lavere samvariasjon med relaterte fenomen enn med ikke-relaterte fenomen

ICC (intraklasse korrelasjonskoeffisient); DIF (differential item functioning)

+ (positive måleegenskaper); ? (uavklarte måleegenskaper); - (negative måleegenskaper)

**Vedlegg IV****Ekskluderte publikasjoner med begrunnelse**

Referanse	Eksklusjonsgrunn
An et al. (2010)	ikke et PROM
Bausewein et al. (2007)	review
Bausewein, Booth and Higginson (2008)	review
Bergman et al. (1994)	ikke rette dimensjoner
Blinderman et al. (2009)	ikke et PROM
Brabo et al. (2006)	ikke rette dimensjoner
Campbell (2008)	ikke et PROM
Centeno et al. (2004)	ikke rette dimensjoner
Chan, Richardson and Richardson (2005)	ikke rette dimensjoner
Chang, Hwang and Feuerman (2000)	ikke rette dimensjoner
Chen et al. (2007)	ikke rette dimensjoner
Chiang et al. (2011)	ikke rette dimensjoner
Chie et al. (2004)	ikke rette dimensjoner
Choi et al. (2011)	ikke rette dimensjoner
Cullen and Rodak (2002)	ekspert review
Damato et al. (2005)	ikke rette dimensjoner
Darbee and Ohtake (2006)	ikke rette dimensjoner
Dorman, Byrne and Edwards (2007)	review
Eakin et al. (1996)	ikke rette dimensjoner
Eakin et al. (1998)	ikke rette dimensjoner
Eakin, Kaplan and Ries (1993)	ikke mulig å oppdrive
Engstrom et al. (2001)	Ikke rette dimensjoner
Fiocchetti et al. (2006)	ikke et PROM
Friis et al. (2012)	ikke et PROM
Fung et al. (2012)	ikke rette dimensjoner
Gabrielii, Vidotto and Bertolotti (2011)	ikke rette dimensjoner
Gaguski et al. (2010)	ikke rette dimensjoner
Haave, Hyland and Engvik (2005)	ikke rette dimensjoner
Hajiro et al. (1998)	ikke rette dimensjoner
Howard et al. (2012)	ikke rette dimensjoner
Hutter and Wurtemberger (1997a)	ikke rette dimensjoner
Hutter and Wurtemberger (1997b)	ikke rette dimensjoner
Hyland et al. (1998)	ikke rette dimensjoner
Jablonski, Gift and Cook (2007)	ikke rette dimensjoner
Jacobs et al. (2004)	ikke rette dimensjoner
Jones et al. (2012)	review
Jones et al. (2014)	ikke rette dimensjoner
Kamal et al. (2011)	review
Karabulu et al. (2010)	ikke et PROM
Kovelis et al. (2008)	ikke rette dimensjoner

Kovelis et al. (2011)	ikke rette dimensjoner
Kuhl et al. (2011)	ikke rette dimensjoner
Lansing, Gracely and Banzett (2009)	Expert review
Lareau et al. (1994)	ikke rette dimensjoner
Lareau, Meek and Roos (1998)	ikke rette dimensjoner
Lopez-Campos et al. (2006)	artikkel på spansk
Lyng et al. (2000)	ikke valideringsstudie
Mahler et al. (2009)	ikke rette dimensjoner
Mazur et al. (2011)	generisk verktøy
Meek et al. (2012)	ikke rette dimensjoner
Miravittles et al. (2009)	ikke rette dimensjoner
Monteiro Dda, Almeida Mde and Kruse (2013)	ikke rette dimensjoner
Moro et al. (2006)	ikke rette dimensjoner
Ninot, Soyez and Prefaut (2013)	ikke rette dimensjoner
Nishimura et al. (2008)	ikke rette dimensjoner
Nowak, Stockler and Byrne (2004)	ikke rette dimensjoner
Opasich et al. (2008)	ikke rette dimensjoner
Parshall (2002)	ikke rette dimensjoner
Perez et al. (2009)	ikke rette dimensjoner
Pommer et al. (2013)	ikke rette dimensjoner
Puhan et al. (2007)	ikke rette dimensjoner
Reddy et al. (2009)	ikke rette dimensjoner
Ries (2005)	ikke rette dimensjoner
Ries (2006)	review
Schunemann et al. (2003)	ikke rette dimensjoner
Stavem, Boe and Erikssen (1999)	ikke rette dimensjoner
Stiel et al. (2010)	ikke rette dimensjoner
Suarez-del-Real et al. (2011)	ikke rette dimensjoner
Tanaka et al. (2002)	ikke relevant referanse standard
Tsang et al. (2001)	ikke mulig å oppdrive
van der Molen (1995)	review
Wang et al. (2013)	Treningsutløst dyspne
Watkins et al. (2013)	ikke rette dimensjoner
Wilcock et al. (1999)	ikke rette dimensjoner
Wilcox et al. (2013)	ikke rette dimensjoner
Williams et al. (2007)	ikke et PROM
Wingardh, Engstrom and Claesson (2007)	ikke rette dimensjoner
Yorke et al. (2010)	ikke rette dimensjoner
Yorke, Horton and Jones (2012)	ikke rette dimensjoner
Yount et al. (2011)	ikke rette dimensjoner
Zeng et al. (2013)	ikke rette dimensjoner

An, L., Lin, Y. X., Yang, T., Zhang, H., Jiao, X., Zhang, S., Chang, X. H., Wang, Z. M. & Wang, C. (2010) Predictive validity of BODE index for anxious and depressive symptoms in patients with chronic obstructive pulmonary disease. *Chinese Medical Journal*, 123 (14), s. 1845-51.

Bausewein, C., Booth, S. & Higginson, I. J. (2008) Measurement of dyspnoea in the clinical rather than the research setting. *Current Opinion in Supportive & Palliative Care*, 2 (2), s. 95-9.

Bausewein, C., Farquhar, M., Booth, S., Gysels, M. & Higginson, I. J. (2007) Measurement of breathlessness in advanced disease: a systematic review. *Respiratory Medicine*, 101 (3), s. 399-410.

Bergman, B., Aaronson, N. K., Ahmedzai, S., Kaasa, S. & Sullivan, M. (1994) The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. *European Journal of Cancer*, 30A (5), s. 635-42.

Blinderman, C. D., Homel, P., Billings, J., Tennstedt, S. & Portenoy, R. K. (2009) Symptom distress and quality of life in patients with advanced chronic obstructive pulmonary disease. *Journal of Pain and Symptom Management*, 38 (1), s. 115-123.

Brabo, E. P., Paschoal, M. E., Biasoli, I., Nogueira, F. E., Gomes, M. C., Gomes, I. P., Martins, L. C. & Spector, N. (2006) Brazilian version of the QLQ-LC13 lung cancer module of the European Organization for Research and Treatment of Cancer: preliminary reliability and validity report. *Quality of Life Research*, 15 (9), s. 1519-24.

Campbell, M. (2008) Psychometric Testing of a Respiratory Distress Observation Scale. *Journal of Palliative Medicine*, 11 (1), s. 44-50.

Centeno, C., Noguera, A., Lopez, B. & Carvajal, A. (2004) Some tools of assessment used in palliative care (I): The Edmonton System Assessment Symptoms (ESAS). [Spanish]

Algunos instrumentos de evaluacion utilizados en cuidados paliativos (I): El cuestionario de Evaluacion de Sintomas de Edmonton (ESAS). *Medicina Paliativa*, 11 (4), s. 239-245.

Chan, C. W., Richardson, A. & Richardson, J. (2005) A study to assess the existence of the symptom cluster of breathlessness, fatigue and anxiety in patients with advanced lung cancer. *European Journal of Oncology Nursing*, 9 (4), s. 325-33.

Chang, V. T., Hwang, S. S. & Feuerman, M. (2000) Validation of the Edmonton Symptom Assessment Scale. *Cancer*, 88 (9), s. 2164-71.



- Chen, L., Antras, L., Duh, M. S., Levy, N., Neary, M., O'Brien, M. E. & von Pawel, J. (2007) Psychometric validation of the Patient Symptom Assessment in Lung Cancer instrument for small cell lung cancer. *Current Medical Research & Opinion*, 23 (11), s. 2741-52.
- Chiang, L. K., Ng, L. V., Fung, L., Tang, R. & Tam, W. W. S. (2011) Translation and validation of the COPD self-efficacy scale (CSES) into Chinese version (CSES-Chi). *Hong Kong Practitioner*, 33 (4), s. 139-145.
- Chie, W.-C., Yang, C.-H., Hsu, C. & Yang, P.-C. (2004) Quality of life of lung cancer patients: Validation of the Taiwan Chinese version of the EORTC QLQ-C30 and QLQ-LC13. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*, 13 (1), s. 257-262.
- Choi, S. W., Victorson, D. E., Yount, S., Anton, S. & Cella, D. (2011) Development of a conceptual framework and calibrated item banks to measure patient-reported dyspnea severity and related functional limitations. *Value in Health*, 14 (2), s. 291-306.
- Cullen, D. L. & Rodak, B. (2002) Clinical utility of measures of breathlessness. *Respiratory Care*, 47 (9), s. 986-93.
- Damato, S., Bonatti, C., Frigo, V., Pappagallo, S., Raccanelli, R., Rampoldi, C. & Rodi, F. (2005) Validation of the Clinical COPD questionnaire in Italian language. *Health & Quality of Life Outcomes*, 3, s. 9.
- Darbee, J. C. & Ohtake, P. J. (2006) Research corner. Outcome measures in cardiopulmonary physical therapy: Medical Research Council (MRC) Dyspnea Scale. *Cardiopulmonary Physical Therapy Journal*, 17 (1), s. 29-37.
- Dorman, S., Byrne, A. & Edwards, A. (2007) Which measurement scales should we use to measure breathlessness in palliative care? A systematic review. *Palliative Medicine*, 21 (3), s. 177-91.
- Eakin, E. G. (1995) Clinical assessment of shortness of breath in chronic obstructive pulmonary disease. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 55 (9-B), s. 4165.
- Eakin, E. G., Kaplan, R. M. & Ries, A. L. (1993) Measurement of dyspnoea in chronic obstructive pulmonary disease. *Quality of Life Research*, 2 (3), s. 181-91.
- Eakin, E. G., Kaplan, R. M., Ries, A. L. & Sassi-Dambron, D. E. (1996) Patients' self-reports of dyspnea: An important and independent outcome in chronic obstructive pulmonary disease. *Annals of Behavioral Medicine*, 18 (2), s. 87-90.

Eakin, E. G., Resnikoff, P. M., Prewitt, L. M., Ries, A. L. & Kaplan, R. M. (1998) Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest*, 113 (3), s. 619-24.

Engstrom, C. P., Persson, L. O., Larsson, S. & Sullivan, M. (2001) Health-related quality of life in COPD: why both disease-specific and generic measures should be used. *European Respiratory Journal*, 18 (1), s. 69-76.

Fiocchetti, L., Bastianelli, P., Civilotti, R., Rossi, M., Sperduti, I., Nelli, F., Moschetti, L. & Pollera, C. F. (2006) The nurse Symptoms Assessment Chart allows a better evaluation of cancer patients needs [Italian]. *International Nursing Perspectives*, 6 (1), s. 7-12.

Friis, S., Kvåle, A., Opjordsmoen, S. & Bunkan, B. H. (2012) The Global Body Examination (GBE). A useful instrument for evaluation of respiration. *Advances in Physiotherapy*, 14 (4), s. 146-54.

Fung, A., Chan, L. L. C., So, C. T., Chau, S. S. L., Chan, T. M., Chan, C., Chu, A. W. Y., Ng, B. H. P., Cheung, B. Y. H., Chan, A. K. K., Wong, W. K. W., Chu, C. W. H. & Fong, K. N. K. (2012) Reliability and validity of the self-administered Chinese version of the shortness of breath questionnaire (C-SOBQ) in patients with chronic obstructive pulmonary disease. *Hong Kong Journal of Occupational Therapy*, 22 (2), s. 75-83.

Gabrieli, E., Vidotto, G. & Bertolotti, G. (2011) [A comparison of three health-related quality of life measures in COPD patients]. *Giornale Italiano di Medicina del Lavoro Ed Ergonomia*, 33 (3 Suppl B), s. B47-54.

Gaguski, M. E., Brandsema, M., Gernalin, L. & Martinez, E. (2010) Assessing dyspnea in patients with non-small cell lung cancer in the acute care setting. *Clinical Journal of Oncology Nursing*, 14 (4), s. 509-13.

Haave, E., Hyland, M. E. & Engvik, H. (2005) Physical and emotional aspects of self-reported health status: a two-factor model of the short-form Breathing Problems Questionnaire. *Chronic Respiratory Disease*, 2 (1), s. 21-6.

Hajiro, T., Nishimura, K., Tsukino, M., Ikeda, A., Koyama, H. & Izumi, T. (1998) Analysis of clinical methods used to evaluate dyspnea in patients with chronic obstructive pulmonary disease. *American Journal of Respiratory & Critical Care Medicine*, 158 (4), s. 1185-9.

Howard, K., Berry, P., Petrillo, J., Wiklund, I., Roberts, L., Watkins, M., Crim, C. & Wilcox, T. (2012) Development of the Shortness of Breath with Daily Activities questionnaire (SOBDA). *Value in Health*, 15 (8), s. 1042-50.

Hutter, B. & Wurtemberger, G. (1997a) Reliability and validity of the German version of the Sickness Impact Profile in patients with chronic obstructive pulmonary disease. *Psychology & Health*, 12 (2), s. 149-159.

Hutter, B. O. & Wurtemberger, G. (1997b) [Quality of life in patients with chronic obstructive lung diseases: psychometric criteria of the German version of the Sickness Impact Profile and initial results of its application]. *Pneumologie*, 51 (2), s. 108-14.

Hyland, M. E., Singh, S. J., Sodergren, S. C. & Morgan, M. P. (1998) Development of a shortened version of the Breathing Problems Questionnaire suitable for use in a pulmonary rehabilitation clinic: a purpose-specific, disease-specific questionnaire. *Quality of Life Research*, 7 (3), s. 227-33.

Jablonski, A., Gift, A. & Cook, K. E. (2007) Symptom assessment of patients with chronic obstructive pulmonary disease. *Western Journal of Nursing Research*, 29 (7), s. 845-63.

Jacobs, J. E., Maille, A. R., Akkermans, R. P., van Weel, C. & Grol, R. P. (2004) Assessing the quality of life of adults with chronic respiratory diseases in routine primary care: construction and first validation of the 10-Item Respiratory Illness Questionnaire-monitoring 10 (RIQ-MON10). *Quality of Life Research*, 13 (6), s. 1117-27.

Jones, P., Miravittles, M., van der Molen, T. & Kulich, K. (2012) Beyond FEV1 in COPD: a review of patient-reported outcomes and their measurement. *International Journal of Copd*, 7, s. 697-709.

Jones, P. W., Beeh, K. M., Chapman, K. R., Decramer, M., Mahler, D. A. & Wedzicha, J. A. (2014) Minimal clinically important differences in pharmacological trials. *American Journal of Respiratory & Critical Care Medicine*, 189 (3), s. 250-5.

Kamal, A. H., Maguire, J. M., Wheeler, J. L., Currow, D. C. & Abernethy, A. P. (2011) Dyspnea review for the palliative care professional: assessment, burdens, and etiologies. *Journal of Palliative Medicine*, 14 (10), s. 1167-72.

Karabulu, N., Erci, B., Ozer, N. & Ozdemir, S. (2010) Symptom clusters and experiences of patients with cancer. *Journal of Advanced Nursing*, 66 (5), s. 1011-1021.

Kovelis, D., Segretti, N. O., Probst, V. S., Lareau, S. C., Brunetto, A. F. & Pitta, F. (2008) Validation of the Modified Pulmonary Functional Status and Dyspnea Questionnaire and the Medical Research Council scale for use in Brazilian patients with chronic obstructive pulmonary disease. *Jornal Brasileiro De Pneumologia: Publicacao Oficial Da Sociedade Brasileira De Pneumologia E Tisiologia*, 34 (12), s. 1008-18.

Kovelis, D., Zabatiero, J., Oldemberg, N., Colange, A. L., Barzon, D., Nascimento, C. H., Probst, V. S. & Pitta, F. (2011) Responsiveness of three instruments to assess self-reported

functional status in patients with COPD. *Copd: Journal of Chronic Obstructive Pulmonary Disease*, 8 (5), s. 334-9.

Kuhl, K., Kuhn, C., Kenn, K. & Rief, W. (2011) [The COPD-Anxiety-Questionnaire (CAF): a new instrument to assess illness specific anxiety in COPD patients]. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 61 (1), s. e1-9.

Lansing, R. W., Gracely, R. H. & Banzett, R. B. (2009) The multiple dimensions of dyspnea: Review and hypotheses. *Respiratory Physiology & Neurobiology*, 167 (1), s. 53-60.

Lareau, S. C., Carrieri-Kohlman, V., Janson-Bjerklie, S. & Roos, P. J. (1994) Development and testing of the Pulmonary Functional Status and Dyspnea Questionnaire (PFSDQ). *Heart & Lung*, 23 (3), s. 242-50.

Lareau, S. C., Meek, P. M. & Roos, P. J. (1998) Development and testing of the modified version of the pulmonary functional status and dyspnea questionnaire (PFSDQ-M). *Heart & Lung*, 27 (3), s. 159-68.

Lopez-Campos, J. L., Failde, I., Leon Jimenez, A., Masa Jimenez, F., Barrot Cortes, E., Benitez Moya, J. M., Ayerbe Garcia, R. & Windisch, W. (2006) [Health-related quality of life of patients receiving home mechanical ventilation: the Spanish version of the severe respiratory insufficiency questionnaire]. *Arch Bronconeumol*, 42 (11), s. 588-93.

Lyng, S. A., Suadicani, P., Guldager, B., Ishöy, T., Appleyard, M. & Gyntelberg, F. (2000) Selvrappporteret dyspnø er stærkt associeret med psykiske og kognitive symptomer. The Danish Gulf War Study. *Ugeskrift for Læger*, 162 (49), s. 6687-91.

Mahler, D. A., Ward, J., Waterman, L. A., McCusker, C., Zuwallack, R. & Baird, J. C. (2009) Patient-reported dyspnea in COPD reliability and association with stage of disease. *Chest*, 136 (6), s. 1473-9.

Mazur, W., Kupiainen, H., Pitkaniemi, J., Kilpelainen, M., Sintonen, H., Lindqvist, A., Kinnula, V. L. & Laitinen, T. (2011) Comparison between the disease-specific Airways Questionnaire 20 and the generic 15D instruments in COPD. *Health & Quality of Life Outcomes*, 9, s. 4.

Meek, P. M., Banzett, R., Parthall, M. B., Gracely, R. H., Schwartzstein, R. M. & Lansing, R. (2012) Reliability and Validity of the Multidimensional Dyspnea Profile. *Chest*, 141 (6), s. 1546-1553.

Miravittles, M., Llor, C., de Castellar, R., Izquierdo, I., Baro, E. & Donado, E. (2009) Validation of the COPD severity score for use in primary care: the NEREA study. *European Respiratory Journal*, 33 (3), s. 519-27.

Monteiro Dda, R., Almeida Mde, A. & Kruse, M. H. (2013) [Translation and cross-cultural adaptation of the Edmonton Symptom Assessment System for use in Palliative Care]. *Revista Gaucha de Enfermagem*, 34 (2), s. 163-71.

Moro, C., Brunelli, C., Miccinesi, G., Fallai, M., Morino, P., Piazza, M., Labianca, R. & Ripamonti, C. (2006) Edmonton symptom assessment scale: Italian validation in two palliative care settings. *Supportive Care in Cancer*, 14 (1), s. 30-7.

Ninot, G., Soyeze, F. & Prefaut, C. (2013) A short questionnaire for the assessment of quality of life in patients with chronic obstructive pulmonary disease: psychometric properties of VQ11. *Health & Quality of Life Outcomes*, 11, s. 179.

Nishimura, K., Oga, T., Ikeda, A., Hajiro, T., Tsukino, M. & Koyama, H. (2008) Comparison of health-related quality of life measurements using a single value in patients with asthma and chronic obstructive pulmonary disease. *Journal of Asthma*, 45 (7), s. 615-620.

Nowak, A. K., Stockler, M. R. & Byrne, M. J. (2004) Assessing quality of life during chemotherapy for pleural mesothelioma: feasibility, validity, and results of using the European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire and Lung Cancer Module. *Journal of Clinical Oncology*, 22 (15), s. 3172-80.

Opasich, C., Gualco, A., De Feo, S., Barbieri, M., Cioffi, G., Giardini, A. & Majani, G. (2008) Physical and emotional symptom burden of patients with end-stage heart failure: what to measure, how and why. *Journal of Cardiovascular Medicine*, 9 (11), s. 1104-8.

Parshall, M. B. (2002) Psychometric characteristics of dyspnea descriptor ratings in emergency department patients with exacerbated chronic obstructive pulmonary disease. *Research in Nursing & Health*, 25 (5), s. 331-44.

Perez, T., Arnould, B., Grosbois, J. M., Bosch, V., Guillemin, I., Bravo, M. L., Brun, M., Tonnel, A. B. & Group, T. S. (2009) Validity, reliability, and responsiveness of a new short Visual Simplified Respiratory Questionnaire (VSRQ) for health-related quality of life assessment in chronic obstructive pulmonary disease. *International Journal of Copd*, 4, s. 9-18.

Pommer, A. M., Prins, L., van Ranst, D., Meijer, J. W., Van't Hul, A., Janssen, J., Pouwer, F. & Pop, V. J. M. (2013) Development and validity of the Patient-centred COPD Questionnaire (PCQ). *Journal of Psychosomatic Research*, 75 (6), s. 563-571.

Puhan, M. A., Guyatt, G. H., Goldstein, R., Mador, J., McKim, D., Stahl, E., Griffith, L. & Schunemann, H. J. (2007) Relative responsiveness of the Chronic Respiratory Questionnaire, St. Georges Respiratory Questionnaire and four other health-related quality of life instruments for patients with chronic lung disease. *Respiratory Medicine*, 101 (2), s. 308-16.

Reddy, S. K., Parsons, H. A., Elsayem, A., Palmer, J. L. & Bruera, E. (2009) Characteristics and Correlates of Dyspnea in Patients with Advanced Cancer. *Journal of Palliative Medicine*, 12 (1), s. 29-36.

Ries, A. L. (2005) Minimally clinically important difference for the UCSD Shortness of Breath Questionnaire, Borg Scale, and Visual Analog Scale. *Copd: Journal of Chronic Obstructive Pulmonary Disease*, 2 (1), s. 105-10.

Ries, A. L. (2006) Impact of chronic obstructive pulmonary disease on quality of life: the role of dyspnea. *American Journal of Medicine*, 119 (10 Suppl 1), s. 12-20.

Schunemann, H. J., Griffith, L., Stubbing, D., Goldstein, R. & Guyatt, G. H. (2003) A clinical trial to evaluate the measurement properties of 2 direct preference instruments administered with and without hypothetical marker states. *Medical Decision Making*, 23 (2), s. 140-9.

Stavem, K., Boe, J. & Erikssen, J. (1999) Health status, dyspnea, lung function and exercise capacity in patients with chronic obstructive pulmonary disease. *International Journal of Tuberculosis & Lung Disease*, 3 (10), s. 920-6.

Stiel, S., Matthes, M. E., Bertram, L., Ostgathe, C., Elsner, F. & Radbruch, L. (2010) [Validation of the new version of the minimal documentation system (MIDOS) for patients in palliative care : the German version of the edmonton symptom assessment scale (ESAS)]. *Der Schmerz*, 24 (6), s. 596-604.

Suarez-del-Real, Y., Allende-Perez, S., Alferez-Mancera, A., Rodriguez, R. B., Jimenez-Toxtle, S., Mohar, A. & Onate-Ocana, L. F. (2011) Validation of the Mexican-Spanish version of the EORTC QLQ-C15-PAL questionnaire for the evaluation of health-related quality of life in patients on palliative care. *Psycho-Oncology*, 20 (8), s. 889-96.

Tanaka, K., Akechi, T., Okuyama, T., Nishiwaki, Y. & Uchitomi, Y. (2002) Prevalence and screening of dyspnea interfering with daily life activities in ambulatory patients with advanced lung cancer. *Journal of Pain & Symptom Management*, 23 (6), s. 484-9.

Tsang, A. H., Carrieri Kohlman, V., Janson, S., Gold, W., Kwiatkowski, M. & Stulbarg, M. (2001) Reliability and validity testing of the self-efficacy and dyspnea strategies scale (SEADS) in patients with COPD. *American Journal of Respiratory and Critical Care Medicine* [Internet], 163 (5 Suppl), s. A813. Tilgjengelig fra:  
<<http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/746/CN-00394746/frame.html>>

van der Molen, B. (1995) Dyspnoea: a study of measurement instruments for the assessment of dyspnoea and their application for patients with advanced cancer. *Journal of Advanced Nursing*, 22 (5), s. 948-56.

Wang, H. Y., Xu, Q. F., Yuan, W., Nie, S., He, X., Zhang, J., Kong, Y. Y., Sperry, A. & Guan, D. (2013) A new method for rating dyspnea during exercise in patients with chronic obstructive pulmonary disease. *Chinese Medical Journal*, 126 (19), s. 3616-20.

Watkins, M. L., Wilcox, T. K., Tabberer, M., Brooks, J. M., Donohue, J. F., Anzueto, A., Chen, W. H. & Crim, C. (2013) Shortness of Breath with Daily Activities questionnaire: validation and responder thresholds in patients with chronic obstructive pulmonary disease. *BMJ Open*, 3 (10), s. e003048.

Wilcock, A., Crosby, V., Clarke, D. & Tattersfield, A. (1999) Repeatability of breathlessness measurements in cancer patients. *Thorax*, 54 (4), s. 375.

Wilcox, T. K., Chen, W. H., Howard, K. A., Wiklund, I., Brooks, J., Watkins, M. L., Cates, C. E., Tabberer, M. M. & Crim, C. (2013) Item selection, reliability and validity of the Shortness of Breath with Daily Activities (SOBDA) questionnaire: a new outcome measure for evaluating dyspnea in chronic obstructive pulmonary disease. *Health & Quality of Life Outcomes*, 11, s. 196.

Williams, M., Sorich, M., Cafarella, P. & Petkov, J. (2007) Thirty-count breathlessness score: reliability, sensitivity, specificity and validity. *Respirology*, 12 (5), s. 724-31.

Wingardh, A., Engstrom, C. P. & Claesson, L. (2007) Test-retest of a Swedish version of the Pulmonary Functional Status & Dyspnea Questionnaire-modified. *Scandinavian Journal of Occupational Therapy*, 14 (3), s. 183-91.

Yorke, J., Horton, M. & Jones, P. W. (2012) A critique of Rasch analysis using the Dyspnoea-12 as an illustrative example. *Journal of Advanced Nursing*, 68 (1), s. 191-8.

Yorke, J., Moosavi, S. H., Shuldham, C. & Jones, P. W. (2010) Quantification of dyspnoea using descriptors: development and initial testing of the Dyspnoea-12. *Thorax*, 65 (1), s. 21-6.

Yount, S. E., Choi, S. W., Victorson, D., Ruo, B., Cella, D., Anton, S. & Hamilton, A. (2011) Brief, valid measures of dyspnea and related functional limitations in chronic obstructive pulmonary disease (COPD). *Value in Health*, 14 (2), s. 307-15.

Zeng, L., Bedard, G., Cella, D., Thavarajah, N., Chen, E., Zhang, L., Bennett, M., Peckham, K., De Costa, S., Beaumont, J. L., Tsao, M., Danjoux, C., Barnes, E., Sahgal, A. & Chow, E. (2013) Preliminary results of the generation of a shortened quality-of-life assessment for patients with advanced cancer: the FACIT-Pal-14. *Journal of Palliative Medicine*, 16 (5), s. 509-15.

## Vedlegg V Dataekstraksjonsskjema for inkluderte publikasjoner

Bokser som ikke er besvart eller ikke er relevant er fjernet fra skjemaet av plasshensyn

Tanaka, K., Akechi, T., Okuyama, T., Nishiwaki, Y. & Uchitomi, Y. (2000) <b>Development and validation of the Cancer Dyspnoea Scale: a multidimensional, brief, self-rating scale.</b> British Journal of Cancer, 82 (4), s. 800-5.	
Dataekstraksjon utført av Kjersti Solvåg 29.09.14	Referanse nr: 327
<b>Kjennetegn ved studien</b>	
Formål	Å utvikle en kortfattet egenvurderingsskala for å vurdere pustebesvær hos pasienter med kreft, samt å validere skalaen.
Studiedesign	Utviklings- og valideringsstudie
Land	Japan
Finansiering	Uklart om det var gitt noen økonomisk støtte, men: «This work was supported in part by a Grant-in-Aid for Cancer research (9-31) and the Second Term Comprehensive 10-Year Strategy for Cancer Control of the Ministry of Health and Welfare»
<b>Kjennetegn ved kartleggingsverktøyet</b>	
Navn (for kortelse)	Cancer Dyspnoea Scale (CDS)
Begrepet (beskrivelse):	Multidimensionale aspekt ved pustebesvær
Konseptuelt rammeverk	Refleksiv
Opprinnelig er utviklet for	Pasienter med kreft
Antall spørsmål	12
Dimensjoner	3
	Sense of effort 5 spørsmål
	Sense of anxiety 4 spørsmål
	Sense of discomfort 3 spørsmål
Tidsreferanse for skår	De siste dagene
Måleskala	5-punkts Likert-skala fra 1 (not at all) til 5 (very much)
Algoritme for skår	Legg sammen skår for hver faktor: Faktor 1= (spm 4+6+8+10+12)·5 = sense of effort Faktor 2= (spm 5+7+9+11) · 4 = sense of anxiety Faktor 3= 15 - (spm 1+2+3) = sense of discomfort Legg sammen totalskårene for hver faktor = total dyspnoe
Administrasjonsform	Selv-administrert
Gjennomføring	Gjennomsnittlig tidsforbruk ved utfylling: 140 s (s.d.=44,1, median = 138)
Tilgjengelighet	Engelsk versjon tilgjengelig i vedlegg til artikkelen
<b>Kjennetegn ved studiens utvalg</b>	
Antall	Valideringsfasen: n= 166, test-retest: n=37
Alder	Median 64, range 27-87
Kjønn (%)	Menn n=123 (74,1 %)

Sykdomskarakteristika	Avansert eller residiverende lungekreft. Mer detaljert informasjon tilgjengelig i artikkelens Tabell 1.	
Performance status: ECOG	0	31 (18,7 %)
	1	125 (75,3 %)
	2	4 (2,4 %)
	3	4 (2,4 %)
	4	2 (1,2 %)
Setting	135 (81,3 %) polikliniske pasienter, 31 inneliggende	
Språkversjon	Japansk	
Metode for utvalgelse	Fortløpende: polikliniske, tversnitt: inneliggende	
Andel manglende respons	Fem pasienter lot ett spørsmål stå ubesvart, tre pasienter markerte doble svar på ett spørsmål (totalt åtte pasienter: 4,8 %)	
<b>COSMIN</b>	<b>Kritisk vurdering</b>	<b>Resultat/kommentar</b>
Statistiske metoder benyttet	CTT	
basert på CTT eller IRT		
Internal consistency	Excellent	(n= 166) Cronbach's alpha: Sense of effort 0,83 Sense of anxiety 0,81 Sense of discomfort 0,94 Total skår 0,64
Reliability	Fair	Test-retest (n = 37) Gjennomsnitt 6,9 dager (median = 7 dager) mellom de to målepunktene. Pearson's correlation coefficient: Sense of effort 0,71 Sense of anxiety 0,69 Sense of discomfort 0,58 Total skår 0,69 P < 0,001 for alle
Content validity	Excellent	179 begrepsuttrykk for å beskrive, som representerer og for å evaluere pustebesvær ble samlet inn ved intervju med 20 pasienter med pustebesvær, brainstorming blant 23 medisinske eksperter og fra publikasjoner om pustebesvær. Deretter ble testledd som ikke tilfredsstilte bestemte krav fjernet. Et preliminært spørreskjema på 24 testledd ble administrert til 117 pasienter med kreft. 6 testledd ble deretter fjernet etter bestemte kriterier. Faktoranalyse på 3 fikserte faktorer ble gjennomført og ytterligere 6 testledd ble fjernet.
Structural validity	Excellent	Utviklingsfasen:



Hypotheses testing	Good Hypoteser vart formulert og forventet størrelse og reining på samvariasjoner ikke oppgitt. Noe mangelfull informasjon om instrumentene til sammenligning er validert for populasjonen i studien.	Principal component analysis (PCA) with varimax rotation. Antall faktorer ble fastsatt ved Scree-test og testledd som ladet mindre enn 0,65 ble fjernet fra hver subskala. 3 faktorer: Sense of effort: 5 spørsmål, 27% av variansen Sense of anxiety: 4 spørsmål, 21% av variansen Sense of discomfort: 3 spørsmål, 14% av variansen <u>Valideringsfasen:</u> Faktoranalyse etterfulgt av varimax rotasjon.
Cross-cultural validity	Ikke rapportert	Engelsk versjon validert på et senere tidspunkt (Uronis et al., 2012)
<b>Notater</b>		
Det er mulig at Tanaka et al. (2000) og Tanaka et al. (2002) har delvis overlappende utvalg. Dette har ikke praktiske konsekvenser da de har sammenlignet med ulike verktøy.		
<b>Notater</b>		
Tanaka, K., Akechi, T., Okuyama, T., Nishiwaki, Y. & Uchitomi, Y. (2002) <b>Factors correlated with dyspnea in advanced lung cancer patients: organic causes and what else?</b> <i>Journal of Pain &amp; Symptom Management</i> , 23 (6), s. 490-500.		
Dataekstraksjon utført av Kjersti Solvåg 29.09.14	Referanse nr. 545	
<b>Kjennetegn ved studien</b>		
Formål	A identifisere faktorer som korrelerer med pustebesvær hos pasienter med kreft blant et bredt spekter av medisinske, psykologiske og sosiodemografiske faktorer.	
Studiedesign	Valideringsstudie	
Land	Japan	
Finansiering	Uklart om studien har fått økonomisk støtte, men: «This work was supported in part by a Grant-in-Aid for Cancer research (9-31) and the Second Term Comprehensive 10-Year Strategy for Cancer Control of the Ministry of Health and Welfare, Japan»	
<b>Kjennetegn ved kartleggingsverktøyet</b>	<b>Beskrivelse</b>	
Navn (for kortelse)	Cancer Dyspnea Scale (CDS)	
Begrepet (beskrivelse):	Multidimensjonale aspekt ved pustebesvær	
Konseptuelt rammeverk	Refleksiv	

Opprinnelig er utviklet for	Pasienter med kreft
Antall spørsmål	12
Dimensjoner	3 Sense of effort: 5 spørsmål Sense of anxiety: 4 spørsmål Sense of discomfort: 3 spørsmål
Måleskala	5-punkts Likert-skala fra 1 (not at all) til 5 (very much)
Algoritme for skår	Maksimal skår: 20 for sense of effort, 16 for sense of anxiety, 12 for sense of discomfort. Jo høyere skår jo mer alvorlig er pasientens pustebesvær
Administrasjonsform	Selv-administrert
<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b>
Antall	n=171
Alder	Gjennomsnitt 63 (SD 9,3)
Kjenn	Menn: 122 (70 %)
Sykdomskarakteristika	Avansert eller residiverende lungekreft Mer detaljert informasjon tilgjengelig i artikkelens tabell 1
Performance status: ECOG	0 36 (21%) 1 126 (74%) 2-4 9 (6%)
Setting	Ambulante pasienter, thorax/onkologisk divisjon
Land	Japan
Språkversjon	Japansk
Metode for utvalgelse	Fortløpende ambulante pasienter
<b>COSMIN</b>	<b>Kritisk vurdering</b> <b>Resultat/kommentar</b>
Statistiske metoder benyttet basert på CTT eller IRT	CTT
Hypotheses testing	Good Lite antall hypoteser formulert a priori, forventet størrelse på samvariasjon eller forskjeller ikke formulert, usikkert om instrumenter til sammenligning er validert for studiepopulasjonen
Sammenligning:	HADS Spørsmål om opplevelse siste dager av varlige symptomer som hoste, slimløpspytt, smerte på en 4-punkts skala fra 1 (absent) til 4 (severe) Demografiske, medisinske og sosiale data
Notater	Se artikkelens Tabell 2, 3 og 4 for data om samvariasjon.
Det er mulig at Tanaka et al. (2000) og Tanaka et al. (2002) har delvis overlappende utvalg. Dette har ikke praktiske konsekvenser da de har sammenlignet med ulike verktøy.	

Henoch, I., Bergman, B., & Gaston-Johansson, F. (2006) **Validation of a Swedish version of the Cancer Dyspnea Scale.** *Journal of Pain & Symptom Management*, 31 (4), s. 353-61.

Henoch, I., Bergman, B., Gustafsson, M., Gaston-Johansson, F. & Danielson, E. (2008) **Dyspnea experience in patients with lung cancer in palliative care.** *European Journal of Oncology Nursing*, 12 (2), s. 86-96.

Dataekstraksjon utført av Kjersti Solvåg 29.09.14 Referanse nr. 1383 og 401

#### Kjennetegn ved studien

**Formål**  
Å beskrive validitets- og reliabilitetstesting av CDS i en svensk populasjon av pasienter med lungekreft, og å sammenligne med den publiserte engelske versjonen. Hoved forskningsspørsmål: Er en svensk versjon av Cancer Dyspnea Scale (CDS-S) valid, reliabel og gjennomførbar for å beskrive pustebesvær hos pasienter med lungekreft når kurativ og livsforlengende behandling ikke er tilgjengelig?

Studiedesign	Valideringsstudie
Land	Sverige
Finansiering	Swedish Society of Nursing (SfN) og Association for the Cancer and Traffic Injured (CTRF)

<b>Kjennetegn ved testen</b>	<b>Beskrivelse</b>
Navn	Cancer Dyspnea Scale
Forkortelse	CDS-S
Begrepet (beskrivelse)	Multidimensjonale aspekt ved pustebesvær
Konseptuelt rammeverk	Refleksiv
Opprinnelig er utviklet for	Pasienter med kreft
Antall spørsmål	12
Dimensjoner	3

Målestaka	5-punkts Likert-skala fra 1 (not at all) til 5 (very much)
Administrasjonsform	Selv-administrert
Gjennomføring	<p>Plotstudie med 10 pasienter:</p> <p>Tid: gjennomsnitt 2 min</p> <p>Ressurser: 2 pasienter trengte assistanse</p>

<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b>
Antall	n = 99
Alder	Median 69, range 36-85
Kjønn	49 menn og 50 kvinner

Sykdomskarakteristika	Avansert lungekreft. 60 personer hadde metastaser
Performance status:	Mer detaljert informasjon i artikkelens tabell 1
ECOG	<p>0 17</p> <p>1 59</p> <p>2 18</p> <p>3 4</p> <p>4 1</p>

Setting	Lunge-onkologisk poliklinikk
Land	Sverige
Språkversjon	Svensk
Andel manglende respons	151 personer ble spurt om å delta. 30 % avsto, i hovedsak fordi de følte seg for dårlig. 7 personer (6,6%) ble ekskludert fra analysen fordi de ikke hadde ferdigstilt CDS-S korrekt. 99 personer ble inkludert i analysen.

<b>COSMIN</b>	<b>Kritisk vurdering</b>	<b>Resultat/kommentar</b>
Statistiske metoder benyttet	CTT	
basert på CTT eller IRT		
Internal consistency	Good	Cronbach's alpha:

	På grunn av fratall: 30 %	Sense of effort	0,81
	av de forespurte takket	Sense of anxiety	0,84
	nei, i hovedsak fordi de	Sense of discomfort	0,88
	følte seg for dårlig, 6,6%	Total skår	0,90
	av spørreskjemaene ble fjernet fordi de ikke var utfyllt korrekt		

Strukturell validitet	Excellent	Faktoranalyse med Varimax med Kaiser Normalization som rotasjons metode.
		Faktor-lastingen korresponderte akseptabelt men ikke komplett med resultatene til Tanaka et al. (2000)
		Varians ikke oppgitt

Hypothesetesting	Good	<b>Sammenslutninger:</b>
	Forventet størrelse på korrelasjoner ikke oppgitt, og noe manglende informasjon om måleegenskapene til instrumentene for sammenligning og om de er validert for studiepopulasjonen	HADS
		VAS av pustebesvær og VAS ubehag
		Deler av EORTC QLQ-30 (skala for pustebesvær, fysisk og følelsesmessig funksjon) og QLQ-LC13 (skala for pustebesvær og hoste)

Cross-cultural validity	Ikke gjennomført	Oversettelse fra engelsk versjon til svensk ble gjennomført med frem-oversettelse av to uavhengige oversettere med svensk som morsmål. En tredje persons samordnet de to versjonene til en. To profesjonelle oversettere med engelsk sykdom/feromen er ikke beskrevet. Bare en frem- og en tilbake-oversettelse ble gjennomført. Det er mange full beskrivelse av hvordan uenighet mellom oversetterne ble løst. Det synes ikke som om oversettelsen ble vurdert av en komité.
Oversettelsesprosessen	Good	Totalt vurderes til good på tross av at det bare var en frem- og en tilbake-oversettelse.

#### Notater

Henoch et al. (2006) og Henoch et al. (2008) ser ut til å omhandle samme studie. Det er noen flere deltagere i Henoch et al. (2008). Jeg har valgt å trekke ut dataene fra Henoch et al. (2006) (ref 1383) da Henoch et al. (2008) ikke tilfører noe mer er aktuelt for min systematiske oversikt.

Uronis, H. E., Shelby, R. A., Currow, D. C., Ahmedzai, S. H., Bosworth, H. B., Coan, A. & Abernethy, A. P. (2012) <b>Assessment of the psychometric properties of an English version of the cancer dyspnea scale in people with advanced lung cancer. <i>Journal of Pain &amp; Symptom Management</i>, 44 (5), s. 741-9.</b>	
Dataekstraksjon utført av Kjersti Solvåg 06.10.14	Referanse nr. 87
<b>Kjennetegn ved studien</b>	
Formål	Vurdere validiteten og reliabiliteten til den engelske oversettelsen av CDS hos pasienter med lungekreft.
Studiedesign	Skaife tilveie et multifaktorrett kartleggingsverktøy for pustebesvær på engelsk
Land	Valideringstudie
Finansiering	USA, Australia og Storbritannia
<b>Kjennetegn ved kartleggingsverktøyet</b>	Ingen
Navn	<b>Beskrivelse</b> Cancer Dyspnea Scale
Forkortelse	CDS-E
Begrepet (beskrivelse)	Multidimensionale aspekt ved pustebesvær

Konseptuelt rammeverk	Refleksiv
Opprinnelig er utviklet for	Pasienter med kreft
Antall spørsmål	12
Dimensjoner	3
	Sense of effort Sense of anxiety Sense of discomfort
Tidreferanse for skår	Siste 7 dager
Måleskala	5-punkts Likert-skala fra 1 (not at all) til 5 (very much)
Algoritme for skår	Legg sammen skår for hver faktor Faktor 1= (spm 4+6+8+10+12) - 5 = sense of effort Faktor 2= (spm 5+7+9+11) - 4 = sense of anxiety Faktor 3= 15 - (spm 1+2+3) = sense of discomfort Legg sammen totalskårene for hver faktor = total dyspne
Administrasjonsform	Selv-administrert
Tilgjengelighet til fulltekst kopi	Kartleggingsverktøyet er vedlagt som Tabell 1
<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b>
Antall	N=109. US: n=78; UK: n=12; AUS: n=22
Alder	Gjennomsnitt 64.5 (SD 11.5)
Kjønn	67 % menn
Sykdomskarakteristika	Avansert lungekreft
Performance status:	Mer detaljert informasjon tilgjengelig i artikkelens tabell 3
ECOG	0 13 (12 %) 1 56 (51 %) 2 22 (20 %) 3 11 (10 %)
Setting	Sykehus, men det er ikke oppgitt om det er inneleggende eller polikliniske pasienter
Land	USA, Storbritannia, Australia
Språkversjon	Engelsk
Andel manglende respons	2 av 111 hadde ikke svart på alle testledd og ble ekskludert (1,8 %)
<b>COSMIN</b>	<b>Kritisk vurdering</b> <b>Resultat/kommentar</b>
Statistiske metoder benyttet basert på CTT eller IRT	CTT
Internal consistency	Excellent Cronbach's alpha (n=109) Sense of effort 0,84 Sense of anxiety 0,80 Sense of discomfort 0,84 Total skår 0,71

Structural validity	Good Eksplorative faktoranalyse (EFA) er utført, men bekrefte faktoranalyse hadde vært mer hensiktsmessig siden faktorene er predefiner fra den japanske studien	EFA viste at flere testledd lastet annerledes enn i den originale studien til Tanaka et al. Varians ikke oppgitt. RMSE = 0.07 (90 % KI = 0.02 til 0.10)
Hypotheses testing	Sammenligninger: Modified BORG, VAS, HADS, FACT-L, funksjonsstatus (ECOG), SpO2. Se artikkelens Tabel 9 for data om samvariasjon. for sammenligningene med VAS/BORG/ECOG	
	Good Ikke beskrevet hvordan manglende spørsmål er håndtert, antar at de bare har tatt med de som har svart på begge skjema innenfor hver parvise sammenligning. Mangelfullt formulerte hypoteser og noe manglende informasjon om instrumentene til sammenligning. Totalskår good på tross av manglende hypoteser apriori.	
	Fair Ikke beskrevet hvordan manglende spørsmål er håndtert, antar at de bare har tatt med de som har svart på begge skjema innenfor hver parvise sammenligning. Mangelfullt formulerte hypoteser	for sammenligningene med HADS/FACT-G/FACT/USpO2

	Og noe manglende informasjon om testene til sammenligning. Risiko for utvalgsskjevhet, på grunn av manglende svar på endel av spørreskjemaene	Ikke utført. Oversettelse til engelsk versjon utført av Tanaka et al. (2000)
Cross-cultural validity		
Notater	Faktoranalysen viste at faktor loadingen ikke helt samsvarer med originalversjonen til Tanaka et al. Flere spørsmål lastet forskjellig i den engelske og svenske studien vs den originale. Dette er vurdert som kulturelt betinget. En r-CDS-E ble også utviklet i studien, men denne har jeg ikke tatt med i min systematiske oversikt	
	Windisch, W., Freidel, K., Schucher, B., Baumann, H., Wiebel, M., Matthys, H., & Petermann, F. (2003) <b>The Severe Respiratory Insufficiency (SRI) Questionnaire A specific measure of health-related quality of life in patients receiving home mechanical ventilation.</b> <i>Journal of clinical epidemiology</i> , 56 (8), s. 752-759.	
	Dataekstraksjon utført av Kjersti Solvåg 07.10.14	Referanse nr. 1452
Kjennetegn ved studien		
Formål	Å utvikle ett spesifikt kartleggingsverktøy for helserelatert livskvalitet hos pasienter med alvorlig respirasjonsvikt som får behandling med home mechanical ventilation (HMV) ved å identifisere relevante livskvalitets dimensjoner hos disse pasientene	
Studiedesign	Utvikling- og valideringsstudie	
Land	Tyskland	
Aktuelt kartleggingsverktøy	Severe Respiratory Insufficiency (SRI) Questionnaire	
Finansiering	Breas® Medical AB	
Kjennetegn ved kartleggingsverktøyet	<b>Beskrivelse</b>	
Navn	Severe Respiratory Insufficiency Questionnaire	
Forkortelse	SRI	
Begrepet (beskrivelse)	Helserelatert livskvalitet	
Konseptuelt rammeverk	Refleksiv	
Opprinnelig er utviklet for	Pasienter med kronisk respirasjonsvikt som følge av ulike underliggende sykdommer og som får behandling med HMV	
Antall spørsmål	49	
Dimensjoner	7 + totalskår	(SRI)RC: 8 testledd
	Respiratory complaints	

Physical functioning	(SRI-PF):	6 testledd
Attendant symptoms and sleep	(SRI-AS):	7 testledd
Social relationships	(SRI-SR):	6 testledd
Anxiety	(SRI-AX):	5 testledd
Psychological well-being	(SRI-WB):	9 testledd
Social functioning	(SRI-SF):	8 testledd
Summary scale	(SRI-SS)	
Tidsreferanse for skår		
Siste uke		
Måleskala	Testledd: 5-punkts Likerts-skala fra «strongly agree» til «strongly disagree»	
Faktor: 0-100, høyere skår indikerer bedre livskvalitet		
Algoritme for skår	Hvert testledd sikres fra 1 til 5, 35 testledd omkodes og gjennomsnittsverdien for hvert domene transformeres til en prosentverdi (prosent av maks) (Duiveman et al., 2008)	
Administrasjonsform	Selv-administrert	
Gjennomføring	Tid: ca 20 minutter Ressurser: en pasient med kols og seks med nevrologiske sykdom måtte ha hjelp av en intervjuer for å fylle ut skjemaet.	
Tilgjengelighet til fulle tekst kopi		
Full tekst med veiledning for skår og analyser kan fås for vitenskapelige formål ved kontakt med forfatter		
<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b>	
Antall	n=226	
Alder	Gjennomsnitt 57,3 (SD 14,0)	
Kjønn	118 kvinner/108 menn	
Sykdomskarakteristika	Pasienter med alvorlig hyperkapnisk respirasjonssvikt, som var stabilisert på HMV, hadde stabil sykdom og som var innlagt til rutinekontroll for HMV kols n=78; kyfoskoliose n= 57; posttuberkulose sekvele n=20; nevrologiske sykdommer n=49; diverse diagnoser n=22 Mer detaljert informasjon tilgjengelig i artikkelens tabell 1, 2 og kap.3	
Setting	Sykehus. Inneliggende til rutinekontroll for HMV	
Land	Tyskland	
Kartleggingsverktøyet	Tysk	
språkversjon		
Andel manglende respons	8 personer ble ekskludert på grunn av manglende ferdigstilte spørreskjema.	
<b>COSMIN</b>	<b>Kritisk vurdering</b>	<b>Resultat/kommentar</b>
Statistiske metoder benyttet	CTT	
basert på CTT eller IRT		
Internal consistency	Data fra Windisch et al. (2008) benyttes	
Content validity	<i>Excellent</i>	Utvikling av testledd var basert på sosiale, psykologiske og fysiske helsedomener. Første versjon bestod av 80 testledd

etterfulgt av et åpent intervju med pasienter behandlet med HMV som gav viktige tilbakemeldinger om egen livskvalitet.	
Testleddene ble utviklet av et eksperipanel som bestod av lungeleger og psykologer som var spesialister på livskvalitet.	
Structural validity	Data fra Windisch et al. (2008) benyttes
Hypotheses testing	Good
Sammenligning: SF-36	
Se artikkelens Tabel 15 for data om samvariasjon.	
<b>Interpretability</b>	
Skår og endret skår for relevante grupper	Gjennomsnittlig totalskår varierte mellom ≈ 50 og 65 mellom ulike diagnosegrupper (kols lavest og kyfoskoliose høyest)
<b>Notater</b>	
Kols utvalget i denne artikkelen er inkludert som en del av kols utvalget i Windisch et al. (2008). Jeg benytter derfor ikke data for Internal consistency og Structural validity fra denne artikkelen, men fra Windisch et al (2008).	
<b>Duiveman, M. L., Wempe, J. B., Bladder, G., Kerstjens, H. A. &amp; Wijkstra, P. J. (2008) Health-related quality of life in COPD patients with chronic respiratory failure. <i>European Respiratory Journal</i>, 32 (2), s. 379-86.</b>	
Dataekstraksjon utført av Kjersti Solvåg 08.10.14	Referanse nr. 605
<b>Kjennetegn ved studien</b>	
Formål	Undersøke om SRI er reliabelt og valid livskvalitetsspørreskjema for pasienter med kols som har kronisk hyperkapnisk respirasjonssvikt.
Studiedesign	Valideringsstudie
Land	Nederland
Finansiering	Ingen
<b>Kjennetegn ved kartleggingsverktøyet</b>	<b>Beskrivelse</b>
Navn	Severe Respiratory Insufficiency Questionnaire
Forkortelse	SRI
Begrepet (beskrivelse)	Helserelatert livskvalitet
Konseptuelt rammeverk	Refleksiv
Opprinnelig er utviklet for	Sammensatt gruppe av pasienter med kronisk respirasjonssvikt
Antall spørsmål	49
Dimensjoner	7 + totalscore
Respiratory complaints	(SRI-RC): 8 testledd
Physical functioning	(SRI-PF): 6 testledd
Attendant symptoms and sleep	(SRI-AS): 7 testledd

Social relationships	(SRI-SR): 6 testledd
Anxiety	(SRI-AX): 5 testledd
Psychological well-being	(SRI-WB): 9 testledd
Social functioning	(SRI-SF): 8 testledd
Summary scale	(SRI-SS)
Måleskala	Testledd: 5-punkts Likerts-skala fra «strongly agree» til «strongly disagree» (Windisch et al., 2003)
Algoritme for skår	Faktor: 0-100, høyere skår indikerer bedre livskvalitet
Hvert testledd skåres fra 1 til 5, 35 testledd omkodes og gjennomsnittsverdien for hvert domene transformeres til en prosentverdi (prosent av maks)	
Administrasjonsform	Selv-administrert
Gjennomføring	Tid: 20 min
<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b>
Antall	n=72
Alder	test-retest n=23
Kjønn	Gjennomsnitt 62 (SD 8.6)
Sykdomskarakteristika	39 menn/33 kvinner
	Kols GOLD III eller IV, kronisk hyperkapnisk respirasjonssvikt, stabil fase, uten rehabilitering for $\geq 18$ måneder og behandlet med medikamenter og langtidsoxygenoterapi (LTOT) hvis nødvendig. Ingen var eller hadde vært behandlet med LTMV.
Land	Mer detaljert informasjon tilgjengelig i artikkelens Tabell 1.
Språkversjon	Nederland
<b>COSMIN</b>	Nederlandsk
Statistiske metoder benyttet basert på CTT eller IRT	<b>Kritisk vurdering</b> <b>Resultat/kommentar</b>
Internal consistency	CTT
	Poor
	Cronbach's alpha
	SRI-RC 0,73
	Utvalget var betydelig mindre enn det som kreves og det er ikke gjennomført faktoranalyse eller referanse til annen studie der dette er gjennomført
	SRI-PF 0,73
	SRI-AS 0,66
	SRI-SR 0,78
	SRI-AX 0,70
	SRI-WB 0,84
	SRI-SF 0,73
Reliability	Poor
	Test-retest intervall 12 uker
	ICC 0.81
	Utvalget ved test-retest er lite

Hypothesetesting	Good	Sammenligninger:
	Totalskår good selv om et spørsmål er fair. De oppgir å ha fjernet en som ikke kunne ferdigstille spørreskjemaene. De er allikevel ikke eksplisitt på hvordan manglende skår er håndtert. Vi velger ikke å la det trekke ned hele boksen	CRQ, HADS, GARS, BDI, MRC HADS anxiety forklarte 56% av variansen for SRI totalskår, mens HADS anxiety, GARS og exacerbasjonsfrefkvens til sammen forklarte 72 % av totalskår Fysiske/objektive mål
		Se artikkelens Tabell 3 og 4 for data om samvariasjoner.
<b>Interpretabilit</b>		
Distribusjon av (total) skår i utvalget		Gjennomsnittlig sum skår 51,3 (SD 13,6) Normalfordelt
% av respondenter med lavest skår		Ingen har laveste sum skår. 1 respondent har laveste skår på SRI-AX og 1 på SRI-SF
% av respondenter med høyest skår		Ingen har høyeste skår for subskalær eller sum skår

Lopez-Campos, J. L., Falde, I., Masa, J. F., Benitez-Moya, J. M., Barrot, E., Ayerbe, R., Leon-Jimenez, A. & Windisch, W. (2008) Transculturally adapted Spanish SRI questionnaire for home mechanically ventilated patients was valid, and reliable. <i>J Clin Epidemiol</i> , 61 (10), s. 1061-6.	
Dataekstraksjon utført av Kjersti Solvåg 21.11.14	Referanse nr. 1456
<b>Kjennetegn ved studien</b>	
Formål	Malet med studien var å validere spansk versjon av SRI.
Studiedesign	Prospektiv multiserierstudie
Land	Spania
Aktuelt kartleggingsverktøy	SRI
Finansiering	Ikke oppgitt
<b>Kjennetegn ved kartleggingsverktøyet</b>	<b>Beskrivelse</b>
Navn	Severe Respiratory Insufficiency Questionnaire
Forkortelse	SRI
Begrepet (beskrivelse)	Syv spesifikke aspekter ved helse relatert livskvalitet
Konseptuelt rammeverk	Refleksiv
Opprinnelig er utviklet for	Pasienter med kronisk respirasjonssvikt som følger av ulike underliggende årsaker og som får behandling med LTMV
Antall spørsmål	49

Dimensjoner	7 + totalscore		
	Respiratory complaints	(SRI-RC):	8 testledd
	Physical functioning	(SRI-PF):	6 testledd
	Attendant symptoms and sleep	(SRI-AS):	7 testledd
	Social relationships	(SRI-SR):	6 testledd
	Anxiety	(SRI-AX):	5 testledd
	Psychological well-being	(SRI-WB):	9 testledd
	Social functioning	(SRI-SF):	8 testledd
	Summary scale	(SRI-SS)	
Tidsreferanse for skår	Siste uke		
Måleskala	Ordinal lengradert skala (fra «totally false» til «totally true»)		
Algoritme for skår	Hvert testledd skåres fra 1 til 5, 35 testledd omkodet og gjennomsnittsverdien for hvert domene transformeres til en prosentverdi (prosent av maks) (Duiverman et al., 2008).		
Administrasjonsform	Selv-administrert. Når det ikke var mulig administrerte forskeren spørreskjemaet.		
Gjennomføring	Tid: 10 +/- 5 minutter Ressurser: 54 % av spørreskjemaene var selv-administrert. Grunner oppgitt for manglende selv-administrering: Ikke i stand til å lese eller skrive flyende (n=17); manglet lesebriller (n=10); forstod ikke hvordan de skulle fullføre spørreskjemaet (n=9); for fysisk funksjonshemmet til å skrive (n=8); pasienten foretrakk å ikke fylle ut selv (n=3); ikke spesifisert (n=5)		
Tilgjengelighet til fulltekst kopi	Gratis fulltekst kopi og veiledning kan fås fra forfatter		
<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b>		
Antall	n=115		
Alder	62 +/- 13 år		
Kjønn	53 (52.4 %) menn		
Sykdomskarakteristika	Pasienter som fikk HMV av ulike årsaker, klinisk stabile og velipasset på HMV i minst tre måneder. Thoracic cage (n=33); OHS (n=37); Nevromuskulær (n=18); Tuberkulose sekvele (n=12); kols (n=15) Mer detaljert informasjon tilgjengelig i artikkelens tabell 1		
Performance status	Oppgitt at Performance status i henhold til British Thoracic Society guidelinees er vurdert, men resultater ikke oppgitt.		
Setting	Poliklinisk		
Land	Spania		
Språkversjon	Spansk		
Metode for utvalgelse	Noe uklart beskrevet, men sannsynlig fortløpende rekruttering/vrutinemessig kontroll på poliklinikk		

Andel manglende respons	30 % manglende svar på spørsmål 31 fordi det ikke var aktuelt (ingen partner). Ellers en responsrate på testledd på 96-100 %		
<b>COSMIN</b>	<b>Kritisk vurdering</b>	<b>Resultat/kommentar</b>	
Statistiske metoder benyttet basert på CTT eller IRT		CTT	
Internal consistency	Poor	Cronbach's alpha	
	Uvalget mindre enn det som kreves i forhold til spørsmål.	SRI-RC 0,81	
		SRI-PF 0,82	
		SRI-AS 0,74	
		SRI-SR 0,63	
		SRI-AX 0,73	
		SRI-WB 0,85	
		SRI-SF 0,83	
		SRI-SS (summary scale) 0,93	
Reliability	Fair	Test-retest	
	Ikke eksplisitt informasjon om alle deltek ved retest.	n for test retest ikke oppgitt, men antar at den er > 100	
	Manglende informasjon om mulig frafall mellom test og retest og om de faktisk er stabile i denne perioden.	Tid mellom test-retest: 2 uker.	
		Kliniske data samlet ved begge målepunktene, men oppgir ikke om det er forskjell mellom disse. Vi antar at pasientene er klinisk stabil.	
		ICC:	
		SRI-RC 0,90	
		SRI-PF 0,96	
		SRI-AS 0,91	
		SRI-SR 0,89	
		SRI-AX 0,84	
		SRI-WB 0,91	
		SRI-SF 0,92	
		SRI-SS (summary scale) 0,95	
		Alle P<0,001	
Structural validity	Poor	PCA gjennomført	
	Uvalget mindre enn det som kreves i forhold til spørsmål.	Faktoranalyse forklarte 60 % av variansen i spørreskjemaet, men resulterte i 13 skalaer.	
		SRI-RC delt i to skalaer	
		SRI-SR delt i to skalaer	
		SRI-AX delt i to skalaer	
		SRI-AS delt i fire skalaer	
Hypotheses testing	Fair	Sammenligning:	
	Mangelfull informasjon om SF-36 som den	SF-36	



<p>sammenlignes mot, men de benytter SF-36 som de tidligere studiene også har benyttet. Velger derfor å ikke gi dårligste skår</p>	<p>Se artikkelens Tabell 2 for data om samvariasjon.</p>
<p><b>Notater</b></p> <p>Finner 13 vs syv skalaer i det originale skjemaet. Begrunner at dette er vanlig ved oversatte og kulturelt tilpassede instrumenter.</p> <p>Oversettelse og tverkkulturell tilpassning er gjennomført i en studie av López-Campos et al. fra 2006. Deme har vi ikke hatt muligheten til å vurdere da vi bare har funnet den spanske versjonen av publikasjonen.</p>	
<p><b>Kjennetegn ved studien</b></p> <p>Formål</p> <p>Studiedesign</p> <p>Land</p> <p>Aktuelt kartleggingsverktøy</p> <p>Finansiering</p> <p><b>Kjennetegn ved kartleggingsverktøyet</b></p> <p>Navn</p> <p>Forkortelse</p> <p>Begrepet (beskrivelse)</p> <p>Konseptuelt rammeverk</p> <p>Opprinnelig er utviklet for</p> <p>Antall spørsmål</p> <p>Dimensjoner</p>	<p>Referanse nr. 1453</p> <p>Undersøke den interne strukturen til SRI for pasienter med kronisk respirasjonssvikt på grunn av kols, for dermed ha et egnet instrument som tillater pålitelig longitudinell evaluering av livskvalitet hos disse pasientene</p> <p>Tyskland</p> <p>Severe Respiratory Insufficiency Questionnaire</p> <p>Ikke oppgitt</p> <p><b>Beskrivelse</b></p> <p>Severe Respiratory Insufficiency Questionnaire</p> <p>SRI</p> <p>Helserelatert livskvalitet</p> <p>Refleksiv</p> <p>Pasienter med kronisk respirasjonssvikt som følge av ulike underliggende sykdommer og som får behandling med HMV.</p> <p>49</p> <p>7 + en totalskår</p> <p>Respiratory complaints (SRI-RC): 8 testledd</p> <p>Physical functioning (SRI-PF): 6 testledd</p> <p>Attendant symptoms and sleep (SRI-AS): 7 testledd</p> <p>Social relationships (SRI-SR): 6 testledd</p> <p>Anxiety (SRI-AX): 5 testledd</p>
<p>Windisch, W., Budweiser, S., Heinemann, F., Pfeifer, M., &amp; Rzehak, P. (2008) <b>The Severe Respiratory Insufficiency Questionnaire was valid for COPD patients with severe chronic respiratory failure.</b> <i>Journal of Clinical Epidemiology</i>, 61 (8), s. 848-853.</p>	

<p>Psychological well-being (SRI-WB): 9 testledd</p> <p>Social functioning (SRI-SF): 8 testledd</p> <p>Summary scale (SRI-SS)</p> <p>Siste uke</p> <p>Testledd: 5-punkts Likerts-skala fra «strongly agree» til «strongly disagree»</p> <p>Faktor: Faktor: 0-100, høyere skår indikerer bedre livskvalitet</p> <p>Hvert testledd skåres fra 1 til 5, 35 testledd omkodes og gjennomsnittsverdien for hvert domene transformeres til en prosentverdi (prosent av maks) (Duiverman et al., 2008)</p> <p>Selv-administrert</p> <p>Tid: ca 20 minutter</p> <p>Full tekst med veiledning for skår og analyser kan fås for vitenskapelige formål ved kontakt med forfatter (Windisch et al., 2003)</p> <p><b>Kjennetegn ved studiens utvalg</b></p> <p>Antall</p> <p>Alder</p> <p>Sykdomskarakteristika</p>	<p>9 testledd</p> <p>8 testledd</p> <p>Siste uke</p> <p>Testledd: 5-punkts Likerts-skala fra «strongly agree» til «strongly disagree»</p> <p>Faktor: Faktor: 0-100, høyere skår indikerer bedre livskvalitet</p> <p>Hvert testledd skåres fra 1 til 5, 35 testledd omkodes og gjennomsnittsverdien for hvert domene transformeres til en prosentverdi (prosent av maks) (Duiverman et al., 2008)</p> <p>Selv-administrert</p> <p>Tid: ca 20 minutter</p> <p>Full tekst med veiledning for skår og analyser kan fås for vitenskapelige formål ved kontakt med forfatter (Windisch et al., 2003)</p> <p><b>Beskrivelse</b></p> <p>n=162. To utvalg: Gr.1 n=78, Gr.2 n=84.</p> <p>Gr. 1: 60,9 (SD 9,6); Gr. 2: 64,3 (SD 7,4); Total gruppe: 62,4 (SD 8,9)</p> <p>Obs: Signifikant forskjell i alder mellom gruppene p=0,02</p> <p>Kronisk hyperkapnisk respirasjonssvikt forårsaket av kols.</p> <p>Langsiktig behandling med NPPV (noninvasive positive pressure ventilation) på grunn av hyperkapni, i tillegg LTO.</p> <p>Stabil fase, innlagt til rutinesjekk for NPPV.</p> <p>Mer detaljert informasjon tilgjengelig i artikkelens tabell 1.</p> <p>Obs: Signifikant forskjell mellom gruppene for FEV<sub>1</sub>/FVC med p= 0.01. Dette kan tyde på en mer alvorlig luftveisobstruksjon i Gr.1</p>
<p>Setting</p> <p>Land</p> <p>Språkversjon</p> <p><b>COSMIN</b></p> <p>Statistiske metoder benyttet</p> <p>basert på CTT eller IRT</p> <p>Internal consistency</p>	<p>Sykehus, innleggende til rutinekontroll for NPPV</p> <p>Tyskland</p> <p>Tysk</p> <p><b>Kritisk vurdering</b></p> <p>CTT</p> <p>Poor</p> <p>Uvalget var betydelig mindre enn det som kreves</p> <p>Cronbach's alpha</p> <p>Gr.1</p> <p>SRI-RC 0,85</p> <p>SRI-PF 0,81</p> <p>SRI-AS 0,70</p> <p>SRI-SR 0,76</p> <p>SRI-AX 0,77</p> <p>SRI-WB 0,82</p> <p>SRI-SF 0,79</p> <p>Gr.2</p> <p>0,80</p> <p>0,82</p> <p>0,74</p> <p>0,72</p> <p>0,73</p> <p>0,82</p> <p>0,82</p> <p>total gruppe</p> <p>0,82</p>



Structural validity	SRI-SS 0.87 0.89 0.88
Poor	PCA bekreftet en enkelt faktor i fem subskalaer, men to faktorer ble funnet i to subskalaer (SRI-AS og SRI-WB). Det er foreslått at disse to faktorene inneholder to ulike aspekter ved livskvalitet, men begge faktorene i to-faktor løsningen for begge subskalaene (SRI-AS og SRI-WB) var korrelert med hverandre. CFA viste at en to-faktor løsning var mer akseptabel enn en en-faktor løsning. Teoretisk sett kan det bety at basert på disse resultatene kan flere faktorer enn de originale 7 innlemmes. Men to-faktor løsningen viste at disse to faktorene var korrelert til hverandre ( $r$ fra 0,43-0,80), noe som indikerer at de ikke er uavhengige av hverandre. En ytterligere inndeling i subskalaer ville gitt flere subskalaer med et betydelig lavt antall testledd. Den originale 7- subskala strukturen ble derfor beholdt.
	En høy andel av variansen kunne forklares av hver subskala (fra 44,6-58,0 %, total gruppe). SRI-SS forklarte 58,5 % av variansen (total gruppe).
	Bare en faktor med en eigenvalue $>1$ ble funnet i SRI-SS (sum skår). De resterende 6 faktorene produserte en eigenvalue $<1$ .
<b>Interpretability</b>	
Distribusjon av (total) skår i utvalget	Gjennomsnittsskår for SS-SRI var 52,3 +/- 16,5
<b>Notater</b>	
Gr 1 er fra kols utvalget fra Windisch et al. (2003). Fordi utvalget i denne studien er størst velger jeg å benytte data om Internal consistency og Structural validity fra denne artikkelen til resultatdelen i min systematiske oversikt istedenfor fra Windisch et al. (2003).	
<b>Validation of the English Severe Respiratory Insufficiency Questionnaire</b>	
Dataekstraksjon utført av Kjersti Solvåg 27.11.14	Referanse nr. 1457
<b>Kjennetegn ved studien</b>	
Formål	A lage og validere en engelsk versjon av SRI
Studiedesign	Tverrsnittstudie
Land	England
Aktuelt kartleggingsverktøy	SRI

Finansiering	Ikke helt klart. Windisch har fått diverse reise- og forskningsstøtte, og undervisningshonorar som oppgis på: <a href="http://erj.ersjournals.com/site/misc/statenen4002.xhtml">http://erj.ersjournals.com/site/misc/statenen4002.xhtml</a>
<b>Kjennetegn ved kartleggingsverktøyet</b>	<b>Beskrivelse</b>
Navn	Severe Respiratory Insufficiency Questionnaire
Forkortelse	SRI
Begrepet (beskrivelse)	Helserelatert livskvalitet
Konseptuelt rammeverk	Refleksiv
Opprinnelig er utviklet for	Sammensatt gruppe av pasienter med kronisk respirasjonssvikt
Antall spørsmål	49
Dimensjoner	7 + totalscore
	Respiratory complaints (SRI-RC): 8 testledd
	Physical functioning (SRI-PF): 6 testledd
	Attendant symptoms and sleep (SRI-AS): 7 testledd
	Social relationships (SRI-SR): 6 testledd
	Anxiety (SRI-AX): 5 testledd
	Psychological well-being (SRI-WB): 9 testledd
	Social functioning (SRI-SF): 8 testledd
	Summary scale (SRI-SS)
Tidsreferanse for skår	Siste uke
Måleskala	Testledd: 5-punkts Likerts-skala fra «completely untrue» til «always true» Faktor: 0-100, høyere skår indikerer bedre livskvalitet
Algoritme for skår	Hvert testledd skåres fra 1 til 5, 35 testledd omkodes og gjennomsnittsverdien for hvert domene transformeres til en prosentverdi (prosent av maks)
Administrasjonsform	Selv-administrert
Tilgjengelighet til fullest kopi	I online tillegg
<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b>
Antall (%)	Total n=152
	kronisk obstruktiv lungesykdom (kols) n=26 (17)
	restriktive bystveggsleidelser (RCWD) n=58 (38)
	nevromuskulære sykdommer (NMD) n=32 (21)
	adipositas hypoventilasjonsyndrom (OHS) n=29 (19)
	andre n=7 (5)
Kjønn	75 kvinner
Sykdomskarakteristika	Kronisk hyperkapnisk respirasjonssvikt av ulike årsaker med HMV for minst en måned. Klinisk stabile uten endring i medikamentell behandling siste fire uker
	142 pasienter fikk ikke-invasiv og 10 fikk invasiv (tracheostomi) ventilasjonsstøtte. Mer detaljert informasjon tilgjengelig i artikkelens tabell 1.
Setting	Poliklinisk

Land	England
Språkversjon	Engelsk
Andel manglende respons	Ikke oppgitt, men n for hver faktor er oppgitt
<b>COSMIN</b>	<b>Kritisk vurdering</b> <b>Resultat/kommentar</b>
Statistiske metoder benyttet basert på CTT eller IRT	CTT
Internal consistency	<p>Poor</p> <p>Cronbach's alpha</p> <p>SRI-RC 0.84</p> <p>SRI-PF 0.77</p> <p>SRI-AS 0.81</p> <p>SRI-SR 0.81</p> <p>SRI-AX 0.87</p> <p>SRI-WB 0.89</p> <p>SRI-SF 0.87</p> <p>SRI-SS (summary scale) 0.93</p>
Structural validity	<p>Poor</p> <p>Principal component analysis (PCA) avdekket en-faktor løsning for fire og to-faktor løsning for tre subskalær.</p> <p>Bekreftende faktoranalyse (CFA) avdekket to-faktor løsning for seks subskalær, men disse faktorene er avhengige av hverandre.</p> <p>Bare en faktor med en egenvalue <math>\geq 1</math> kunne ekstraheres fra de syv subskalæne til den oppsummerende skalaen (SRI-SS). SRI-SS forklarte 70 % av variansen.</p> <p>Mer detaljert informasjon tilgjengelig i artikkelens tabell 5 og 6.</p>
Hypotheses testing	<p><u>Sammenligning:</u></p> <p>SF-36</p> <p>Får good på tross av at to av spørsmålene er skåret fair, fordi hypotesene er sannsynlige om de ikke er eksplisitt formulert. SF-36 er et mye brukt og anerkjent validert verktøy på mange språk og for mange pasient populasjoner.</p>
Cross-cultural validity	<p>Ikke gjennomført</p> <p><u>Oversettelsesprosessen</u></p> <p>Alle oversetterne var tospråklige profesjonelle oversettere og informert om konseptet i spørreskjemaet. En oversetter frem-oversatte spørreskjemaet. Initiativtakeren til</p>
Oversettelsesprosessen	Fair

Oversettelses prosessen	oversettelsen ble konsultert ved behov, og de to ble til slutt enige om en preliminær frem oversettelse. En annen oversetter gjennomførte tilbake-oversettelsen og denne versjonen ble evaluert av initiativtakeren. Testledd med tilsvarende mening, men litt annen formulering ble diskutert, men bare endret dersom en helt sammenfallende ordlyd var oppnåelig. Ellers ble beste tilpassning akseptert. Testledd som potensielt kunne gi en annen mening sammenlignet med originalversjonen ble reformulert av initiativtakeren inntil enighet ble oppnådd. Prosessen var ikke formalisert, men alle testleddene kunne tilsnitt oversettes tilfredsstillende og med hensyn til at de skulle være korte og enkle å forstå. Den endelige engelske versjonen ble evaluert for lesbarhet av initiativtakeren til det britiske studien og pilottestet.
Interpretability	
Distribusjon av (total) skår i utvalget	Fem av åtte skalaer er normalfordelt

<p>Struik, F. M., Kerstjens, H. A., Bladder, G., Sprooten, R., Zijnen, M., Asin, J., van der Molen, T. &amp; Wijkstra, P. J. (2013) <b>The Severe Respiratory Insufficiency Questionnaire scored best in the assessment of health-related quality of life in chronic obstructive pulmonary disease.</b> <i>Journal of Clinical Epidemiology</i>, 66 (10), s. 1166-74.</p>	
Dataekstraksjon utført av	Kjersti Solvåg 08.10.14      Referanse nr. 1212
<b>Kjennetegn ved studien</b>	
Formål	Vurderer og sammenligner reliabilitet og validitet for en rekke ulike instrumenter deriblant SRI hos pasienter med kols med kronisk hyperkapnisk respirasjonssvikt ved en innleggesels som krever ventilasjonsstøtte.
Studiedesign	Valideringsstudie
Land	Nederland
Aktuelt kartleggingsverktøy	Severe Respiratory Insufficiency Questionnaire
Finansiering	Ikke oppgitt
<b>Kjennetegn ved kartleggingsverktøyet</b>	<b>Beskrivelse</b>
Navn	Severe Respiratory Insufficiency Questionnaire
Forkortelse	SRI
Begrepet (beskrivelse)	Helserelatert livskvalitet
Konseptuelt rammeverk	Refleksiv
Opprinnelig er utviklet for	En sammensatt gruppe av pasienter med respirasjonssvikt som får NVVP

Antall spørsmål	49
Dimensjoner	7 + totalscore Respiratory complaints (SRI-RC): 8 testledd Physical functioning (SRI-PF): 6 testledd Attendant symptoms and sleep (SRI-AS): 7 testledd Social relationships (SRI-SR): 6 testledd Anxiety (SRI-AX): 5 testledd Psychological well-being (SRI-WB): 9 testledd Social functioning (SRI-SF): 8 testledd + en summary score (SRI-SS)
Tidreferanse for skår	Ikke oppgitt
Måleskala	Testledd: 5-punkts Likerts-skala fra «strongly agree» til «strongly disagree» (Windsch et al., 2003) Faktor: 0-100, høyere skår indikerer bedre livskvalitet
Algoritme for skår	Hvert testledd skåres fra 1 til 5. 35 testledd omkodes og gjennomsnittsverdien for hvert domene transformeres til en prosentverdi (prosent av maks) (Duiverman et al., 2008)
Administrasjonsform	Selv-administrert
<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b>
Antall	n=170
Alder	Gjennomsnitt 64 (SD 8,0) år
Kjønn	42 % menn
Sykdomskarakteristika	Hospitaliserte kols pasienter med kronisk hyperkapnisk respirasjonssvikt som fortsatt er hyperkapnisk 48 timer etter avsluttet ventilasjonstøtte for en akutt excacerbasjon. Mer detaljert informasjon tilgjengelig i artikkelens tabell 1.
Setting	Sykehus: inneliggende
Land	Nederland
Språkversjon	Nederlandsk
Metode for utvalgelse	Ikke oppgitt. Randomisert? Studie er en del av et større RCT
Andel manglende respons	180 inkludert, 170 ferdigstilte spørreskjemaet.
<b>COSMIN</b>	<b>Kritisk vurdering</b> <b>Resultat/kommentar</b>
Statistiske metoder benyttet basert på CTT eller IRT	CTT
Internal consistency	Poor Cronbach's alpha Lite utvalg og manglende faktoranalyse eller referanse til annen studie der dette er gjennomført SRI-RC 0,82 SRI-PF 0,80 SRI-AS 0,77 SRI-SR 0,68 SRI-AX 0,80 SRI-WB 0,85

Hypothesetesting	Good Får good selv om det ikke er helt klart hvordan manglende svar er håndtert.	SRI-SF 0,81 Sammenligninger: CCQ, CRQ, HADS anxiety, HADS depression, GARS, MRC. HADS anxiety forklarte 62%, mens HADS anxiety, GARS og MRC til sammen forklarte 73% av variansen i SRI total skår. Se artikkelens Tabell 3 og 4 for data om samvariasjon.
<b>Interpretability</b>		
Distribusjon av (total) skår i utvalget	Normalfordelt	
% av respondenter med lavest skår	For subskalaer: 0 – 6.5% Sum skår: 0	
% av respondenter med høyest skår	For subskalaer: 0-1.8% Sum skår: 0	
<b>Notater</b>		
	HADS anxiety, GARS og MRC forklarte 73% av variansen i SRI total skår. Dette antyder at SRI kan være et godt mål for disse spesielle faktorene. Svak samvariasjon mellom antall excacerbasjoner siste år og sumskår. Andre fysiske parametre viste svak eller ingen samvariasjon med dimensjoner og sumskår	
<b>Markussen, H., Lehmann, S., Nilsen, R. M., &amp; Natvig, G. K. (2014) The Norwegian version of the Severe Respiratory Insufficiency Questionnaire. <i>International Journal of Nursing Practice</i>.</b>		
Dataekstraksjon utført av Kjersti Solvåg 17.11.14	Referanse nr. 1458	
<b>Kjennetegn ved studien</b>		
Formål	Målene med studien var å oversette og kulturelt tilpasse SRI til norsk, og teste reliabilitet, intern konsistens og construct validity	
Studiedesign	Tverrsnittstudie	
Land	Norge	
Aktuelt kartleggingsverktøy	SRI	
Finansiering	Økonomisk støtte fra Nasjonalt kompetansesenter for hjemmrespiratorbehandling i Norge	
<b>Kjennetegn ved kartleggingsverktøyet</b>	<b>Beskrivelse</b>	
Navn	Severe Respiratory Insufficiency Questionnaire	
Forkortelse	SRI	
Begrepet (beskrivelse)	Helserelatert livskvalitet	
Konseptuelt rammeverk	Refleksiv	

Opprinnelig er utviklet for	Pasienter med kronisk respirasjonssvikt som følge av ulike underliggende årsaker og som får behandling med LTMV.		
Antall spørsmål	49 testledd		
Dimensjoner	7 + totalscore		
	Respiratory complaints	(SRI-RC):	8 testledd
	Physical functioning	(SRI-PF):	6 testledd
	Attendant symptoms and sleep	(SRI-AS):	7 testledd
	Social relationships	(SRI-SR):	6 testledd
	Anxiety	(SRI-AX):	5 testledd
	Psychological well-being	(SRI-WB):	9 testledd
	Social functioning	(SRI-SF):	8 testledd
	Summary scale	(SRI-SS)	
Måleskala	Testledd: 5-punkts Likerts-skala fra «strongly agree» til «strongly disagree» (Windsch et al., 2003)		
Algoritme for skår	Faktor: 0-100, høyere skår indikerer bedre livskvalitet		
Administrasjonsform	Selv-administrert		
<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b>		
Antall (%)	n=127		
	NMD n=54 (42.5); COPD n=26 (20.5); OHS n=37 (29.9); Chest wall disease n=9 (7.1)		
Alder (SD)	Gjennomsnittsalder 61.5 (15.6)		
	Gjennomsnittsalder med SD for de ulike sykdomsgruppene		
	NMD 57.2 (17.8); COPD 67.7 (9.2); OHS 65.9 (12.2); Chest wall disease 51.6 (18.3)		
Kjønn (%)	Menn n=68 (53.5)		
	Nevromuskulære sykdommer (NMD) n=23 (18.1)		
	Kronisk obstruktiv lungesykdom (COPD) n=16 (12.6)		
	Adipositas hypoventilasjonssyndrom (OHS) n=23 (18.1)		
	Brystveggslidelse n=6 (4.7)		
Sykdomskarakteristika	Pasienter med alvorlig hyperkapnisk kronisk respirasjonssvikt som har vært etablert på LTMV i minst tre måneder.		
Setting	Merdelajert informasjon i artikkelens tabell 1.		
Land	Hjemmeboende pasienter		
Språkversjon	Norge		
Metode for utvalgelse	Norsk		
	Alle pasienter over 18 år som var registrert Nasjonalt register for langtids mekanisk ventilasjon LTMV ble invitert til å delta		

Andel manglende respons	211 tilfredsstilte inklusjonskriteriene, av disse var 193 potensielle respondenter, mens 127 ferdigstilte spørreskjema. Dette gav en responsrate på 65.8 %		
<b>COSMIN</b>	<b>Kritisk vurdering</b>	<b>Resultat/kommentar</b>	
Statistiske metoder benyttet basert på CTT eller IRT	CTT		
Internal consistency	Poor	Cronbach's alpha	
	Uvalget var mindre enn det som anbefales og faktoranalyse var ikke gjennomført eller referanse til annen studie	SRI-RC	0,81
		SRI-PF	0,76
		SRI-AS	0,68
		SRI-SR	0,82
		SRI-AX	0,81
		SRI-WB	0,88
		SRI-SF	0,79
		SRI-SS (summary scale)	0,94
Content validity	Good	Det overstatte spørreskjemaets face validity og feasibility ble vurdert i en pilot av 6 medlemmer fra den norske interesseorganisasjonen for brukere av LTMV. Alle de 6 hadde brukt LTMV i minimum 3 mnd. De ble bedt om å fylle ut spørreskjemaet og ble spurt om spørreskjemaet var klart, lett å forstå, og dekket emner av interesse, og om noen testledd hadde vært vanskelig å besvare. De ble også spurt om spørreskjemaet var relevant for deres liv som LTMV brukere.	
	Uvalget som ble benyttet til å vurdere relevans var 6 personer	En gruppe sykepleiere og leger med erfaring i LTMV behandling ble spurt om testleddene var relevant for å bruke i LTMV populasjonen.	
		Den tilbake-oversatte versjonen ble også sendt til forfatteren av det originale SRI som kommenterte på likeverdiget mellom original og oversettelse.	
		En konsensusgruppe sammenlignet tilsatt den oversatte versjonen og originalversjonen for likeverdiget, face validity og feasibility.	
Hypothesetesting	Good	<b>Sammensligning:</b>	
	SF-36 er for dårlig beskrevet, men er det samme verktøyet som er benyttet til hypotesetesting i de fleste andre studiene. Velger derfor å gi	SF-36	
		Se studiens Tabell 4 for data om samvariasjon	

totaliskår <i>good</i> for å være konsekvent i forhold til de andre studiene om SRI	
Cross-cultural validity	<i>Poor</i> Oversettelises prosessen Antall respondenter er for lite i forhold til antall spørsmål, og fordi det ikke er gjennomført faktoranalyse.
Oversettelse	<i>Good</i> tilslutt enige om en versjon for pre-testing.
<b>Interpretability</b>	
Distribusjon av (total) skår i utvalget	Gjennomsnittskår med SD for de ulike pasientgruppene og for total gruppe er oppgitt for hver faktor og for SRI-SS, se artikkelen Tabell 5.
<p>Migliore Nonweg, A., Whiteson, J., Demetis, S. &amp; Rey, M. (2006) <b>A new functional status outcome measure of dyspnea and anxiety for adults with lung disease: the dyspnea management questionnaire.</b> <i>Journal of Cardiopulmonary Rehabilitation</i>, 26 (6), s. 395-404.</p>	
Dataekstraksjon utført av Kjersti Solvåg 09.10.14	Referanse nr. 828
<b>Kjennetegn ved studien</b>	
Formål	Utvikle og psykometrisk teste Dyspnea Management Questionnaire (DMQ), et nytt multidimensjonalt måleverkøy for pustebesvær hos voksne med kols
Studiedesign	Valideringsstudie
Land	USA
Aktuelt kartleggingsverktøy	Dyspnea Management Questionnaire
Finansiering	The National Institute of Disability and Rehabilitation Research, U.S
<b>Kjennetegn ved kartleggingsverktøyet</b>	
Navn	Dyspnea Management Questionnaire
Forkortelse	DMQ-30
Begrepet (beskrivelse)	Multidimensjonale aspekt ved dyspnea
Konseptuelt rammeverk	Refleksiv
Opprinnelig er utviklet for	Pasienter med kols i rehabilitering
Antall spørsmål	30
Dimensjoner	5
	Dyspnea intensitet 7 testledd
	Dyspnea-related anxiety 7 testledd

Fearful activity avoidance	7 testledd
Self-efficacy for activity	5 testledd
Satisfaction with strategy use	4 testledd
Tidsreferanse for skår	Siste 2 uker
Måleskala	7-punkts Likert-skala fra 0 til 6.
Algoritme for skår	Høyere skår representerer bedre funksjonsstatus. For å kalkulere DMQ subskala skår for enkelt individ, ble testleddene gitt en numerisk verdi fra 0 til 6, og disse verdiene ble summert. Hver summerte rå-skår ble deretter delt på antall testledd i en subskala for å få en gjennomsnittlig skår for subskalaen. En totalskår skår ble kalkulert ved å summere subskalaene Dyspnea intensitet, Dyspnea-related anxiety og Fearful activity avoidance.
Administrasjonsform	Selv-administrert
Tilgjengelighet til fullest kopi	Vedlegg til artikkelen
<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b>
Antall	85 deltagere (kols n=73, astma n=12)
	Studie 2: Psykometrisk testing: n=73 (KOLS)
	Studie 3: n=22 (52 % kols, 48 % astma)
Alder	Gjennomsnitt 75 (SD 9.6, n=76)
	Studie 2: Psykometrisk testing: gjennomsnitt 76 (SD 8.8, n=67)
Kjønn (%)	65 % kvinner
	Studie 2: Psykometrisk testing: 60 % menn
Sykdomskarakteristika	Kols og astma
	Studie 2: Psykometrisk testing: kols
Setting	Rekuttert fra Rusk Institute of Rehabilitation Medicine ved hjelp av deres database og undersøkelsesmetoder
Land	USA
Språkversjon	Ikke oppgitt, men må være engelsk
Metode for utvalgelse	Se setting
<b>COSMIN</b>	<b>Kritisk vurdering</b> <b>Resultat/kommentar</b>
Statistiske metoder benyttet	CTT
basert på CTT eller IRT	
Internal consistency	<i>Good</i> Cronbach's alpha Andelen manglende svar er ikke oppgitt. Utvalget er mindre enn anbefalt
	Dyspnea intensitet 0,93 Dyspnea-related anxiety 0,92 Fearful activity avoidance 0,93 Self-efficacy for activity 0,87 Satisfaction with strategy use 0,90 DMQ totalskår (21 testledd) 0,96

Reliability	Poor	Test-retest	ICC	n
	Andelen manglende svar er ikke oppgitt. Utvalget var lite	Dyspnea intensity	0.88	26
		Dyspnea-related anxiety	0.92	25
		Fearful activity avoidance	0.90	24
		Self-efficacy for activity	0.71	23
		Satisfaction with strategy use	0.72	21
		DMQ totalskår (21 testledd):	0.95	24
Content validity	Excellent	5 konseptuelt utledede dimensjoner: En samling av testledd ble generert ved hjelp av kvalitative intervju data og litteraturgjennomgang på områdene dyspne, angst, unngåelsesadferd, funksjonsstatus, helserelatert livskvalitet, brukertilfredshet og lungerehabilitering. I tillegg ble 14 publiserte kartleggingsverktøy gjennomgått. 74 testledd som representerte de fem subskala dimensjonene ble forberedt og pilotestet av 3 voksne med kols. Content validity ble evaluert av 12 eksperter (bestående av 8 interdisiplinære spesialister i lungerehabilitering og 4 voksne med kols. Etter content validity fasen var det 66 testledd igjen.		
Hypotheses testing	Good for sammenligninger: DMQ/SOLQ, DMQ/SF-12 Poor for sammenligninger: DMQ/HADS-Anxiety	Sammenligninger: SF-12, SOLQ, HADS anxiety Se artikkelens Tabel 3, 4 og 6 for data om samvariasjoner		
	Andelen manglende svar er ikke oppgitt. Minimalt med hypoteser og forventet størrelse og reining ikke oppgitt. Utvalget var mindre enn anbefalt, men noe større for sammenligningene DMQ/SOLQ, DMQ/SF-12			
Interpretability	Distribusjon av (total) skår i utvalget	Dyspnea intensity	Gjennomsnitt	SD
			3,13	1,57

	Dyspnea-related anxiety	3,94	1,39
	Fearful activity avoidance	3,39	1,76
	Self-efficacy for activity	2,77	1,55
	Satisfaction with strategy use	3,33	1,44
	DMQ totalskår (21 testledd)	10,94	3,97
<b>Notater</b>			
Utvalgene i Norge et al. (2011) og Norweg et al. (2006) er delvis overlappende, men validerer ulike versjoner av DMQ.			

Norweg, A., Jette, A. M., Ni, P., Whithson, J. & Kim, M. (2011) <b>Outcome measurement for COPD: reliability and validity of the Dyspnea Management Questionnaire</b> . <i>Respiratory Medicine</i> , 105 (3), s. 442-53.			
Dataekstraksjon utført av Kjersti Solvåg 10.10.14 Referanse nr. 869			
<b>Kjennetegn ved studien</b>			
Formål	Evaluere reliabilitet og validitet ved en utvidet DMQ spørsmåssamling, som et ledd i utviklingen av en elektronisk versjon.		
Studiedesign	Valideringsstudie		
Land	USA		
Aktuelt kartleggingsverktøy	Dyspnea Management Questionnaire		
Finansiering	National Institute on Disability and Rehabilitation Research U.S.		
<b>Kjennetegn ved kartleggingsverktøyet</b>			
Navn	Dyspnea Management Questionnaire		
Forkortelse	DMQ-56		
Begrepet (beskrivelse)	Multidimensionale aspekt ved dyspne		
Konseptuelt rammeverk	Refleksiv		
Opprinnelig er utviklet for	Pasienter med kols		
Antall spørsmål	56		
Dimensjoner	5		
	Dyspnea intensity	12 testledd	
	Dyspnea-related anxiety	16 testledd	
	Fearful activity avoidance	15 testledd	
	Self-efficacy for activity	6 testledd	
	Satisfaction with strategy use	7 testledd	
Måleskala	7-punkts Likerts-skala fra 0 til 6		
Algoritme for skår	Høyere skår representerer bedre dyspne relatert funksjon For å kalkulere hver dimensjon blir råverdier for spørsmål summert og delt på antall spørsmål i dimensjonen og gir dermed en gjennomsnitt skår.		
<b>Kjennetegn ved studiens utvalg</b>			
Antall	Beskrivelse n=107		

Utvalget er sammensatt av 73 pasienter med kols fra utvalget til Norweg et al. (2006) + ytterligere 34 pasienter med kols	
Alder	Gjennomsnittsalder 75,1 (SD 8,6)
Kjønn (%)	58,9 % kvinner
Sykdomskarakteristika	Kols, stadler ikke oppgitt
Setting	Rekruttert fra Rusk Institute of Rehabilitation Medicine
Land	USA
Språkversjon	Engelsk
Metode for utvalgelse	n= 73 rekruttert fra Rusk Institute of Rehabilitation Medicine ved hjelp deres database og undersøkelsesmetoder. n=34 med legebekreftet kols henvisst Rusk Institute of Rehabilitation Medicine for lungerehabilitering.
Andel manglende respons	Oppgir n for de ulike analysene som er gjort
<b>COSMIN</b>	<b>Kritisk vurdering</b> <b>Resultat/kommentar</b>
Statistiske metoder benyttet basert på CTT eller IRT	CTT for de data som er aktuelle for oss
Internal consistency	Good Cronbach's alpha
Reliability	Fair Test-retest
Structurell validitet	Poor CFA støttet fem-faktor modellen. Tre av faktorene var moderat til høyt korrelert (r= 0.69 til 0.80). Modelltilpassning viste at den beste tilpassningen ble oppnådd med en løsning som omfattet fem distinkte og meningsbærende domener.
	Dyspnea intensity    n 0,71    103
	Dyspnea-related anxiety    n 0,31    99
	Fearful activity avoidance    n 0,20    91
	Self-efficacy for activity    n 0,09    92
	Satisfaction with strategy use    n 0,69    32
	p verdi <0,001 for alle

Interpretability	Satisfaction with strategy use	0,31	92
% av respondenter med lavest skår	Fra 0 til 1 % (n=1) for de fem subskalaene		
% av respondenter med høyest skår	Fra 0 til 2 % (n=2) for de fem subskalaene		
<b>Notater</b>			
Uthvalgene i Norweg et al. (2011) og Norweg et al. (2006) er delvis overlappende, men validerer ulike versjoner av DMQ.			
<b>Norweg, A., Ni, P., Garshick, E., O'Connor, G., Wilke, K. &amp; Jette, A. M. (2011) A multidimensional computer adaptive test approach to dyspnea assessment. Archives of Physical Medicine &amp; Rehabilitation, 92 (10), s. 1561-9.</b>			
Dataekstraksjon utført av Kjersti Solvåg 11.10.14      Referanse nr. 809			
<b>Kjennetegn ved studien</b>			
Formål	Uvikle å teste en prototype på en elektronisk tilpasset dyspneatest (CAT)		
Studiedesign	Prospektiv studie		
Land	USA		
Aktuelt kartleggingsverktøy	Dyspnea Management Questionnaire		
Finansiering	Ingen av betydning		
<b>Kjennetegn ved kartleggingsverktøyet</b>	<b>Beskrivelse</b>		
Navn	Dyspnea Management Questionnaire		
Forkortelse	DMQ-CAT		
Begrepet (beskrivelse)	Multidimensionale aspekt ved dyspnea		
Konseptuelt rammeverk	Refleksiv		
Opprinnelig er utviklet for	Pasienter med kols		
Antall spørsmål	20		
Dimensjoner	4		
	Dyspnea intensitety		
	Dyspnea anxiety		
	Activity avoidance		
	Activity self-efficacy		
Tidsreferanse for skår	Siste 2 uker		
Målestøle	6-punkt Likert-skala		
Administrasjonsform	Elektronisk, selv-administrert		
Tilgjengelighet til fulle tekst kopi	Samlingen av testledd er tilgjengelig som vedlegg til artikkelen.		
<b>Kjennetegn ved studiepopulasjonen (utvalget)</b>	<b>Beskrivelse</b>		
Antall	n=292 voksne med kols fra to medisinske senter		
Alder	Gjennomsnitt 68,4 år, range fra 40 til 92 år		



Kjønn (%)	71,6 % menn
Sykdomskarakteristika	Kols med pustebesvær ved aktivitet GOLD I 5,48 % II 43,49% III 39,4 % IV 11,99 % Mer detaljert informasjon tilgjengelig i artikkelens tabell 1.
Setting	Polikliniske
Land	USA
Språkversjon	Engelsk
Måte for utvelgelse	Bekvemmelighetsutvalg
Andel manglende respons	945 personer ble kontaktet. 119 var ikke kvalifisert for studiet. Undersøkelsens responsrate var 35,4 %
<b>COSMIN</b>	<b>Kritisk vurdering</b> <b>Resultat/kommentar</b>
Statistiske metoder benyttet basert på CTT eller IRT	Både CCT og IRT
Generelle krav til studier som har benyttet IRT modeller	<i>Excellent</i>
Internal consistency	Good Item internal consistency reliability ble undersøkt ved item-to-scale correlation ranges. Andel manglende svar ikke beskrevet. Ikke helt klart hvordan manglende svar er håndtert. Alle testledd korrelerte > 0.50 med sin foreslåtte skala. Interskala korrelasjoner var moderat til høy, med range fra 0,58 til 0,82.
Content validity	<i>Excellent</i> Den opprinnelige samlingen av 56 testledd ble revidert og utvidet til et kjernesett på 121 testledd, basert på 4 fokusgrupper med klinikere fra flere disipliner som var spesialister i lungemedisin og personer med kols, 2 kognitive testgrupper med voksne med kols, og en omfattende gjennomgang av litteratur. IRT ble benyttet for å utvikle nye og reviderte DMQ testledd.
Strukturell validitet	<i>Fair</i> EFA ble utført på de 121 testleddene. Deretter CFA på hver hypotiseret subskala og testledd reduksjon ble utført. Deretter ble EFA utført på de gjenværende 71 testledd for å sjekke at den foreslåtte strukturen ble opprettholdt. Tre ulike IRT modeller ble testet på de 71 testleddene. Modellen som passet dataene best var en multidimensjonal nonhierarkisk modell med 4 faktorer.

Får fair selv om utvalget er litt lite i forhold til det store antall spørsmål. Det er allikevel et stort utvalg på denne populasjonen.	4 faktor løsningen forklarte en samlet varians på 63,06 %.
Hypotheses testing	<i>Fair</i> Andel manglende svar ikke beskrevet. Ikke helt klart hvordan manglende svar er håndtert. Mangelfulle hypoteser.
Criterion validity	Good DMQ-CAT vs DMQ item pool Pearson's 0,94- 0,97
Interpretability	Distribusjon av (total) skår i utvalget <b>Notater</b> Skalaen ble endret fra en 7-punkts til en 6-punkts Likert skala. Dimensjonert «satisfaction with strategy use» fra det originale DMQ ble ikke tatt med i den reviderte samlingen av spørsmål. Det var benyttet simuleringsteknikk ved testing av DMQ-CAT.



Han, J. N., Xiong, C. M., Yao, W., Fang, Q. H., Zhu, Y. J., Cheng, X. S. & Van de Woestijne, K. P. (2011) <b>Multiple dimensions of cardiopulmonary dyspnea.</b> <i>Chinese Medical Journal</i> , 124 (20), s. 3220-6.	
Dataekstraksjon utført av Kjersti. Solvåg 09.10.14 Referanse nr. 815	
<b>Kjennetegn ved studien</b>	
Formål	Identifisere ulike dimensjoner av dyspne
Studiedesign	Valideringsstudie
Land	Kina
Aktuelt kartleggingsverktøy	«respiratorisk symptom sjekkliste»
Finansiering	Ingen av betydning
<b>Kjennetegn ved kartleggingsverktøyet</b>	<b>Beskrivelse</b> Beskrives bare som en «respiratorisk symptom sjekkliste»
Navn	
Forkortelse	
Begrepet (beskrivelse)	Multidimensjonale aspekt ved dyspne
Konseptuelt rammeverk	Refleksiv
Opprinnelig er utviklet for	Pasienter med cardiopulmonar dyspne
Antall spørsmål	54 testledd
Dimensjoner	3 dimensjoner: Sensory qualities of dyspnea (4 faktorer) Affective aspects of dyspnea (2 faktorer) Behavioral impact of dyspnea (1 faktor) I tillegg 2 faktorer under Andre symptomer
Tidsreferanse for skår	Siste måned
Måleskala	Hvor ofte har du opplevd... siste måned? 5 punkt Likert-skala (0= never, 1= rare, 2= sometimes, 3=often, 4=very often)
Administrasjonsform	Selv-administrert
Gjennomføring	Tid: flest mellom 10-15 minutter
<b>Kjennetegn ved studiens utvalg</b>	<b>Beskrivelse</b> 396 pasienter Astma n=65; kols n=58; DPLD n=53; primære hjertesykdommer med kronisk hjertesvikt n=62; MUD (medisinsk uforklart dyspne) n=105
Antall	
Alder	Varierer for de ulike diagnosegruppene. Informasjon finnes i artikkelens tabell 1.
Kjønn (%)	Varierer for de ulike diagnosegruppene. Informasjon finnes i artikkelens tabell 1.
Sykdomskarakteristika	Dyspne som følge av ulike baktenforliggende sykdommer Mer detaljert informasjon tilgjengelig i artikkelen under kap. Subjects.
Setting	Sykehus. Sannsynligvis en blanding av inneliggende og polikliniske
Land	Kina
Språkversjon	Antatt kinesisk
<b>COSMIN</b>	<b>Kritisk vurdering</b> <b>Resultat/kommentar</b>

Statistiske metoder benyttet CTT  
basert på CTT eller IRT

Strukturell validitet

Excellent

Faktoranalyse med orthogonal varimax rotasjon ble utført på 54 av de 61 utsagn om dyspne og viste 12 sammenhengende og ukorreleerte faktorer basert på minimum egenvalue kriteriet (>1) som forklarer 64 % av total varians. Minimum egenvalue kriteriet beholder noen ganger for mange faktorer. Ved å benytte ytterligere interpretability kriterier ble 9 faktorer identifisert ved høye ladinger av et unikt sett med karakteristika. 7 faktorer synes å måle tre dimensjoner av dyspne: sensoriske kvaliteter (4 faktorer), affektive aspekt (2 faktorer) og atferdsmessig påvirkning (1 faktor). De resterende to faktorene «tingling» og «cough» sorterte under andre symptomer.  
De 7 karakteristika som ikke er tatt med i FA har i tidligere studier blitt funnet irrelevant i forhold til pustebesvær.

## Vedlegg VI Oversikt over metodisk kvalitet og måleegenskaper fra hver studie

PROM	Studie	Intern konsistens		Reliabilitet		Innholdsvaliditet		Strukturell validitet		Hypotesetesting		Kriterievaliditet		Tverrkulturell validitet	
		M	R	M	R	M	R	M	R	M	R	M	R	M	R
<b>CDS</b>	Tanaka et al., 2000	Utmerket	+	Rimelig	-	Utmerket	+	Utmerket	?	God	+				
	Tanaka et al., 2002									God	-				
	Henoch et al., 2006	God	+					Utmerket	?	God	+				
	Uronis et al., 2012	Utmerket	+					God	?	God/Rimelig'	-				
<b>SRI</b>	Windisch et al., 2003					Utmerket	+			God	+				
	Duierveman et al., 2008	Svak	+	Svak	+					God	+				
	Lopez-Campos et al., 2008	Svak	+	Rimelig	+			Svak	+	Rimelig	-				
	Windisch et al., 2008	Svak	+					Svak	+						
	Ghosh et al., 2012	Svak	+					Svak	+	God	+				
	Struik et al., 2013	Svak	+							God	+				
	Markussen et al., 2014	Svak	+			God	+			God	+			Svak	-
<b>DMQ-30</b>	Norweg et al., 2006	God	+	Svak	+	Utmerket	+			God/Svak''	+				
<b>DMQ-56</b>	Norweg et al., 2011a	God	+	Rimelig	+			Svak	+						
<b>DMQ-CAT</b>	Norweg et al., 2011b	God	+			Utmerket	+	Rimelig	+	Rimelig	+	God	+		
<b>RSC</b>	Han et al., 2011							Utmerket	+						

PROM (Patient-reported Outcome Measures); M (metodisk kvalitet); R (måleegenskapens resultat); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); ' (sammenligningene med VAS (Visual Analog Scale), BORG, ECOG (Eastern Cooperative Oncology Group) / HADS (Hospital Anxiety and Depression Scale) ,FACT (Functional Assessment of Cancer Therapy), SpO2 (perifer oksygenmetning); '' (sammenligningene med SOLQ (Seattle Obstructive Lung Disease Questionnaire), SF-12 (Short Form Health Survey) / HADS-angst.

## Vedlegg VII

## Oversikt over tilgjengelige samvariasjoner for hvert kartleggingsverktøy

PROM	CDS	SRI	DMQ-30	DMQ-56	DMQ-CAT
BDI		x			
CCQ		x			
CRQ		x			x
CSES					x
EORTC QLQ-LC13**	x				
EORTC QLQ-C30*	x				
FACT-L	x				
GARS		x			
HADS	x	x	x		x
Modifisert Borg Skala	x				
MRC		x			
SF-12			x		
SF-36		x			
SOLQ			x		
STAI	x				
USCD SOBQ					x
VAS-D	x				
VAS -Dis	x				
Fysiske/objektive mål	x	x			
Sosio-demografiske variabler	x				

PROM (Patient-reported Outcome Measures); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); BDI (Baseline Dyspnoea Index); CCQ (Clinical COPD Questionnaire); CRQ (Chronic Respiratory Disease Questionnaire); CSES (COPD Self-Efficacy Scale); EORTC QLQ (European Organization for Research and Treatment of Cancer)-LC13; EORTC QLQ-C30; FACT-L (Functional Assessment of Cancer Therapy-Lung); GARS (Groningen Activity and Restriction Scale); HADS (Hospital Anxiety and Depression Scale); MRC (Medical Research Council dyspnoea scale); SF (Short Form Health Survey)-12; SF-36; SOLQ (Seattle Obstructive Lung Disease Questionnaire); STAI (State-Trait Anxiety Inventory); USCD SOBQ (University of California, San Diego Shortness of Breath Questionnaire); VAS-D (Visual Analogue Scale of Dyspnea); VAS-Dis (VAS Discomfort).

## Vedlegg VIII      **Beskrivelse av kartleggingsverktøy brukt til sammenligninger**

Forkortelse	Fullt navn	Beskrivelse
<b>Borg skala</b>		Måler pustebesværets intensitet på en tolv-punkt skala (Tanaka et al., 2000).
<b>CRQ</b>	Chronic Respiratory Disease Questionnaire	Kartleggingsverktøy som måler fysisk og følelsesmessig funksjon hos pasienter med kroniske lungesykdommer (Struik et al., 2013; <a href="http://www.proqolid.org/">http://www.proqolid.org/</a> ).
<b>CSES</b>	COPD Self-Efficacy Scale	CSES kartlegger tiltro til egen mestringsevne hos pasienter med kols (Wigal, Creer & Kotses, 1991).
<b>GARS</b>	Groningen Activity Restriction Scale	Generisk kartleggingsverktøy som kartlegger funksjonssvikt i daglige aktiviteter (ADL) og mer komplekse daglige aktiviteter (instrumentell ADL) ( <a href="http://www.proqolid.org/">http://www.proqolid.org/</a> ).
<b>HADS</b>	Hospital Anxiety and Depression Scale	Et screeningverktøy som måler symptomer på angst og depresjon ( <a href="http://www.proqolid.org/">http://www.proqolid.org/</a> ).
<b>MRC</b>	Medical Research Council dyspnoea scale	MRC graderer effekten av pustebesvær på daglige aktiviteter (Bestall et al., 1999).
<b>SF-36</b>	Short Form Health Survey	Generisk kartleggingsverktøy som måler ulike aspekt av helsestatus relevant på tvers av alder, sykdom og behandlingsgrupper. (Ghosh et al., 2012; <a href="http://www.proqolid.org/">http://www.proqolid.org/</a> )
<b>SF-12</b>	Short Form Health Survey	Kortform av SF-36 ( <a href="http://www.proqolid.org/">http://www.proqolid.org/</a> ).
<b>UCSD SOBQ</b>	University of California, San Diego Shortness of Breath Questionnaire	Kartlegger alvorlighetsgraden av pustebesvær ved ulike daglige aktiviteter (ADL) (Eakin et al., 1998)
<b>SOLQ</b>	Seattle Obstructive Lung Disease Questionnaire	SOLQ er et kartleggingsverktøy som måler helserelatert livskvalitet hos pasienter med kols ( <a href="http://www.proqolid.org/">http://www.proqolid.org/</a> )
<b>VAS-dyspné</b>	Visual Analogue Scale	Måler pustebesværets intensitet på en 100-mm linje (Tanaka et al., 2000)

Kols (kronisk obstruktiv lungesykdom)

## Vedlegg IX Oversikt over samvariasjoner mellom SRI og SF-36

	SF-36 PF	SF-36 RP	SF-36 BP	SF-36 GH	SF-36 VT	SF-36 SF	SF-36 RE	SF-36 MH	SF-36 PHC	SF-36 MHC
<b>SRI-RC</b>	0.38 <sup>'''</sup>	0.52 <sup>'''</sup>	0.36 <sup>'''</sup>	0.63 <sup>'''</sup>	0.52 <sup>'''</sup>	0.52 <sup>'''</sup>	0.40 <sup>'''</sup>	0.29 <sup>'''</sup>	0.53 <sup>'''</sup>	0.40 <sup>'''</sup>
	0.26 <sup>***</sup>	0.54 <sup>***</sup>	0.49 <sup>***</sup>	0.61 <sup>***</sup>	0.63 <sup>***</sup>	0.59 <sup>***</sup>	0.48 <sup>***</sup>	0.43 <sup>***</sup>		
	0.37 <sup>''</sup>	0.43 <sup>''</sup>	0.16 <sup>''</sup>	0.47 <sup>''</sup>	0.54 <sup>''</sup>	0.49 <sup>''</sup>	0.16 <sup>''</sup>	0.24 <sup>''</sup>	0.48 <sup>''</sup>	0.28 <sup>''</sup>
	0.34 <sup>*</sup>	0.41 <sup>*</sup>	0.38 <sup>*</sup>	0.51 <sup>*</sup>	0.50 <sup>*</sup>	0.41 <sup>*</sup>	0.36 <sup>*</sup>	0.42 <sup>*</sup>	0.42 <sup>*</sup>	0.45 <sup>*</sup>
<b>SRI-PF</b>	0.73 <sup>'''</sup>	0.66 <sup>'''</sup>	0.25 <sup>'''</sup>	0.54 <sup>'''</sup>	0.35 <sup>'''</sup>	0.40 <sup>'''</sup>	0.43 <sup>'''</sup>	0.27 <sup>'''</sup>	0.61 <sup>'''</sup>	0.29 <sup>'''</sup>
	0.56 <sup>***</sup>	0.41 <sup>***</sup>	0.35 <sup>***</sup>	0.57 <sup>***</sup>	0.53 <sup>***</sup>	0.57 <sup>***</sup>	0.26 <sup>***</sup>	0.38 <sup>***</sup>		
	0.75 <sup>''</sup>	0.52 <sup>''</sup>	0.09 <sup>''</sup>	0.50 <sup>''</sup>	0.60 <sup>''</sup>	0.51 <sup>''</sup>	0.21 <sup>''</sup>	0.32 <sup>''</sup>	0.66 <sup>''</sup>	0.28 <sup>''</sup>
	0.65 <sup>*</sup>	0.48 <sup>*</sup>	0.32 <sup>*</sup>	0.54 <sup>*</sup>	0.55 <sup>*</sup>	0.50 <sup>*</sup>	0.38 <sup>*</sup>	0.41 <sup>*</sup>	0.54 <sup>*</sup>	0.44 <sup>*</sup>
<b>SRI-AS</b>	0.17 <sup>'''</sup>	0.21 <sup>'''</sup>	0.49 <sup>'''</sup>	0.36 <sup>'''</sup>	0.44 <sup>'''</sup>	0.33 <sup>'''</sup>	0.25 <sup>'''</sup>	0.32 <sup>'''</sup>	0.34 <sup>'''</sup>	0.27 <sup>'''</sup>
	0.19 <sup>***</sup>	0.41 <sup>***</sup>	0.61 <sup>***</sup>	0.54 <sup>***</sup>	0.62 <sup>***</sup>	0.54 <sup>***</sup>	0.48 <sup>***</sup>	0.57 <sup>***</sup>		
	0.38 <sup>''</sup>	0.40 <sup>''</sup>	0.48 <sup>''</sup>	0.50 <sup>''</sup>	0.53 <sup>''</sup>	0.40 <sup>''</sup>	0.24 <sup>''</sup>	0.34 <sup>''</sup>	0.56 <sup>''</sup>	0.28 <sup>''</sup>
	0.25 <sup>*</sup>	0.24 <sup>*</sup>	0.44 <sup>*</sup>	0.32 <sup>*</sup>	0.48 <sup>*</sup>	0.35 <sup>*</sup>	0.21 <sup>*</sup>	0.49 <sup>*</sup>	0.34 <sup>*</sup>	0.35 <sup>*</sup>
<b>SRI-SR</b>	0.27 <sup>'''</sup>	0.43 <sup>'''</sup>	0.48 <sup>'''</sup>	0.48 <sup>'''</sup>	0.58 <sup>'''</sup>	0.66 <sup>'''</sup>	0.42 <sup>'''</sup>	0.52 <sup>'''</sup>	0.47 <sup>'''</sup>	0.58 <sup>'''</sup>
	0.14 <sup>***</sup>	0.48 <sup>***</sup>	0.45 <sup>***</sup>	0.59 <sup>***</sup>	0.62 <sup>***</sup>	0.61 <sup>***</sup>	0.41 <sup>***</sup>	0.71 <sup>***</sup>		
	0.29 <sup>''</sup>	0.35 <sup>''</sup>	-0.09 <sup>''</sup>	0.34 <sup>''</sup>	0.40 <sup>''</sup>	0.49 <sup>''</sup>	0.28 <sup>''</sup>	0.54 <sup>''</sup>	0.17 <sup>''</sup>	0.53 <sup>''</sup>
	0.25 <sup>*</sup>	0.24 <sup>*</sup>	0.32 <sup>*</sup>	0.37 <sup>*</sup>	0.51 <sup>*</sup>	0.56 <sup>*</sup>	0.28 <sup>*</sup>	0.61 <sup>*</sup>	0.21 <sup>*</sup>	0.57 <sup>*</sup>
<b>SRI-AX</b>	0.30 <sup>'''</sup>	0.41 <sup>'''</sup>	0.45 <sup>'''</sup>	0.50 <sup>'''</sup>	0.44 <sup>'''</sup>	0.58 <sup>'''</sup>	0.44 <sup>'''</sup>	0.49 <sup>'''</sup>	0.42 <sup>'''</sup>	0.55 <sup>'''</sup>
	0.20 <sup>***</sup>	0.54 <sup>***</sup>	0.50 <sup>***</sup>	0.58 <sup>***</sup>	0.66 <sup>***</sup>	0.59 <sup>***</sup>	0.49 <sup>***</sup>	0.63 <sup>***</sup>		
	0.15 <sup>''</sup>	0.40 <sup>''</sup>	0.10 <sup>''</sup>	0.48 <sup>''</sup>	0.37 <sup>''</sup>	0.43 <sup>''</sup>	0.22 <sup>''</sup>	0.37 <sup>''</sup>	0.29 <sup>''</sup>	0.37 <sup>''</sup>
	0.26 <sup>*</sup>	0.35 <sup>*</sup>	0.31 <sup>*</sup>	0.51 <sup>*</sup>	0.56 <sup>*</sup>	0.45 <sup>*</sup>	0.38 <sup>*</sup>	0.51 <sup>*</sup>	0.28 <sup>*</sup>	0.52 <sup>*</sup>
<b>SRI-WB</b>	0.28 <sup>'''</sup>	0.48 <sup>'''</sup>	0.53 <sup>'''</sup>	0.63 <sup>'''</sup>	0.72 <sup>'''</sup>	0.70 <sup>'''</sup>	0.58 <sup>'''</sup>	0.64 <sup>'''</sup>	0.43 <sup>'''</sup>	0.71 <sup>'''</sup>
	0.22 <sup>***</sup>	0.54 <sup>***</sup>	0.52 <sup>***</sup>	0.66 <sup>***</sup>	0.68 <sup>***</sup>	0.66 <sup>***</sup>	0.50 <sup>***</sup>	0.78 <sup>***</sup>		
	0.30 <sup>''</sup>	0.51 <sup>''</sup>	0.19 <sup>''</sup>	0.59 <sup>''</sup>	0.57 <sup>''</sup>	0.60 <sup>''</sup>	0.37 <sup>''</sup>	0.74 <sup>''</sup>	0.35 <sup>''</sup>	0.64 <sup>''</sup>
	0.35 <sup>*</sup>	0.34 <sup>*</sup>	0.41 <sup>*</sup>	0.57 <sup>*</sup>	0.67 <sup>*</sup>	0.62 <sup>*</sup>	0.39 <sup>*</sup>	0.79 <sup>*</sup>	0.31 <sup>*</sup>	0.69 <sup>*</sup>
<b>SRI-SF</b>	0.42 <sup>'''</sup>	0.61 <sup>'''</sup>	0.45 <sup>'''</sup>	0.59 <sup>'''</sup>	0.52 <sup>'''</sup>	0.66 <sup>'''</sup>	0.45 <sup>'''</sup>	0.37 <sup>'''</sup>	0.59 <sup>'''</sup>	0.49 <sup>'''</sup>
	0.33 <sup>***</sup>	0.61 <sup>***</sup>	0.48 <sup>***</sup>	0.62 <sup>***</sup>	0.69 <sup>***</sup>	0.80 <sup>***</sup>	0.53 <sup>***</sup>	0.60 <sup>***</sup>		
	0.65 <sup>''</sup>	0.55 <sup>''</sup>	0.21 <sup>''</sup>	0.54 <sup>''</sup>	0.62 <sup>''</sup>	0.65 <sup>''</sup>	0.23 <sup>''</sup>	0.51 <sup>''</sup>	0.60 <sup>''</sup>	0.46 <sup>''</sup>
	0.45 <sup>*</sup>	0.45 <sup>*</sup>	0.35 <sup>*</sup>	0.51 <sup>*</sup>	0.63 <sup>*</sup>	0.65 <sup>*</sup>	0.39 <sup>*</sup>	0.56 <sup>*</sup>	0.41 <sup>*</sup>	0.62 <sup>*</sup>
<b>SRI-SS</b>	0.45 <sup>'''</sup>	0.62 <sup>'''</sup>	0.58 <sup>'''</sup>	0.70 <sup>'''</sup>	0.65 <sup>'''</sup>	0.74 <sup>'''</sup>	0.56 <sup>'''</sup>	0.54 <sup>'''</sup>	0.62 <sup>'''</sup>	0.61 <sup>'''</sup>
	0.70 <sup>''</sup>	0.62 <sup>''</sup>	0.26 <sup>''</sup>	0.66 <sup>''</sup>	0.76 <sup>''</sup>	0.69 <sup>''</sup>	0.26 <sup>''</sup>	0.60 <sup>''</sup>	0.67 <sup>''</sup>	0.54 <sup>''</sup>
	0.47 <sup>*</sup>	0.49 <sup>*</sup>	0.46 <sup>*</sup>	0.63 <sup>*</sup>	0.73 <sup>*</sup>	0.65 <sup>*</sup>	0.45 <sup>*</sup>	0.70 <sup>*</sup>	0.46 <sup>*</sup>	0.67 <sup>*</sup>

SRI (Severe Respiratory Insufficiency Questionnaire); RC (respiratory complaints); PF (physical functioning); AS (attendant symptoms and sleep); SR (social relationships); AX (anxiety); WB (psychosocial well-being); SF (social functioning); SS (summary scale); SF (Short Form Health Survey )-36; PF (physical functioning); RP (role-physical); BP (body pain); GH (general health); VT (vitality); SF (social functioning); RE (role-emotiona); MH (mental health); PHC (physical health component); MHC (mental health component); <sup>'''</sup>(Markussen et al.,2014); <sup>\*\*\*</sup> (Gosh et al.,2012); <sup>''</sup> (Lopez-Campos et al.,2008); <sup>\*</sup>(Windisch et al.,2003).

## Vedlegg X Oversikt over samvariasjoner mellom SRI og CRQ, HADS, GARS og MRC

	CRQ-Dyspne	CRQ-Fatigue	CRQ-Emotion	CRQ-Mastery	CRQ-Total	HADS-anxiety	HADS-depression	GARS	MRC
<b>SRI-RC</b>	0.57** 0.56'	0.59** 0.60'	0.40** 0.45'	0.34** 0.50'	0.54** 0.62'	0.50** 0.56'	0.41** 0.43'	0.49** 0.54'	0.43** 0.41'
<b>SRI-PF</b>	0.46** 0.57'	0.61** 0.49'	0.47** 0.37'	0.42** 0.42'	0.59** 0.54'	0.51** 0.46'	0.54** 0.50'	0.86** 0.74'	0.59** 0.49'
<b>SRI-AS</b>	0.42'	0.39** 0.46'	0.38** 0.48'	0.27** 0.43'	0.39** 0.53'	0.39** 0.56'	0.26** 0.42'	0.26'	0.27'
<b>SRI-SR</b>	0.44** 0.28'	0.57** 0.33'	0.59** 0.56'	0.45** 0.53'	0.63** 0.56'	0.56** 0.59'	0.61** 0.55'	0.53** 0.21'	0.30**
<b>SRI-AX</b>	0.35** 0.49'	0.57** 0.52'	0.65** 0.67'	0.60** 0.70'	0.67** 0.74'	0.69** 0.74'	0.54** 0.59'	0.40** 0.38'	0.26** 0.33'
<b>SRI-WB</b>	0.38** 0.45'	0.60** 0.39'	0.79** 0.77'	0.70** 0.64'	0.79** 0.72'	0.79** 0.78'	0.68** 0.71'	0.36** 0.26'	0.28'
<b>SRI-SF</b>	0.42** 0.52'	0.57** 0.52'	0.51** 0.51'	0.41** 0.49'	0.60** 0.62'	0.54** 0.60'	0.58** 0.65'	0.66** 0.54'	0.43** 0.51'
<b>SRI-SS</b>	0.53** 0.63'	0.74** 0.63'	0.73** 0.71'	0.62** 0.71'	0.81** 0.82'	0.75** 0.80'	0.68** 0.72'	0.65** 0.56'	0.41** 0.46'

SRI (Severe Respiratory Insufficiency Questionnaire); RC (respiratory complaints); PF (physical functioning); AS (attendant symptoms and sleep); SR (social relationships); AX (anxiety); WB (psychosocial well-being); SF (social functioning); SS (summary scale); CRQ (Chronic Respiratory Disease Questionnaire); HADS (Hospital Anxiety and Depression Scale); GARS (Groningen Activity and Restriction Scale); MRC (Medical Research Council dyspnoea scale); \*\* (Duiverman et al., 2008); '(Struik et al., 2013).