

Grundliggende principper for valg og anvendelse af test og målemetoder i fysioterapi

Mette Aadahl, fysioterapeut MPH, Hans Lund, fysioterapeut ph.d.

Aadahl M, Lund H (2003, 20. oktober).

Grundliggende principper for valg og anvendelse af test og målemetoder i fysioterapi. *Forskning i Fysioterapi(online)*, (1. årg.), s. 1-9. URL: <http://www.ffy.dk/sw599.asp>

Hvorfor måle fysioterapi?

I daglig praksis har vi brug for systematisk at vurdere, beskrive og kommunikere relevante fund og resultater i forbindelse med alle former for fysioterapi-interventioner. For den enkelte patient har det groft sagt størst betydning, om det går "godt" i en specifik behandlingsseance, dvs. om et behandlingsresultat er tilfredsstillende og fører til f.eks. smertefrihed og forventet funktionsniveau. Fysioterapeuten har derimod brug for stringent at indsamle information, hvis hun skal være i stand til at anvende sine erfaringer og til løbende at vurdere og justere en behandlingsindsats.

Som faggruppe har vi traditionelt arbejdet efter en høj grad af "mental rutine" dvs. vi har vurderet, justeret og intervenseret uden systematisk at formulere og redegøre skriftligt for fund og resultater. Et udsagn som "Patienten har færre smerter og bedre funktionsniveau" er uspecifikt og rummer meget lidt information.

Den behandlende fysioterapeut ved måske nok, hvad udsagnet rummer, men enhver anden modtager af informationen tolker udsagnet efter egne forudsætninger. Den faktiske behandling kan være af nok så høj kvalitet, men som ud-

gangspunkt for f.eks. planlægning af opfølgende indsats er kommunikationen af behandlingsresultatet utilstrækkelig.

Den erfaring, vi indsamler i arbejdet med patienter som enkelte fagpersoner, såvel som den kollektive erfaring vi som faggruppe akkumulerer, kan kun anvendes til at udvikle faget, såfremt vi gør informationerne målbare og til at kommunikere.

Med andre ord er vi ofte nødt til at kvantificere og kategorisere for at kunne generalisere. Generalisering eller det at gøre informationerne almenne er et udtryk for, i hvor høj grad specifik information om enkelte patienter er alment gældende for grupper af patienter. Vil vi kunne udsige noget om gruppen af knæpatienter, må vi indsamle og gemme vores erfaringer med enkelte knæpatienter på en struktureret måde. Derfor må vi ofte ty til kvantificering af informationerne.

En standardisering og kvantificering af udsagnet "Patienten har færre smerter" kunne eksempelvis se ud som i figur 1, hvor informationerne om behandlingsresultatet er entydigt kommunikeret, systematisk registreret og dokumenteret. Erfaringen med den specifikke patient kan lagres og anvendes som systematisk indsamlet viden, der

Figur 1

	Før behandling	Efter behandling
Visual analog skala (smerte score)	67 mm	12 mm

kan danne grundlag for kollektiv faglig erfaring med en specifik type behandling til en bestemt type patient.

Målemetoders begrænsninger

For patienten er det stadig mest relevant, om behandlingsresultatet er "godt", og standardiserede test og målemetoder må aldrig skygge for en subjektiv helhedsvurdering af patient og behandlingsindsats. Et testsvar erstatter aldrig den interpersonelle kontakt og den unikke patients behov for empati og individuelt tilpasset behandling. Dybest set vil vi aldrig kunne måle det, vi egentlig ønsker at måle: Hvordan har patienten det? Dvs. hvordan er patientens fysiske funktion, livskvalitet, livssituation, almene velbefindende osv.?

Helheden rummer så komplekse dele, at vi aldrig vil kunne måle eller redegøre for alle elementerne. Vi kan kun måle "toppen af isbjerget" ved at udvælge enkelte aspekter af helheden og gøre disse målbare. Visse aspekter er velegnede til kvantificering, mens andre er vanskelige eller umulige at kvantificere. F.eks. kan ledbevægelighed forholdsvis let kvantificeres, mens træthed eller vævsforskydelighed er langt vanskeligere at kvantificere.

Tre formål med at måle

Som hovedregel findes tre overordnede formål med at anvende test og målemetoder. Vi kan ønske at diagnosticere, dvs. at be- eller afkræfte tilstedeværelsen af en bestemt sygdom eller tilstand, vi kan ønske at prædikere, dvs. forudsige en udvikling eller udtale os om en sandsynlig prognose. Endelig kan vi ønske at effektmåle, dvs. vurdere en ændring over tid, som oftest i form af vurdering af behandlingseffekt ved hjælp af måling før og efter et fysioterapiforløb. Principielt bestemmer undersøgelses- eller testsammenhængen, om der er tale om en diagnostisk, en prognostisk eller en effektmålstest. Ofte vil en fysioterapeutisk test eller et spørgeskema dog være udviklet med et af ovenstående formål for øje, og vil ikke umiddelbart kunne anvendes til andre formål. Et teknisk instrument derimod vil ofte i forskellig kontekst kunne anvendes med alle tre formål for øje. F.eks. vil måling af nerveledningshastighed kunne diagnosticere en perifer nervelæsion. Målingen vil kunne anvendes prognostisk: Vil

tilstanden sandsynligvis remittere eller er det en permanent skade? Endelig vil gentagne målinger i et behandlingsforløb kunne anvendes til at måle effekten af behandlingen.

Diagnostisk test

Diskriminere mellem personer.

Man kan bruge en målemetode til at diagnosticere eller klassificere en patient inden for en bestemt sygdom eller tilstand. Som eksempel kan nævnes klassificering af rygpatienter ad modum McKenzie, hvor klassificeringen af patientens ryglidelse i enten Posturalt syndrom, Dysfunktions syndrom eller Derangement anvendes som rettesnor for valg af fysioterapeutisk behandling.

Prognostisk test

Prædiktion af en udvikling.

Her er formålet med målemetoden at kunne forudsige en sandsynlig udvikling i en tilstand eller sygdom på baggrund af et testresultat. Som eksempel kan nævnes vurdering af sværhedsgrad af klumpfod. Vurdering af sværhedsgrad af klumpfod hos en nyfødt giver en sandsynlig prognose for, hvordan tilstanden vil udvikle sig: Hvad forventer vi, der vil ske med den nyfødte? Skal han opereres? Skal han have skinner? Fysioterapi? Hvordan vil den motoriske udvikling være? Kan han hoppe, når han bliver syv år gammel?

Effektmål

Vurdering af forandring over tid (effekt af behandling).

Som eksempel på anvendelsen af en målemetode til at vurdere effekten af en behandling kan nævnes en muskelstyrketest. Ved starten af en fysioterapibehandling måles muskelstyrken i f.eks. m. quadriceps hos en knæpatient for at kunne vurdere effekt af behandlingen på muskelstyrken løbende og ved afslutning af behandlingen.

Inddeling af målemetoder

Overordnet kan man inddele målemetoder alt efter om metoden er udviklet til en specifik gruppe af patienter (f.eks. reumatoid artrit patienter), i så fald er der tale om en *sygdomsspecifik målemetode*, om metoden er udviklet til at måle en bestemt *dimension* (f.eks. smerte eller dagligdags funktion) eller om metoden er udviklet til mere generelt at måle et sammensat begreb (f.eks. generel sygdomsopfattelse eller helbredsrelateret livskvalitet). I sidstnævnte tilfælde

er der tale om en såkaldt *generisk målemetode*. Hovedpointen med denne inddeling er, at man som bruger er opmærksom på begrænsningerne og mulighederne ved den enkelte målemetode.

En anden inddeling, som også kan være til stor hjælp i vurderingen af om en målemetode skal anvendes eller ikke, er en inddeling set i forhold til ICF klassifikationen (International Classification of Functioning, Disability and Health). Her kan det være af stor betydning at kunne skelne mellem såkaldte surrogatmål og real endpoints. Et surrogatmål er, holdt op imod ICF, en målemetode, der befinder sig på impairment niveauet (f.eks. muskelstyrke, ledbevæglighed o.l.), mens et real endpoint befinder sig på enten et activity eller et participation niveau. Et real endpoint er et mål for noget, som er interessant set fra patientens synsvinkel (f.eks. smerte, tilbagevenden til arbejde o.l.), mens et surrogatmål er interessant set fra terapeutens synsvinkel. Det er typisk sådan, at vi som behandlere interesserer os for surrogatmålene, fordi de hjælper os til at forstå mekanismerne bag sygdommen eller smerten.

Generelle forudsætninger

Terminologien i forbindelse med test og målemetoder er mangfoldig, og det er vigtigt at kende ordenes specifikke betydning, når man skal finde, vurdere og anvende en relevant test eller målemetode. I tabel 1 ses definition af nøgleord og begreber inden for området, som de anvendes i denne artikel.

Skalatyper

Når vi tester eller anvender målemetoder, producerer vi standardiseret information eller data, og disse data måles på forskellige måleskalaer eller måleniveauer. Anvender vi et målebånd som målemetode, vil vores data være målt i centimeter og tilhører én bestemt type skala: interval- eller ratioskalaen. Måler vi derimod muskulær stivhedsgrad ad modum Bunkan, vil data være målt i rangordnede kategorier, og data tilhører en anden type skala: ordinalskalaen.

Typen af data afhænger af den valgte målemetode, og der skelnes mellem tre typer måleskalaer eller måleniveauer: nominalskala, ordinal-

Tabel 1: Definition af nøgleord og begreber

Målemetode	Et specifikt redskab til at måle et eller andet
Test	En specifik og standardiseret målemetode
Måling	Selve handlingen hvor testen anvendes som målemetode.
Variabel	Den størrelse eller det aspekt som testen anvender
Data	Det resultat der kommer ud af at anvende en test med sine specifikke variabler

Tabel 2: Måleskalaer

Nominalskala	Data repræsenterer kategorier, dvs. de repræsenterer ikke en egentlig numerisk værdi og kategorierne er ikke rangordnede i forhold til hinanden. F.eks: Fordeling af hårfarve blandt en gruppe kvinder
Ordinalskala	Data kan rangordnes, dvs. data repræsenterer kategorier, som er rangordnede i forhold til hinanden. Vi kender dog ikke afstanden fra et punkt på skalaen til et andet. F.eks.: Vurdering af grad af tilfredshed med behandling: Meget tilfreds, tilfreds, mindre tilfreds, ikke tilfreds.
Intervalskala og ratioskala	Intervalskalaen beskriver data, som ikke alene kan rangordnes, men hvor vi desuden kan antage, at afstanden mellem to scores er forskellig fra eller lig med afstanden mellem to andre scores. F.eks.: Måling af temperatur på celsius temperaturskala. Denne skala har ikke et egentligt 0-punkt. Temperaturen 0 beskriver ikke fraværet af varme, men blot den temperatur hvor vand fryser til is. På Ratioskalaen er der endvidere tale om et egentligt 0-punkt. F.eks.: Måling af benlængde eller rækkeindeks i centimeter.

skala og intervallskala/ratioskala (Tabel 2). Som huskeregel kan man tænke på ordet "NOIR", hvor forbogstaverne i de forskellige skalatyper indgår: (Nominalskala, Ordinalskala, Intervallskala og Ratioskala).

Interval- og rationskalaen repræsenterer det højeste målniveau, fordi denne type data tillader talmæssig bearbejdning i et større omfang end data målt på de øvrige skalaer. Det er vigtigt at være opmærksom på, hvilken type skala man har at gøre med, da typen af data er afgørende for valg af statistisk metode ved eventuelt senere statistisk bearbejdning af de indsamlede informationer.

Validitet

En test, et spørgeskema eller et andet måleinstruments validitet eller gyldighed er et udtryk for den pågældende målemetodes evne til at måle det, den tilstræber at måle. Validitetsbegrebet kan inddeles i en række underbegreber, som tilsammen er bestemmende for et givent måleinstruments samlede validitet. Vi foreslår her anvendelse af de engelske validitetsbetegnelser. Dels er en stor del af den eksisterende litteratur om test og målemetoder på engelsk, dels ses en tendens til, at validitetsbegreberne oversættes forskelligt i danske tekster, hvilket bidrager til begrebsforvirringen på området. Construct validity oversættes f.eks. både med "konstruktionsvaliditet" og "begrebsvaliditet".

Validitetsbegreber

Criterion Validity betragtes ofte som det klassiske validitetsbegreb (dansk: kriterievaliditet). Måleinstrumentet valideres ved at sammenstille det med et afgørende kriterium, en såkaldt "guld standard", der kan opfattes som facitlisten eller den målemetode, der regnes for den bedste metode til vurdering af det fænomen, vi ønsker at måle med den aktuelle test eller målemetode. I praksis er det sjældent muligt at validere en målemetode mod en "guld standard", og vi vil i stedet være nødt til at undersøge concurrent validitet.

Concurrent Validity (samstemmende validitet) er en form for kriterievaliditet, hvor det aktuelle måleinstrument vurderes i forhold til et andet tilsvarende instrument. Anvendes især, når der ikke

findes en egentlig "guld standard", eller når denne er for dyr, for invasiv eller for kompliceret at administrere.

Predictive Validity (prædiktiv validitet) refererer til et instruments evne til at forudsige noget centralt vedrørende det fænomen vi ønsker at måle.

Face Validity (overflade-validitet) bruges især om spørgeskemaer og refererer til, hvordan spørgeskemaet umiddelbart ser ud og virker. Kan spørgeskemaet ud fra en umiddelbar betragtning udsige noget om det, vi ønsker at undersøge? Ser spørgsmålene fornuftige og rimelige ud? Face validity kan ikke måles, men vurderes ved f.eks. interviews med potentielle respondenter/brugere.

Construct Validity kaldes på dansk konstruktionsvaliditet eller mere præcist begrebsvaliditet. Begrebet er især relevant, når det drejer sig om måling af komplekse fænomener eller tilstande som f.eks. træthed, depression, angst osv., der typisk vil søges målt med et spørgeskema. Begrebsvaliditet refererer til, om de svar, vi indhenter, kan siges at repræsentere "den latente variabel", dvs. den underliggende tilstand vi egentlig ønsker at undersøge. Denne type validitet undersøges bedst ved at opstille og afprøve hypoteser på baggrund af eksisterende teorier på området.

Vi kunne f.eks. undersøge en hypotese om at en person, der scorer lavt på en skala, der måler eksamensskræk også scorer lavt på en skala, der måler generel angst. Vi ville forvente at finde overensstemmelse mellem de to skalaer. Finder vi ikke overensstemmelse, betyder det dog ikke nødvendigvis, at vores skala har lav validitet. Det kan også betyde, at vi ikke har opstillet den rigtige hypotese, og at der i realiteten ikke findes en sammenhæng mellem eksamensskræk og generel angst.

Tæt knyttet til begrebsvaliditet findes begrebet Content Validity eller på dansk indholdsvaliditet. Begrebet bruges om spørgeskemaer eller test. Det refererer til, om skemaet eller testen dækker alle relevante aspekter af det emne eller område, vi ønsker at måle med skemaet eller testen. Dvs. er udvalget af spørgsmål eller elementer i

Table 3: Validitetsbegreber

Criterion-based validity	Kriterievaliditet
Concurrent validity	Samstemmende validitet
Predictive validity	Prædiktiv validitet
Face validity	Overfladevaliditet
Construct validity	Begrebs- eller konstruktionsvaliditet
Content validity	Indholdsvaliditet

testen relevant og repræsentativt for det område, vi faktisk ønsker at afdække og belyse? Indholdsvaliditet undersøges f.eks. ved at inddrage et ekspertpanel af forskere, klinikere og potentielle brugere til vurdering af alle tænkelige aspekter ved det, område vi ønsker at måle.

Reliabilitet

En test, et spørgeskema eller et andet måleinstrumentets reliabilitet eller pålidelighed refererer til metodens evne til at udvise en høj grad af overensstemmelse ved gentagne målinger under ensartede eller tilsvarende omstændigheder. Der er således tale om et måleinstrumentets grad af stabilitet og reproducerbarhed. Reliabilitet er derfor en forudsætning for et instruments validitet. Omvendt kan det synes uvæsentligt, om en målemetode har høj grad af reliabilitet, dvs. måler ens ved gentagne målinger, hvis den "rammer forkert" hver gang og måler noget andet, end det vi har til hensigt at måle. Validitet og reliabilitet kan derfor betragtes som indbyrdes afhængige, komplementære størrelser.

Det er desuden vigtigt at understrege, at en målemetodes validitet og reliabilitet principielt kun gælder for den specifikke sammenhæng, hvor målemetoden er valideret eller reliabilitetstestet.

En specifik test kan være valid til måling af funktionsniveau hos hemiplegikere, men er ikke nødvendigvis valid til måling af funktionsniveau hos sclerosepatienter. Tilsvarende kan en målemetode være reliabel "i hænderne" på trænede testere, men mindre reliabel når den udføres af utrænede testere.

Reliabilitetsbegreber

Interrater reliabilitet udtrykker graden af overensstemmelse mellem to eller flere "rater", bedømmere eller testere. Får to eller flere kolleger det samme svar ved anvendelsen af den samme test, under ensartede omstændigheder. Intrarater reliabilitet udtrykker graden af overensstemmelse mellem målinger udført af den samme "rater" i samme måleseance. I praksis kun muligt f.eks. ved brug af videooptagelse. Ofte vil der være tale om gentagne målinger udført af den samme "rater", forudsat det målte fænomen er stabilt i den mellemliggende tid.

Test-retest reliabilitet siger noget om en test eller et spørgeskemas stabilitet over tid. Måler et spørgeskema det samme ved gentagne målinger, forudsat det målte fænomen er stabilt i den mellemliggende tid? Test-retest reliabilitet svarer til intrarater reliabiliteten, men anvendes især om spørgeskemaer, hvor der ikke er tale om en "rater" eller testere.

Table 4: Reliabilitet

Inter-/intrarater reliabilitet	Inter-reliabilitet: graden af overensstemmelse mellem to bedømmere i samme måleseance. Intra-reliabilitet: graden af overensstemmelse mellem målinger udført af den samme bedømmer i samme måleseance.
Test-retest reliabilitet	Graden af overensstemmelse mellem to målinger af det samme fænomen under den forudsætning af at fænomenet ikke har ændret sig.

En målemetode kan være reliabel i større eller mindre grad. Der er sjældent tale om fuld reliabilitet. Graden af reliabilitet kan ses som et kontinuum fra ingen reliabilitet til fuld reliabilitet. Jo større grad af subjektivitet målingen indebærer, jo mindre reliabilitet vil der være tale om. Jo større grad af objektivitet, jo højere grad af reliabilitet.

Ud over validitet og reliabilitet findes der en række andre begreber, som refererer til egenskaber ved en given test eller målemetode:

Appropriateness

Appropriateness dækker over testens egnethed og relevans til formålet. Hvis vi f.eks. ønsker at undersøge balance hos ældre hofteopererede, vil en 6 minutters gangtest ikke være et velegnet valg. Det kunne Bergs balance skala derimod være. Appropriateness bestemmes således af den underliggende problemstilling og formålet med at teste.

Responsiveness

Responsiveness siger noget om testens kliniske følsomhed. Vi vil ofte være interesserede i at se effekten af en behandling ved at teste en patient før og efter behandlingen. Den test, vi vælger, skal være tilstrækkelig følsom til at opfange og vise en ændring, når der faktisk er sket en ændring. F.eks. vil en grov 1-5 muskeltest ikke nødvendigvis opfange og vise en lille, men klinisk relevant ændring i muskelstyrke, og vi vil derfor have brug for en mere følsom og detaljeret målemetode. Responsiveness er således et spørgsmål om, hvorvidt en test kan vise de relevante ændringer i klinikken, vi er interesserede i.

Precision

Precision er den tekniske præcision ved testen, altså målemetodens tekniske præstation og detaljeringsgrad i teknisk måleforstand. F.eks. hvor mange decimaler måles der ved anvendelse af et bestemt måleudstyr.

Interpretability

Direkte oversat betyder interpretability for tolkbarhed og refererer til, om målemetoderesultaterne eller testsvarene er umiddelbart forståelige set udfra den kliniske situation og patientens behov, eller om svarene først må oversættes til det klinisk interessante. Det er alt-

så et spørgsmål om, hvor let det er at tolke testresultatet. F.eks. vil svarene fra en EMG-måling skulle bearbejdes, inden disse er klinisk relevante og forståelige.

Acceptability

Acceptability siger noget om, hvorvidt patienten kan tåle eller tolerere testen. Acceptability stiller spørgsmålene: Er testen "venlig" overfor patienten? Gør det ondt at blive målt? Er det besværligt eller ubehageligt? F.eks. vil en test med fremliggende udgangsstilling have lav acceptability hos ældre hoftearthrosepatienter. Graden af acceptability vil selvsagt have stor betydning for, hvor stor tillid man kan have til testresultatet.

Feasibility

Feasibility kan oversættes med brugbarhed og refererer til, hvor omstændig eller enkel anvendelsen af en test eller målemetode er. Hvor let er det at gennemføre den pågældende test? Kræver det flere timers forberedelse, eller kan det gøres på 5 min.? Hvor meget udstyr er nødvendigt? Kræves der certificering af testeren?

Reaktivitet

Reaktivitet bliver ofte forbundet med Hawthorne effekten, som peger på det faktum, at hvis folk ved, de deltager i et studie, vil de reagere anderledes eller stærkere end i den kliniske hverdag. Reaktivitet siger altså noget om, i hvor høj grad selve det at blive testet påvirker testresultatet. Der kan desuden være reaktivitet i form af læringseffekt, når en patient måles flere gange med samme test eller målemetode.

Vurdering af en specifik test/ målemetodes anvendelighed

Når du har fundet en specifik test eller målemetode, bør du kritisk gennemgå testen for at vurdere om den passer til og opfylder dit formål med at teste eller måle. I det følgende gennemgås en checkliste (Tabel 5A), som kan anvendes til en systematisk gennemgang af specifikke test og målemetoder. Checklisten er tænkt som en hjælp til detaljeret gennemgang og vurdering af målemetoden. Hermed sikrer du også optimal udnyttelse af målemetoden.

I denne artikel har vi valgt ikke at behandle oversættelse af test og målemetoder. Ved oversættelse af udenlandske målemetoder bør man

anvende særlige standardiserede og velbeskrevne oversættelsesprocedurer, som det ville føre for vidt at uddybe her. Det samme gælder udvikling af egne test og målemetoder, som

også repræsenterer et stort og omfattende område, der ligger udenfor denne artikels ærinde og rammer. Der henvises i stedet til listen af litteratur over målemetoder.

Tabel 5A: Målemetode checkliste

Målemetode checkliste

Måleprocessen	
1. Baggrund for udvikling af målemetoden <i>Hvor, hvornår og af hvem blev testen udviklet?</i>	
2. Formål med målemetoden <i>Er målemetoden udviklet med henblik på at stille en diagnose, give en prognose eller måle en effekt?</i>	
3. ICF <i>På hvilket/hvilke niveauer kan svarene indplaceres? (Impairment, activity, participation)</i>	
4. Condition specific/dimension specific/generic instrument <i>Hvilken overordnet type målemetode er der tale om?</i>	
5. Dataindsamlingsmetode <i>Er målemetoden et spørgeskema, et interview, en observation, et instrument eller en funktionstest?</i>	
6. Terapeutens forudsætninger <i>Kræves der særlige forudsætninger for at anvende målemetoden eller kan alle terapeuter umiddelbart anvende den?</i>	
7. Målemetodens ressourceforbrug <i>Hvilke ressourcer kræver det at anvende målemetoden? (Tid/udstyr/personer)</i>	
8. Manual/instruktion <i>Findes der detaljerede manualer/instruktioner på dansk for anvendelse af målemetoden?</i>	
9. Instruktion til respondenter <i>Hvilken instruktion skal der gives til respondenterne?</i>	
10. Fysiske rammer <i>Hvilke fysiske rammer kræves for gennemførelse af målemetoden?</i>	
11. Fejlkilder ved dataindsamlingen <i>Hvilke forhold kan kompromittere målingen?</i>	
12. Kalibreringsprocedure <i>Er en valid kalibreringsprocedure beskrevet? (gælder kun for målemetode der anvender eet eller flere instrumenter)</i>	
13. Måleskala <i>På hvilken skala måles resultaterne?</i>	
14. Validitet <i>Er der redegjort for, hvilken type af validitet målemetoden er undersøgt for?</i>	
15. Reliabilitet <i>Er der redegjort for, hvilken type af reliabilitet målemetoden er undersøgt for?</i>	

Etik og sikkerhed	
16. Integritet <i>Varetages respondentens integritet i hele undersøgelsesforløbet?</i>	
17. Anonymitet <i>Sikres respondentens anonymitet i hele undersøgelsesforløbet, når dette er relevant?</i>	
18. Informeret samtykke <i>Sikres respondenten mulighed for når som helst, frit og uden konsekvenser at kunne trække sig fra undersøgelsesforløbet?</i>	
19. Information <i>Gives respondenten mulighed at blive informeret om undersøgelsens resultater?</i>	
20. Acceptability <i>Gennemføres undersøgelsen på en tålelig og sikker måde for respondenterne?</i>	
21. Opbevaring af data <i>Opbevares resultaterne, så disse ikke forandres under opbevaring, mistes eller lignende.</i>	

Når du har vurderet en tests anvendelighed og måletekniske egenskaber og har overvejet de etiske aspekter i forbindelse med anvendelse af testen, bør du videre forholde dig til dels databearbejdning, dels en eventuel senere rapportering af dine testresultater. Punkter i checklisten om data-

bearbejdning og rapportering er tænkt som en hjælp til at overveje disse aspekter (Tabel 5B). Endvidere kan punkterne også anvendes ved gennemgang og vurdering af en artikel, som omhandler udviklingen, valideringen eller anvendelsen af en specifik test eller målemetode.

Tabel 5B: Målemetode checkliste

Databearbejdning	
22. Fejlkilder ved dataanalyse <i>Er der ved dataanalysen taget højde for mulige fejlkilder ved tolkning af resultaterne?</i>	
23. Dataanalyse <i>Er principperne for processen fra rådata til færdige resultater korrekte, entydige og velbeskrevne?</i>	
24. Normalområder <i>Findes der beskrevne normalområder for målemetodens resultater?</i>	
Rapportering	
25. Begrundelse for valg af målemetode <i>Er der ved rapportering gjort rede for valget af målemetoden?</i>	
26. Afvigelser i udførelse <i>Hvis udførelsen af målemetoden afviger fra den oprindelige målemetode, er afvigelsen og begrundelsen da beskrevet?</i>	
27. Datapræsentation <i>Er data præsenteret som anbefalet af målemetodens udvikler? Hvis ikke, er datatransformeringen da beskrevet?</i>	

Taksigelse:

Artiklen er skrevet med støtte fra bl.a. "The Oak Foundation" og Helsefonden.

Litteratur

- 1) *Beyer N, Magnusson P: Målemetoder i fysioterapi. Munksgaard Danmark 2003.*
- 2) *Rothstein JM, Echternach JL: Primer on Measurement: An Introductory Guide to Measurement Issues. Alexandria, Va: APTA 1993.*
- 3) *Cole B et al: Physical Rehabilitation Outcome Measures, Williams and Wilkins, 1993.*
- 4) *Domholdt E: Physical Therapy Research. W. B. Saunders Company, Pennsylvania 2000.*
- 5) *McDowell et al: Measuring Health, Oxford University Press, 1996.*
- 6) *Kirshner B, Guatt G: A Methodological Framework for Assessing Health Indices. J Chronic Disease 1985; 38(1): 27-36.*
- 7) *Bartels EM, Winkel A, Lund H: Hvordan kan jeg finde relevante målemetoder til at måle effekt af fysioterapi? Nyt om Forskning 2000; 1: 10-17.*
- 8) *Guide to Physical Therapist Practice/ Chapter 2: What Types of Test and Measures do Physical Therapists Use? APTA 2000.*
- 9) *McEwen I: Writing a Case Report. A How-to Manual for Clinicians/ Chapter 4: Defining and Measuring.*
- 10) *Zachariae B: Det Vellykkede Eksperiment. Munksgaard, København 1998.*
- 11) *Bork CE: Research in Physical Therapy. Lippincott Company, Pennsylvania, 1993.*
- 12) *Black N, Brazier J, Fitzpatrick R, Reeves B: Health Services Research Methods – A guide to best practice. BMJ Books, London, 1998.*