

VIDENSKAB OG PRAKSIS | OVERSIGTSARTIKEL

Litteratur

- World Health Organization. The first ten years. Geneva: World Health Organization, 1958.
- Bech P. Rating scales for psychopathology, health status and quality of life. Berlin: Springer Verlag, 1993.
- Hagerty MR, Cummins RA, Ferriss AL et al. Quality of Life Indexes for national policy: review and agenda for research. Soc Indic Res 2001;55:1-96.
- Salvaris M. Community and social indicators: how citizens can measure progress. Hawthorn: Institute for Social research, Swinburne University of Technology, 2000.
- Veenhoven R. The four qualities of life: ordering concepts and measures of the good life. J Happiness Studies 2000;1:1-39.
- Calman KC. Quality of life in cancer patients – an hypothesis. J Med Ethics 1984;10:124-7.
- Festinger L. A theory of cognitive dissonance. Stanford, CA: Stanford University Press, 1957.
- Pedersen AF, Zachariae R, Mainz J. Livskvalitet som sundhedsfagligt effektivt mål. Ugeskr Læger 2005;167:4545-9.
- Birkwood G. Quality of life – how can it be assessed and improved? Basel: John Wiley & Sons Ltd., 1987.
- Zachariae R, Zachariae H, Blomqvist K et al. Quality of life in 6497 Nordic patients with psoriasis. Br J Dermatol 2002;146:1006-16.
- Morrow GR, Roscoe JA, Hickok JT et al. Nausea and emesis: evidence for a biobehavioral perspective. Support Care Cancer 2002;10:96-105.
- Amtsrådsforeningen, Center for Små Handicapgrupper, Dansk Sygeplejeråd et al. Patientens møde med sundhedsvæsenet – de mellem menneskelige relationer – anbefalinger for kommunikation, medinddragelse og kontinuitet. Århus: Kvalitetsafdelingen i Århus Amt, 2003.
- World Health Organization. International Classification of impairments, disabilities, and handicaps (ICDI/WHO). Geneva: World Health Organization, 1980.
- Sartorius N. Cross-cultural comparisons of data about quality of life: a sample of issues. I: Aaronson N, Beckman J, red. Quality of life in cancer patients. New York: Raven Press, 1987:19-24.
- Croog SH, Levine S, Testa MA et al. The effects of antihypertensive therapy on the quality of life. N E J M 1986; 314:1657-64.
- Bech P, Olsen LR, Kjoller M et al. Measuring well-being rather than the absence of distress symptoms: a comparison of the SF-36 Mental Health subscale and the WHO-Five Well-Being Scale. Int J Methods Psychiatr Res 2003;12:85-91.
- Shephard RJ, Franklin B. Changes in the quality of life: a major goal of cardiac rehabilitation. J Cardiopulm Rehabil 2001;21:189-200.
- Pedersen AF, Zachariae B. Livskvalitetsmåling i sundhedsvæsenet – en introduktion. Århus: Kvalitetsafdelingen, Århus Amt, 2003.
- Leong KP, Yeak SC, Saurajen AS et al. Why generic and disease-specific quality-of-life instruments should be used together for the evaluation of patients with persistent allergic rhinitis. Clin Exp Allergy 2005;35:288-98.
- Aaronson NK, Ahmedzai S, Bergman B et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85:365-6.
- Rector TS, Cohn JN. Assessment of patient outcome with the Minnesota Living with Heart Failure questionnaire: reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan. Pimobendan Multicenter Research Group. Am Heart J 1992;124:1017-25.
- Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI) – a simple practical measure for routine clinical use. Clin Exp Dermatol 1994;19:210-6.
- Hammond GS, Aoki TT. Measurement of health status in diabetic patients. Diabetes impact measurement scales. Diabetes Care 1992;15:469-77.
- Meenan RF, Mason JH, Anderson JJ et al. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum 1992;35:1-10.
- Juniper EF, Guyatt GH, Ferrie PJ et al. Measuring quality of life in asthma. Am Rev Respir Dis 1993;147:832-8.
- Groenwold M, Klee MC, Sprangers MA et al. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. J Clin Epidemiol 1997;50:441-50.
- Zachariae R, Zachariae C, Ibsen H et al. Dermatology life quality index: data from Danish inpatients and outpatients. Acta Derm Venereol 2000;80:272-6.
- WHOQOL SRPB Group. A cross-cultural study of spirituality, religion, and personal beliefs as components of quality of life. Soc Sci Med 2006;62:1486-97.
- Peterman AH, Fitchett G, Brady MJ et al. Measuring spiritual well-being in people with cancer: the functional assessment of chronic illness therapy – Spiritual Well-being Scale (FACIT-Sp). Ann Behav Med 2002;24:49-58.
- Smets EM, Garssen B, Bonke B et al. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995;39:315-25.
- La Puma J, Lawlor EF. Quality-adjusted life-years. Ethical implications for physicians and policymakers. JAMA 1990;263:2917-21.
- Egan TM. QALYs or quackery? The quagmire of quantifying the cost of breathing. J Thorac Cardiovasc Surg 2002;123:406-8.
- Rapley M. Quality of life research – a critical introduction. London: Sage Publishers, 2003.

Metoder i livskvalitetsforskning

Overlæge, lektor Mogens Grønvold

Bispebjerg Hospital, Palliativ Medicinsk Afdeling,
Forskningsenheden, og
Københavns Universitet, Institut for Folkesundhedsvidenskab,
Afdeling for Sundhedstjenesteforskning

Resume

Der er udviklet en omfattende forskningsmetodologi til belysning af den videnskabelige kvalitet af spørgeskemaer om helbredsrelateret livskvalitet. Da denne metodologi kan udforske vigtige aspekter af spørgeskemaers validitet og reliabilitet, er det centrale spørgsmål ikke længere, om spørgeskemaundersøgelser er valide, men om, hvor tilstrækkeligt deres validitet ved givne anvendelser

er undersøgt. Med nyere statistiske metoder ses et skift i retning af interaktive, individualiserede spørgeskemaer.

Der forskes i vidt omfang i patienters helbredsrelaterede livskvalitet. Forskningen omfatter både kvantitative (typisk spørgeskemaer) og kvalitative metoder (typisk interview). I denne artikel fokuseres der på spørgeskemaundersøgelser. Er denne slags data videnskabeligt solide? Det er de ikke altid – som på alle andre forskningsfelter er der store variationer i kvalitet – men der eksisterer en velbeskrevet forskningsmetodologi, som kan anvendes til at opnå høj videnskabelig kvalitet.

Faktaboks

Spørgeskemaers videnskabelige kvalitet kan udforskes ved at undersøge bl.a.

Indholdsvaliditet: Belyser spørgeskemaet de rette emner, så det giver udtømmende svar på forskningsspørgsmålet?

Kriterievaliditet: Er svarene korrekte, når de sammenholdes med kriterier for de pågældende begreber?

Begrebsvaliditet: Kan der via teoretiske modeller skabes evidens, der belyser spørgeskemaets validitet?

Skalavalidering: Fungerer multiitemskalaer korrekt således, at de sammenfatter informationen fra enkeltspørgsmålene?

Reliabilitet: Er målingerne præcise og reproducerbare?

Responsiveness og sensitivitet: Kan spørgeskemaet opfange relevante ændringer over tid eller relevante forskelle mellem grupper?

Begreberne som validitet, reliabilitet mv. knytter sig til anvendelser af spørgeskemaer, ikke til selve spørgeskemaerne, idet livskvalitetsspørgeskemaer som andre forskningsmetoder kan være velegnede til nogle formål, men uegnede til andre

Metoderne belyser bl.a. validitet, altså om et spørgeskema måler det, det forventes at måle [1], og reliabilitet (pålidelighed), altså om resultaterne er konsistente og reproducerbare [1, 2]. Denne artikel vil give en introduktion til denne forskningsmetodologi.

Indholdsvaliditet

Det ses nogle gange, at der planlægges kliniske undersøgelser, der har til formål at »forbedre patienters livskvalitet«, uden at dette er defineret nøjere, eller hvor arbejdet med at udvælge livskvalitetsendemål overlades til nærmeste, uerfarne reserve-læge. Det svarer til, at en klinisk forsker skriver i sin forskningsprotokol, at en ny behandling skal »forbedre patienternes blodprøver«, og overlader udvælgelsen af disse til en tilfældig medarbejder.

Et livskvalitetsspørgeskemas indholdsvaliditet [1] i en given undersøgelse handler om, hvorvidt det indeholder de rette spørgsmål til at give svar på forskningsspørgsmålet [1-4]. Vil man sammenligne to behandlinger, skal et indholdsvalidt spørgeskema indeholde spørgsmål, som belyser både de mulige gevinster og de mulige ulemper ved begge behandlinger.

Indholdsvaliditeten sikres ved først at gennemgå den relevante litteratur på feltet: Hvad ved man om, hvordan patienterne har det ved den sygdom, behandling eller i den livssituation, man vil undersøge (fysisk, psykisk, socialt og i form af

påvirkning af funktionsniveau)? Hvilke teorier er der udarbejdet for de begreber, der skal undersøges? Det kan ofte betale sig at supplere litteraturgennemgangen med en mindre interviewundersøgelse med f.eks. 10-15 patienter og et tilsvarende antal fagpersoner, der interviewes åbent med udgangspunkt i forskningsspørgsmålet [2].

Ud fra dette materiale udarbejdes der en liste over de emner, der skal være indeholdt i spørgeskemaet. Bliver listen for lang, eller er der tvivl om, hvorvidt alle temaerne er relevante, kan udvælgelsen understøttes af yderligere ekspert- og patientinterview om emnernes relative vigtighed [2, 5].

Ud fra denne liste kan der foretages en kvalificeret analyse af de eksisterende spørgeskemaer. Analysen vil vise, om der er standardspørgeskemaer, der har det ønskede indhold, og om der er behov for at udvikle nye spørgsmål.

Kriterievaliditet

Når man har sikret sig, at der i undersøgelsen stilles de rette spørgsmål til at give svar på forskningsspørgsmålet, drejer den næste overvejelse sig om, hvorvidt spørgsmålene giver korrekte svar. Besvarelse af et spørgeskema er en kompliceret proces, der er beskrevet kognitionspsykologisk og bl.a. omfatter, at personen skal forstå spørgsmålet, finde relevant information frem, gennemgå en række beslutnings- og vurderingsprocesser, og vælge et svar [6, 7]. Man kan absolut ikke tage det for givet, at alle trin i denne proces fungerer optimalt.

Er der tale om observerbare fænomener, der kan måles uafhængigt af spørgeskemaet, kan der opstilles eksperimenter, hvor spørgeskemaet anvendes parallelt med en ekstern måling af de samme fænomener, der således opfattes som kriterier for »den sande værdi«.

Kriterievaliditet (*criterion validity*) [1, 2] kan f.eks. belyses ved spørgsmål om alder og køn, hvis man har adgang til CPR-numre. Problemet er imidlertid, at det ofte ikke er muligt at finde kriterier for gyldigheden af svar vedrørende subjektive fænomener som smerter, psykisk velbefindende og social støtte [2].

Begrebsvaliditet

Tankegangen ved begrebsvaliditet (*construct validity*) er, at når det ikke er muligt at måle validiteten direkte ved at sammenligne med et kriterium, må validiteten belyses indirekte. Begrebsvaliditet drejer sig om at undersøge, om en metode måler det begreb, den formodes at måle [2], og undersøgelser af begrebsvaliditet bygger på at opstille teoretiske modeller, der kan belyse validiteten, og teste disse modeller [1].

F.eks. kan validiteten af et spørgsmål om »dit samlede helbred« belyses ved at se på, om svarene forudsiger svarpersonernes fremtidige forbrug af sundhedsydelser eller deres levetid. En stærk sammenhæng taler for, at svarene er i hvert fald delvist korrekte. Undersøgelser, hvor svarene sammenlignes med fremtidige hændelser, betegnes som studier af prædiktiv validitet [1].

VIDENSKAB OG PRAKSIS | OVERSIGTSARTIKEL

Tilsvarende kan man analysere sammenhænge mellem spørgsmål. Man må forvente, at forskellige spørgsmål om psykisk velbefindende har en relativt høj korrelation med hinanden. Dette betegnes som analyser af *convergent validity* [1]. Omvendt forventes deres korrelation med spørgsmål om fysisk helbred at være mindre (*divergent validity* [1]). En korrelationsmatrix, hvor spørgsmålenes interne korrelationer vises, kan således anvendes til en grov vurdering af, om mønstrene er, som man ville forvente.

I undersøgelser af *known groups validity* er tankegangen, at man udvælger grupper af personer, som man ved er forskellige [2, 8]. Ved validering af et spørgeskema om aldersrelaterede sygdomme kunne man sammenligne svar fra ældre plejehjemsboere med svar fra unge mennesker. Finder man store forskelle, har man en vis evidens for, at spørgeskemaet kan være validt.

Validering af et spørgeskema ud fra tankegangen bag begrebsvaliditet drejer sig således om at designe eksperimenter, der kan skabe evidens for, at svarene er valide.

I en anden tilgang til begrebsvaliditet udfyldte kræftpatienter et spørgeskema og lagde det i en lukket kuvert. Dernæst gennemførtes et interview, hvor spørgsmålene i spørgeskemaet blev stillet ordret, men hvor patienterne svarede med deres egne ord. Dernæst sammenlignedes patienternes svar i spørgeskemaet forud for interview med et spørgeskema baseret på det åbne interview. I undersøgelsen fandt man (med visse undtagelser) en høj grad af overensstemmelse, der støttede antagelsen om, at spørgsmålene blev opfattet og besvaret som forventet og dermed gav valide svar [9].

Skalavalidering

I en særlig disciplin inden for undersøgelser af begrebsvaliditet beskæftiger man sig med at konstruere såkaldte multiitem-skalaer (hvor svarene fra to eller flere spørgsmål samles til en samlet score) og med at undersøge, om multiitemskalaer, der er konstrueret tidligere, giver korrekte svar. Dette kaldes skalavalidering eller psykometriske analyser.

Scoren fra en multiitemskala kan være misvisende. For eksempel kan et af de spørgsmål, der indgår i skalaen, rumme vigtig information, der ikke afspejles i den samlede score, fordi effekten bliver fortyndet, eller fordi et andet spørgsmål viser modsat tendens [10].

Skalavalidering udføres med statistiske metoder, hvoraf en del bygger på korrelationer mellem spørgsmål i og uden for hver enkelt multi-itemskala. *Multitrait scaling* [2, 11] er relativt simple teknikker baseret på korrelationskoefficienter, der udforsker, om sammenhænge er som forventet (dvs. belyser *convergent/divergent validity* som nævnt ovenfor).

Faktoranalyse er mere komplekse metoder. Eksplorativ faktoranalyse bruges ofte til at danne multiitemskalaer med. Konfirmatorisk faktoranalyse anvendes til at teste, om multiitemskalaer fungerer hensigtsmæssigt [2]. De eksplorative faktoranalyser kan anvendes til at »skabe« overordnede begre-

ber, idet man ud fra resultaterne beslutter, hvilke spørgsmål der slås sammen [2]. Der er dog en risiko for, at der herved skabes meningsløse begreber, hvis ikke resultaterne fortolkes ud fra teorien på området. Konfirmatoriske faktoranalyser, hvor modeller, der er opstillet på basis af et kendskab til teorier og mekanismer, afprøves statistisk, er i bedre overensstemmelse med princippet om at teste hypoteser, snarere end at bruge statistikken til at opspore sammenhænge.

Med de nyere psykometriske metoder, der bygger på såkaldt *item response theory* (IRT), kan man bearbejde data fra multiitemskalaer på en langt mere avanceret måde, end man gør i de traditionelle sumskaler, hvor man blot lægger svarene på de enkelte spørgsmål sammen [2]. IRT-modeller kan undersøge sammenhænge mellem sandsynligheden for hver enkelt svarmulighed og svarpersonens beregnede »niveau« for den pågældende egenskab, og derved udregne de enkelte spørgsmåls »sværhedsgrad«. F.eks. kan man for en fysisk funktionsskala finde, at tærsklen for, at mindst halvdelen svarer, at de kan gå selv, ligger relativt lavt på en skala for fysisk funktion. Tilsvarende vil et spørgsmål om, hvorvidt man kan løbe et maratonløb ligge i den øvre ende.

IRT-modellerne giver helt andre analytiske muligheder end de klassiske metoder, idet de gør det muligt at drage mere information ud af hvert enkelt svar. Således vil svaret om, at en person kan løbe et maratonløb, indikere et meget højt niveau for fysisk funktion, og med denne viden er det overflødig at spørge den samme person, om han eller hun kan gå selv. Metoderne kan derfor udnyttes til udvikling af »intelligente«, computerbaserede spørgeskemaer, såkaldt *computerized testing*, hvor svarpersonen stilles de spørgsmål, der ud fra de hidtidige svar vil være mest informative [12]. Fordelene er således, at man kan undgå at stille »dumme« spørgsmål, ligesom man med det samme antal spørgsmål kan måle mere præcist og derved gennemføre undersøgelser med færre deltagere.

Disse åbenlyse fordele og muligheder har medført, at det amerikanske National Institutes of Health med en meget stor forskningsbevilling har igangsat projektet Patient-Reported Outcomes Measurement Information System (PROMIS) [13]. PROMIS forventes sammen med andre projekter i løbet af de kommende få år at føre til en metodemæssig revolution, hvor fremtidens spørgeskemaer vil være interaktive og computerbaserede og bygge på IRT.

Med mindre komplicerede metoder som *analyser for differential item functioning* (»DIF-analyser«) (tidligere kaldet test for item bias) undersøger man også samspillet mellem spørgsmålene på en mere informativ måde end med de klassiske metoder, der er baseret på korrelationer.

Således kan man belyse sammenhænge mellem spørgsmålene i en multiitemskala og eksterne variabler og bl.a. se, om de trækker i samme retning, og om spørgsmålene fungerer på samme måde i forskellige undergrupper af den undersøgte population [10, 14-16]. Er det ikke tilfældet, kan der opstå tab

VIDENSKAB OG PRAKSIS | OVERSIGTSARTIKEL

Figur 1. Forskellige kombinationer af høj og lav validitet og reliabilitet. Måleresultater (»skud«) i centrum af »skydeskiven« angiver, at metoden måler det, den forventes at måle (er valid). Omvendt svarer mere perifert placerede måleresultater til, at metoden helt eller delvis måler noget andet. Tæt placerede måleresultater ses ved høj reliabilitet. Til venstre ses høj validitet og høj reliabilitet, i midten lav validitet kombineret med høj reliabilitet og til højre lav validitet og lav reliabilitet. Illustration: Lise V. Grønvold.



eller fordrejning af information og bias ved sammenligninger på tværs af undergrupper. Med DIF-metoder kan man også udforske oversættelser af spørgeskemaer [17, 18] og kulturelle forskelle [19].

Reliabilitet

Hvor høj validitet svarer til fravær af systematisk målefejl, svarer høj reliabilitet til fravær af tilfældig (usystematisk) målefejl (Figur 1). Lav reliabilitet kan opvejes ved at øge deltagerantallet, mens dette ikke gælder for manglende validitet.

Gentagelighedsreliabilitet

Opnås der samme svar ved gentagne målinger? Dette kan undersøges ved en *test-retest*-undersøgelse, hvor et spørgeskema besvares to gange med et passende interval [1, 3]. Intervallet skal være langt nok til, at personen ikke blot reproducerer sine svar ud fra hukommelsen, men samtidig så kort, at egenskaben, der undersøges, ikke har ændret sig [1, 3]. Da mange af de egenskaber, der indgår i undersøgelser af helbredsrelateret livskvalitet ændrer sig konstant, er det sidste krav ikke let at opfylde. Anvendeligheden af *test-retest*-studier afhænger derfor af, hvor konstant fænomenet er.

Cronbachs alpha

Cronbachs alpha er en hyppigt anvendt koefficient med værdier fra nul til en [1, 3, 20]. Værdien afspejler både antallet af items og korrelationen mellem items. Mange items og høj korrelation (altså stærk sammenhæng mellem svarene) fører til en høj værdi [2], som tolkes som høj reliabilitet eller høj »intern konsistens«. Cronbachs alpha for en skala bestående af mange vidt forskellige spørgsmål kan være høj, på trods af at mange af spørgsmålene er lavt korreleret [2]. Det fortolkes somme tider som værende udtryk for, at det er meningsfuldt at samle spørgsmål, der måler vidt forskellige emner, i en multiitemskala, men dette er en misforståelse af metoden, der har som forudsætning, at spørgsmålene måler det samme. Der er yderligere problemer med tolkning og anvendelse af Cronbachs alpha, hvilket, som citeret af *Fayers*, førte til, at Cronbach selv kort før sin død advarede mod ukritisk brug af den [2]. *Fayers* anbefaler i stedet beregning af *standard error of the mean* (SEM) [2].

Andre relevante begreber

Begrebet *responsiveness* beskriver et spørgeskemas egnethed til at opfange ændringer over tid og er således en praktisk afprøvning af, om metoden giver tilstrækkelige udslag [2, 21]. En sammenligning af forskellige skemaers anvendelighed til at opfange klinisk betydningsfulde hændelser, såsom f.eks. reduktionen i symptomatologi fra syg til rask, kan være en god hjælp ved valg mellem metoder.

Når spørgeskemaers multiitemskalaer bearbejdes til samlede scorer, står brugeren med det problem, at man har fået nye, abstrakte talskalaer, der ikke giver nogen umiddelbar mening. Dette problem har affødt forskning i kliniske fortolkninger af scorer, hvor de bl.a. kobles til kliniske scenarier [22]. Tilsvarende udforskes den minimale betydningsfulde forskel [22-24]: Hvor stor en ændring skal der til for, at forskellen har en reel betydning?

Begrebet *response-shift* beskriver den problematik, at patienternes opfattelse af de fænomener, som måles med livskvalitetsspørgeskemaer, kan ændre sig over tid. Dette må medtænkes i forskningsdesign [25-27].

Diskussion

I hvor høj grad skal et spørgeskema valideres og hvordan? Valget afhænger bl.a. af: 1) undersøgelsens ambitionsniveau – er det en lille, indledende *survey*, eller en stor, klinisk kontrolleret undersøgelse? 2) Er det nyudvikling af et spørgeskema eller brug af et standardspørgeskema? og 3) Er det konkrete, veldefinerede begreber eller vage, subjektive og teoretisk uklare begreber?

Det betaler sig ofte at bruge kræfter på den indholdsmæssige validitet. En grundig pilottestning af såvel spørgeskema som dataindsamlingsmetode er i reglen også anbefalelsesværdig. Mere dybtgående validitetsundersøgelser kan eksempelvis anbefales til ph.d.-projekter. Konstruktion af multiitemskalaer forudsætter tilstrækkelig metodemæssig/statistisk kompetence. Er den ikke tilgængelig, er det bedre at foretage analyser på enkeltspørgsmålsniveau.

Det kan stærkt anbefales at bruge velundersøgte spørgeskemaer, f.eks. det meget udbredte almene helbredsstatusspørgeskema Short Form (SF)-36 [28-30] eller sygdomsspecifikke skemaer. Man bør kun udvikle egne spørgeskemaer, så-

VIDENSKAB OG PRAKSIS | OVERSIGTSARTIKEL

fremt man ikke kan finde et godt standardspørgeskema, og man har tilstrækkelig tid til at gøre det grundigt. Det kan det dog være hensigtsmæssigt at supplere et standardskema med enkelte studiespecifikke spørgsmål.

Selv anvendelse af det mest velvaliderede standardspørgeskema vil aldrig kunne garantere valide resultater. Måske er fokus helt forkert, eller måske fungerer spørgsmålene ikke i den konkrete sammenhæng. Brugen af et standardspørgeskema fritager således ikke forskeren for arbejdet med bl.a. at sikre, at indholdet er det rette til at besvare forskningsspørgsmålet.

Spørgeskemaer bør ikke udelukkende udvælges ud fra deres dokumenterede videnskabelige kvalitet. Det er også vigtigt at vurdere, hvor meget skemaet er brugt af andre forskere på området, om der er relevante referencematerialer, om der er udviklet standardiserede scoringsprocedurer og medfølgende computerprogrammer, og om der er lavet kvalitetssikrede, officielle oversættelser.

Den videnskabelige kvalitet af livskvalitetsundersøgelser afhænger af meget andet end selve spørgeskemaet. Det største problem er formentlig ufuldstændige data som følge af manglende spørgeskemaer, hvilket kan besværliggøre analyserne så meget, at de må opgives.

Ser man overordnet på livskvalitetsforskningen, er der sket en massiv udvikling gennem de seneste 30 år. Der er nu en velbeskrevet og relativt avanceret forskningsmetodologi, mindst et metodelidsskrift (Quality of Life Research) og årlige metodeorienterede konferencer i International Society of Quality of Life Research. Senest har den amerikanske Food and Drug Administration for spørgeskemaer til anvendelse ved medicinafprøvning udgivet et udkast til detaljerede metodologiske retningslinjer [31], som formentlig vil få stor gennemslagskraft.

Samlet kan man sige, at forskningsmetoderne i livskvalitetsforskningen nu er velbeskrevne. Generelt er kvaliteten opadgående [32]. Der er stadig mange udfordringer, men det vigtigste er fortsat, at forskningsspørgsmålene er klare og væsentlige, at spørgeskemaerne giver svar på forskningsspørgsmålene, og at spørgeskemaerne bliver besvaret.

Korrespondance: *Mogens Grønvald*, Forskningsenheden, Palliativ Medicinsk Afdeling, Region H, Bispebjerg Hospital, DK-2400 København NV. E-mail: mg02@bbh.regionh.dk

Antaget: 9. januar 2008
Interessekonflikter: Ingen

Litteratur

- Nunnally JC, Bernstein IH. Psychometric Theory. 3rd ed. New York: McGraw-Hill, Inc., 1994.
- Fayers PM, Machin D. Quality of life: the assessment, analysis, and interpretation of patient-reported outcomes. 2nd ed. Chichester, UK: Wiley, 2007.
- Streiner DL, Norman GR. Health measurement scales – a practical guide to their development and use. 2. ed. Oxford: Oxford University Press, 1995.
- Feinstein AR. Clinimetrics. New Haven: Yale University Press, 1987.
- Watt T, Feldt-Rasmussen U, Rasmussen ÅK et al. Måling af helbredsrelateret livskvalitet hos patienter med thyroideasygdomme. Ugeskr Læger 2008;170:850-2.
- Willis G, Reeve BB, Barofsky I. The use of cognitive interviewing techniques in quality-of-life and patient-reported outcomes assessment. I: Lipscomb J, Gotay CC, Snyder C, red. Outcomes assessment in cancer: measures, methods, and applications. Cambridge: Cambridge University Press, 2005:610-22.
- Tourangeau R, Rips LJ, Rasinski K. The psychology of survey response. Cambridge: Cambridge University Press, 2000.
- Aaronson NK, Ahmedzai S, Bergman B et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85:365-76.
- Grønvald M, Klee M, Sprangers MAG, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. J Clin Epidemiol 1997;50:441-50.
- Grønvald M, Bjørner JB, Klee MC et al. Test for item bias in a quality of life questionnaire. J Clin Epidemiol 1995;48:805-16.
- Bjørner JB, Damsgaard MT, Watt T et al. Tests of data quality, scaling assumptions, and reliability of the Danish SF-36. J Clin Epidemiol 1998;51:1001-11.
- Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. Qual Life Res 1997;6:595-600.
- Reeve BB, Hays RD, Bjørner JB et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 2007;45:S22-S31.
- Grønvald M, Petersen MA. The role and use of differential item functioning (DIF) analysis of quality of life data from clinical trials. I: Fayers P, Hays R, red. Assessing quality of life in clinical trials. 2nd ed. Oxford: Oxford University Press, 2005:195-208.
- Avlund K, Era P, Davidsen M et al. Item bias in self-reported functional ability among 75-year-old men and women in three Nordic localities. Scand J Soc Med 1996;24:206-17.
- Kreiner S. Validation of index scales for analysis of survey data: the Symptom Index. I: Dean K, red. Population health research: linking theory and methods. London: SAGE Publications, 1993:116-44.
- Bjørner JB, Kreiner S, Ware JE et al. Differential item functioning in the Danish translation of the SF-36. J Clin Epidemiol 1998;51:1189-202.
- Petersen MA, Grønvald M, Bjørner JB, et al. Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. Qual Life Res 2003;12:373-85.
- Scott NW, Fayers PM, Aaronson NK et al. The use of differential item functioning analyses to identify cultural differences in responses to the EORTC QLQ-C30. Qual Life Res 2007;16:115-29.
- Grønvald M. Validation of a quality of life questionnaire for breast cancer patients [ph.d.-afhandling]. København: Københavns Universitet, 1996.
- Guyatt GH, Deyo RA, Charlson M et al. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol 1989;42:403-8.
- Lydick E, Epstein RS. Interpretation of quality of life changes. Qual Life Res 1993;2:221-6.
- Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chron Dis 1987;40:171-8.
- Osoba D, Rodrigues G, Myles J et al. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998;16:139-44.
- Breetvelt IS, van Dam FSAM. Underreporting by cancer patients: the case of response-shift. Soc Sci Med 1991;32:981-7.
- Sprangers MA, Schwartz CE. The challenge of response shift for quality-of-life-based clinical oncology research. Ann Oncol 1999;10:747-9.
- Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999;48:1507-15.
- Bjørner JB, Damsgaard MT, Watt T et al. Dansk manual til SF-36 – et spørgeskema om helbredsstatus. København: Lif, 1997.
- Ware JE, Snow KK, Kosinski M et al. The SF-36 health survey. Manual and interpretation guide. Boston: The Health Institute, New England Medical Center, 1993.
- Ware JE, Gandek B. The SF-36 health survey: Development and use in mental health research and the IQOLA project. Int J Ment Health 1994;23:49-73.
- www.fda.gov/CDER/GUIDANCE/5460dft.pdf /feb 2008
- Efficace F, Osoba D, Gotay C et al. Has the quality of health-related quality of life reporting in cancer clinical trials improved over time? Ann Oncol 2007;18:775-81.