

# Udvikling og validering af patientrapporterede spørgeskemaer – del 1

Cecilie Balslev Willert<sup>1</sup>, Lisbet Rosenkrantz Hölmich<sup>1</sup> & Kristian Thorborg<sup>2</sup>

## STATUSARTIKEL

1) Plastikkirurgisk Afdeling V, Herlev Hospital  
2) Ortopædkirurgisk Afdeling og Fysioterapien, Amager-Hvidovre Hospital

Ugeskr Læger  
2015;177:V08140450

Ved måling af eksempelvis smerte og livskvalitet i klinisk praksis og forskning inddrages patienternes egen vurdering i tiltagende grad. Det skyldes, at sådanne begreber ikke kan måles fysisk, og derfor anvender man som regel patientrapporterede data, der er enhver form for estimering af et givent helbredsaspekt, som kommer direkte fra patienten [1]. På engelsk anvendes ofte udtrykket *patient-reported outcome measurement* (PROM) om et spørgeskema, hvormed man på baggrund af patientens egen oplevelse og eget perspektiv kan evaluere helbredsstatus og effekten af en specifik behandling [2, 3]. På dansk betegnes det et patientrapporteret spørgeskema. PROM-data indgår mere og mere i klinisk og sundhedspolitisk beslutningstagen, og derfor er det afgørende, at der er videnskabeligt belæg for, at denne type data er gyldige. Det opnås ved, at anvendte spørgeskemaer udvikles og valideres korrekt, hvilket desværre ikke altid er tilfældet. F.eks. fandt man i et canadisk studie, at 65 ud af 227 spørgeskemaer (29%), som omhandlede tilfredshed og/eller livskvalitet hos patienter, der havde fået foretaget kosmetisk eller rekonstruktiv brystkirurgi, var konstrueret til det pågældende studie uden at være validerede [4]. I et andet studie fandt man, at kun to af 14 skalaer brugt i

fase III- og IV-forsøg om multipel sklerose var validerede [5]. At validering er vigtig, viste et studie af *Marshall et al*, som dokumenterede, at der i kliniske studier, hvor man anvendte upublicerede måleinstrumenter, f.eks. PROM, var en større tendens til at måle positive effekter af en given behandling end i studier, hvor man anvendte publicerede måleinstrumenter [6]. Anvendelse af reliable og validerede spørgeskemaer er derfor afgørende for at opnå gyldige resultater.

Inden for sundhedsvidenskab anvender man i dag såkaldte psykometriske metoder til både udvikling og validering af spørgeskemaer. Psykometri er en videnskab, der har eksisteret siden 1879 [7]. I psykometrien beskæftiger man sig med at måle psykologiske egenskaber og træk, som ikke kan måles gennem observation, f.eks. personlighed og intelligens, ved at applicere matematik og statistik på psykologiske data [8-10].

I denne artikel og i den følgende [11] om patientrapporteret spørgeskemaer er formålet at opdatere læseren i forhold til de mest anvendte metoder ved udvikling og oversættelse samt validering af spørgeskemaer og gennemgå de væsentligste psykometriske begreber i denne sammenhæng.

## UDVIKLING AF ET PATIENTRAPPORTERET SPØRGESKEMA

At udvikle et patientrapporteret spørgeskema fra bunden er en omfattende proces, der kan vare flere år. Inden man går i gang, er det derfor vigtigt at undersøge, hvilke spørgeskemaer der allerede er tilgængelige inden for det område, man ønsker at undersøge [8, 12]. Hvis man finder et spørgeskema, som er skrevet på et andet sprog, og som har relevans for den patientgruppe eller den kliniske problemstilling, man ønsker at måle på, og som ser ud til at have tilstrækkelige psykometriske egenskaber, kan man vælge at lave en tværkulturel validering i stedet, hvilket er langt mindre tidskrævende om end langt fra enkelt (**Figur 1**) [13]. Der findes efterhånden flere systematiske review om tilgængelige PROMs inden for mange forskellige patientkategorier og kliniske problemstillinger, og et sådant systematisk review giver et godt overblik over de målemæssige egenskaber ved forskellige tilgængelige instrumenter.

En god metode ved udviklingen af et spørgeskema er essentielt for at sikre spørgeskemaundersøgelers vi-

## FAKTABOKS

- ▶ I patientrapporteret spørgeskemaer belyses kliniske problemstillinger, som ikke kan måles fysisk, fra patientens synspunkt, og de anvendes i stigende grad i klinisk praksis og forskning.
- ▶ Et patientrapporteret spørgeskema kan være diagnostisk, prognostisk eller evaluerende.
- ▶ En god metode ved udviklingen af et patientrapporteret spørgeskema er vigtig for at sikre spørgeskemaresultaternes videnskabelige styrke og gyldighed.
- ▶ Udvikling af et patientrapporteret spørgeskema er en omfattende proces, der kan inddeles i seks faser.
- ▶ Det er essentielt at inddrage patienter fra målgruppen under udviklingen af spørgeskemaet, f.eks. ved fokus-gruppe- eller enkeltpersoninterview.
- ▶ Hvis der allerede eksisterer et spørgeskema på et andet sprog, kan man foretage en tværkulturel validering i stedet for at udvikle et nyt spørgeskema.
- ▶ En tværkulturel validering består af en tværkulturel oversættelse og en kulturel adaptation, hvor spørgeskemaets psykometriske egenskaber undersøges i det nye land/den nye kultur.
- ▶ Fordelen ved at lave en tværkulturel validering i stedet for et nyt spørgeskema er, at det er mindre tidskrævende, og at man bedre kan sammenligne resultatet med lignende undersøgelser med samme udfald.

denskabelige gyldighed og styrke. Processen kan inddeles i seks faser [12] (Figur 2), hvilket hjælper til en systematisk og grundig tilgang, sikrer en god metode og gør processen mere overskuelig for den, der udvikler spørgeskemaet.

### Definition og udarbejdelse af begrebet

Først skal man gøre sig klart, præcist hvad man vil måle, hos hvilken patientgruppe og til hvilket formål [8, 12]. Skal spørgeskemaet være diagnostisk, prognostisk eller evaluerende? Med et diagnostisk spørgeskema skal man kunne differentiere mellem patienterne, mens man med et evaluerende skal kunne måle ændringer over tid. Denne proces handler om at indkredse og danne et begreb, og det er vigtigt at være så præcis og detaljeret som muligt.

Under dannelse af begrebet bør man også planlægge, om spørgeskemaet skal opbygges efter en refleksiv eller en formativ model (Figur 3). Ved en refleksiv model er itemsene (spørgsmål og tilhørende svarkategori) manifestationer af det begreb, der måles på, og vil ændre sig, hvis begrebet ændres [12]. Items kaldes her effektindikatorer [8, 12]. Eksempelvis giver en depression en række symptomer, f.eks. træthed/nedsat energi, og hvis patienten får det bedre, vil det være forventeligt, at symptomerne også bliver mindre udtalte. Alle itemsene i en refleksiv model er korrelerede, fordi de afspejler samme begreb [12].

I en formativ model har forskellige faktorer betydning for begrebet, og items kaldes kausale indikatorer [8, 12]. Et eksempel er livskvalitet, hvor f.eks. både evnen til at bevæge sig frit uden hjælpemidler og økonomisk status har indflydelse herpå uden at være korrelerede. I den formative model er det vigtigt at inkludere items inden for alle de områder, som har betydning for det begreb, der måles på.

Om man anvender en refleksiv eller formativ model har betydning ved den senere validering af spørgeskemaet: Man kan kun anvende statistiske beregninger ved en refleksiv model.

### Valg af målemetode

Det skal besluttes, hvordan det pågældende PROM skal opbygges. Jo flere spørgsmål, der stilles, jo mere information får man, og des mere specifikt kan man spørge patienterne. Til gengæld er der mere information at administrere og tolke på, og risiko for at patienterne mister koncentrationen ved udfyldelse af lange spørgeskemaer. Til komplekse begreber, som f.eks. livskvalitet, er det tilrådeligt at bruge multiitemskemaer for at få flere aspekter med [12]. Man kan også opdele spørgeskemaet i domæner, dvs. man underopdeler spørgeskemaet i forskellige emneområder, der alle antages at være vigtige for det samlede begreb. Hvert emneområde har sit eget spørgeskema. F.eks. kan man ved må-

ling af livskvalitet måle på fysisk smerte og psykisk stress i hvert sit domæne. Herved får man en såkaldt profil, altså et overblik over patientens tilstand inden for forskellige områder, og derved hvilke områder der påvirker det overordnede begreb negativt eller positivt [12].

### Udvælgelse og formulering af items

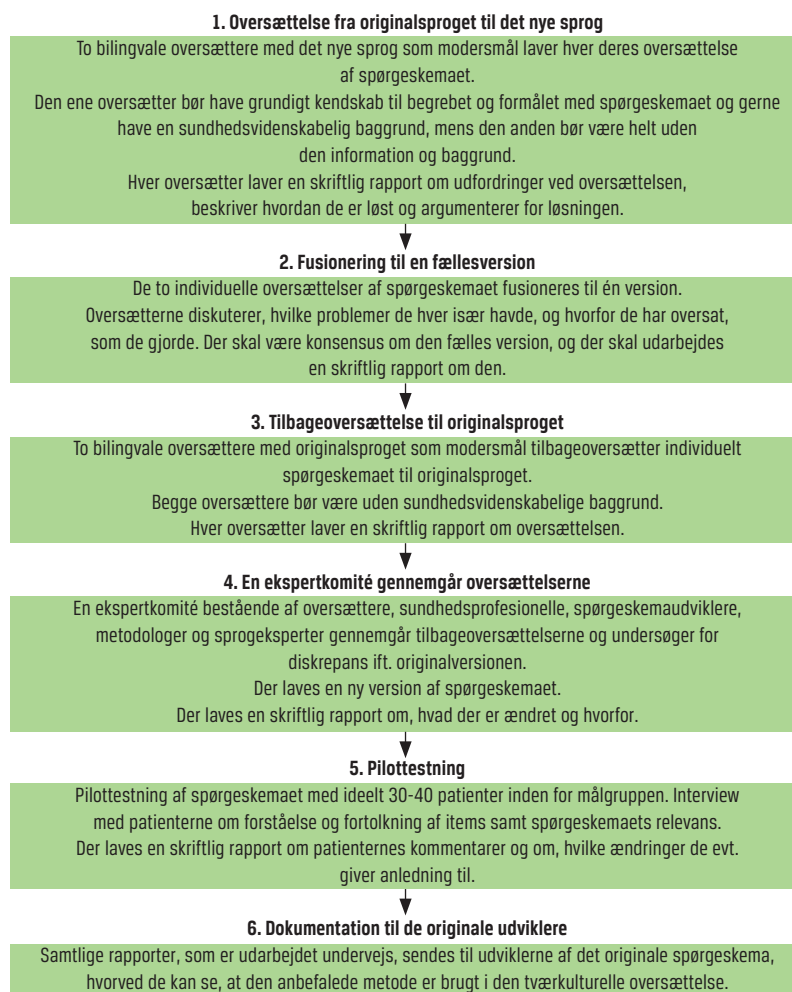
Der skal nu udarbejdes individuelle spørgsmål og/eller udsagn samt svarkategorier, og disse benævnes items [2, 9]. Nye items udvikles på baggrund af patienternes egne oplevelser, klinisk observation, teori/litteratur, forskning og ekspertvurdering, og det er tilladt at genbruge items fra eksisterende spørgeskemaer [3, 8, 12].

Inddragelse af patienter fra målgruppen er afgørende for at sikre overflade- og indholdsvaliditet (se [11] for uddybning), og det gøres ved brug af fokusgrupper og/eller enkeltpersoninterview [1-4, 14, 15].



FIGUR 1

Oversigt over processen ved tværkulturel oversættelse af et spørgeskema, som er første trin i en tværkulturel validering [12].



Der er fordele og ulemper ved de to interviewformer. I fokusgrupperne får man meget information på én gang, og patienterne kan blive inspirerede af hinanden til at udtrykke eller uddybe deres synspunkter. Omvendt kan én dominerende person i gruppen betyde, at andres mening ikke kommer frem [3]. Med enkeltpersoninterview kan man ofte gå mere i dybden med patientens oplevelser, og disse interview kan være brugbare til udledning af sensitiv information, som det kan være ubehageligt at tale om i en gruppe. Dog er det mere tidskrævende, og man er begrænset til en enkelt patients holdning ad gangen uden at kunne sammenligne med andres [3]. Det er i begge tilfælde vigtigt, at der foreligger en guide til interviewene, så de bliver semistrukturerede og ensartede, og at interviewerne er erfarne og godt inde i begrebet og formålet med spørgeskemaet [3]. Interviewene optages enten på bånd eller video og transskriberes efterfølgende til tekst. Herefter analyseres de, og nye items formuleres, gerne med brug af patienternes egne ord [3, 15]. Man bør interviewe en heterogen gruppe patienter for at dække hele spektret af det begreb, som undersøges [3].

Når man skal formulere items, er der flere hensyn at

tage for at maksimere forståelsen af det, der spørges om. F.eks. bør sprogniveauet svare til en 12-årigs, itemsene bør ikke være tvetydige eller indeholde to spørgsmål i ét, man bør undgå fagord, negative ord bør undlades, og itemsene bør ikke være for lange [8, 12, 14]. Hvis ikke man er kritisk i ovenstående vurdering, kan spørgeskemaet blive behæftet med større fejl. Et eksempel på tvetydighed kunne være: »Har du haft lændesmerter for nylig?«. Her afhænger svaret af, hvordan den enkelte patient opfatter og tolker »for nylig«, og resultaterne vil være behæftet med fejl. Man bør i stedet være præcis ved angivelse af tidsperioder, f.eks. »Har du haft lændesmerter inden for de sidste to uger?« [8].

### Problemstillinger ved dannelse af en samlet score

Når spørgsmålene er formuleret, skal der udarbejdes tilhørende svarmuligheder. Her er der flere muligheder: kategoriske eller kontinuerte svarskalaer. Kategoriske skalaer har f.eks. ja-/nej-svarmuligheder, mens udsagn gradbøjes i kontinuerte skalaer. Svarskalaer inddeles yderligere i nominale og ordinale (begge kategoriske) eller interval- og ratioskalaer (begge kontinuerte) [8, 12]. Forskellen på ordinale og intervallskalaer er vigtig, idet der ved den ordinale skala, en skala med ordnede kategorier såsom forskellige grader af uddannelse, ikke med sikkerhed er lige stor forskel på de enkelte svarmuligheder og evt. tilhørende værdier. Det er der derimod ved intervallskalaer, f.eks. temperaturgrader, og de giver derfor resultaterne fra et spørgeskema større statistisk styrke samt mulighed for at udføre parametriske beregninger [8, 16].

Ofte benyttes en Likert-svarskala, som er en skala, der består af tre, fem eller syv svarmuligheder, hvor den midterste svarmulighed er neutral [12]. Hvis man vil fremtvinge et positivt eller negativt svar, kan man vælge at bruge et lige antal svarmuligheder i stedet [12].

Hvordan laver man en samlet score? Skal alle point fra itemsene blot adderes, eller er nogle items vigtigere end andre og bør udløse flere point? Man kan vælge at vægte itemsene, dvs. prioritere nogle items, som er vigtigere for begrebet og derfor skal udløse flere points. En vægtning af items kan foretages af både patienter og klinikere, eller man kan tage matematiske metoder i brug [12]. Hvis man har opbygget spørgeskemaet i domæner, scores hvert domæne for sig, og man får altså ikke én samlet score [12].

### Pilottestning

Når man har dannet sit spørgeskema, skal der foretages en pilottestning med en mindre gruppe patienter. Formålet er at afklare forståeligheden af itemsene, bekræfte spørgeskemaets relevans og sikre, at patienterne synes, at det er nemt at forstå og logisk at udfylde [2, 12, 17]. Det er vigtigt, at det er patienter fra målgrup-

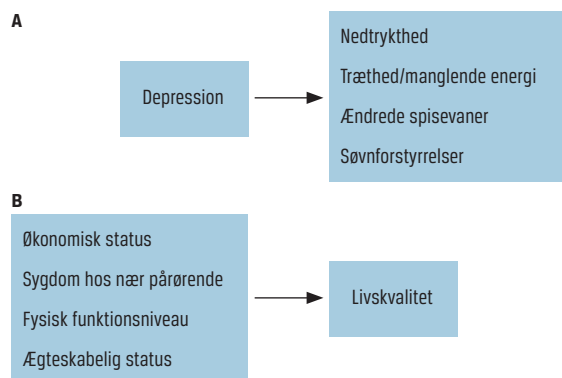
 FIGUR 2

Oversigt over de seks faser i udviklingen af et patientrapporteret spørgeskema.




**FIGUR 3**

**A.** En reflektiv model, hvor itemsene er såkaldte effektindikatorer. Et begreb (depression) resulterer i en række symptomer, som vil ændres, hvis begrebet ændres; f.eks. hvis tilstanden bedres. Effektindikatorerne (i eksemplet er det symptomerne på depression) er internt korrelerede. **B.** Det omvendte forhold med en formativ model, hvor itemsene er kausale faktorer. Her har mange forskellige faktorer betydning for begrebet, der undersøges (i eksemplet er det livskvalitet). De kausale faktorer er ikke internt korrelerede.



pen, som indgår i pilottestningen, da det kun er de patienter, som kan vurdere relevansen, samt om der mangler nogle vigtige items, og om de f.eks. synes, at spørgeskemaet er vanskeligt at besvare [12, 15]. Man bør inkludere forskellige typer patienter fra målgruppen for at øge kvaliteten af undersøgelsen [15].

Man kan vælge at lave kognitive interview med patienterne, hvor de ved udfyldelsen af spørgeskemaet skal tænke højt om, hvad de forstår ved det enkelte item, og hvorfor de vælger at svare, som de gør [12, 15]. Interviewene bør være semistrukturerede, og en erfaren interviewer er at foretrække [15]. På denne måde kan man undersøge, om itemsene fortolkes og fungerer efter hensigten. Hvis ikke, må man omformulere de pågældende items og gentage pilottestningen [15].

## KONKLUSION

Denne artikel har til formål at gennemgå, hvordan man udvikler et patientrapporteret spørgeskema og sætte fokus på god metode i forbindelse hermed. At udvikle et nyt spørgeskema er en omfattende proces, men også en glimrende måde at inddrage patientens eget helbreds-mæssige perspektiv på i såvel klinisk praksis som forskning. Vi har her beskrevet de første fem faser i selve udviklingen. I »Udvikling og validering af patientrapporteret spørgeskemaer i klinisk forskning og praksis – del 2« gennemgås feltundersøgelse og validering af et patientrapporteret spørgeskema og dets psykometriske egenskaber.

## SUMMARY

Cecilie Balslev Willert, Lisbet Rosenkrantz Hölmich & Kristian Thorborg:

Developing and validating of patient-reported questionnaires – part 1  
 Ugeskr Læger 2015;177:V08140450

Patient-reported outcome measurements (PROMs) are often questionnaires which provide and rate the patient's point of view in the measurement of subjective clinical phenomena such as pain or quality of life. In the first of two articles we describe five of the six developmental phases of a patient-reported outcome questionnaire. Adherence to good methodology throughout the process is of utmost importance to ensure reliable and valid results from studies using PROM. Patients from the target population should be involved in the process to obtain content and face validation, i.e. through focus groups.

**KORRESPONDANCE:** Cecilie Balslev Willert. E-mail: ceciliebwi@gmail.com

**ANTAGET:** 9. september 2015

**PUBLICERET PÅ UGESKRIFTET.DK:** 12. oktober 2015

**INTERESSEKONFLIKTER:** Forfatterens ICMJE-formularer er tilgængelige sammen med artiklen på Ugeskriftet.dk

## LITTERATUR

1. U.S. Department of Health, Human Services, FDA Center for Drug Evaluation Research, U.S. Department of Health and Human Services, FDA Center for Biologics Evaluation Research, U.S. Department of Health and Human Services, FDA Center for Devices Radiological, Health.: Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79.
2. Cano SJ, Klassen A, Pusic AL. The science behind quality-of-life measurement: a primer for plastic surgeons. Plast Reconstr Surg 2009; 123:98e-106e.
3. Patrick DL, Burke LB, Gwaltney CJ et al. Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1-eliciting concepts for a new PRO instrument. Value Health 2011;14:967-77.
4. Pusic AL, Chen CM, Cano S et al. Measuring quality of life in cosmetic and reconstructive breast surgery: a systematic review of patient-reported outcomes instruments. Plast Reconstr Surg 2007;120:823-37.
5. Cano SJ, Hobart JC. The problem with health measurement. Patient Prefer Adherence 2011;5:279-90.
6. Marshall M, Lockwood A, Bradley C et al. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. Br J Psychiatry 2000;176:249-52.
7. Bech P. Clinical psychometrics. 1st ed. Chichester, West Sussex: Wiley-Blackwell, 2012.
8. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. Oxford: Oxford University Press, 2008;xvii:431.
9. McDowell I. Measuring health: a guide to rating scales and questionnaires. 3rd ed. Oxford: Oxford University Press, 2006:10-46.
10. Guilford JP. Psychometric methods. 2d ed. New York: McGraw-Hill, 1954.
11. Willert CB, Hölmich LR, Thorborg K. Udvikling og validering af patientrapporteret spørgeskemaer i klinisk forskning og praksis – del 2. Ugeskr Læger 2015;177:V04150290.
12. Vet HCWd. Measurement in medicine: a practical guide. Cambridge: Cambridge University Press, 2011.
13. Beaton DE, Bombardier C, Guillemin F et al. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000;25:3186-91.
14. Terwee CB, Bot SD, de Boer MR et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34-42.
15. Patrick DL, Burke LB, Gwaltney CJ et al. Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2-assessing respondent understanding. Value Health 2011;14:978-88.
16. Bowling A. Measuring health: a review of quality of life measurement scales. 3rd ed. Maidenhead: Open University Press, 2005.
17. Pusic AL, Lemaine V, Klassen AF et al. Patient-reported outcome measures in plastic surgery: use and interpretation in evidence-based medicine. Plast Reconstr Surg 2011;127:1361-7.