

Spørgeskemaer i klinisk forskning

Fokus på oversættelse, tilpasning og afprøvning af fremmedsprogede spørgeskemaer

Hanne Thorsen, seniorforsker, fysioterapeut, læge, ph.d., Institut for Folkesundhedsvidenskab, Afdeling for Almen Medicin, Københavns Universitet

Thorsen H (2005, 15. november). Spørgeskemaer i klinisk forskning – fokus på oversættelse, tilpasning og afprøvning af fremmedsprogede spørgeskemaer. *Forskning i Fysioterapi (online)*, 2. årg.) s. 1-10: URL: www.ffy.dk/sw6351asp

Spørgeskemaer i sundhedsfaglig sammenhæng

Spørgeskemaer har deres naturlige plads i undersøgelser af befolkningens sundhedstilstand, og resultaterne fra sådanne befolkningsundersøgelser har i mere end et halvt århundrede haft stor betydning for initiativer på sundhedsområdet. I klinisk forskning er det gennem de sidste årtier blevet mere og mere almindeligt at inddrage patienters mening om f.eks. effekten af en behandling. Til dette formål har spørgeskemaer også holdt deres indtog. Når nye medikamenter, som ikke forventes at kurere patienter, afprøves, anbefaler f. eks. *The Food and Drug Administration (FDA)* i USA, at der som ét blandt flere effektmål anvendes *Patient Reported Outcome*, det vil sige patientbesvarede spørgeskemaer.

I klinisk praksis anvendes spørgeskemametoden også. Det sker enten ved, at patienten stilles standardiserede spørgsmål eller ved, at man beder patienterne selv besvare et spørgeskema. Fremgangsmåden startede inden for psykiatrien (1) og har bredt sig til andre sundhedsfaglige områder, inklusive fysioterapien.

Pålideligheden og gyldigheden af de standardiserede spørgeskemaer, som anvendes f.eks. inden for psykiatrien, kan drages i tvivl (2). Man kan også i mange andre tilfælde sætte spørgsmålstegn ved værdien af de resultater, man opnår, og dermed ved de følgeslutninger, der drages på grundlag af spørgeskemabesvarelser.

Denne artikel vil herefter:

- gennemgå krav til spørgeskemaer som måleinstrumenter;
- gennemgå nogle af de overvejelser, man bør gøre sig før valg af et spørgeskema som effektmål;
- kritisere de almindeligt forekommende oversættelsesmetoder for spørgeskemaer;
- gennemgå en velafprøvet metode til oversættelse af et fremmedsproget spørgeskema, hvori der indgår en kvalitetskontrol, som sikrer, at det oprindelige formål med spørgeskemaet bevares;
- nævne sproglige faldgruber ved formulering af skrivelser til patienter, herunder spørgeskemaer, og understrege betydningen af høj indholdsvaliditet af disse;
- kort redegøre for nyere statistiske metoder, som er nødvendige at anvende, hvis resultater af en spørgeskemabesvarelse skal udregnes som en score.

Krav til spørgeskemaer som måleinstrumenter

Der er forskel på de krav, man bør stille til et spørgeskema, som anvendes i en befolkningsundersøgelse, sammenlignet med de krav, man bør stille, hvis det skal anvendes i et klinisk forsøg eller i klinisk praksis. Når mange personer besvarer et spørgeskema i en befolkningsundersøgelse, vil mindre usikkerheder, som følge af spørgeskemaets udformning sandsynligvis blive udjævnet. Kravet til et spørgeskema er langt større ved klini-

ske forsøg end i befolkningsundersøgelser, og de er specielt store, hvis man med et spørgeskema monitorerer enkelte patienter i klinisk praksis og sammenligner gentagne målinger for at vurdere, om der er forskel på resultaterne over tid.

I klinisk forskning og klinisk praksis kommer et spørgeskema til at fungere som et måleinstrument og kan i virkeligheden sammenlignes med et apparat til måling af f.eks. blodtrykket.

Et måleinstrument skal først og fremmest have begrebsvaliditet, være gyldigt, dvs. at det skal måle det, man ønsker at måle og være behæftet med så få systematiske fejl som muligt (*construct validity*¹). Det skal også, hvis man vil bruge det til at sammenligne gentagne målinger, f.eks. ved en opfølgning over tid, fungere pålideligt og stabilt fra gang til gang. Det vil sige, at resultater skal kunne reproducere ved gentagen brug af instrumentet under de samme betingelser (*test-retest reliability*)(3,4).

Et eksempel fra "det virkelige liv" kunne være en gammel, slidt og i sin tid billigt indkøbt badevægt til at vurdere ens egen vægt. Den kan være blevet rusten i maskineriet og dermed måle forkert hver gang, altså være behæftet med en systematisk fejl. Den kan også den ene dag være løs i én skrue og en anden dag løs i en anden skrue. Hvis det er tilfældet, kan man ikke vide, om den måler for meget den ene dag og måske lidt for lidt den næste. Den er behæftet med tilfældige fejl og er upålidelig. Hvem ville bruge en badevægt af den kvalitet til at drage følgeslutninger om en eventuel slankekur ud fra måleresultaterne?

Valg af et egnet spørgeskema

Når man ønsker at inddrage patienter i vurdering af f.eks. en behandlingseffekt, må man gøre sig klart, om de patienter, som skal indgå i den aktuelle undersøgelse, er i stand til selv at besvare et spørgeskema. Hvis der er tvivl om dette, bør man overveje at anvende et spørgeskema, som udfyldes af andre end patienterne selv, eventuelt en projektmedarbejder. Når dette er overvejet, og man mener, at et patientbesvaret spørgeskema er den bedste metode, er næste overvejelse, om man skal anvende et allerede eksisterende pålideligt og gyldigt spørgeskema, eller man skal udvikle og validere et nyt.

Udvikling af et spørgeskema til brug som et måleinstrument over for en særlig patientgruppe, f.eks. en gruppe med en afgrænset sygdom, er ressourcekrævende og kan tage flere år. En genvej er derfor, at man gennem en litteratursøgning i relevante databaser finder frem til et allerede udviklet spørgeskema, som umiddelbart ser ud til at være egnet til et aktuelt formål. En optimal fremgangsmåde, som alt for sjældent udnyttes, er, at projektmapperne fra allerførste færd gennemfører interviews med repræsentanter for målgruppen. Resultater af en sådan forundersøgelse danner så grundlag for litteratursøgningen efter et egnet spørgeskema.

De fleste spørgeskemaer til brug både i kliniske forsøg og klinisk praksis stammer fra USA eller Storbritannien og er derfor udviklede og validerede på engelsk.

Ved læsning af artikler i faglige tidsskrifter om resultater af danske kliniske forsøg, hvori der er anvendt spørgeskemaer til måling af en behandlingseffekt, opleves ikke sjældent, at der, hvad angår det anvendte spørgeskema, er henvist til en engelsksproget originalartikel. Det er sjældent, at der henvises til en artikel, som rapporterer, hvordan den danske version af et oprindeligt fremmedsproget spørgeskema er oversat og afprøvet i en dansk population (6).

Mange redaktører af sundhedsfaglige tidsskrifter er efterhånden blevet opmærksomme på sådanne mangler, og der er ingen tvivl om, at det i fremtiden vil blive krævet, at der i en artikel redegøres for, hvordan et spørgeskema er oversat fra originalsproget, og hvordan den oversatte version er afprøvet.

Hvis en fintfølede vægt bliver flyttet fra et laboratorium til et andet, vil man ikke finde på at anvende den, før den atter er blevet kalibreret. Samme omhu bør iagttages, inden et fremmedsproget spørgeskema efter oversættelse til dansk bliver anvendt som et måleinstrument.

Oversættelse og kvalitetskontrol af oversættelsen

Der er flere måder at gribe en oversættelse af et spørgeskema an på. Den simpleste er, at projektmapperne selv oversætter, eller at de overgiver oversættelsen til professionelle translatører og

¹ Construct validity: engelsk for begrebsvaliditet. Bemærk at trykket på ordet construct i denne sammenhæng ligger på første stavelse, og ordet kan oversættes til dansk med 'begreb'. Hvis trykket lægges på anden stavelse, betyder ordet på dansk 'konstruere' (5).

derefter anvender spørgeskemaet uden at kontrollere oversættelsens kvalitet. Hermed forstås blandt andet, at det kontrolleres, om det oversatte er formuleret i et letforståeligt dansk, som ikke kan misforstås af lægfolk, og om de personer/patienter, som skal besvare spørgeskemaet, virkelig forstår meningen med det skrevne.

Hvis projektmagerne selv oversætter, kan oversættelsen blive uheldig, fordi de mangler en dybere indsigt i det aktuelle fremmedsprog. Hvorfor skulle han eller hun være lige så god til at oversætte som til sit eget fag? Hvis en professionel translator oversætter, kan oversættelsen blive uheldig, fordi han eller hun mangler indsigt, enten på fremmedsproget eller på dansk, i den helt nøjagtige betydning af udtryk, som er specifikke for det aktuelle fagområde. Hvis der ikke foretages en kvalitetskontrol af oversættelsen, er det u hensigtsmæssige i denne fremgangsmåde indlysende.

Kritik af den almindeligste metode til oversættelse af spørgeskemaer

I adskillige oversigtsartikler anbefales en metode til oversættelse og kvalitetskontrol af spørgeskemaer, som er blevet internationalt udbredt inden for det sundhedsfaglige område (7).

Metoden består af en primær oversættelse, som foretages af flere af hinanden uafhængige oversættere. Disse skal sammen blive enige om én version af oversættelsen, som derefter tilbageoversættes til det oprindelige sprog af en uafhængig oversætter. De to versioner af spørgeskemaet, den oprindelige og den tilbageoversatte, sammenlignes derefter for at vurdere, om der er uoverensstemmelser mellem dem.

Hvordan ideen om tilbageoversættelse som kvalitetskontrol af en oversættelse er opstået, fortøner sig i det uvisse, og undrer i høj grad sprogkyndige. Antagelig skyldtes det i sin tid et behov for en eller anden form for *videnskabelig* kvalitetskontrol af den oversatte version. Der findes intet klart videnskabeligt grundlag for metoden, og dens udbredte anvendelse inden for det sundhedsfaglige område er derfor betænkelig (8).

Målet for en oversættelse af et spørgeskema må være, at al tekst, både instruktioner i hvordan spørgsmålene skal besvares, de enkelte spørgs-

mål og deres svarkategorier, oversættes så *det, der oprindeligt er ment, bevares i den oversatte version*. Det har især betydning, hvis man over landegrænser vil sammenligne resultater, opnået med oversatte versioner af et spørgeskema. Det behøver ikke nødvendigvis betyde, at den primære version og tilbageoversættelsen er enslydende.

Faldgruberne ved at anvende en tilbageoversættelse som en kvalitetskontrol af en primær oversættelse kan illustreres med tre forskellige situationer.

1. Hvis den primære oversættelse er god, og indholdet i hver enkelt sætning er korrekt overført til det nye sprog, så er det højst sandsynligt, at tilbageoversættelsen slet ikke ligner originalen (se eksempler på side 4). Herved giver tilbageoversættelsen ingen brugbar information, medmindre sammenligningen bliver foretaget af en person, som har et så grundigt kendskab til det sprog, der oversættes til, at vedkommende kan gennemskue, at der er overensstemmelse mellem de to versioner. Da denne forudsætning sjældent forekommer, er øvelsen af tvivlsom værdi.
2. Der er ikke grund til at antage, at den person, som tilbageoversætter, er bedre til sit job end den primære oversætter. Det er i virkeligheden det, man antager, når man beder en tilbageoversætter vurdere om en primær oversættelse er god. I et sådant tilfælde kan en tilbageoversætter foreslå, at en primær oversættelse er dårlig. I bedste tilfælde er de informationer, man opnår i denne situation, forvirrende.
3. Den primære oversættelse kan også være foretaget alt for bogstaveligt eller være alt for tæt på formelle aspekter i originalsproget, hvad angår sætningsstruktur eller valg af ord. Det værste, der kan ske i den situation, er, at tilbageoversætteren gør, hvad der er muligt for at få tilbageoversættelsen til at blive så læselig som mulig og finder frem til de korrekte ord, som var brugt i den originale version. En dårlig oversættelse er let for en sprogkyndig at gennemskue, og en tilbageoversætter vil eventuelt bevidst eller ubevidst finde frem til den oprindelige ordlyd. På den måde vil tilbageoversættelsen, som skulle være en kontrol af en oversættelse - ved en misforståelse - komme til at indikere, at den primære oversættelse er god.

To eksempler på oversættelse, hvor det oprindelige indhold ikke kommer frem, og på, at en tilbageoversættelse ikke ville ligne det oprindelige:

I en ikke-autoriseret oversættelse af et udtryk fra *Nottingham Health Profile (NHP)*, som på engelsk lyder: *I wake up in the early hours of the morning*, blev det på dansk til: *Jeg vågner tidligt om morgenen*. På engelsk betyder det imidlertid, at man vågner klokken 1 eller 2 om natten, hvilket eventuelt kan være et tegn på en begyndende depression. Den ikke-autoriserede danske oversættelse fangede ikke dette problem, og den blev derfor ændret til *Jeg vågner alt for tidligt*.

Et andet udtryk fra NHP lyder på engelsk *I feel on edge*. Dette udtryk kan ikke umiddelbart oversættes til dansk og blev derfor for at sikre meningen med udtrykket til *Jeg føler mig nervøs*. En tilbageoversættelse ville i ingen af tilfældene ligne det oprindelige engelske udtryk (6).

Hvordan opnår man god kvalitet ved oversættelse af et spørgeskema?

Det er nærliggende at foreslå, at kvaliteten i oversættelsesproceduren søges opnået i den primære oversættelse og ikke som en *efterfølgende* kontrol af en allerede gennemført oversættelse. *The European Group for Quality of Life Assessment and Health Measurement* (9) anvender en metode, som i højere grad end den ovenfor nævnte kontrollerer kvaliteten under hele processen. Metoden fokuserer tillige på en reel tilpasning af spørgeskemaet til det nye sprog og/eller kultur, og ikke kun på oversættelsen.

For at opnå dette, inddrages rigtige mennesker/patienter for at sikre sig, at spørgeskemaet fungerer efter hensigten. Baseret på mange års arbejde med metoden og oversættelse af utallige spørgeskemaer til mere end 30 sprog, herunder til dansk (10-12), anbefaler gruppen, at der ned sættes to paneler, et oversættelsespanel og et panel bestående af lægfolk.

Fremgangsmåden er,

- at der rekrutteres 4-6 oversættere til et *oversættelsespanel*. Gruppen skal have en varieret sammensætning, men fælles for paneldelta-

gerne er, at de er dobbeltsprogede, hvad angår fremmedsproget og dansk;

- at deltagerne i oversættelsespanelet ved et gruppemøde oplyses om en eventuel tilgrundliggende model for formulering af spørgeskemaet, hvordan det er udviklet, og oplyses om karakteristika for den gruppe personer, som efterfølgende skal besvare spørgeskemaet;
- at de oplyses om, at oversættelsen sprogligt ikke nødvendigvis skal ligne den originale version, men at den indholdsmæssigt skal være i overensstemmelse med den;
- at oversættelsespanelet arbejder under supervision af en eller flere personer, som skal sikre, at metoden overholdes. Det ideelle er, at den person, som har udviklet det originale spørgeskema, deltager i denne del af oversættelsesprocessen;
- at der i tilfælde, hvor oversættelsespanelet ikke kan nå til enighed om oversættelse af enkelte mindre dele, sætninger eller spørgsmål og svarkategorier, efterlader alternative forslag, som det efterfølgende panel, bestående af lægfolk, skal tage stilling til;
- at den foreløbige oversættelse præsenteres for et lægpanel, der arbejder som en fokusgruppe. Dette panel har til opgave at vurdere, om spørgeskemaet sprogligt repræsenterer almindeligt, let forståeligt dansk. Gruppen ser ikke den oprindelige fremmedsprogede version af spørgeskemaet. Dens opgave er nemlig ikke at finde frem til, hvad de oprindelige spørgsmål kunne betyde, men hvad den danske version af spørgeskemaet reelt betyder.
- at mindst en af koordinatorene fra oversættelsespanelet skal deltage i lægpanelet for at sikre, at den originale betydning af spørgsmålene bevares.
- hvis oversættelsespanelet har efterladt alternativer at vælge imellem, er det lægpanelets opgave at beslutte hvilken, der er bedst, det vil sige lettest af forstå og bedst repræsenterer almindelig dansk.

Da et spørgeskema skal besvares af ganske almindelige mennesker, er det vigtigt, at sproget er almindeligt. Da personer med et relativt højt uddannelsesniveau er overrepræsenterede blandt dobbeltsprogede personer, kan der være en tendens til, at oversættelsen bliver lidt for akademisk. Dette problem udjævnes ved, at deltagere til det efterfølgende lægpanel rekrutteres blandt

personer med en mellemlang uddannelse, som et maksimum.

I oversættelsespanelet er det en fordel, at der er mindst én, som har det oprindelige sprog som modersmål.

Det er unødvendigt og kan endog være uhen-sigtsmæssigt, at der er sundhedsfaglige perso-ner eller personer, som er specialister i design og layout af spørgeskemaer til stede under paneler-nes arbejde. De tekniske detaljer skulle der, på oversættelsestidspunktet, være taget stilling til.

Det er en fordel, at begge køn er repræsenteret i lægpanelet, medmindre spørgeskemaet henven-der sig udelukkende til det ene køn. Aldersforde-lingen af paneldeltagerne bør i begge paneler af-spejle målgruppen for spørgeskemaet, og en rimelig fordeling med hensyn til høj og lav social status og beskæftigelse bør også tages med i overvejelserne ved udvælgelse af deltagere til lægpanelet.

Deltagere, som har den sygdom, som spørge-skemaet er udviklet til at dække, bør ikke deltage i nogle af panelerne. Grunden er, at de vil være alt for koncentrerede om indholdet i spørgeske-maet (som ikke kan ændres, da der jo er tale om et allerede udviklet spørgeskema), i stedet for at koncentrere sig om en vurdering af den sproglige formulering.

Feltundersøgelse af oversættelsen

Når oversættelsesproceduren er gennemført, hører det med i metoden, at testversionen af spørgeske-maet skal vurderes i en *feltundersøgelse*. I denne gennemgås spørgeskemaet med ca. 10 personer/patienter svarende til dem, som senere skal besva-re det. Feltundersøgelsen foregår bedst som indi-viduelle interviews. De interviewede skal tage stil-ling til, om spørgeskemaet er let at overskue og let at besvare samt, om det indholdsmæssigt er rele-vant og dækkende for deres egen situation.

Feltundersøgelsen gennemføres ved, at man be-der de interviewede besvare spørgeskemaet og samtidig *tænke højt*. Hvis man under processen oplever en tøven under markering af et svar, spørger man ind til, om den skyldes problemer med at beslutte, hvor krydset skal sættes, proble-mer med at forstå, hvad spørgsmålet eller en in-

struktion går ud på, eller andre problemer. Inter-vieweren, som sidder med et eksemplar af spørgeskemaet, gør notater ud for de steder i skemaet, hvor der opstår problemer.

Når denne del er overstået, beder man patienter vurdere, om der er områder, relevante for den en-kelte, som *ikke* er med i skemaet og noterer sig dette (se nedenfor under *Indholdsvaliditet*). Even-tuelle rettelser, som måtte ske efter de første 4-5 interviews i feltundersøgelsen, indføres i den næ-ste version af spørgeskemaet, som så afprøves i de efterfølgende interviews.

At rekruttere et tilstrækkeligt stort antal patienter til mindre kliniske forsøg kan være en vanskelig op-gave. Det er videnskabeligt acceptabelt, at de pa-tienter, som indgår i feltundersøgelsen, også ind-går i det efterfølgende kliniske forsøg (se oversigt over procedure på side 6).

Sproget i et spørgeskema og indholdsvaliditet

Hvis en høj motivation for at besvare et spørge-skema skal sikres, ligegyldigt om det er et over-sat spørgeskema eller et nyudviklet, er det afgø-rende, at spørgsmålene, svarkategorierne og eventuelle instruktioner er forståelige for lægfolk og ikke kan misforstås.

Det sprog, som fagfolk anvender, behøver ikke nødvendigvis være forståeligt for lægfolk. Der er evidens for, at ganske almindelige ord, anvendt i en - for læseren - usædvanlig sammenhæng, let bliver misforstået (13,14). Der findes også evi-dens for, at fagfolk har svært ved at frigøre sig fra deres eget *stamme-sprog*, når de formulerer sig skriftligt (15).

En hyppigt forekommende kilde til misforståelse af skriftlige meddelelser til patienter (f.eks. spør-geskemaer), er, at sundhedsfagligt personale tror, at danske ord for fagudtryk forstås af læg-folk. En aktuel undersøgelse har afsløret, at et ud-tryk som "hjertekarsygdomme", der er en sam-mensætning af tre almindelige danske ord, misforstås af størstedelen af danskere på tiende klasses læseniveau (16). En anden undersøgel-se, ligeledes med deltagelse af danskere på tien-deklasses læseniveau, viste, at et ord som "akut" blev misforstået af 25 procent i den sammen-hæng, det blev anvendt (17).

Oversigt over proceduren		
Trin	Deltagere	Fremgangsmåde og formål
Ekspertpanel	1) 4-6 personer, som er dobbeltsprogede i dansk og spørgeskemaets oprindelige sprog Mindst én af personerne skal have det oprindelige sprog som modersmål 2) Mindst én koordinator, som kender baggrunden for spørgeskemaet og som kan forklare formålet med dette og metoden i oversættelse for panelet	Gruppen samles og oversætter indtil der er enighed om items, og eventuel tekst, så alt <i>indholdsmæssigt</i> svarer til det oprindelige spørgeskema. Enkelte dele af oversættelsen, som panelet ikke kan blive enige om, overlades til det efterfølgende panel at tage stilling til
Lægpanel	1) 4-6 personer uden særligt kendskab til det sprog, som spørgeskemaet oprindeligt er udviklet på. Ingen af deltagerne må have den sygdom, som spørgeskemaet handler om 2) Mindst én af koordinatorerne fra ekspertpanelet	Paneler skal tage stilling til om den oversættelse, de præsenteres for, er let forståeligt dansk, og hvad items angår, om svarkategorierne hænger naturligt sammen med spørgsmålene. Panelet ser ikke det oprindelige spørgeskema og har <i>ikke</i> til formål at kommentere indholdet
Feltundersøgelse	1) 10-15 personer/patienter svarende til den gruppe, som spørgeskemaet henvender sig til. Om muligt skal deltagerne i køn og alder svare til de personer, som fremtidigt skal besvare spørgeskemaet. 2) En af de personer, som har deltaget i de to første paneler som koordinator, gennemfører feltundersøgelsen .	Personerne interviewes individuelt eller i 2-3 grupper. Ved individuelle interviews besvarer deltagerne spørgeskemaet samtidig med, at de tænker højt. Det søges afdækket om en eventuel tøven med at besvare et spørgsmål skyldes problemer med at forstå spørgsmålene og svarkategorierne. I gruppeinterviews besvarer deltagerne først spørgeskemaet og derefter gennemgås de enkelte spørgsmål for forståelighed og for hvordan indholdet opfattes. I begge interviewformer kommenterer deltagerne til slut indholdet af spørgeskemaet (dækningsgrad)*).

*) Hvis feltundersøgelsen afslører væsentlige mangler i indholdet af spørgeskemaet, kan der tilføjes nye items. Disse testes derefter for forståelighed i de sidst planlagte individuelle interviews eller gruppeinterviews, og de behandles senere statistisk som enkeltstående items.

Ikke sjældent vil en patient få et spørgeskema tilsendt og blive bedt om at returnere det efter besvarelsen. I sådanne situationer har patienterne kun få muligheder for at få forklaret sproglige uklarheder. Det bør derfor undgås, at det sprog, der anvendes, skaber grundlag for usikkerhed, og resultaterne af ovennævnte undersøgelser viser betydningen af, at lægfolk inddrages i vurderingen af sproget i et spørgeskema.

Hvis indholdsvaliditeten (*content validity*) er lav, kan det også påvirke motivationen for besvarelse (3,4). Selv om patienter, som besvarer et spørgeskema, finder mange af spørgsmålene irrelevante, svarer de, som de pæne mennesker de er, uden at kunne identificere sig med det, der bliver spurgt om. De spilder derfor deres tid på noget, som ikke kan bruges til ret meget. Hvis spørgeskemaet på den anden side ikke dækker alle væsentlige aspekter af det begreb, som det er udviklet til at dække, får projektmageren ikke svar på det, der er værd at vide.

Når meningen med spørgeskemaet er vanskelig at gennemskue for den enkelte patient, hvad enten det skyldes sproglige uklarheder, irrelevans eller væsentlige mangler, kan det betyde, at deltagerne i en spørgeskemaundersøgelse ikke gider besvare skemaet eller endnu værre, at de bliver så irriterede, at de besvarer det på skrømt eller med *bind for øjnene*.

Gennem primære interviews, dvs. før man vælger, om man vil oversætte et spørgeskema eller udvikle et nyt, får man indsigt i, hvad patienterne har af meninger om det emne, man gerne vil udforske. Gennem en feltundersøgelse som inddrager patienter i vurderingen af en foreløbig version af et spørgeskema, får man undersøgt dækningsgrad, relevans og eventuel irrelevans. På den måde kan man eventuelt undgå uheldige situationer, f.eks. en lav svarprocent. Hvis mange af de interviewede i en feltundersøgelse af et oversat spørgeskema sammenfaldende giver udtryk for, at et bestemt område burde være bredere dækket, må det overvejes at formulere et eller flere lokale tillægsspørgsmål. I de efterfølgende statistiske analyser behandles sådanne tilføjelser uafhængigt af det oprindelige spørgeskema.

Statistisk afprøvning af et oversat spørgeskema

Selvom oversættelsen og tilpasningen af det fremmedsprogede spørgeskema, og afprøvnin-gen af den danske version for forståelighed, gennemføres uden de store problemer, er arbejdet ikke slut. I en fuld tilpasningsproces af et spørgeskema fra et sprog til et andet bør det indgå, at spørgeskemaets *psykometriske egenskaber*, det vil sige blandt andet dets pålidelighed og gyldighed, undersøges i den nye kultur. Man kan ikke, hvis man vil holde en høj videnskabelig standard, umiddelbart tillade sig at antage, at disse egenskaber kan overføres, selv om der er gjort alt for at sikre en ordentlig sproglig tilpasning (18).

Ved undersøgelse af de psykometriske egenskaber anvendes både klassiske og nyere statistiske metoder. Gode beskrivelser af de klassiske metoder findes blandt andet hos *Streiner & Norman* og *McDowell & Newell* (se Supplerende litteratur).

Hvis spørgeskemaresultaterne skal rapporteres som en score på en eller flere skalaer, som det så ofte gøres, kræver det, at det er undersøgt, om et spørgeskema indeholder en eller flere skalaer/dimensioner, der hver for sig dækker ét begreb. Hvis dette *ikke* er undersøgt, og man alligevel anvender spørgeskemaet som et måleinstrument og afrapporterer scores, svarer det til, at man adderer små og store æbler med små og store gulerødder og derefter udregner en middelværdi af denne blanding.

De moderne statistiske metoder, som går under betegnelsen *Item Response Theory* (IRT), f.eks. Rasch analyser, kan i dag bruges til at undersøge mange af de psykometriske egenskaber ved et spørgeskema (19).

Georg Rasch var dansk matematiker. Han skabte sin statistiske model til brug ved udvikling af redskaber til evaluering inden for undervisningssektoren, f. eks. af læsefærdigheder. Modellen bygger på sandsynligheden for, at en elev, som har et givent færdighedsniveau, vil være i stand til at løse opgaver op til en sværhedsgrad, som svarer til elevens aktuelle kompetence. Modellen er siden 1980'erne kommet i brug ved validering af spørgeskemaer inden for sundhedssektoren og psykologien over hele verden.

Ved at anvende Rasch-modellen i analyser af data fra et relativt stort antal besvarede spørgeskemaer, er det muligt at finde frem til

- hvilke *items*, der hører sammen og derved danner en dimension, som *ikke består af både æbler og gulerødder* (dimensionalitet);
(et *item* udgøres af de enkelte elementer i et spørgeskema, det vil sige et spørgsmål med tilhørende svarkategorier, og scoren for ét item er værdien af den svarkategori, der er afkrydset).
- hvilke items som *ikke* kan indgå i en skala, fordi de måler noget, som afviger fra resten af items i den pågældende skala;
(hvis sådanne items identificeres, men har høj relevans for deltagerne i en spørgeskemaundersøgelse, bør det overvejes nøje, om de kan fjernes, eller om de skal besvares og behandles statistisk som enkelt-items).
- hvor de enkelte items i én dimension er placeret på et kontinuum fra, f. eks. mindst handicappet til mest handicappet eller ingen smertepåvirkning til ekstrem smertepåvirkning, naturligvis afhængig af, hvad skalaen er udviklet til at måle (skalérbarhed).

Hvis analyser baseret på Item Response Theory (IRT) er anvendt i validering af det oprindelige spørgeskema, er det muligt ved gentagelse af IRT-analyser på data fra den tilpassede version, at få at vide, om de enkelte items i det tilpassede spørgeskema er placeret i samme rækkefølge som i det oprindelige. Resultater af sådanne analyser kan indirekte fortælle noget om kvaliteten af oversættelsen, men også om eventuelle kulturelle forskelle (20).

Undersøgelser af dimensionalitet og skalérbarhed bør altså gennemføres for et spørgeskema, hvis man ønsker at benytte sig af det som et måleinstrument og udregne score. Med lidt god vilje er det acceptabelt at gennemføre analyser baseret på IRT, når resultaterne fra anvendelse af spørgeskemaet foreligger. Hvis sådanne analyser *ikke* gennemføres, må man nøjes med at rapportere de faktuelle oplysninger, som kan udledes fra analyser af svarene på de enkelte items.

For uddybning af den teoretiske baggrund og detaljer vedrørende disse analysemetoder henvises til *Embretson & Steven* samt *Bond & Fox* (se under Supplerende litteratur).

Afslutning

Blot 20 år tilbage blev spørgeskemametoden meget lidt anvendt inden for fysioterapien, men den er sandsynligvis kommet for at blive. Hvis man vil kunne stole på sine resultater af målinger ved hjælp af et spørgeskema, må instrumentet/målebåndet være både pålideligt og gyldigt, og hvis den enkelte patient ikke skal blive træt af eller opgive at besvare de spørgsmål, som er så vigtige for os at få svar på, må sproget i skemaet være forståeligt og spørgsmålene være relevante og dækkende for patienten i den givne sammenhæng. Der er heldigvis noget, der tyder på, at der er ved at blive sat fokus på en høj grad af relevans af et spørgeskema for de enkelte grupper af personer/patienter, som indgår i en spørgeskemaundersøgelse (21).

Redaktørerne på internationale sundhedsfaglige tidsskrifter kræver efterhånden, at de metoder, som er anvendt til oversættelse og kvalitetskontrol af oversættelsen beskrives i artikler, der handler om kliniske forsøg, hvor der er anvendt et oversat spørgeskema.

Hvis man i fremtiden vil afrapportere score på grundlag af spørgeskemadata, vil det uden tvivl blive en forudsætning for godkendelse af en artikel om et projekt, hvori der er anvendt et spørgeskema, at der er anvendt analyser baseret på IRT i valideringen af spørgeskemaet.

Referencer

1. Bech P, Kastrup M, Rafaelsen OJ. Mini-compendium of rating scales for states of anxiety depression mania schizophrenia with corresponding DSM-III syndromes *Acta Psychiatr Scand Suppl.* 1986; 326:1-37.
2. Thorsen H, la Cour P, Brodersen J, Waldorff FB. En kritisk gennemgang af to skalaer til måling af depression i almen praksis. *Månedsskrift for Praktisk Lægegering.* 2003; 3:361-71.
3. Streiner DL, Norman GR. Health measurement scales: A practical guide to their development and use. Oxford: Oxford University Press, 1995.
4. McDowell I., Newell C. Measuring Health. A guide to rating scales and questionnaires. Oxford: Oxford University Press, 1996.
5. Kjærulff Nielsen B. Engelsk-Dansk Ordbog. 2202. Gyldendals Røde Ordbøger, 6 Udgave, 4. oplag. Nordisk Forlag
6. Hanne Thorsen: Helbredsstatusmål, Erfaringer med Edinburgh Rehabilitation Status Scale og Nottingham Health Profile. 1995; FADL's Forlag. København - Århus - Odense (Ph.d.-afhandling).
7. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol.* 1993 Dec;46:1417-32.
8. Swaine-Verdier A, Doward LC, Hagell P, Thorsen H, McKenna SP. Adapting Quality of Life Instruments. *Value in Health* 2004 7 S1: S27-S30.
9. European Group for Quality of Life Assessment and Health Measurement. European Guide to the Nottingham Health Profile. Brookwood, Surrey: Brookwood Medical Publication 1993
10. Doward LC, McKenna SP, Kohlmann T, Niero N, Patrick D, Spencer B, Thorsen H. The International Development of the RGHQoL: A Quality of Life Measure for Recurrent Genital Herpes. *Quality of Life Research.* 1998; 7: 143-153.
11. Thorsen H, Hansen TM, McKenna SP, Sørensen SF, Whalley D. Adaptation into Danish of the Stanford Health Assessment Questionnaire (HAQ) and the Rheumatoid Arthritis Quality of Life Scale (RAQoL). *Scandinavian Journal of Rheumatology.* 2001; 30:103-9.
12. McKenna SP, Doward LC, Kohlmann T, Mercier C, Niero M, Paes M, Patrick D, Ramirez N, Thorsen H, Whalley D. International development of the Quality of Life in Depression Scale (QLDS). *Journal of Affective Disorders* .2001;63:189-199.
13. Elbro C. Møller S. Nielsen EM. Danskernes læsefærdigheder. En undersøgelse af 18-67-åriges læsning af dagligdags tekster. Projekt Læsning og Undervisningsministeriet. København 1991.
14. Dalby MA et al (ed) Bogen om læsning - set i et udviklingsperspektiv. Munksgaard. Danmarks Pædagogiske Institut. 1989.
15. Larsen PH: Skriv sundere - en kritisk brugsbog for sygeplejersker om skriftlig kommunikation. Teknisk Forlag. 1987.
16. Thorsen H, Witt K. Afdeling for Almen Medicin, Københavns Universitet. Foreløbige upublicerede resultater fra projektet: *Forbedring af skriftlig kommunikation fra sundhedsvæsenet til patienter.*
17. Eriksen J-M. Møller B. Svære ord - og lette. En undersøgelse af, hvordan folk forstår - og ikke forstår - en række udvalgte ord, som bruges i skrivelser fra det offentlige og andre steder. Statens Information. 1997.
18. McKenna SP, Doward LC. The translation and cultural adaptation of patient-reported outcome measures. *Value in Health.* 2005, 8: 89-91.

19. Tennant A, McKenna SP, Hagell P. Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Value in Health* 2004; 7:S22-S26.
20. Petersen MA, Groenvold M, Bjorner JB, Aaronson N, Conroy T, Cull A, Fayers P, Hjermstad M, Sprangers M, and Sullivan M. Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Qual.Life Res.* 2003; 12:373-385.
21. Zarins B. Are Validated Questionnaires Valid? *J. Bone Joint Surg. Am.* 2005; 87: 671 - 1672.

Supplerende litteratur

Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Lawrence Erlaum Associates Publishers. Mahwah, New Jersey, London 2001.

Embretson SE, Steven PR. Item Response Theory for Psychologists Lawrence Erlaum Associates Publishers. Mahwah, New Jersey, London 2000.

McDowell I, Newell C. Measuring Health. A guide to rating scales and questionnaires. Oxford: Oxford University Press, 1996.

Streiner DL, Norman GR. Health measurement scales: A practical guide to their development and use. Oxford: Oxford University Press, 1995.