

Bedømmerbias i speciallægeuddannelsen

Kirsten Müller Bested¹, Bente Malling², Karen Skjelsager³, Doris Østergaard⁴, Helle Thy Østergaard⁵ & Charlotte Ringsted⁵

UDVIKLING OG UDDANNELSE

- 1) Anæstesiologisk Afdeling, Sygehus Lillebælt, Vejle Sygehus,
- 2) HR-afdelingen, Aarhus Universitetshospital,
- 3) Anæstesiologisk Afdeling, Næstved Hospital,
- 4) Dansk Institut for Medicinsk Simulation, Herlev Hospital,
- 5) Anæstesiologisk Afdeling, Herlev Hospital, og
- 6) Center for Klinisk Uddannelse, Rigshospitalet

I den lægelige videreuddannelse foretages der løbende kompetencevurdering af den enkelte uddannelsesstagende læge. På baggrund af kompetencevurderingerne vurderes det, om den uddannelsesstagende læge progredierer tilfredsstillende i sit uddannelsesforløb, og om vedkommende opnår de forventede kompetencer. Kompetencevurderingerne foretages ved hjælp af forskellige metoder, såsom struktureret observation, vejledersamtale, audit, skriftlig opgave, 360-graders-feedback m.fl. [1, 2]. Den enkelte kompetencevurdering foretages af en læge med højere uddannelsesniveau end den uddannelsesstagende læges. Kompetencevurderingerne foretages af mange forskellige læger. Dette giver risiko for uensartede bedømmelser og har dermed indflydelse på både pålidelighed og validitet. I denne artikel beskrives kilder til bedømmerbias, og der gives et bud på, hvorledes bias kan reduceres for at opnå bedre pålidelighed og validitet i bedømmelserne.

KILDER TIL BEDØMMERBIAS

Due eller høj og relationer mellem bedømmer og uddannelsesstagende

En bedømmer er sjældent klar over, om han/hun er mild eller streng (due/høj) i sin bedømmelse sammenlignet med andre bedømmere. En mild bedømmer er som oftest mild i bedømmelsen af enhver uddannelsesstagende, ligesom en streng bedømmer oftest er streng i bedømmelsen af enhver uddannelsesstagende [3]. I flere studier beskrives der en generel tendens til, at bedømmere er for milde i bedømmelsen af uddannelsesstagende og dermed giver for positive kompetencevurderinger [1, 3, 4]. Samtidig er der modstand hos bedømmeren mod at beskrive og især handle på ikketilfredsstillende resultater. Bedømmeren antager ofte uberettiget, at klarer den uddannelsesstagende læge sig godt på et område, klarer vedkommende sig også godt på andre områder. Relationer mellem bedømmeren og den uddannelsesstagende kan, ofte ubevidst, påvirke kompetencevurderingen i både positiv og negativ retning. F.eks. kan etniske forhold give uerkendte påvirkninger i bedømmelsessituationen [3].

Nøjagtige observationer og genkaldelse

Observationer af en uddannelsesstagende er svære at genkalde præcist. Jo længere tid, der går fra observa-

tion, til vurderingen gives til den uddannelsesstagende, des mere upræcis bliver denne. F.eks. har man i et studie [3] fundet, at kun 30% af den uddannelsesstagendes præsentationer kunne huskes, når der ikke blev anvendt noter.

Den kliniske situation

Bedømmelse af komplekse kompetencer som samarbejde og interpersonelle kommunikative færdigheder foretages i kliniske daglige patientkontakter, hvor personer fra forskellige personalegrupper er involveret. Bedømmerens relationer til aktørerne kan påvirke bedømmelsen i både positiv og negativ retning. Da direkte observationer i dagligt arbejde påvirkes af patienter, omgivelser og samarbejdspartnere, kan bedømmerens vurdering farves af stemningen i situationen. Under en planlagt vurdering forstyrres bedømmeren ofte af andre udfordringer end de forventede, hvilket kan hæmme fokus på den planlagte bedømmelse. Tidspres og afbrydelser er ofte en hæmsko for gennemførelse af en planlagt observation og vurdering [1, 3]. Ligeledes kan sværhedsgraden af den enkelte situation påvirke vurderingen både negativt og positivt [5].

Norm for bedømmelse

Norm for bestået/ikke bestået i den enkelte kompetencevurdering sættes ofte af bedømmeren på baggrund af niveauet i den gruppe uddannelsesstagende læger, som bedømmeren p.t. arbejder sammen med, og kan således variere over tid [6].

MULIGHEDER FOR AT REDUCERE BIAS

Struktur, definition af kompetence og niveau

Kompetencevurderingsprogrammerne for speciallægeuddannelserne bør være udviklet således, at de afspejler målbeskrivelserne og giver en sufficient stikprøve [5, 7]. Der vil være dels summative, punktvis vurderinger med et eksamenspræg (f.eks. direkte observation og bedømmelsen »god nok«/»ikke god nok«) og formative, mere generelle vurderinger, der afspejler den uddannelsesstagende læges generelle adfærd (f.eks. 360-graders-feedback), og tilsammen giver disse et helhedsindtryk af den bedømte. Pålideligheden af kompetencevurdering er som ovenfor anført afhængig af bedømmelsesmetode og -redskab, om der vurderes ud fra klare kriterier, om bedømme-

ren er trænet i at foretage kompetencevurdering, og om der er specifikke normer for bedømmelse, som er kendt af bedømmeren [7, 8].

Observation, noter og brug af bedømmelsesredskaber

Som bedømmer bør man overveje sine kognitive egenskaber; hvordan registrerer og husker man eksakte data om den bedømtes præstation, hvilken form for noter og optegnelser hjælper en bedst. Det er påvist, at bedømmere, der anvender noter, giver lavere karakterer og har større variation på bedømmelseskalaen, når forskellige kompetencer vurderes, og har større variation i karaktergivning til enkeltpersoner end bedømmere, der ikke anvender noter, alle tre er parametre, der er udtryk for en mere specifik og præcis vurdering [9].

Anvender man et struktureret bedømmelsesredskab, hvor enkelte elementer af den kliniske kompetence er beskrevet, øges præcisionen i vurderingen, ligesom det medvirker til at fastholde niveauet for bedømmelsen. *Williams et al* beskrev en stigning fra 30% til 60% i præcision i bedømmelse af journaloptagelse, når der blev anvendt et struktureret skema. Skemaer bevirker, at tendensen til primært at kommentere positive præsentationer mindskes, og konstruktiv feedback fremmes. Kendes redskabet forud for bedømmelsen, øges præcisionen yderligere.

Opmærksomhed på kompleksitet i klinisk kontekst

Den lægelige videreuddannelse er en arbejdsbaseret uddannelse, hvor både oplæring og vurdering foregår i en klinisk kontekst. Dette giver en enestående mulighed for at bedømme den uddannelsestagende læge, men forudsætter at bedømmeren erkender kompleksiteten i den enkelte situation. Bedømmeren bør overveje, om kompleksiteten passer til niveauet for den planlagte bedømmelse af netop denne uddannelsestagende læge: Er patientens sygdomsforløb for enkelt eller for komplekst? Er samarbejdspartnerne få eller mange? Støtter eller modarbejder samarbejdspartnerne den uddannelsestagende? Er situationen rolig eller hektisk?

Opmærksomhed på påvirkning af relationen mellem bedømmer og uddannelsestagende samt omgivelsernes indflydelse på situationen forbedrer vurderingen [1].

Kendskab til vurderingsprogram og standard for bedømmelse

Den enkelte kompetencevurdering skal ses som en stikprøve. De enkelte, konkrete kompetencevurderinger kombineres med vurderinger af mere generelle kompetencer foretaget som f.eks. 360-graders-feedback og generel vurdering, hvor 8-12 personer giver



Kompetencevurdering
(Fotograf:
Hans Mandøe).

bedømmelse på en vurderingsskala. Kombinationen giver en helhedsvurdering af den uddannelsestagende. Diskussion i gruppen af bedømmere om anvendelse af vurderingsprocedure og fastsættelse af kriterier for »beståelse« på den enkelte afdeling virker fremmende på gennemførelse af vurderingen og øger præcisionen [3].

Ved at lade bedømmere vurdere kompetencer inden for deres eget arbejdsområde får man en mere præcis og vedkommende vurdering, bedre feedback til den uddannelsestagende, når resultatet af vurderingen ikke er tilfredsstillende, og større chancer for, at der bliver udarbejdet en plan for det videre forløb [2, 4].

Træning i kompetencevurdering og kendskab til vurderingsprogrammet

Bedømmeren bør som et minimum være bekendt med de redskaber, der skal anvendes til vurderingerne. At træne på cases eller videoseancer giver mulighed for diskussion af metodens anvendelighed og mulighed for konsensus om standard for niveau. I et studie har man påvist, at træning i bedømmelse og træning i anvendelse af pålidelige vurderingsmetoder modvirker, at bedømmeren falder i ovennævnte faldgruber [10]. Feedback til bedømmeren fra kollegaer om dennes vurderinger mindsker unødigt mildhed eller hårdhed i vurderingerne [3].

Den enkelte bedømmer kan mangle helhedsindtrykket i den enkelte bedømmelse både i direkte observationer, audits, strukturerede vejledersamtale og enkeltelementer af 360-graders-feedback. Helhedsindtrykket opstår, når tilstrækkelig mange vurderinger med passende afspejling af målbeskrivelse samles over en tidsperiode.

Både den uddannelsestagende læge og bedømmerne bør have kendskab til såvel den enkelte kompetencevurderingsmetode som til det samlede kompetencevurderingsprogram for den pågældende uddannelse, ikke mindst da kendskabet motiverer til gennemførelse af vurderingerne [1, 5].

KONKLUSION

Kendskab til kompetencevurderingsprogrammet for det pågældende uddannelsesniveau motiverer til gennemførelse af vurderinger. Diskussion af norm for vurdering og fastsatte kriterier blandt bedømmere øger pålideligheden af den enkelte kompetencevurdering. Vurderes tillige sværhedsgrad og kompleksitet af den enkelte situation i forhold til bedømmelsen, og anvendes der noter, øges pålideligheden yderligere. Bedømmeren bør fortage vurderinger inden for eget arbejdsfelt og bør erkende interpersonelle positive eller negative forhold mellem bedømmer og uddannelsestagende.

KORRESPONDANCE: Kirsten Müller Bested, Anæstesiologisk Afdeling, Sygehus Lillebælt, Vejle Sygehus, Kappeltoft 25, 7100 Vejle.
E-mail: kirsten.muller.bested@slb.regionsyddanmark.dk

ANTAGET: 16. august 2011

FØRST PÅ NETTET: 26. september 2011

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

LITTERATUR

- Govaerts MJ, van der Vleuten CP, Schuwirth LW et al. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Educ* 2007;12:239-60.
- Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance rating. *Teach Learn Med* 2003;15:270-92.
- Harasym PH, Woloschuk W, Cuning L. Undesired variance due to examiner stringency/leniency effect in communication skill scores assessed in OSCEs. *Adv Health Sci Educ Theory Pract* 2008;13:617-32.
- Ringsted C, Østergaard D, Ravn L et al. A feasibility study comparing checklist and global rating forms to assess resident performance in clinical skills. *Med Teach* 2003;25:654-8.
- Schuwirth LW, van der Vleuten CP. A plea for new psychometric models in educational assessment. *Med Educ* 2006;40:296-300.
- Norcini JJ, Blank LL, Duffy FD et al. The Mini-CEX: a method for assessing clinical skills. *Ann Intern Med* 2003;138:476-81.
- Ringsted C, Østergaard D, Scherpbier A. Embracing the new paradigm of assessment in residency training: an assessment programme for first-year residency training in anaesthesiology. *Med Teach* 2003;25:54-62.
- Shumway JM, Harden RM. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach* 2003;25:569-84.
- DeNisi AS, Peters LH. Organization of information in memory and the performance appraisal process: evidence from the field. *J Appl Psychol* 1996;81:717-37.
- Malling B, Bested KM, Skjelsager K et al. Long-term effect of a course on in-training assessment in postgraduate specialist education. *Med Teach* 2007;29:966-71.

Manipulation af hvirvelsøjlen ved kronisk rygsmerter – en gennemgang af et Cochranereview

Tom Bendix

EVIDENSBASERET MEDICIN

Videncenter for Rygsygdomme, Glostrup Hospital

RESUME

Et nyt Cochranereview om manipulation af ryggen ved kronisk rygsmerter er udkommet. Som i tidligere review påvises effekten at være beskedent og i almindelighed ikke bedre end ved træning og andre anvendte behandlinger af denne tilstand. Udbredelsen af manipulation til personer med kroniske rygsmerter bør reduceres. Dog har visse undergrupper givetvis positiv effekt af manipulation, mens andre har en negativ effekt. Hos specielle undergrupper kan manipulation stadig være berettiget, og behandlingen kan ligeledes være indiceret i udvalgte tilfælde som en kortvarig lindring ved opblussen af kroniske smerter.

Behandlinger af rygsygdomme omfatter et utal af forskellige metoder. Manipulation betegner den sidste fase af manuel mobilisering, hvor bevægelsen hurtigt føres forbi den passive yderstilling, som regel med den kendte »knæklyd«. Det gøres i et led, der er låst. Man har tidligere troet, at der kunne være tale om en mindre sublaksation, eller at en kapselfold eller en lille meniskoid struktur sad i klemme i facitleddet, sådan som man kender det fra knæet. I dag synes en anden forklaring at være mest acceptabel: Et smertende segment har en betydelig spænding – iblandt

giver det sig udtryk i krampe – ofte initieret af de følere på discus' overflade, som man ved stimulerer de små rygmuskler. Det er påvist, at denne spændingsmekanisme hæmmes ved udspilning af facitleddet [1]. Man ved, at det er facitleddet, der giver knæklyden, og da der i en del tilfælde kommer en umiddelbar lindring, har man selvfølgelig troet, at det har været facitleddet, der forårsagede smerten. Men sandsynligvis er det oftere discus, der er synderen, og manipulationens udspilning af facitleddet er blot midlet til at bryde spændingen/krampen (**Figur 1**).

Denne behandling udføres langt hyppigst af kiropraktorer, men en del fysioterapeuter og læger har også lært sig teknikken.

Det er lidt trist at måtte erkende, at de fleste rygbehandlinger ikke ændrer meget på forløbet af almindelige, uspecifikke rygsmerter [2]. I den ene metaanalyse efter den anden (kliniske retningslinjer/ Cochranereview) har man som regel kun kunnet finde beskedent effekt, men i de få, hvor man har påvist en sådan effekt, er det netop ved manipulation, (symptombaseret) fysisk træning, smertestillende medicin og ikke mindst tillidsskabende information, hvor personen med rygsmerter får forståelse af, at