

## VIDENSKAB OG PRAKSIS | SEKUNDÆRPUBLIKATION

3. Stadberg E, Mattsson LÅ, Milsom I. The prevalence and severity of climacteric symptoms and the use of different treatment regimens in a Swedish population. *Acta Obstet Gynecol Scand* 1997;76:442-8.
4. Gass MLS, Taylor MB. Alternatives for women through menopause. *Am J Obstet Gynecol* 2001;185:S47-56.
5. Vashisht A, Domoney CL, Cronje W et al. Prevalence of and satisfaction with complementary therapies and hormone replacement therapy in a specialist menopause clinic. *Climacteric* 2001; 4:250-256.
6. Hundrup YA. Livsstil og risikofaktorer for osteoporose hos sygeplejersker, der har valgt substitution med kvindelige kønshormoner. København: Statens institut for Folkesundhed, 2001.
7. Gupta S, Forbes N, Kirkman R. Attitudes to menopause and hormone replacement therapy among Asian and Caucasian women general practitioners. *Maturitas* 2001;39:169-75.
8. Isaacs AJ, Britton AR, McPhearson K. Why do women doctors in the UK take hormone replacement therapy? *J Epidemiol Community Health* 1997;51:373-7.
9. Andersson K, Pedersen AT, Mattson L-Å, Milsom I. Swedish gynecologists' and general practitioners' views on the climacteric period: knowledge, attitudes and management strategies. *Acta Obstet Gynecol Scand* 1998;77:909-16.
10. Bardel A, Wallander MA, Svardsudd K. Hormone replacement therapy and symptom reporting in menopausal women: a population-based study of 35-65-year-old women in mid-Sweden. *Maturitas* 2002;41:7-15.
11. McKinlay SM, Brambilla DJ, Posner JG. The normal menopause transition. *Am J Hum Biol* 1992; 4:37-46.
12. Kvindehormon og overgangsalder. Roskilde Amt: Sundhedsafdelingen, 1999.

# Selektiv rapportering af effektmål i randomiserede forsøg – sekundærpublikation

## En sammenligning af protokoller med publicerede artikler

Seniorforsker Asbjørn Hróbjartsson, læge An-Wen Chan, stud.med. Mette T. Haahr, overlæge Peter C. Gøtzsche & professor Douglas G. Altman

H:S Rigshospitalet, Det Nordiske Cochrane Center, University Health Network Toronto, Department of Medicine og Cancer Research UK/NHS, Centre for Statistics in Medicine, Oxford

### Resume

**Introduktion:** Selektiv rapportering af positive effektmål giver bias.

**Materiale og metoder:** Vi foretog et kohortestudie af 102 protokoller til randomiserede forsøg godkendt af De Videnskabetiske Komiteer for København og Frederiksberg Kommuner i 1994-1995 og tilhørende 122 artikler.

**Resultater:** Den mediane frekvens af ufuldstændigt rapporterede effektmål pr. forsøg var på 50%. Statistisk signifikante effektmål var oftere fuldt rapporteret, odds-ratio var 2,4. I 62% af forsøgene var et primært effektmål ændret.

**Konklusion:** Effektmål i randomiserede forsøg er ofte selektivt rapporteret.

Randomiserede kliniske forsøg med statistisk signifikante fund publiceres oftere end – og før – forsøg med negative fund [1]. I modsætning til denne velbeskrevne publikationsbias er selektiv rapportering af positive effektmål inden for hvert forsøg ikke undersøgt tilsvarende grundigt. Man har længe haft mistanke om eksistensen af en sådan bias [2], men de hidtidige anekdotiske rapporter er ikke nødvendigvis repræsentative [3] og kan selv være et resultat af publikationsbias.

Vort studie havde tre formål: 1) at bestemme prævalensen af urapporterede og ufuldstændigt rapporterede effektmål i publicerede artikler om randomiserede forsøg, 2) at undersøge om rapportering af effektmål var relateret til disses statistiske signifikans og 3) at vurdere konsistensen mellem primære effektmål angivet i forsøgsprotokoller og primære effektmål angivet i publicerede artikler.

### Metode

En udførlig beskrivelse af metoderne findes i primærpublikationen [4]. Materialet bestod af en kohorte af protokoller (og tillæg til protokoller) til randomiserede forsøg godkendt af De Videnskabetiske Komiteer for København og Frederiksberg Kommuner i 1994-1995 og tidsskriftartikler, der beskrev forsøgsresultaterne. Artiklerne blev fundet ved at kontakte protokolforfattere og ved at søge i MEDLINE, EMBASE og Cochrane Controlled Trials Register.

For hvert forsøg gennemlæste vi forsøgsprotokollen, alle tillæg til protokollen og alle publicerede artikler. Vi noterede forsøgets karakteristika og informationer om effektmål, inklusive antal og type af effektmål, hvor detaljeret effektmålet var rapporteret, om det var angivet som primært eller sekundært effektmål, og om resultatet var statistisk signifikant.

Vi definerede et effektmål som en variabel til vurdering af den gavnlige eller skadelige effekt af en behandling. Et effektmål blev betragtet som fuldt rapporteret, hvis tilstrækkelige data til en metaanalyse kunne udtrages fra artiklen, f.eks. hvis der for hver gruppe var angivet antallet af patienter, et mål for effekten (f.eks. et gennemsnit) og dennes præcision (f.eks. standarddeviation). Ufuldstændigt rapporterede effektmål

## VIDENSKAB OG PRAKSIS | SEKUNDÆRPUBLIKATION

**Tabel 1.** Vægtet gennemsnitlig odds-ratio for bias relateret til selektiv rapportering af effektmål (komplet vs. inkomplet rapporterede effektmål).

	Gavnige effekter		Skadelige effekter	
	n	OR (95% SI)	n	OR (95% SI)
Alle forsøg . . . . .	50	2,4 (1,4-4,0)	18	4,7 (1,8-12,0)
Parallelgrupper . . . . .	38	2,8 (1,5-5,4)	17	5,3 (2,0-14,0)
Overkrydsning . . . . .	11	1,4 (0,63-3,2)	0	Ingen data

n = antal forsøg, SI = sikkerhedsintervaller, OR = odds-ratio.  
Forsøg blev ekskluderet, hvis en odds-ratio ikke kunne beregnes, fordi hele kolonner eller rækker var tomme i 2 × 2-tabellen. Odds-ratio >1 betyder, at statistisk signifikante effektmål (p<0,05) havde en højere odds for at blive fuldt rapporteret end ikkesignifikante effektmål.

**Tabel 2.** Andel af forsøg med klar diskrepans mellem primære effektmål i protokoller og artikler.

Diskrepans mellem primære effektmål: protokol vs. artikel	n (%)
<i>Primært effektmål specificeret i protokol (n = 76)</i>	
Alle ændringer . . . . .	40 (53)
Rapporteret som ikkeprimært i artikler . . . . .	26 (34)
Optræder ikke i artikler . . . . .	20 (26)
<i>Primært effektmål specificeret i artikler (n = 63)</i>	
Alle nye primære effektmål rapporteret i artikler . . . . .	21 (33)
Ændret fra ikkeprimært i protokol til primært i artiklen . . . . .	12 (19)
Ikke nævnt i protokollen . . . . .	11 (17)
Samlet antal diskrepanser for primære effektmål (n = 82) . . . . .	51 (62)

n = antal forsøg.  
I flere forsøg defineredes mere end et primært effektmål både i protokoller og i artikler.

kunne ikke anvendes til metaanalyse, men i artiklen rapporteredes dog om nogen information om resultatet, f.eks. en p-værdi. Vi betragtede effektmål som urapporterede, hvis de var anført i protokollen eller nævnt i en tidsskriftsartikels metodeafsnit, uden at artiklen indeholdt nogen som helst information om resultatet.

Analyserne blev gennemført på forsøgsniveau og var stratificeret efter forsøgsdesign. Effektmål, der drejede sig om henholdsvis gavnlige og skadelige effekter, blev undersøgt separat. For hvert forsøg anførte vi alle effektmål i en 2 × 2-tabel, der relaterede graden af rapportering (fuld vs. ufuldstændig) til statistisk signifikans (p<0,05 vs. p≥0,05). En odds-ratio blev derefter beregnet for hvert forsøg. Odds-ratio over 1 betyder, at statistisk signifikante effektmål har højere odds for at blive fuldt rapporteret end ikkesignifikante effektmål. Et overordnet mål for bias blev beregnet ved at kombinere odds-ratioerne med *random effects*-metaanalyse.

Dernæst vurderede vi konsistensen mellem primære effektmål angivet i forsøgsprotokoller og primære effektmål rapporteret i publicerede artikler. Vi definerede et primært effektmål som et, der eksplicit var karakteriseret som primært i protokollen eller i den publicerede artikel. Når der ikke var anført noget primært effektmål, valgte vi det, der var anvendt i styrkeberegningen ved skønnet over den nødvendige materialestørrelse. Vi vurderede, at der var en klar diskrepans, hvis:

1) et primært effektmål i protokollen i artiklen blev rapporteret som et sekundært eller et uspecificeret effektmål eller var urapporteret, 2) et primært effektmål, der var rapporteret i den publicerede artikel, ikke optrådte i protokollen, eller 3) der var forskellige effektmål i protokollens og artiklens styrkeberegninger.

## Resultater

Der var 1.403 ansøgninger til De Videnskabetiske Komiteer for København og Frederiksberg Kommuner i 1994-1995. Vi fandt 274 protokoller for randomiserede forsøg, hvoraf 172 (63%) aldrig var påbegyndt, færdiggjort eller publiceret. Den endelige kohorte bestod derfor af 102 forsøg med i alt 122 publicerede artikler.

Det typiske forsøg var et industrisponsoreret parallelgruppesforsøg. I artiklerne for 39% af forsøgene oplystes der en kontaktforfatteradresse uden for Danmark. Mediant antal patienter i parallelgruppesforsøg var 151 (10-90 percentil 28-935). Publikationsåret for artiklerne gik fra 1995 til 2003. Vi fandt i alt 3.736 effektmål (mediant 27 pr. forsøg; 10-90 percentil 7-79).

Blandt de forsøg, hvor man målte gavnlige eller skadelige effekter, havde hhv. 71% (70 ud af 99) og 60% (43 ud af 72) mindst et urapporteret effektmål. I disse forsøg var medianen af urapporterede effektmål 4 (10-90 percentil 1-25) for gavnlige effektmål, og 3 (10-90 percentil 1-18) for skadelige effektmål.

Mindst et ufuldstændigt rapporteret gavnligt effektmål forekom i 92% (91 ud af 99) af forsøgene, mens 81% (58 ud af 72) havde mindst et ufuldstændigt rapporteret skadeligt effektmål. Den mediane andel af ufuldstændigt rapporterede effektmål pr. forsøg var 50% (10-90 percentil 4-100%) for gavnlige effektmål og 65% (10-90 percentil 0-100%) for skadelige effektmål.

Den gennemsnitlige odds-ratio for rapporteringsbias (95% konfidensinterval) var 2,4 (1,4-4,0) for gavnlige effekter og 4,7 (1,8-12,0) for skadelige effekter (**Tabel 1**). Med andre ord var odds for fuld rapportering af et givet effektmål mere end dobbelt så stor, hvis effektmålet var statistisk signifikant.

Primære effektmål blev defineret for 82 ud af de 102 forsøg (80%) enten i protokollen, tillægsprotokoller eller i de publicerede artikler. I alt 51 ud af 82 forsøg (62%) havde klare diskrepanser mellem det primære effektmål angivet i protokoller og de primære effektmål angivet i de publicerede arbejder (**Tabel 2**). I ingen af de publicerede artikler var det anført, at en ændring i de primære effektmål havde fundet sted i forhold til protokollen.

## Diskussion

Vi fandt, at der er adskillige urapporterede og ufuldstændigt rapporterede effektmål i de fleste randomiserede forsøg, at positive fund favoriseres, og at der er en diskrepans mellem primære effektmål i publicerede artikler og i forsøgsprotokoller i næsten to tredjedele af tilfældene.

## VIDENSKAB OG PRAKSIS | SEKUNDÆRPUBLIKATION

Formålet med a priori-definerede primære effektmål er at vælge de klinisk mest relevante effektmål og at beskytte mod »fisketure« efter positive resultater. Denne mekanisme sættes ud af funktion, hvis prædefinerede primære effektmål efterfølgende ændrer status.

Selv om ændringer af primære effektmål efter påbegyndelse af et forsøg giver risiko for bias, er rationalet for sådanne ændringer ikke altid klart. En præference for statistisk signifikante resultater er en indlysende forklaring. En anden forklaring kunne være, at resultater fra andre effektmål i forsøget kunne have influeret på, om et statistisk signifikant resultat blev anset for at være væsentligt. For eksempel kunne et effektmål være blevet bortcensureret, hvis det var inkonsistent med andre effektmål.

Nogle af de observerede ændringer kunne skyldes ufuldstændige protokoller. Det er muligt, at nogle forskere ikke har været klar over, at en protokol angiver, hvordan data vil blive analyseret, og at termen »primær« kun refererer til a priori-definerede primære effektmål, og ikke til effektmål valgt post hoc som dem, der har størst betydning eller interesse. Imidlertid var mange protokoller ganske detaljeret udarbejdet af store farmaceutiske virksomheder, hvorfor manglende kendskab til de krav, der stilles til en protokol, næppe kan være forklaringen. Det er også muligt, at diskrepansen skyldes, at logistiske problemer umuliggør det protokollerede effektmåls implementering, eller at ny forskning sandsynliggør, at effektmålet er upålideligt. Imidlertid var det ikke anført i en eneste af artiklerne, at primære effektmål var blevet ændret.

Uanset årsagen til ændringerne i de primære effektmål er der risiko for bias, når ændringerne sker efter påbegyndelse af forsøget. Denne uheldige praksis strider mod samarbejdsaftalen om kliniske forsøg mellem lægestand og lægemiddelindustri, hvori det påpeges, at »De dataanalyser, der ligger til grund for publiceringen, skal være i overensstemmelse med forsøgets protokol, hvori den statistiske metode skal være beskrevet« [5].

Halvdelen af forsøgene var multicenterstudier, de fleste med inddragelse af centre uden for Danmark, og i 39% af forsøgene var den korresponderende forfatter tilknyttet et udenlandsk center. Vore fund er i overensstemmelse med resultaterne fra to andre kohorter af randomiserede forsøg [6, 7]. Det er derfor sandsynligt, at vore fund ikke kun afspejler et dansk fænomen.

Selektiv rapportering af effektmål har omfattende implikationer. Det øger prævalensen af tilfældige fund, og både i randomiserede forsøg og i systematiske oversigter vil der være en tendens til at overvurdere effekter af behandlinger. Der er derfor behov for væsentlige forbedringer i rapporteringen af effektmål i randomiserede forsøg. Protokoller bør være offentligt tilgængelige, ikke kun som kilde til urapporterede effektmål og post hoc-ændringer, men også for at forebygge bias [8].

Ideelt set bør protokoller publiceres online efter forsøgets registrering og før påbegyndelse af forsøget. Afvigelser fra for-

søgsprotokollen bør beskrives i de publicerede artikler, så læseren kan vurdere risikoen for bias. Endelig bør tidsskrifts-redaktører ikke kun overveje et rutinemæssigt krav om, at den originale protokol (samt tillægsprotokoller) indsendes sammen med forsøgsmanuskripter, men også at dette materiale benyttes i *peer review* og helst gøres offentligt tilgængeligt på tidsskriftets hjemmeside [8].

Korrespondance: *Asbjørn Hróbjartsson*, Det Nordiske Cochrane Center, H:S Rigshospitalet, DK-2100 København Ø. E-mail: a.hrobjartsson@cochrane.dk

Antaget: 27. maj 2005

Interessekonflikter: Ingen angivet

## Litteratur

1. Song F, Eastwood AJ, Gilbody S et al. Publication and related biases. *Health Technol Assess* 2000;4:1-115.
2. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials: a survey of three medical journals. *N Engl J Med* 1987;317:426-32.
3. Chan AW. Outcome reporting bias in randomised trials: implications for systematic reviews [DPhil-thesis]. Oxford: Oxford University, 2003.
4. Chan AW, Hróbjartsson A, Haahr MT et al. Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to publications. *JAMA* 2004;291:2457-65.
5. Samarbejdsaftale om kliniske forsøg mellem lægestand og lægemiddelindustri. [www.dadlnet.dk/2laegeforeningen/retoetik/vejledninger/dadlvejledninger/klinisk.htm](http://www.dadlnet.dk/2laegeforeningen/retoetik/vejledninger/dadlvejledninger/klinisk.htm) /maj 2005.
6. Chan AW, Krleza-Jeric K, Schmid I et al. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004; 171:735-40.
7. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330: 753.
8. Krleza-Jeric K, Chan AW, Dickersin K et al. Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). *BMJ* 2005;330:956-8.