



Unngå signifikanstesting av bakgrunnsvariabler i randomiserte studier

MEDISIN OG TALL

STIAN LYDERSEN

E-post: stian.lydersen@ntnu.no

Stian Lydersen er dr.ing. og professor i medisinsk statistikk ved Regionalt kunnskapssenter for barn og unge – psykisk helse og barnevern (RKBU Midt-Norge) ved Institutt for psykisk helse, NTNU. Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

I en randomisert kontrollert studie skal det ikke være noen systematiske forskjeller mellom gruppene før behandling. Noen velger imidlertid å signifikansteste for eventuelle forskjeller i bakgrunnsvariabler. Men dette er overflødig, og som regel også misvisende.

En vesentlig styrke ved randomiserte kontrollerte studier er at bakgrunnsvariabler, som for eksempel alder og kjønn, er tilfeldig fordelt mellom behandlingsgruppene. Ifølge CONSORT-retningslinjene for rapportering av randomiserte studier skal man rapportere demografiske og kliniske variabler separat for hver gruppe (1). Et eksempel er vist i tabell 1, som er et utdrag av en lengre tabell med 23 variabler hentet fra en studie hvor vi sammenliknet to forløp for behandling av hoftebrudd (2). I artikkelen skrev vi at «bakgrunnsvariablene var ikke forskjellige i gruppene» (2, s. 1629). Dette var basert på klinisk skjønn. I tråd med CONSORT-retningslinjene (1) hadde vi ikke signifikanstestet om det var forskjeller.

Tabell 1

Utdrag av en tabell med bakgrunnsvariabler i en randomisert kontrollert studie (2). Gjengitt med tillatelse fra Elsevier. Antall (%) dersom ikke annet er angitt.

	Bred geriatrisk utredning og behandling (N = 198)	Standard ortopedisk behandling (N = 199)
Alder i år, gjennomsnitt (standardavvik)	83,4 (5,4)	83,2 (6,4)
Kvinne	145 (73)	148 (74)
Sivilstatus: bor alene	115 (58)	124 (62)
Bruddtype: intrakapsulært	119 (60)	127 (64)

Det vil alltid være noen forskjeller mellom bakgrunnsvariablene i en randomisert

kontrollert studie. Disse forskjellene er vanligvis små. Og fordi man har randomisert, vet man at eventuelle forskjeller vil skyldes tilfeldigheter. Hvis man signifikanstester for hver bakgrunnsvariabel, vil man forvente statistisk signifikante forskjeller i ca. 5 % av tilfellene, altså ca. hver tjuende variabel. I den aktuelle studien kunne man altså forventet et slikt «signifikant» funn i én av de 23 variablene.

Imidlertid ser vi fremdeles at slik signifikanstesting gjøres i enkelte rapporter fra randomiserte kontrollerte forsøk. Hva kan motivasjonen være for å gjennomføre slik testing? Man kan tenke seg to grunner: å teste om randomiseringen ble gjort riktig, og å identifisere ubalanserte bakgrunnsvariabler.

Ble randomiseringen gjort riktig?

Hvis man har grunn til å mistenke at randomiseringen ikke er gjort riktig, kan man teste dette. Men da bør man bruke et signifikansnivå som er vesentlig lavere enn 5 %. Fayers og King beskriver en slik situasjon: I et bestemt forsøk var det en sterk overvekt av yngre deltakere i den ene gruppen, og forskjellen var høysignifikant, med $p < 0,0005$. Ved nærmere ettersyn viste det seg at randomiseringsprotokollen ikke var fulgt (3).

Ubalanserte bakgrunnsvariabler?

Det vanligste er antakelig at man vil identifisere eventuelle bakgrunnsvariabler som ikke er balansert mellom gruppene. Og deretter kan man justere for disse i analysene. Men statistisk signifikans uttrykt ved en p-verdi avhenger av både antall observasjoner og grad av ubalanse, så en slik fremgangsmåte frarådes (4, 5). I en liten studie kan en variabel være ganske ubalansert uten at denne ubalansen er statistisk signifikant. Det synes mer fornuftig å vurdere størrelsen av observert ubalanse og deretter justere for variabler som både er ubalansert og vurderes som klinisk viktige, enn å basere justerte analyser på en p-verdiredret seleksjon. Men denne metoden er også omdiskutert, da det igjen blir en data-drevet seleksjon av variabler til analysen (6, s. 417–8). Og dette bør i så fall være sensitivitetsanalyser gjort etter den primære analysen.

Ingen hensikt å signifikansteste

Det har ingen hensikt å signifikansteste om det er forskjeller i bakgrunnsvariabler i en randomisert studie, med mindre man har grunn til å tro at randomiseringen ikke er gjort korrekt. De Boer og medarbeidere skriver at slik testing er en usunn praksis som er vanskelig å bli kvitt (5). I noen randomiserte kliniske studier kan det være hensiktsmessig å justere for enkelte bakgrunnsvariabler, men disse må være forhåndsdefinert før man ser på data. Dette skal vi ta for oss i neste artikkel i Medisin og tall.

LITTERATUR:

1. Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869. [PubMed][CrossRef]
2. Prestmo A, Hagen G, Sletvold O et al. Comprehensive geriatric care for patients with hip fractures: a prospective, randomised, controlled trial. *Lancet* 2015; 385: 1623–33. [PubMed][CrossRef]
3. Fayers PM, King M. A highly significant difference in baseline characteristics: the play of chance or evidence of a more selective game? *Qual Life Res* 2008; 17: 1121–3. [PubMed][CrossRef]
4. Pocock SJ, Assmann SE, Enos LE et al. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002; 21: 2917–30. [PubMed][CrossRef]
5. de Boer MR, Waterlander WE, Kuijper LD et al. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *Int J Behav Nutr Phys Act* 2015; 12: 4. [PubMed][CrossRef]

6. Vittinghoff E, Glidden DV, Shiboski SC et al. Regression methods in biostatistics linear, logistic, survival, and repeated measures models. 2. utg. New York, NY: Springer, 2012.

Publisert: 9. mars 2020. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.19.0684
© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no