



Kan man estimere effekt av legemidler i observasjonelle studier?

MEDISIN OG TALL

EVA SKOVLUND

E-post: eva.skovlund@ntnu.no

Eva Skovlund er professor i medisinsk statistikk ved Institutt for samfunnsmedisin og sykepleie, NTNU, og seniorforsker ved Folkehelseinstituttet.

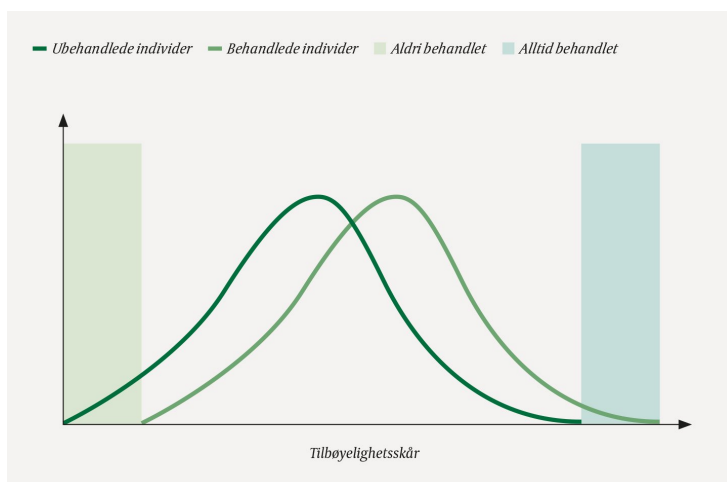
Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

For å estimere effekt av legemidler i observasjonelle studier benyttes ofte tilbøyelighetsanalyser.

Den beste metoden for å unngå systematiske skjevheter i effektestimater er randomiserte kontrollerte studier, men slike studier er kostbare, og det kan være etiske problemer knyttet til å randomisere. En viktig begrensning ved observasjonelle studier er at det er systematiske forskjeller mellom pasienter som får en gitt behandling, og kontrollgruppen. Slik ubalanse vil lede til et skjevt estimat av effekt (bias). Tradisjonelt har man forsøkt å gjøre gruppene sammenlignbare ved å justere for alle pasientkarakteristika som kan ha betydning for både eksponering og utfall (konfunderende faktorer) i en multivariabel analyse. En annen strategi som nå er mye brukt i farmakoepidemiologisk forskning, baserer seg på å utnytte en tilbøyelighetsskår (på engelsk: *propensity score*) for hvert individ.

Hva er en tilbøyelighetsskår?

Tilbøyelighetsskår er sannsynligheten for å få behandling (være eksponert) gitt observerte utgangskarakteristika (1). Skåren er et tall mellom 0 og 1 og estimeres vanligvis ved hjelp av en logistisk regresjonsmodell der alle karakteristika/kovariater som er målt før behandling og kan ha betydning for utfallet, inkluderes. Noen individer kan få en høy skår selv om de ikke har fått den aktuelle behandlingen, mens andre individer er behandlet selv om den estimerte sannsynligheten for at de skulle få behandling, gitt deres karakteristika, er liten. Individer med samme skår anses som «like». Figur 1 viser en tenkt fordeling av skår i to grupper.



Figur 1 Eksempel på fordeling av skårer i ubehandlede og behandlede individer. Ved matching på tilbøyelighetsskår må individer med skår i et av de fargelagte områdene ekskluderes fordi det ikke finnes et individ med matchende skår i den andre gruppen.

I en tilstrekkelig spesifisert modell er pasientkarakteristika likt fordelt mellom individer i behandlings- og kontrollgruppe med lik skår. Mange benytter signifikanstester for å vurdere dette, men en anbefalt metode er å heller verifisere at standardisert differanse er $< 0,1$ (2). Ubalanse tyder på at det finnes umålte konfunderende faktorer som burde vært inkludert.

Effekt av behandling

Når en tilbøyelighetsskår er estimert for hvert individ, er neste trinn selve analysen av utfallet. Denne kan utføres ved matching, stratifisering, justering eller invers sannsynlighetsvekting.

Ved matching sammenlignes individer med samme eller tilnærmet samme skår. Det vanligste er å matche én-til-én, men data kan utnyttes bedre ved å matche flere til én. Matching er den metoden som best reduserer skjevheter, og den er mest brukt (2). Den største ulempen med matching er at individer kan mangle en «match» og derfor ikke kan inkluderes i analysen. Dette reduserer presisjon i estimatene og, kanskje enda viktigere, ekstern validitet. Figur 1 viser en situasjon der det er stor forskjell i fordelingen av skår mellom de to gruppene, slik at mange individer i begge gruppene må ekskluderes.

Stratifisering ligner matching, men her deles skårene inn i et antall strata som avhenger av antall individer som er med i studien. Fordelen med stratifisering er at man kan inkludere alle individene, men metoden er ikke like god til å redusere systematiske skjevheter som matching.

En enkel tilnærming ville være å justere for tilbøyelighetsskår i en vanlig regresjonsanalyse, men med en slik strategi kunne man like gjerne ha gjort en multivariabel analyse der man justerte for alle de konfunderende faktorene samtidig. Det er altså lite å vinne, bortsett fra hvis utfallet er sjeldent og utvalgsstørrelsen for liten til at man kan justere for tilstrekkelig mange kovariater. Den viktigste grunnen til ikke å velge justering er likevel at den i mindre grad enn de andre strategiene vil redusere systematiske skjevheter (2).

En ytterligere metode er å vekte pasienter i de to gruppene med den inverse verdien av skårene. Slik vekting svarer til metoder for spørreundersøkelser der man vekter deltagere slik at utvalget skal bli representativt for populasjonen man studerer. Et individ med høy skår vil vektet mindre i analysen enn et individ med lav skår. Fordelene er at hele datasettet utnyttes, og at metoden reduserer skjevhet i estimatene bedre enn både stratifisering og justering. Ulempen er at den er mer sensitiv for feilspesifisering av modellen (2).

Diskusjon

Uansett hvilken metode man velger, er en viktig forutsetning at alle konfunderende faktorer er inkludert i modellen. Hvis viktige konfunderende faktorer ikke er målt, vil man ikke unngå skjevhet i effektestimatet. Umålt konfundering er alltid en utfordring i observasjonelle studier. Kanskje er det urealistisk å tro at man skal kunne fortolke sammenhenger som kausale dersom man ikke har kunnet randomisere.

LITTERATUR:

1. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70: 41-55. [CrossRef]
2. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011; 46: 399-424. [PubMed][CrossRef]

Publisert: 17. februar 2021. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.20.0778
© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no