



Trekantdrama

MEDISIN OG TALL

JO RØISLIEN

jo@joroislien.no

Jo Røislien er professor i medisinsk statistikk ved det helsevitenskapelige fakultet, Universitetet i Stavanger, og profilert vitenskapsformidler.

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

JAN TERJE KVALØY

Jan Terje Kvaløy er professor i statistikk ved institutt for matematikk og fysikk, Universitetet i Stavanger, og seniorrådgiver i medisinsk statistikk ved Stavanger universitetssjukehus.

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

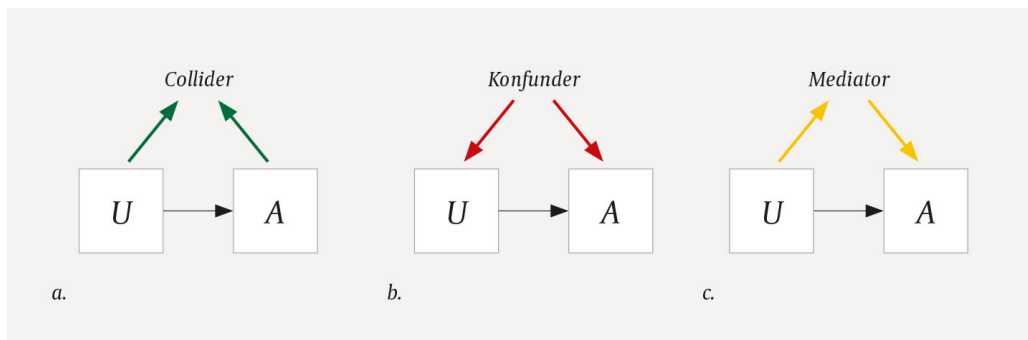
Årsak, virkning. Det er få ting mennesker er mer opptatt av enn det. At noe fører til at noe *annet* skjer. For forfattere er det konstruksjonen av årsakskjeder det hele handler om. Tilfeldigheter er ikke noe spennende. «Everything happens for a reason» er et Hollywood-mantra. Årsak, virkning.

En statistisk analyse kan ikke nødvendigvis avdekke kausale sammenhenger, men den kan skaffe indisier. Med tallinformasjon er bivariat analyse det første skrittet mot å avdekke strukturer og sammenhenger. To og to variabler, der ulike typer tall og fordelinger peker på bestemte statistiske framgangsmåter. I en randomisert kontrollert studie er en bivariat analyse gjerne nok. Hele den virkelige verden er nulltet ut ved studiens design, og man kan uforstyrret fokusere på de to variablene man vil studere. Én uavhengig variabel og én avhengig variabel. Årsak, virkning. Men kommer en *tredje* variabel inn i bildet blir det drama. Trekantdrama.

«Two's company, three's a crowd»

Når to variabler blir til tre kan alt skje, og hva vi bør gjøre avhenger av situasjonen. Det er som når en tredje person entrer et rom. Hva slags relasjon har hen til de to første? Og hva er formålet med møtet? Ikke rart kammerspill er en egen dramaturgisk sjanger. Så også i statistikk. De potensielle forviklingene står i kø, og spesielt må vi være på vakt når den tredje variabelen har en relasjon til begge de to første (Figur 1): Den kan være en *avhengig*

variabel for begge (1a), en *uavhengig* variabel for begge (1b), eller den kan være en *avhengig* variabel for den ene og en *uavhengig* variabel for den andre (1c). Tre situasjoner, tre trekanter. Alle med hvert sitt unike interne drama.



Figur 1 Tre vanlige trekantdramaer i statistisk analyse mellom en uavhengig (U) og en avhengig (A) variabel.

Tre situasjoner

Figur 1 er grafiske beskrivelser av årsak og virkning heftet sammen. Det statistikere kaller en DAG (Directed Acyclic Graph). Et tankekart for statistisk analyse.

I situasjon 1a kalles den tredje variabelen en *collider*. Her påvirker den ikke sammenhengen mellom den uavhengige og den avhengige variabelen, og skal ikke tas med i den statistiske analysen. Dersom man vil undersøke om diabetes påvirker risikoen for hjertesykdom, og tar med sykehusopphold i analysen – som er en *virkning* av begge faktorene, ikke en *årsak* – vil det gi et feilaktig bilde av sammenhengen mellom diabetes og hjertesykdom. *Collidere* er skumle fordi man ubevisst kan ha tatt dem med i analysen, for eksempel ved å kun samle inn data fra personer innlagt på sykehus. Da har man i praksis feilaktig justert analysen for en *collider*. Dette er en bekymring i mange covid-19-studier (1).

Situasjon 1b kalles *konfundering*. En konfunderende variabel påvirker den avhengige variabelen, og er samtidig assosiert med, men ikke påvirket av, den uavhengige variabelen. Dette må vi justere for. Dersom man vil studere sammenhengen mellom trening og lungekreft vil røyking være en konfunderende variabel (2). Denne tredje variabelen – røyking – påvirker risikoen for lungekreft, samtidig som røyking er assosiert med trening (de som røyker mye trener gjerne lite). Uten å justere for røykestatus kan man trekke gale konklusjoner (2).

Situasjon 1c kalles ofte *mediering*. Hva som er rett å gjøre avhenger av formålet med analysen: Er vi på jakt etter direkte, indirekte eller total effekt? Om vi vil studere effekten av røyking på risiko for hjerte- og karsykdommer, vil høyt blodtrykk være av interesse. Med blodtrykk som del av analysen finner vi hvor mye røyking påvirker hjerte- og karsykdommer *direkte*, og hvor mye røyking påvirker hjerte- og karsykdommer *indirekte* via forhøyet blodtrykk. Dersom blodtrykk ikke tas med i analysen, får vi estimert den *totale* effekten av røyking. Ofte vil man gjøre begge analysene for å få det fulle bildet. «Low birth weight paradox» – at røyking har en beskyttende effekt for dødelighet hos små nyfødte – er et klassisk eksempel på hvor galt det kan gå om man ikke tar høyde for mediering (3).

Vit hva du vil

Statistisk analyse forteller hvordan tallene skal regnes på, men ikke hvilken vei pilene mellom årsak og virkning peker. Det må forskeren vite selv. Er den tredje variabelen en *collider* skal vi *ikke* justere, er den en *konfunder* må vi justere. Er den en *mediator* avhenger rett analyse av hva vi vil med analysen. God statistisk analyse kan ikke gjøres uavhengig av kunnskap om den kliniske situasjonen som skal analyseres.

LITTERATUR

1. Griffith GJ, Morris TT, Tudball MJ et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun* 2020; 11: 5749. [PubMed][CrossRef]
2. Lederer DJ, Bell SC, Branson RD et al. Control of confounding and reporting of results in causal inference studies. *Ann Am Thorac Soc* 2019; 16: 22–8. [PubMed][CrossRef]
3. Hernández-Díaz S, Schisterman EF, Hernán MA. The birth weight "paradox" uncovered? *Am J Epidemiol* 2006; 164: 1115–20. [PubMed][CrossRef]

Publisert: 23. april 2021. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.21.0075

© Tidsskrift for Den norske legeforening 2023. Lastet ned fra tidsskriftet.no 4. februar 2023.