

Avhengige målefeil i observasjonsstudier

Sammendrag

Målefeilene i to variabler er avhengige dersom graden av feil i den ene variabelen korrelerer med graden av feil i den andre. Dersom målt eksponering og målt helseutfall er rammet av avhengige feil, blir resultatet oftest en falskt forhøyet sammenheng mellom de to. Slik resultat-skjevhet er sannsynligvis ikke uvanlig i tverrsnittstudier, hvor data om eksponering og utfall skaffes til veie gjennom spørreskjema. På tross av at studier med resultat-skjevhet fra avhengige feil ofte publiseres, synes det som om oppmerksomheten omkring problemet er liten.

Den grunnleggende kilden til avhengige feil er vanligvis normal populasjonsvariasjon med hensyn til bestemte personlighetstrekk, men avhengige feil kan også genereres av situasjonsbestemte stemninger og av kvaliteten på måleinstrumentene.

Det viktigste tiltaket for å eliminere resultat-skjevhet fra avhengige feil er å bryte bindingen mellom informasjon omkring eksponering og utfall – det kan gjøres ved å innhente data fra to ulike kilder. Vi bør også erkjenne at ikke alle typer studiedesign og data egner seg til årsaksrettet analyse.

Engelsk sammendrag finnes i artikkelen på www.tidsskriftet.no

Oppgitte interessekonflikter: Ingen

Petter Kristensen

petter.kristensen@stami.no
Statens arbeidsmiljøinstitutt
Postboks 8149 Dep
0033 Oslo

Observasjonsstudier avviker fra eksperimentelle studier ved å ha en design som er langt mer sårbar for resultat-skjevhet (bias, falske assosiasjoner) (1). Spørsmålet om eventuell årsakssammenheng mellom forklaringsvariabler (her benevnt eksponeringsvariabler) og utfallsvariabler vil derfor være avhengig av studiens interne validitet: Kan vi stole på den observerte sammenhengen (assosiasjonen)? Ved siden av seleksjonsproblemer anses målefeil som den viktigste kilden til validitetsproblemer i epidemiologi. Skjevheter i sammenhengen mellom en eksponeringsfaktor og et utfall antas å være mest alvorlig når vi har *differensielle* målefeil (ramme 1). Det er vel kjent at vi kan få alvorlig resultat-skjevhet dersom kvaliteten på utfallsdata er forskjellig for eksponerte og ikke-eksponerte, eller vice versa. Ikke-differensielle målefeil regnes ofte som mindre alvorlige fordi man antar at feilene har en forutsigbar og konservativ virkning på resultatet: Er det en sann årsakssammenheng, blir vårt estimat en undervurdering av denne.

Denne antakelsen om ikke-differensielle målefeil stemmer ofte ikke dersom det er *avhengige* målefeil i en studie. Avhengige feil forutsetter at det er målefeil av både eksponering og utfall, og at det er korrelasjon mellom feilene for den ene og feilene for den andre (ramme 1). Resultat-skjevhet på grunn av avhengige målefeil omtales vanligvis ikke i lærebøker i epidemiologi. Rothman & Greenlands lærebok omtaler problemet over fem linjer (2). Det er bare få eksempler på at resultat-skjevhet grunnet avhengige feil er tema i epidemiologisk litteratur for øvrig (3, 4). Grunnen til dette kan være at man forutsetter at eksponeringsdata er uten målefeil i kohortstudier og at utfallsdata er uten målefeil i pasient-kontroll-studier. En konsekvens av dette blir at muligheten for resultat-skjevheter på grunn av avhengige målefeil sjelden erkjennes, verken av forskerne bak en studie, av fagvurderere eller av lesere av vitenskapelige artikler. Temaet dukker imidlertid opp nå og da, blant annet i en artikkel i *BMJ*, hvor Macleod og medarbeidere reiste spørsmålet om deres funn med sammenheng

mellom stress og hjertesykdom var falskt grunnet målefeil (5).

Hensikten med denne artikkelen er å gjøre Tidsskriftets lesere mer oppmerksom på avhengige målefeil og å klarlegge årsaker, karakteristika og virkninger av dem. Jeg vil også diskutere hvorledes man kan unngå resultat-skjevhet som skyldes avhengige feil. Eksemplet fra *BMJ* og diskusjonen som fulgte artikkelen vil bli brukt i tillegg til et eget eksempel med et hypotetisk datasett.

To eksempler

Eksemplet til Macleod og medarbeidere (5)

Studien ble gjennomført i Skottland i 1970-årene blant menn som ble rekruttert fra ulike bedrifter. Ved første undersøkelse i 1970–73 ble stressnivå registrert og kodet i tre kategorier ved hjelp av fire spørsmål. Koronar hjertesykdom ble kartlagt ved hjelp av spørreskjema og EKG-undersøkelse. Menn uten tegn til koronar hjertesykdom i utgangspunktet (N = 2 623) ble fulgt opp i 1977 med ny undersøkelse som inkluderte samme spørreskjema og EKG. Dette var utgangspunktet for å beregne forekomst av angina pectoris og iskemi i 1977 og sammenhengene med tidligere stressnivå. Stressnivået var klart assosiert med angina, men ikke med iskemi: Med lavt stress som referansenivå var middels og høyt stress assosiert med angina med oddsratioer på henholdsvis 1,37 (95 % konfidensintervall (KI) 0,91–2,08) og 2,66 (95 % KI 1,61–4,41), de tilsvarende



Hovedbudskap

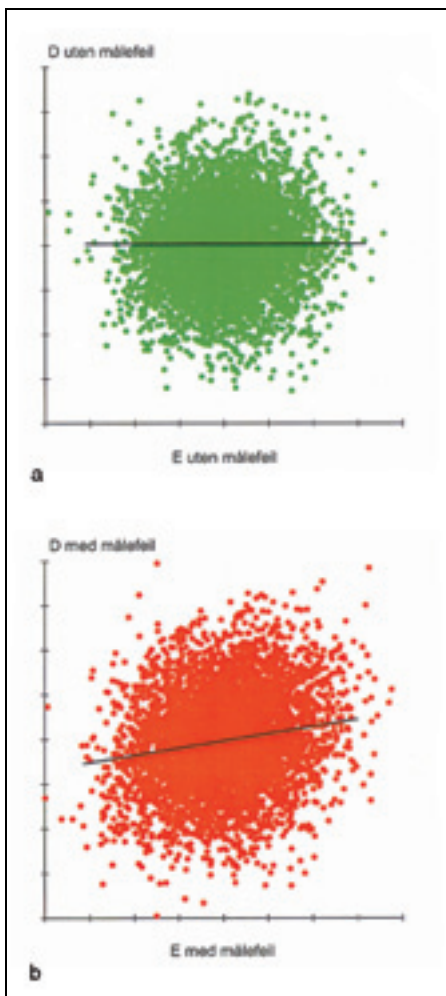
- Målefeil i to variabler er avhengige dersom graden av feil i den ene variabelen korrelerer med graden av feil i den andre
- Avhengige feil kan føre til en falskt forhøyet sammenheng mellom en forklaringsvariabel og en sykdomsindikator, spesielt i tverrsnittstudier og spørreskjemaundersøkelser
- Kilden til avhengige feil er vanligvis populasjonsvariasjon med hensyn til bestemte personlighetstrekk
- Resultat-skjevhet som skyldes avhengige feil, kan elimineres ved å bryte bindingen og innhente data fra ulike kilder

Ramme 1

- Differensielle og ikke-differensielle målefeil:
Feilene i målinger av en variabel er ikke-differensielle dersom de ikke er assosiert med den sanne verdi av de(n) andre variabelene (variabelene) i analysen; ellers er feilene differensielle (1).
- Uavhengige og avhengige målefeil:
Feilene i to variabler er uavhengige dersom feilene i den ene variabelen ikke korrelerer med feilene i den andre; ellers er feilene avhengige (1).

oddsratioene for iskemi var 1,03 (95 % KI 0,71–1,49) og 0,67 (95 % KI 0,36–1,26).

For forfatterne var det et hovedpoeng at sammenhengen mellom selvrapportert eksponering (stress) og selvrapportert utfall (angina) var klart positiv, mens analysen



Figur 1 Plott for verdier av uavhengig variabel E og avhengig variabel D i et hypotetisk datasett. Verdiene for E og D er normalfordelt med middelvei lik 0 og SD lik 1. a) Uten målefeil. Regresjonskoeffisient 0,002 (95 % KI –0,004 til +0,008). b) Med ikke-differensielle, avhengige målefeil. Regresjonskoeffisient 0,168 (95 % KI 0,162–0,175)

med et objektivt utfall (EKG-forandringer) ikke lenger viste positiv sammenheng. Tolkingen deres var at noen kan ha en generell tendens til å overrapportere både psykososiale problemer og sykdomssymptomer, noe som vil føre til falske resultater når informasjonen for begge variabler baseres på egenrapportering. De antydte også at resultatene i andre studier som var basert på selvrapporterte mål, kunne være rammet av samme problem. Artikkelen var ledsaget av en kommentar (6) og førte til debatt på hjemmesiden til *BMJ* (7).

Et hypotetisk datasett

Statistikkpakken SPSS genererte et datasett med 100 000 registreringer på variablene E og D. Verdiene av E og D var normalfordelt med middelvei lik 0 og standardavvik lik 1. Som ventet gir plottet av E og D ingen indikasjon på noen sammenheng mellom de to variablene (fig 1a). Lineær regresjon med E som uavhengig og D som avhengig variabel gir en regresjonskoeffisient nær 0.

Datasettet ble deretter manipulert ved å introdusere «målefeil». Et tilfeldig utvalg på 10 % (10 000) ble gitt «falskt» høye verdier for både E og D, mens 10 000 tilfeldig utvalgte ble gitt «falskt» lave verdier for E og D. De endrede verdiene utgjorde en økning, henholdsvis reduksjon, på ett standardavvik for både E og D. Gitt at dette var feilmålinger, fyller de kriteriene for ikke-differensielle målefeil: Feil i måling av E er ikke assosiert med sann verdi av D, og feil i måling av D er ikke assosiert med sann verdi av E. På den annen side er avhengigheten i feilmålingene total: E er feil når D er feil, og E er korrekt når D er korrekt. Det manipulerede datasettet viste en tydelig positiv sammenheng mellom E og D (fig 1b). Det skal bemerkes at eksemplet er ekstremt i den forstand at avhengigheten er total. I et mer realistisk eksempel med en viss, men ikke total avhengighet, ville resultatet blitt mindre skjevt (8).

Problemets essens

Målinger av flere variabler hvor målefeilene korrelerer, er velkjent i psykologien (8) og sosiologien (9). I sosiologien benevnes dette responssett eller faste svarmønstre, noe som kommer til uttrykk ved at samme datakilde har en tendens til enten «jasiing» eller «nei-siing» (9). Kjernen i avhengige målefeil er at ulike data i en studie kommer fra én felles kilde (common method bias (8)), og at kvaliteten på disse dataene har en felles følsomhet for forhold som er knyttet til denne kilden. Resultatskjevhet (falske assosiasjoner) på grunn av avhengige målefeil får vi når målefeilene omfatter en eksponeringsvariabel og en utfallsvariabel i en analyse.

Kildene til avhengige målefeil

Avhengige målefeil kan knyttes opp til tre kilder: Stabile personlighetstrekk hos deltakerne i en studie, situasjonsbetingede for-

hold ved datainnsamlingen og kvaliteter ved måleinstrumentene (8).

Personlighetstrekk klassifiseres på ulike måter (10). Flere av disse dimensjonene er knyttet opp til forskjeller i tolking av omgivelsene og oppfatninger om egne helseforhold. Nevrotisisme (11, 12) og negativ affektivitet (13, 14) er to dimensjoner hvor høy skåre er knyttet opp til negative og truende oppfatninger hos personen, både av omgivelser og av egen kropp. En lav skåre er knyttet til trygghet i oppfatninger og tolking. Slike personlighetstrekk synes å ha en stabil og langvarig innvirkning (11). Spektret av variasjon med hensyn til nevrotisisme eller negativ affektivitet på gruppeplanet er en plausibel årsak til avhengige målefeil av faktorer som kan tjene som eksponeringsmål og sykdomsmål i observasjonsstudier. Watson & Pennebaker (13) dokumenterte at tersklene for å rapportere stress og helseutfall påvirkes av nivået på negativ affektivitet. Dette kan forklare sammenhengen mellom stress og angina pectoris flere år senere (5): De som har høy negativ affektivitet, har lav terskel for å rapportere stress og angina, mens de med lav skåre på dette personlighetstrekket, har høy rapporteringsterskel for begge. Dalton og medarbeidere (14) studerte selvrapportert persepsjon av kjemisk eksponering (lukt, slimhinneirritasjon) og symptomer og plager under forsøk med kontrollert lav eksponering for løsemidler. Sammenliknet med en gruppe med høy skåre hadde en gruppe med lav skåre på negativ affektivitet høyere terskel for å rapportere persepsjon av kjemikalier og samtidig færre symptomer og plager.

I tverrsnittundersøkelser kan situasjonsbestemte oppfatninger være en kilde til avhengige målefeil. Dalton og medarbeidere viste også dette i studiene av kjemisk persepsjon og selvrapporterte helseplager (14). Tre grupper fikk informasjon om forsøket og kjemikalierne som var henholdsvis nøytralt, positivt ladet og negativt ladet. Det viste seg da at informasjonen i forkant hadde virkninger i ventet retning både på opplevd eksponering og på selvrapporterte symptomer.

Måleinstrumenter kan også være en kilde til avhengige målefeil (8). Spørreskjemaer som er følsomme for positive eller negative responssett, vil generere feil og resultatskjevhet dersom de omfatter spørsmål om både eksponering og utfall. Observator kan også være kilde til avhengige feil som kan klassifiseres som instrumentelle. Utilsiktet overlappning av begrepsinnholdet i ulike spørsmål kan også introdusere avhengige feil. Sammenhengen mellom stress og angina (5) kan delvis skyldes instrumentelt betingede målefeil. Stress ble vurdert med Reeders stressskjema, som helt åpenbart ikke bare måler stress som eksponering (stresspåvirkning) men også belastende kroppsreaksjoner (distress, stressreaksjon) (5). Roses spørreskjema måler antakelig ikke bare angina pectoris, men kan også inkludere et ele-

ment av stressreaksjon (15). Den observerte sammenhengen mellom Reeders stress og Roses angina kan dermed være uttrykk for sammenheng mellom stressreaksjoner og stressreaksjoner.

Resultatskjevhet som skyldes avhengige feil

Avhengige målefeil av eksponering og utfallsvariabel vil som regel føre til for høye estimater på sammenheng (8). Det vil derfor være viktigst å ha avhengige feil i tankene når validiteten av observerte positive sammenhenger skal vurderes. Problemet trenger ikke å være knyttet til forskningsspørsmålet eller hypotesemistanker hos deltakere eller forskere, men er heller betinget av normal variabilitet av karakteregenskaper i studiepopulasjonen.

Falske sammenhenger grunnet avhengige feil er sannsynligvis vanlig i observasjonsstudier, og spesielt i tverrsnittstudier, hvor deltakernes svar i spørreskjemaer er kilden til data om både eksponering og utfall. Dette underbygges av eksempler fra organisasjonspsykologi (8), arbeidsmiljøforskning (14) og inneklimateforskning (16). Man kan også mistenke at problemet forekommer på felter hvor både forklaringsvariabler og utfall kan være følsomme for personlighetstrekk som negativ affektivitet, for eksempel i psykiatrisk epidemiologi, sosialepidemiologi og muskel- og skjelettepidemiologi. Det er imidlertid viktig å presisere at avhengige feil ikke er universelt forekommende, men forutsetter at innholdet oppfattes som ladet og dermed kan feiltolkes (17).

Diskusjonen etter artikkelen i *BMJ* (5) dreide seg i stor grad om målefeil knyttet til høy stresskåre. Det er imidlertid hele variasjonsspekteret i personlighetsegenskaper som teller, både de som svarer ja og de som svarer nei bidrar til resultatskjevhet. Lynch (6) viste til sin australske bestemor. Hennes devise var «musn't grumble». I en studie omkring sammenhengen mellom to variabler kan man tenke seg at hun ville kunne bidra til sterkere sammenheng dersom de målte variablene var påvirkelige av hennes stoisisme. Det hypotetiske datasettet viser også

at avhengige feil i begge retninger bidrar like sterkt til en falskt positiv sammenheng for kontinuerlige data. Det samme vil i prinsippet også være tilfelle for kategoriske data, selv om det er misklassifisering til eksponert og syk kategori som bidrar sterkest til resultatskjevhet når forekomsten av eksponering og sykdom er lav.

Både Macleod og medarbeideres studie og det hypotetiske datasettet gir indikasjoner på at resultatskjevheten kan bli betydelig selv om de avhengige målefeilene er små. Roses spørreskjema er et nyttig og velbrukt verktøy som blant annet predikerer koronar hjertedød (15). I det hypotetiske eksemplet ville det manipulerede datasettet antakelig være fullt brukbart til deskriptive formål. I det manipulerede datasettet (fig 1b) har E og D uendret middelerverdi, og standardavviket er økt med beskjedne 10%. Det betyr at et datasett som er velegnet til beskrivelse, kan være helt uegnet til årsaksrettet analyse. Videre indikerer det at selv nøye validerte spørreskjemaer og andre verktøy ikke nødvendigvis er brukbare i enhver årsaksrettet analyse.

Hvordan unngå problemet?

Resultatskjevhet som skyldes avhengige feil, forebygges lettest ved at man er tilbakeholden med å undersøke på sammenhenger i studier hvor avhengige målefeil er sannsynlige. En slik varsomhet synes å være mest aktuell når data som undersøkes stammer fra et selvutfylt spørreskjema, og spesielt i tverrsnittundersøkelser. Det er ikke alle studiedesign som egner seg til årsaksrettet forskning. Det viktigste tiltaket for å unngå resultatskjevhet er å bryte bindingen mellom de to variablene ved å skaffe til veie data fra ulike kilder. I studiet av sammenhengen mellom stresspåvirkning og koronar hjertesykdom vil man bryte bindingen ved å bruke objektive mål på sykdom. Som et supplement til å innhente data fra ulike kilder kan forbedring av spørreskjemaer eller andre måleinstrumenter være av verdi. I psykologien er det også en tradisjon for å løse problemer med avhengige feil ved å kontrollere for stabile personlighetstrekk i analysen (8).

Slike prosedyrer forutsetter at man har data på personlighetstype, og man kan da også få indikasjoner på om man har et problem med avhengige feil og tilhørende resultatskjevhet.

Litteratur

- Greenland S. Modeling and variable selection in epidemiologic analysis. *Am J Public Health* 1989; 79: 340–9.
- Rothman KJ, Greenland S, red. *Modern epidemiology*. 2. utg. Philadelphia: Lippincott-Raven, 1997: 131.
- Chavance M, Dellatolas G, Lellouch J. Correlated nondifferential misclassifications of disease and exposure: application to a cross-sectional study on the relation between handedness and immune disorders. *Int J Epidemiol* 1992; 21: 537–46.
- Kristensen P. Bias from nondifferential but dependent misclassification of exposure and outcome. *Epidemiology* 1992; 3: 210–5.
- Macleod J, Davey Smith G, Heselop P et al. Psychological stress and cardiovascular disease: empirical demonstration of bias in a prospective observational study of Scottish men. *BMJ* 2002; 324: 1247–51.
- Lynch J. Commentary: psychosocial factors and health – strengthening the evidence base. *BMJ* 2002; 324: 1251.
- <http://bmj.com/cgi/eletters/324/7348/1247>
- Podsakoff PM, MacKenzie SB, Lee J-Y et al. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol* 2003; 88: 879–903.
- Hellevik O. *Forskningsmetode i sosiologi og statsvitenskap*. 4. utg. Oslo: Universitetsforlaget, 1989: 135.
- Larsen RJ, Buss DM. *Personality psychology: domains of knowledge about human nature*. Boston: McGraw Hill, 2002.
- Costa PT, McCrae RR. Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people. *J Pers Soc Psychol* 1980; 38: 668–78.
- Costa PT, McCrae RR. Neuroticism, somatic complaints, and disease: is the bark worse than the bite? *J Pers* 1987; 55: 299–316.
- Watson D, Pennebaker JW. Health complaints, stress, and distress: exploring the central role of negative affectivity. *Psychol Rev* 1989; 96: 234–54.
- Dalton P, Hummel T. Chemosensory function and response in idiopathic environmental intolerance. *Occup Med* 2000; 15: 539–56.
- Bulpitt CJ, Shipley MJ, Demirovic J et al. Predicting death from coronary heart disease using a questionnaire. *Int J Epidemiol* 1990; 19: 899–904.
- Ostroff C, Kinicki AJ, Clark MA. Substantive and operational issues of response bias across levels of analysis: an example of climate-satisfaction relationships. *J Appl Psychol* 2002; 87: 355–68.
- Constans JI, Penn DL, Ihen GH et al. Interpretive biases for ambiguous stimuli in social anxiety. *Behav Res Ther* 1999; 37: 643–51.