
Hva er bias?

MEDISIN OG TALL

BEN RENÉ BJØRSVIK

Ben René Bjørsvik er medisin- og forskerlinjestudent ved Universitetet i Bergen og matematikkstudent på Honours-programmet ved Universitetet i Oslo.

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

MATS JULIUS STENSRUD

mats.stensrud@epfl.ch

Mats Julius Stensrud er lege og professor ved Avdeling for matematikk, École Polytechnique Fédérale de Lausanne (EPFL) i Sveits.

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

Begrepet *bias* har ulike fortolkninger i ulike disipliner. I statistikk betyr bias at en metode ikke besvarer forskningsspørsmålet av interesse. Men er bias en egenskap til *estimander*, *estimatorer* eller *estimer*? Og hva betyr disse begrepene?



Estimatoren vil sjelden treffe estimanden perfekt. Illustrasjon: South_agency/iStock. Tilpasset av Tidsskriftet

Anta at vi ønsker å studere effekten av koloskopiscreening versus ingen screening på tiårsrisikoen for død av tarmkreft hos personer over 55 år. Svaret på dette forskningsspørsmålet, for eksempel kvantifisert som en risikodifferanse eller -ratio, er et eksempel på en *estimand*. Vi analyserer data for å finne en tilnærming til dette svaret, altså til estimanden. For eksempel kan dataene komme fra en randomisert studie, slik som NordICC-studien [\(1\)](#), hvor deltakerne gjennomgikk koloskopiscreening eller ingen screening. Med data fra en slik studie kan vi estimere tiårsrisikoen for tarmkreftdød med for eksempel en Kaplan-Meier-estimator. *Estimatoren* er en algoritme som bruker dataene til å beregne et *estimat*. I screeningeksempelet kan estimatet være en risikodifferanse mellom de som ble screenet, og de som ikke ble screenet.

Bias

Bias sier noe om hvor godt estimatoren sikter på estimanden. Rent teknisk er bias definert som den forventede differansen mellom estimatoren og estimanden. Bias er altså ikke et kjennetegn ved et estimat. Derimot forteller størrelsen oss hvorvidt vi bør forvente at estimatoren vil gi over- eller underestimerer for estimanden, gitt de tilgjengelige dataene. Antakelser om de innsamlede dataene, for eksempel at observasjonene er normalfordelte, påvirker valget av estimator og dermed risikoen for bias.

De fleste medisinske studier er forløpsstudier, hvor det er vanlig å miste informasjon om noen deltakere før studien er fullført, for eksempel ved at individer melder seg ut av studien. Vi sier at slike individer er *sensurerte*. Kaplan-Meier-estimatoren er en av estimatorene som håndterer sensurerte observasjoner. I forløpsanalyser er det vanlig å anta *uavhengig sensurering*,

altså at individer sensureres uavhengig av når hendelsen av interesse inntreffer i fremtiden (2, s. 57). I tarmkrefteksempelet betyr dette at vi bruker data fra individene inntil de dør av tarmkreft eller vi ikke lenger har oppfølgingsdata. Uavhengig sensurering er ikke alltid plausibelt (3). Hva om det viste seg at personer som var friskere ved studiestart, hadde høyere sannsynlighet for å følge hele studieløpet? Hvis helsestatus ved studiestart påvirker både hendelsen av interesse, her risikoen for tarmkreftdød, og sensureringshendelsen, så holder ikke lenger antakelsen om uavhengig sensurering. Baserer vi analysen på en feilaktig antakelse om uavhengig sensurering, introduserer vi en type *seleksjonsbias* (4).

I tillegg til uavhengig sensurering er det vanlig å anta at vi kan justere for alle felles årsaker – *konfundere* – til behandling og utfall. Felles årsaker til behandling og utfall kan fjernes ved å randomisere behandlingen i et kontrollert forsøk. Hvis studien av tarmkreftscreening var observasjonell istedenfor randomisert, kunne vi ikke utelukket felles årsaker til screening og tarmkreftdød. Vi kan tenke oss at individer med høy helsesøkende atferd oftere deltar i screeningprogrammer. Siden risikoen for død av tarmkreft påvirkes av helsesøkende atferd, er det konfundering mellom screening og tarmkreftdød. Behandler vi de observasjonelle dataene som om de var fra en randomisert studie, uten å bruke informasjon om helsesøkende atferd, vil vi ende opp med en estimator med bias. Hvis helsesøkende individer generelt er friskere enn andre, vil vi forvente at effekten av screening overestimeres.

Bias sier *noe* om kvaliteten på estimatoren, *men ikke alt*. Siden estimatoren avhenger av data, vil den også påvirkes av den naturlige variabiliteten i dataene. Dette kan tallfestes som estimatorens *varians*, noe vi ofte refererer til når vi snakker om usikkerhet i estimer. En estimator uten bias kaller vi *forventningsrett*, men det betyr ikke at estimatoren vil treffe estimanden perfekt, nettopp på grunn av variansen.

Konklusjon

Formuleringen av estimanden avhenger av hvilket spørsmål vi vil besvare, og valg av estimator avhenger av estimanden og hvilke antakelser vi tillater oss å gjøre. Bias er en størrelse som forteller oss om vi bør forvente at estimatoren vil gi et over- eller underestimat av estimanden. Kilder til bias inkluderer, men er ikke begrenset til, dårlig konstruerte estimators og feilaktige antakelser om dataene.

REFERENCES

1. Bretthauer M, Kaminski MF, Løberg M et al. Population-Based Colonoscopy Screening for Colorectal Cancer: A Randomized Clinical Trial. JAMA Intern Med 2016; 176: 894–902. [PubMed][CrossRef]
2. Aalen OO, Borgan Ø, Gjessing HK. Survival and Event History Analysis. New York, Ny: Springer New York, 2008.

3. Gran JM, Stensrud MJ. Avhengig sensurering. Tidsskr Nor Legeforen 2024; 144. doi: 10.4045/tidsskr.23.0685. [PubMed][CrossRef]
 4. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004; 15: 615–25. [PubMed][CrossRef]
-

Publisert: 6. februar 2025. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.24.0595

Opphavsrett: © Tidsskriftet 2025 Lastet ned fra tidsskriftet.no 25. desember 2025.