
Hvilke variabler bør være med i en regresjonsmodell?

MEDISIN OG TALL

STIAN LYDERSEN

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.

EVA SKOVLUND

Eva Skovlund er professor i medisinsk statistikk ved Institutt for samfunnsmedisin og sykepleie, NTNU og spesialrådgiver ved Direktoratet for medisinske produkter.

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

Valg av variabler som skal inkluderes i en regresjonsmodell, avhenger i stor grad av forskningsspørsmålet og studiedesignen.

Det kan være hensiktsmessig å dele observasjonelle studier i tre kategorier: deskriptive, prediktive og kausale (1). Eksempler på forskningsspørsmål i de tre kategoriene er vist i tabell 1. I de to sistnevnte er det ofte aktuelt å benytte en regresjonsmodell i den statistiske analysen.

Tabell 1

Eksempler på forskningsspørsmål i ulike kategorier observasjonelle studier

Deskriptiv	Prediktiv	Kausal
Prevalens av atrieflimmer blant individer > 65 år	Sannsynlighet for iskemisk hjerneslag for individer med gitte karakteristika	Effekt av antikoagulasjonsbehandling på hjerneslagrisiko for individer med gitte karakteristika

Prediksjonsmodeller

Formålet med en prediktiv studie kan for eksempel være å lage en prediksjonsmodell som kan brukes til å estimere sannsynligheten for hjerneslag innen 10 år for en person med visse karakteristika (kjønn, alder, blodtrykk, m.m.). I en prediksjonsmodell inkluderer man de variablene som samlet sett gir en god prediksjon, basert på data. Man gjør ingen kausal fortolkning av effekten av de inkluderte variablene [\(2\)](#).

En praktisk anvendelig prediksjonsmodell bør kun inkludere variabler som er enkle å måle i det kliniske arbeidet. En variabel som baseres på observasjoner gjort av klinikere, bør i tillegg ha høyt intra- og interobservatørsamsvar. Variabelen bør være sterkt assosiert med det som skal predikeres, men bør ikke inkluderes kun basert på *p*-verdi. Størrelsen av estimatet er heller ikke alltid det viktigste. For eksempel vil en binær variabel med oddsratio på 2,0 være mer relevant enn en (også binær) variabel med oddsratio på 2,5, dersom den første variabelen har en prevalens på 50 % og den andre en prevalens på 1 % [\(3\)](#). Nyttige retningslinjer for slike kliniske prediksjonsmodeller er beskrevet av Collins og medarbeidere [\(3\)](#).

Prediksjonsmodeller bør valideres. Dette kan for eksempel gjøres ved å dele opp det aktuelle datasettet i et treningssett (som man bruker til å bygge modellen) og et testsett, eller ved å bruke tilfeldige utvalg fra datasettet, som ved *bootstrapping* [\(4\)](#). I tillegg er det ønskelig å teste modellen på et eksternt datasett.

Kausale modeller

Randomiserte kontrollerte studier er ideelle for å trekke kausale slutninger. Siden slike studier ikke alltid kan gjennomføres, blir observasjonelle studier også benyttet til å estimere effekt av en eksponering eller behandling. Det er ofte nyttig å tegne en rettet asyklisk graf (directed acyclic graph, DAG) for å skaffe oversikt over hvilke variabler som kan påvirke eksponering og utfall. Disse kalles konfunderende variabler, og må tas med i modellen. Her bør man både bruke klinisk skjønn og basere seg på publisert litteratur. Under strenge antagelser, blant annet ingen umålt konfundering, er det mulig å estimere kausale effekter fra observasjonelle data. Slike antakelser kan imidlertid ikke verifiseres ut fra observerte data, og i praksis kjenner vi ikke nødvendigvis alle mulige konfundere. Likevel kan det noen ganger være rimelig å anta at sammenhengen er kausal dersom studien er godt planlagt og gjennomført [\(5, 6\)](#).

I en kausal studie er det eksponeringen som er det vesentlige, og effekten av de andre variablene i analysen fortolkes ikke. I en prediksjonsstudie er derimot alle de inkluderte variablene relevante [\(7\)](#).

Interaksjoner

Noen ganger avhenger sammenhengen mellom eksponering og utfall av en annen variabel, for eksempel kan effekten av en behandling være avhengig av nivået av en biomarkør. Da er det rimelig å inkludere et interaksjonsledd i modellen. Igjen er et statistisk signifikansnivå på 5 % neppe det beste kriteriet for å vurdere hvorvidt det er en interaksjon, blant annet fordi interaksjonstester har lav styrke. Dersom det er en tydelig interaksjonseffekt, bør estimatene presenteres separat for de aktuelle undergruppene.

Følg analyseplanen

Før man gjennomfører analysene, bør man skrive en statistisk analyseplan [\(8\)](#). Et sentralt punkt i planen vil være hvilke variabler som skal inkluderes i analysen, og en begrunnelse for valget.

LITTERATUR

1. Hernán MA, Hsu J, Healy B. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *Chance* 2019; 32: 42–9. [[CrossRef](#)]
2. van Diepen M, Ramspek CL, Jager KJ et al. Prediction versus aetiology: common pitfalls and how to avoid them. *Nephrol Dial Transplant* 2017; 32 (suppl_2): ii1–5. [[PubMed](#)][[CrossRef](#)]
3. Collins GS, Moons KGM, Dhiman P et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024; 385: e078378. [[PubMed](#)][[CrossRef](#)]
4. Steyerberg EW. Clinical prediction models. A practical approach to development, validation, and updating. 2. utg. New York, NY: Springer, 2019.
5. Gran JM, Stensrud MJ. Målforsøk. *Tidsskr Nor Legeforen* 2021; 141: 466. [[PubMed](#)][[CrossRef](#)]
6. Hernán MA. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *Am J Public Health* 2018; 108: 616–9. [[PubMed](#)][[CrossRef](#)]
7. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol* 2013; 177: 292–8. [[PubMed](#)][[CrossRef](#)]
8. Lydersen S, Simpson MR. Statistisk analyseplan. *Tidsskr Nor Legeforen* 2026; 146. doi: 10.4045/tidsskr.25.0749. [[CrossRef](#)]

Publisert: 20. februar 2026. *Tidsskr Nor Legeforen*. DOI: 10.4045/tidsskr.25.0750

