
Hva kjennetegner en god prediksjonsmodell?

MEDISIN OG TALL

ARE HUGO PRIPP

apripp@ous-hf.no

Are Hugo Pripp er forsker og biostatistiker ved Oslo senter for biostatistikk og epidemiologi, Forskningsstøtteavdelingen, Oslo universitetssykehus. Han er professor II ved Fakultet for helsevitenskap, OsloMet – storbyuniversitetet.

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

Det finnes mange kliniske prediksjonsmodeller som hevder å forutsi risikoen for nåværende eller fremtidig sykdom. Men hvilke egenskaper kjennetegner en god modell?

Kliniske prediksjonsmodeller som vil si noe om pasientens diagnose, prognose eller valg av behandlingsforløp, utvikles fortløpende. For å vurdere kvaliteten og relevansen av disse modellene bør egenskapene til modellen være godt kartlagt (1, 2). Et illustrativt eksempel kan være beskrivelsen av en prediksjonsmodell om tilbakefall ved multippel sklerose (3).

Forskningsspørsmålet, dataene og den statistiske analysen

Det første steget bør være å formulere forskningsspørsmålet og å undersøke om tilgjengelige data inneholder relevante prediktor- og utfallsvariabler.

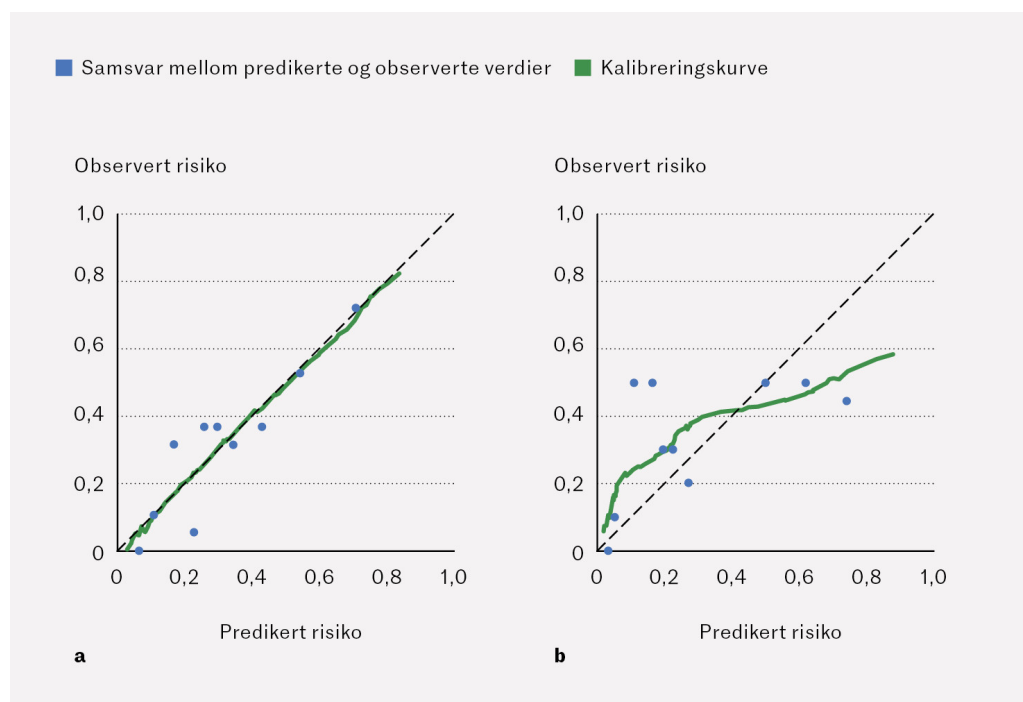
Forskningsgruppen som utviklet modellen om tilbakefall ved multippel sklerose, gjennomførte systematiske litteratursøk for å finne andre modeller og relevante prediktorer. Deretter bør man vurdere valg av statistiske analyser og avgjøre om man vil bruke klassiske statistiske metoder, som lineær, logistisk eller Cox-regresjon, utforske nyere metoder innen maskinlæring og kunstig intelligens eller en kombinasjon av flere metoder.

Det er særlig viktig å vurdere utvalgsstørrelsen med tanke på valg av statistiske analyser (4). Manglende data reduserer den effektive utvalgsstørrelsen, og erstatninger som multiple imputasjoner kompliserer analysene, valideringen og tolkningen. Den nevnte studien om multipel sklerose valgte en avansert variant av logistisk regresjon, men kunne likevel bruke pakken `pmsampsize` som finnes i både R og Stata for å anslå nødvendig utvalgsstørrelse.

Diskriminering og kalibrering

Egenskapene til prediksjonsmodeller vurderes ofte med hensyn til diskriminering og kalibrering. Diskriminering angir hvor godt modell skiller mellom individer med og uten utfallet. Et vanlig statistisk mål for dette er C-indeksen, også kjent som «arealet under kurven» (AUC). En verdi på 0,5 indikerer at modellen ikke er bedre enn tilfeldig gjetning, mens en perfekt klassifikasjon av modellen har en C-indeks på 1,0. I modellen for multipel sklerose var C-indeksen litt under 0,7, noe som er ganske vanlig for mange kliniske prognostiske modeller.

Kalibrering beskriver hvor godt de predikerte verdiene samsvarer med de observerte. Et kalibreringsplott, der observerte utfall plottes mot predikerte verdier, gir nyttig informasjon. Dersom de predikerte verdiene er svært like de observerte, vil de ligge nær diagonalen i plottet (figur 1a).



Figur 1 Eksempel på kalibreringsplott laget med Stata-pakken `pmcalplot`. Plottet er basert på et datasett om lav fødselsvekt fra Stata-manualen (5) og en logistisk prediksjonsmodell. Modellen er a) utviklet i hele datasettet og b) gjentatt på et mindre utvalg av dataene og deretter validert på de resterende dataene.

Validering

En god klinisk prediksjonsmodell bør være nyttig også i nye pasientutvalg. Større avvik mellom observerte og predikerte verdier gir en kalibreringskurve som avviker fra diagonalen (figur 1b). En intern validering der ulike statistiske metoder brukes for å teste modellen på de samme dataene som den ble utviklet på, er et minimumskrav. *Bootstrapping*, der datasettet stokkes om, modellen beregnes på nytt og deretter testes på de opprinnelige dataene, er en anbefalt metode for intern validering (6). Dette gjør det mulig å korrigere for overtilpasning og justere modellen (ofte referert til som *krymping*) for å forbedre dens egenskaper på nye data.

På grunn av en kompleks analyse i den prognostiske studien om multippel sklerose, ble det utviklet et spesialtilpasset program for intern validering. For enklere regresjonsmodeller finnes det ferdige verktøy, som pakkene rms og psfmi i R, eller den brukerutviklete funksjonen bsvalidation i Stata. Ideelt sett bør modellen også valideres eksternt, det vil si på uavhengige datasett fra ulike pasientpopulasjoner, geografiske områder eller tidsperioder.

Presentasjon, rapportering og bruk

En fullstendig og tydelig dokumentasjon, presentasjon og rapportering av kliniske prediksjonsmodeller er avgjørende for å vurdere resultatene og studiekvaliteten. TRIPOD-erklæringen, utarbeidet av en spesialistgruppe innen forskningsmetodikk, statistikk og medisin, gir anbefalinger om dette (7). TRIPOD inkluderer en sjekkliste over essensielle elementer for å sikre systematisk og klar rapportering.

Utvikling av enkle skåringssystemer, skalaer eller kliniske kalkulatorer kan øke modellens nytteverdi og bruk. Sist, men ikke minst, bør prediksjonsmodellen helst testes ut på lik linje med nye behandlinger for å undersøke om den faktisk hjelper leger og pasienter. Dessverre når få prediksjonsmodeller dette stadiet, men nært samarbeid mellom leger og statistiskere i utviklingsprosessen er et avgjørende steg for å oppnå dette.

REFERENCES

1. Efthimiou O, Seo M, Chalkou K et al. Developing clinical prediction models: a step-by-step guide. *BMJ* 2024; 386: eo78276. [PubMed][CrossRef]
2. Hoesseini A, van Leeuwen N, Sewnaik A et al. Key Aspects of Prognostic Model Development and Interpretation From a Clinical Perspective. *JAMA Otolaryngol Head Neck Surg* 2022; 148: 180–6. [PubMed][CrossRef]
3. Chalkou K, Steyerberg E, Bossuyt P et al. Development, validation and clinical usefulness of a prognostic model for relapse in relapsing-remitting

- multiple sclerosis. *Diagn Progn Res* 2021; 5: 17. [PubMed][CrossRef]
4. Pripp AH. Utvalgsstørrelse for en prediksjonsmodell. *Tidsskr Nor Legeforen* 2024; 144. doi: 10.4045/tidsskr.24.0313. [PubMed][CrossRef]
 5. STATA Press. Datasets for Stata Base Reference Manual, Release 18, lbw.dta. <https://www.stata-press.com/data/r18/lbw.dta> Lest 11.12.2024.
 6. Steyerberg EW, Harrell FE, Borsboom GJ et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001; 54: 774–81. [PubMed][CrossRef]
 7. Moons KG, Altman DG, Reitsma JB et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; 162: W1–73. [PubMed][CrossRef]
-

Publisert: 18. februar 2025. *Tidsskr Nor Legeforen*. DOI: 10.4045/tidsskr.24.0570
Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 28. mars 2026.