
Utvalgsstørrelse for en 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.

Pasienter spør ofte legen om risikoen for nåværende eller fremtidig sykdom. I tillegg til medisinsk kunnskap og klinisk erfaring, er kliniske prediksjonsmodeller et viktig verktøy for både diagnose og prognose.

Kliniske prediksjonsmodeller hjelper til med å forutse helseutfall eller helsetilstander. Disse modellene kan variere fra enkle trafikklyslignende metoder til komplekse «svart boks»-matematiske modeller og maskinlæring (1). Ofte brukes klassiske statistiske metoder som lineær, logistisk eller Cox-regresjon. Prediksjonsmodeller utviklet fra datasett med få deltagere, kan medføre ustabile, usikre og feilaktige prediksjoner, selv om p-verdiene er signifikante. Hvis vi bruker slike prediksjonsmodeller på nye data eller pasienter, kan de fungere dårlig (2).

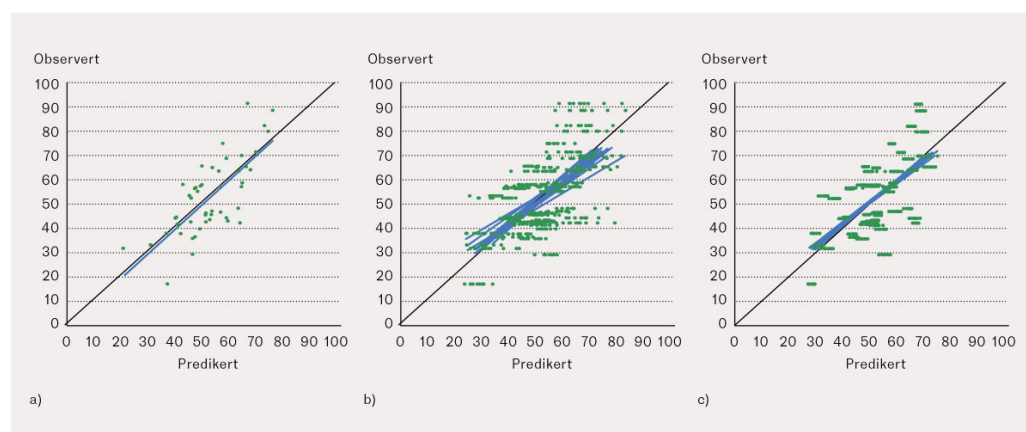
Metoder for å beregne antall deltagere

For prediksjonsmodeller med et kontinuerlig utfall, er den effektive utvalgsstørrelsen lik antallet deltagere. For binære utfall eller tid-til-hendelse-utfall, er den effektive utvalgsstørrelsen det laveste antall deltagere med eller uten hendelser for binære utfall, eller det totale antallet ikke-sensurerte hendelser for overlevelsesanalyse. Det betyr at i et datasett med 1000 deltagere, der bare 10 av dem har et utfall eller en hendelse, vil den effektive utvalgsstørrelsen til prediksjonsmodellen være 10, og ikke 1000. En populær tommelfingerregel er å ha minst 10 deltakere for lineær regresjon eller 10 hendelser for binære utfall eller overlevelsesanalyse per beregnet parameter (koeffisient eller stigningstall) i prediksjonsmodellen. Denne regelen er enkel, men den er basert på simuleringsstudier og er omdiskutert (3).

En mer avansert tilnærming er å beregne den nødvendige utvalgsstørrelsen basert på grunnleggende prinsipper for en valid og stabil prediksjonsmodell. Dette innebærer et presist estimat av den samlede utfallsrisikoen eller gjennomsnittlige utfallsverdien, en lav feilmargen for estimerte verdier til alle individ og lav grad av modellens såkalte overtilpasning (4). Det er utviklet programpakker der man kan beregne nødvendig utvalgsstørrelse ved ulike betingelser og antagelser (5).

Validitet og stabilitet

Utvalgsstørrelsen er avgjørende for stabiliteten og validiteten til kliniske prediksjonsmodeller (2). La oss illustrere dette med et enkelt eksempel. Vi antar at det i pasientpopulasjonen er fem kontinuerlige prediksjonsvariabler som alle har samme effekt (likt stigningstall) på et kontinuerlig klinisk utfall, og at de sammen i en lineær multippelregresjonsmodell vil forklare 50 % av den totale variasjonen til utfallet. Sammenhengen mellom observerte og predikerte utfallsverdier fra en multippel regresjonsmodell i et tilfeldig utvalg med 50 pasienter er vist i figur 1a. Modellen gir best mulig tilpasning til dataene i utvalget, og avvik mellom observerte og predikerte verdier skyldes naturlig statistisk variasjon. Figur 1b viser 10 prediksjonsmodeller i 10 tilfeldige utvalg av 50 personer fra samme pasientpopulasjon, og disse er brukt til å estimere utfallet til de opprinnelige 50 pasientene i figur 1a. De predikerte utfallene varierer betydelig mellom disse 10 prediksjonsmodellene, noe som indikerer at modellene er ganske ustabile. De tilpassete linjene mellom observerte og predikerte verdier ligger heller ikke på diagonalen, noe som skyldes det statistiske fenomenet overtilpasning (6). Når utvalgsstørrelsen økes til 500 pasienter i hver de av nye 10 prediksjonsmodellene, blir både variasjonen mellom predikerte verdier fra disse modellene mindre og prediksjonen av de opprinnelige observerte utfallsverdiene bedre (figur 1c). Prediksjonsmodeller utviklet med mange deltagere er mer nøyaktige og stabile.



Figur 1 Sammenhengen mellom observerte og predikerte verdier med tilpasset lineær kurve fra a) en lineær multippel regresjonsmodell med 50 pasienter, b) 10 prediksjonsmodeller utviklet fra nye tilfeldige utvalg av 50 pasienter og prøvd ut på de opprinnelige 50 pasientene i figur a, og c) 10 prediksjonsmodeller utviklet fra nye tilfeldige utvalg av 500 pasienter prøvd ut på de opprinnelige deltagerne i figur a). Dataene er simulerte.

REFERENCES

1. van Smeden M, Reitsma JB, Riley RD et al. Clinical prediction models: diagnosis versus prognosis. *J Clin Epidemiol* 2021; 132: 142–5. [PubMed][CrossRef]
2. Riley RD, Collins GS. Stability of clinical prediction models developed using statistical or machine learning methods. *Biom J* 2023; 65: e2200302. [PubMed][CrossRef]
3. van Smeden M, de Groot JAH, Moons KGM et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med Res Methodol* 2016; 16: 163. [PubMed][CrossRef]
4. Riley RD, Ensor J, Snell KIE et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020; 368: m441. [PubMed][CrossRef]
5. Ensor J. PMSAMPSIZE: Stata module to calculate the minimum sample size required for developing a multivariable prediction model.
<https://ideas.repec.org/c/boc/bocode/s458569.html> Lest 28.6.2024.
6. Lever J, Krzywinski M, Altman N. Model selection and overfitting. *Nat Methods* 2016; 13: 703–4. [CrossRef]

Publisert: 14. august 2024. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.24.0313

Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 15. juli 2026.