



Tidsskriftet
DEN NORSKE LEGEFORENING

Analyse av longitudinelle data

MEDISIN OG TALL

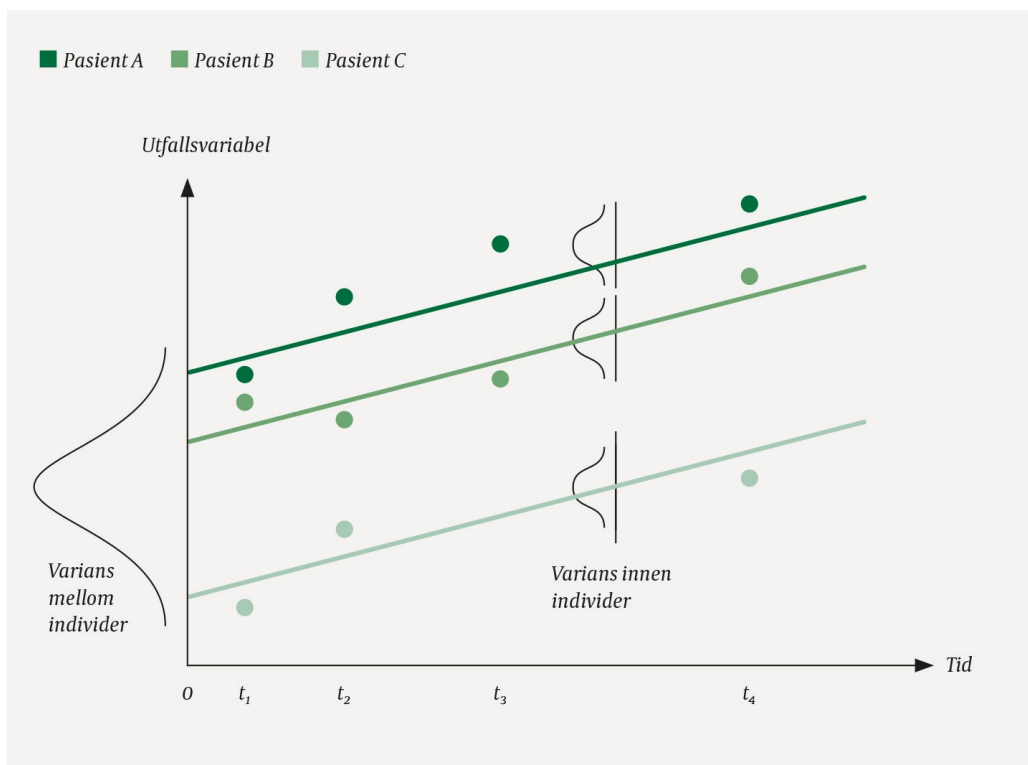
STIAN LYDERSEN

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.

I enkelte studier blir utfallsvariabelen målt flere ganger på samme person, for eksempel i en studie der pasientene undersøkes ved flere oppfølgingstidspunkt. Da må man bruke en statistisk metode for longitudinelle data.

La oss starte med et eksempel: I en randomisert kontrollert studie sammenliknet vi to behandlingsforløp for pasienter med hoftebrudd (1). Den primære utfallsvariabelen var mobilitet, målt med screeningtesten *Short Physical Performance Battery* (SPPB), som gir en skalavariabel som går fra 0 til 12. Pasientenes mobilitet ble undersøkt ved fire tidspunkt: fem dager, én måned, fire måneder og tolv måneder etter kirurgi. Figur 1 viser, noe forenklet, hvordan forløpet kunne ha blitt for tre pasienter målt ved fire tidspunkt.



Figur 1 Utvikling over tid for tre tenkte pasienter, A, B og C, målt ved fire tidspunkt.

Lineær blandet effekt-modell

Det finnes ulike statistiske metoder for analyse av longitudinelle data. Det er naturlig å ta utgangspunkt i en regresjonsmodell. Vi skal her fokusere på en lineær blandet effekt-modell (eng. *linear mixed effects model*), hvor vi tenker oss at hvert individ har sin egen regresjonslinje, som illustrert i figur 1. Enkeltmålingene varierer rundt linjen, med en varians innen individer (*within subjects variance*). Regresjonslinjene varierer også i forhold til hverandre. Dette kvantifiseres med en varians mellom individer (*between subjects variance*). I figur 1 er regresjonslinjene parallelle. Dette er naturlig nok en forenklet modell av virkeligheten, men den viser seg å passe ganske godt i mange studier.

Grunnen til at dette kalles en blandet effekt-modell, er at den inneholder minst én fast effekt, her effekten av tid, altså stigningstallet, og minst én tilfeldig effekt, her en tilfeldig effekt av individet, som er uttrykt ved varians mellom individene. Hvis utviklingen over tid ikke følger en rett linje som i figur 1, kan man for eksempel bruke indikatorvariabler for tidspunktene t_1 , t_2 osv., med tilhørende stigningstall for hvert tidspunkt. Dette ble gjort i den nevnte studien (1).

En av fordelene med en blandet effekt-modell er at den inkluderer data fra alle individene i beregningene, også de som mangler data på ett eller flere tidspunkt, slik som pasient C i figur 1. Og resultatene vil være forventningsrette (*unbiased*) selv om data bare mangler betinget tilfeldig (2). Hvis individer som har lave verdier i utgangspunktet, slik som pasient C i figur 1, har større andel manglende verdier ved senere tidspunkt, så mangler ikke data helt tilfeldig, men kan mangle betinget tilfeldig.

En viktig anvendelse av en blandet effekt-modell er å undersøke om endringen over tid er avhengig av en eksponering, som f.eks. behandling. Da inkluderer man også interaksjonsleddet mellom eksponering og tid for å kunne undersøke om stigningstallet er forskjellig i eksponeringsgruppene.

I en lineær blandet effekt-modell kan man også inkludere flere uavhengige variabler, på samme måte som i en vanlig regresjonsanalyse. Det kan for eksempel være aktuelt å inkludere alder og kjønn som konfundere i en observasjonell studie, eller som viktige prediktorer av utfallsvariabelen i en randomisert studie (3), noe som ble gjort i studien

nevnt ovenfor (1). Men merk at utgangsverdien av utfallsvariabelen normalt ikke skal være med som en uavhengig variabel, men som avhengig variabel ved starttidspunktet. Dette kommer vi tilbake til i en senere artikkel om randomiserte kontrollerte studier.

En alternativ metode

En metode som kalles *repeterte målingers ANOVA*, var tidligere mye brukt. Denne metoden har mange ulemper: Den matematiske modellen er lite transparent, resultatene er vanskelige å tolke, bare individer med komplette data blir med i analysen, og resultatene er forventningsrette bare hvis data mangler helt tilfeldig. I dag frarådes vanligvis repeterte målingers ANOVA (4).

En lineær blandet effekt-modell kan generelt anbefales for longitudinelle data med kontinuerlig utfallsvariabel.

REFERENCES

1. Prestmo A, Hagen G, Sletvold O et al. Comprehensive geriatric care for patients with hip fractures: a prospective, randomised, controlled trial. *Lancet* 2015; 385: 1623–33. [PubMed][CrossRef]
2. Lydersen S. Manglende data – sjelden helt tilfeldig. *Tidsskr Nor Legeforen* 2019; 139. doi: 10.4045/tidsskr.18.0809. [PubMed][CrossRef]
3. Lydersen S. Bør man justere for bakgrunnsvariabler i en randomisert studie? *Tidsskr Nor Legeforen* 2020; 140. doi: 10.4045/tidsskr.19.0685. [PubMed][CrossRef]
4. McCulloch CE. Repeated Measures ANOVA, R.I.P.? *Chance* 2005; 18: 29–33. [CrossRef]

Publisert: 15. mars 2022. *Tidsskr Nor Legeforen*. DOI: 10.4045/tidsskr.21.0740

© Tidsskrift for Den norske legeforening 2022. Lastet ned fra tidsskriftet.no 11. august 2022.