

Uten bruk av statistiske metoder ville ikke klinisk medisin vært der den er i dag

Et uunnværlig verktøy

Biologisk forståelse er selvsagt hjørnesteinen i klinisk medisin. Men selv når vi mener vi har en god mekanistisk forståelse av en sykdom, er det ikke sjelden at et behandlingsprinsipp som i teorien burde virke, viser seg å ha liten eller marginal effekt i praksis. Et gammelt og velkjent eksempel er CAST-studien (1), som overraskende viste at antiarytmisk behandling med enkainid og flekainid ikke reduserte mortaliteten etter hjerteinfarkt sammenliknet med placebo, men tvert imot økte den. Innen legemiddelutvikling er det mange fase 3-studier som mislykkes i å vise effekt av behandling på tross av at resultater fra tidlige studier i begrensede utvalg av friske frivillige og pasienter (fase 1 og fase 2) har vist aktivitet av virkestoffet (2, 3). En mekanistisk forståelse er med andre ord ofte utilstrekkelig – trolig fordi vi kun forstår en liten flik av helheten. Det er da statistiske metoder viser seg å være uerstattelige. Når man gjør randomiserte intervensjonsstudier, trenger ikke engang analysemetodene å være spesielt kompliserte.

I dette nummer av Tidsskriftet er det tre artikler som på hver sin måte omhandler faget medisinsk statistikk. Disse illustrerer at statistikk er blitt et uunnværlig verktøy i medisinsk forskning (4, 5) og at sannsynlighetsregning og statistikk er en del av den kliniske hverdagen (6). Vår økende innsikt i biologiske mekanismer har ikke nødvendigvis gjort det enklere å forstå sykdom, identifisere risikofaktorer eller utvikle nye behandlingsstrategier. Når kompleksiteten øker, blir det tydelig at det er nødvendig å benytte statistiske metoder for å forklare og forstå.

Som kjent er medisinsk forskning i stor grad basert på å trekke slutninger ved hjelp av p-verdier. Såfremt ikke p-verdier misbrukes og overtolkes, er det en helt grei tilnærming som (i kombinasjon med estimater av effektstørrelse med tilhørende konfidensintervall) har vist seg å ha stor praktisk nytte, spesielt i analysen av eksperimentelle studier. Men som Pripp (4) påpeker i sin artikkel, er p-verdien for mange et orakelsvar man trenger kompetanse til å tolke. Selv enkle analyser kan lede til feilslutninger, for eksempel når man utfører mange signifikanstester.

For eksperimentelle studier som er basert på randomisering og en fornuftig forsøksplan, er det gjerne tilstrekkelig å benytte elementære statistiske metoder til å gjøre sammenlikninger og estimere effekt. Når man randomiserer pasienter til forskjellige behandlingsgrupper, sørger man implisitt for at en eventuell observert forskjell i effekt enten skyldes faktisk effektforskjell eller tilfeldighet. Dermed er p-verdiene direkte tolkbare.

Observasjonelle studier er ofte en større utfordring fordi resultatene kan være forstyrret av systematiske skjevheter eller konfunderende (bakenforliggende) faktorer man ikke har målt. Dermed kan det vise seg vanskelig å påvise kausale sammenhenger. Gullstandarden for å etablere kausalitet er de randomiserte studiene, men i de senere år har *kausal inferens* vokst frem som et viktig forskningsfelt, og såkalt rettede asykliske grafer (directed acyclic graph, DAG) blir stadig oftere benyttet innen epidemiologisk forskning. Slike grafer er nyttige til å tydeliggjøre sammenhenger og som verktøy til å velge hvilke variabler som bør være med i en regresjonsmodell, men kan ofte bli kompleks og blir selvsagt ikke bedre enn vår (til tider begrensede) forståelse av sammenhenger og mekanismer. Det fremgår tydelig i artikkelen til Stensrud & Aalen (5) at kausalitet er et utfordrende felt.

En tilbakevendende debatt er valget mellom frekventistisk og bayesiansk metode. Den frekventistiske tradisjonen, som er klart mest brukt i medisinen, er basert på at vi formulerer en nullhypotese

(den påstanden vi ønsker å motbevise) og benytter data til eventuelt å forkaste denne dersom p-verdien er lav.

Dersom vi ikke kan gjøre randomiserte studier og det er vanskelig å skaffe nye data om tilstrekkelig mange pasienter, kan en bayesiansk tilnærming være et alternativ. Bayesiansk statistikk oppsummerer det man på forhånd vet (eller antar) om en ukjent parameter i en såkalt apriorifordeling. Fordelingen av de nye dataene fra forsøket sammenfattes med denne forhåndskunnskapen i en posteriorifordeling, og sannsynligheten for en hypotese kan tallfestes fordi vi beregner «sannsynligheten for hypotesen gitt data» og ikke «sannsynligheten for data gitt hypotesen» som i en frekventistisk tilnærming. Dette er intuitivt attraktivt, og bayesiansk tenkning utnyttes, som Brakedal (6) viser, mer eller mindre ubevisst i for eksempel diagnostisk øyemed.

En grunn til at bayesianske metoder har vært mindre brukt enn frekventistiske innen medisinsk forskning, er at de er beregningsmessig tunge og var nesten umulige å anvende i praksis før vi fikk tilgang til kraftige datamaskiner. En annen og enda viktigere grunn til at de er lite benyttet i klinisk forskning, er at vi i eksperimentelle situasjoner nettopp *ikke* ønsker å være påvirket av forhåndsoppfatninger. For eksempel er det vanlig at et legemiddelfirma er svært optimistisk når det gjelder effekten av et nytt legemiddel, men virkeligheten viser dessverre at effekten i randomiserte studier ofte er mer begrenset enn det som var forventet – eller til og med fraværende. For å unngå at en altfor positiv forhåndsoppfatning skal få lov til å påvirke konklusjonen, synes det mest hensiktsmessig å benytte en frekventistisk tilnærming. Man kan for øvrig legge inn en såkalt ikke-informativ apriorifordeling når man benytter en bayesiansk tilnærming. Da gjøres estimeringen kun på grunnlag av data, og konklusjonen blir dermed den samme som når vi benytter tradisjonelle frekventistiske metoder.

Uansett tilnærming er den økende viktigheten av statistikk som verktøy i alle deler av medisinen slående. Modellering av biologiske mekanismer som forklarer sykdom samt effekt av forebygging og behandling øker i kompleksitet, og konklusjoner kan sjelden trekkes kun ut fra en mekanistisk forståelse av biologien. Medisinsk forskning er dermed ofte helt avhengig av til dels avansert statistisk analyse.

Eva Skovlund
eva.skovlund@ntnu.no

Eva Skovlund (f. 1959) er professor ved Institutt for samfunnsmedisin, Norges teknisk-naturvitenskapelige universitet. Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Litteratur

1. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989; 321: 406–12.
2. Arrowsmith J. Trial watch: phase III and submission failures: 2007–2010. *Nat Rev Drug Discov* 2011; 10: 87.
3. Hay M, Thomas DW, Craighead JL et al. Clinical development success rates for investigational drugs. *Nat Biotechnol* 2014; 32: 40–51.
4. Pripp AH. Hvorfor p-verdien er signifikant. *Tidsskr Nor Legeforen* 2015; 135: 1462–4.
5. Stensrud MJ, Aalen OO. Hva kan vi si om kausalitet? *Tidsskr Nor Legeforen* 2015; 135: 1465–7.
6. Brakedal B. En kliniker og en bayesianer. *Tidsskr Nor Legeforen* 2015; 135: 1468–70.

 Engelsk oversettelse på www.tidsskriftet.no