

# Statistikk og Tidsskriftet

Statistiske metoder er nyttige, men komplekse hjelpemidler for å forstå observasjoner og målinger. En kokebokoppskrift fra Tidsskriftet kan legitimere overfladisk bruk av statistikk.

*Oppgitte interessekonflikter:* Ingen

> Se også side 189

Teoretisk matematikk og statistikk er ikke uten videre et anvendbart forskningsredskap i et praktisk fag som medisin. I andre naturvitenskaper hvor man må løse praktiske problemer, benyttes slik matematikk i bare liten utstrekning. Innenfor medisin har statistiske metoder gitt dokumenterte feil som systematisk misbruk av p-verdier, for eksempel forutsetningen for Cochranedatabasen. Svak innsikt i slik matematikk, selv hos akademisk kvalifiserte leger, gir muligheten for misforståelser og misbruk. «Kokebokoppskriften» for bruk av statistikk i Tidsskriftet fra Geir Aamodt og medarbeidere i nr. 16/2005 (1) er betenkelig, slik også Kunnskapssenterets ensidige bruk av de randomiserte tester som gullstandard for måling av effekter (2) er det. Risikoen er at man legitimerer feil og unnlater å gjøre en dypere utforskning av egne data.

En beskrivelse av observasjoner og målinger bør omfatte både tyngdepunkt og spredning, men i medisin anvendes ofte ikke den informasjon som spredning gir. Denne praksis slutter dessverre Tidsskriftet seg til (1). Det er imidlertid av interesse både om et medikament gir en gjennomsnittlig senkning av blodtrykket og om variasjonen i blodtrykk mellom personene er den samme før og etter behandling. Økt spredning taler sterkt for en varierende individuell effekt av behandlingen. En analyse av spredningen (variansanalyse) viser at alle som tar alendronat som foreskrevet, har effekt av medikamentet, selv hos dem som får målt en lavere beinmengde etter behandling. Hvis man også viser at variasjonen er skjev etter behandling, er det rimelig å postulere at pasientene består av grupper med forskjellig respons.

## Sannsynlighet og spredning

Den statistiske p-verdien og 95 % konfidensintervall angir usikkerheten til estimatet av

utvalgets tyngdepunkt (1). Det blir feil når det anføres tre alternative mål på medianens usikkerhet, nemlig ekstremverdier (spredning), kvartiler og 95 % konfidensintervaller. 95 % konfidensintervall beskriver det intervallet som tyngdepunktet vil kunne forventes å befinne seg innenfor i 95 av 100 tilfeller. Tilsvarende gjelder for standardfeil (standard error of the mean, SEM). Konfidensintervallet og SEM minker med økt antall observasjoner. For normalfordelte målinger endres p-verdien 100 % lineært i forhold til konfidensintervallet. Hvis man måler alle aktuelle objekter, er tyngdepunktet målt – ikke estimert – og intervallet er da definisjonsmessig lik null. Spredning og kvartiler samt standardavvik (SD) og varians ( $SD^2$ ) har en helt annen funksjon. De fortsetter å øke med økt antall objekter som måles, og har maksimal størrelse når alle objekter er målt. Dette er spredningsparametere som beskriver variasjonen, ikke statistisk usikkerhet. Måleusikkerhet forårsaket av apparatinstabilitet eller definisjonspresisjon, ofte benevnt målefeil, skal beskrives med spredningsparametere, ikke med konfidensintervall eller SEM.

Å ikke skille mellom statistisk sannsynlighetsberegning og spredningsparametere kan ha katastrofale følger. Ofte ser vi målefeil underestimert ved at man har benyttet konfidensintervall eller SEM som et spredningsparameter. Det samme ser vi ved vurdering av behandlingseffekter hvor pasienter med målt respons nedenfor et konfidensintervall oppfattes som ikke å ha effekt av behandlingen. Dette er et gjentakende problem ved vurdering av individuell effekt av osteoporosebehandling. Analogt ser vi ofte den feiloppfatning at når målefeil er beskrevet med standardavvik, oppfattes avvik større enn 1 SD som sanne eller biologiske variasjoner, altså utenfor den angitte feilmargin. I sin ytterste konsekvens fører en overfladisk anvendelse av matematikk helt unødvendig til deprimerede pasienter som tror tingene bare forverrer seg. Konfidensintervall og p-verdier er som smør på flesk. Det bør anbefales kun å anvende p-verdien der det er legitimt. Da unngår man den type sammenblanding som er beskrevet over.

Som mål på effekt anfører Aamodt og medarbeidere at resultater fra regresjonsanalyser skal presenteres med regresjonskoeffisienten eller med en ratio (1). Regresjonskoeffisienten er ofte ubegripelig, fordi den ikke viser standardiserte eller direkte sammen-

liknbare helningsvinkler på regresjonslinjen, men kun tilnærmet følgende ratio:

Spredning av effektmål
Antall hele intervaller som forklaringsvariablene inndeles i

Røyking har vanligvis ett intervall (ja/nei). Følgelig kan koeffisienten bli stor, selv ved lav effekt. Voksen alder vil kunne inndeles i fem intervaller (30–80 år) hvis man deler inn alder i tiår, eller 50 intervaller hvis man anvender år. Dette vil gi 10 ganger forskjell i regresjonskoeffisienten. At man ikke har tatt hensyn til ulikt antall intervaller ved risikofaktorer for brudd, forklarer hvorfor de kliniske faktorene for brudd gis ubegrunnet høy betydning. Hvorfor dropper Tidsskriftet  $r^2$ ?

Absolutt endring og ikke relativ endring anbefales av Aamodt og medarbeidere, fordi absolutt endring er mer tolkbar (1). Dette avviker fra praksis i internasjonale fagtidsskrifter. Absolutt og relativ risiko gir opplysning om to helt forskjellige forhold, og begge er nødvendig til hvert sitt bruk. Her må man beholde vitenskapeligheten og se bort fra helsepolitisk begrunnede synspunkter på slike parametere.

Det finnes ingen «kongelig vei» til matematikk. Snarere enn å lage en kokebokoppskrift som kan legitimere overfladisk bruk av matematikk, bør Tidsskriftet bidra til en forståelse av at statistikk dreier seg om nyttige, men komplekse hjelpemidler. Snarere enn å forenkle bør parolen være det motsatte. Men å forlange prinsippkomponentanalyser og faktoranalyser ved visse problemstillinger, vil neppe bidra til bedret medisinsk praksis. Slike metoder viser ikke noe mer enn enkle og forståelige regnemøter.

## Arne Høiseid

ah@sri.no  
Sentrum røntgeninstitutt  
Stenersg. 1A  
0050 Oslo

*Manuskriptet ble godkjent 24.11. 2005.*

## Litteratur

1. Aamodt G, Gulbrandsen P, Laake P et al. Presentasjon av statistiske analyser i Tidsskriftet. Tidsskr Nor Lægeforen 2005; 125: 2183–7.
2. Smith GCS, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: Systematic review of randomised controlled trials. BMJ 2003; 327: 1459–61.