

---

---

# Statistiske metoder i epidemiologi

---

---

Hva er epidemiologi uten statistikk? Man kommer ikke langt i epidemiologiske studier før behovet melder seg for statistiske metoder, enten av beskrivende eller analytisk art. Et enkelt eksempel er den klassiske 2 x 2 tabellen som kan settes opp når eksponerte (ja, nei) følges opp i forhold til utviklet sykdom (ja, nei). Er de observerte antallene fordelt på en slik måte at det faktisk er flere syke blant de eksponerte eller kan det ekstra antallet syke tilskrives en ren tilfeldighet?

Moderne epidemiologi beskjeftiger seg med risikofaktorer for sykdom og død, og i studier av dette spiller statistiske metoder en betydelig rolle. Utviklingen av datamaskiner de siste tretti år har gjort det mulig å analysere data uansett hvor store de er i form av personer og variable. Til dette benyttes vesentlig statistiske programpakker, egne epidemiologiske programpakker er knapt nok eksisterende. Statistikk har fått en stadig større rolle, og Douglas Altman opplyste nylig på den internasjonale biostatistikerkonferansen at andelen artikler i *New England Journal of Medicine* hvor enten logistisk regresjon eller Cox-analyse ble benyttet, nå er hele 65 prosent.

Et betydelig antall statistikere arbeider med å tilpasse og å videreutvikle statistiske metoder til anvendelser innen medisinsk forskning og epidemiologi. Dette er tilfelle også her til lands, og i dette temanummeret presenterer statistikere og epidemiologer noe av aktiviteten.

Jan F. Bjørnstad presenterer Statistisk sentralbyrås generelle utvalgsplan som bygger på teori for stikkprøver. Planen brukes i ulike studier i SSB, men vil også være nyttig på andre områder innen epidemiologisk overvåking.

To artikler tar for seg nye metoder innen deskriptiv epidemiologi. Romlig epidemiologi tar hensyn til hvordan sykdommer og ytre miljøfaktorer varierer med sted (rom) og tid. Faget er også rettet mot å finne og definere klynger eller opphopninger. Geir Aamodt presenterer noen statistiske modeller innen feltet. Metodene omfatter visualisering ved tematiske kart, statistiske tester for å vurdere økt forekomst i avgrensede områder eller om det er romlige trender i sykdomsforekomst. Resultater fra slike analyser brukes både av epidemiologer og politikere. I geomedisin studeres hvordan naturlige ytre miljøfaktorer påvirker helse. Korrelasjonsanalyser blir ofte brukt for å kaste lys over mulige årsaker til sykdommer med uforklart etiologi. For studier som omfatter store geografiske områder, kan slike korrelasjoner variere fra sted til sted. Bjørn Bølviken beskriver geografisk løpende korrelasjoner

med eksempler fra sammenhengen mellom multippel sklerose og to miljøfaktorer, innendørs forekomst av radon og nedfall av atmosfærisk magnesium.

I epidemiologiske studier følges ofte grupper av individer over tid, og ulike hendelser, slik som første hjerteinfarkt, komplikasjoner og død, registreres. Den statistiske metodikken for analyse av individer som følges over tid der flere hendelser kan inntreffe for hver person, kalles forløpsanalyse. Et spesialtilfelle er overlevelsesanalyse der fokus er på tiden frem til en bestemt hendelse for hver person. En av de mest kjente modellene er Cox-modellen. Odd O. Aalen og medarbeidere gir en kort innføring, oversikt over den historiske bakgrunn og utvikling av feltet i Norge. Noen pågående prosjekter omtales. Ørnulf Borgan og Sven Ove Samuelsen har bidratt til utviklingen av ulike design for trekking av utvalg i Cox-modellen. I dette nummeret presenterer de en oversikt over ulike utvalgsmetoder: "nested case-control" og "case-cohort sampling". Disse metodene er nyttige når det er for kostbart å innhente informasjon om kovariable for alle individer som deltar i studien.

Grethe Albrektsen og Ivar Heuch beskriver et kolinearitetsproblem i studier av sammenhengen mellom kvinners reproduksjonshistorie og kreftrisiko. Problemet er å skille effekten av alder ved fødsel fra effekten av kvinnens oppnådde alder og tid siden barnefødsel. Løsningsforslaget deres er basert på estimering av alderseffekten på grunnlag av kvinner uten barn.

Magne Thoresen ser på effekter av målefeil i ulike regresjonsmodeller aktuelle i epidemiologien. Han viser at effektene av målefeil generelt er vanskelig forutsigbare og kommer inn på metoder for å korrigere for målefeil.

Plenumsforedragene ved den siste norske epidemiologikonferansen viser at moderne epidemiologi har et mer helhetlig perspektiv enn tidligere. Man ønsker svar på spørsmål av typen: hva betyr individuelle risikofaktorer, miljøfaktorer og gener sammen for utvikling og forekomst av sykdom og død? Epidemiologien befatter seg videre med hele livsløpet fra unngangelse til grav. Dette betyr at det vil være nødvendig å trekke inn informasjon på flere nivåer i en analyse. De statistiske analysene må videre ta hensyn til at observasjoner innen de ulike nivåene er korrelerte. Anders Skrondal og Sophia Rabe-Hesketh presenterer de nye metodene i et generelt rammeverk, "generalized linear latent and mixed models" (GLLAMM). Eksempler på anvendelse innen epidemiologi omfatter bl.a. repeterte målinger, modeller for målefeil og fler-

nivåanalyser. GLLAMM har imidlertid et stort potensial ut over dette innen epidemiologien. Alle modellene kan implementeres i programmet gllamm som er skrevet i STATA. Artikkelen vil være en viktig referanseartikkel for metodisk orienterte forskere innen epidemiologi.

Det er utviklet mange spørreskjemaer og instrumenter for måling av mentale lidelser i befolkningen. Anne Johanne Søgaard og medarbeidere sammenligner et sett av spørsmål brukt i de store befolkningsundersøkelsene innen CONOR (COhort NORway), med tidligere validerte instrumenter (HSCL-10 og HADS). De konkluderer med at den nye indeksen, CONOR Mental Health Index, viser god overensstemmelse med de validerte instrumentene og at indeksen er et viktig og valid verktøy i epidemiologisk forskning. Artikkelen er videre et interessant eksempel på metoder for sammenligning av ulike spørreskjemainstrumenter.

Økt risiko for sykdom innen visse familier er et viktig tema innen årsaksforskning. Når det gjelder familieoverført risiko for hjerte-karsykdom, påpeker Tormod Brenn at sofistikerte scores for familiehistorie ikke har vist seg mer prediktive i befolkningsstudier enn det enkle spørsmålet "Har en eller flere av foreldre eller søsken hatt hjerteinfarkt eller angina pectoris?" Det enkle spørsmålet fremstår også som en uavhengig risikofaktor når man har justert for kardiovaskulære risikofaktorer. Brenn presenterer andre metoder for å studere familieoverført sykdom når man har informasjon fra flere typer slektninger, fra familiemedlemmer som bor sammen og adskilt og informasjon om felles miljø. Fra varianskomponentmodeller og stianalyser oppnås estimater for betydningen av arv og miljø.

For å finne årsaker til sykdom og utvikle behandlingsmetoder, er man på jakt etter gener og hvordan disse uttrykker seg. Marit Holden og Anders Løland gir en introduksjon til analyse av cDNA mikromatrise-data. Ved slike analyser kan man studere rollen til tusenvis av gener samtidig og f.eks. få svar på hvilke gener eller grupper av gener som er aktive i ulike typer frisk og sykt vev. Det ligger store statistiske utfordringer i å trekke ut informasjon fra de enorme mengder data som kommer fra mikromatriser. Dette er bare et eksempel på de mange utfordringene man står overfor innen feltet bioinformatikk og genetisk statistikk.

Trygdemedisin og samfunnsmedisin har nære tilknytningspunkter til epidemiologi. Harald Reiso og medarbeidere foreslår en ny metode for å estimere når i forløpet av en sykmeldingsperiode en pasients selvrapporterte arbeidsevne er mest prediktiv for varigheten av gjenværende sykmeldingsperiode.

Er det interessant informasjon gjemt i "støy"? En ny ide i epidemiologi er å benytte fraktalanalyser av tidsserier. Dette er tema hos Jan F. Nygård og Eystein Glattre.

Ulike epidemiologiske studier gir oss ikke bare større innsikt i sykdommers utbredelse og årsaker, men danner også grunnlag for de beslutninger som tas av f.eks. helsepolitikere. Oppgaven til Senter for medisinsk metodevurdering (SMM) ved Sintef Unimed, er å bidra til et godt beslutningsgrunnlag ved å foreta vurderinger basert på systematisk litteraturgjennomgang. Lise Lund Håheim og Berit Mørland beskriver den systematiske metoden. Metoden har gjerne vært anvendt på nye medisinske metoder eller tiltak, men er nyttig på flere felt, f.eks. ved innføring av "Grønn resept" eller ved vurdering av effekt av ulike risikofaktorer på sykkelighet og dødelighet, slik som passiv røyking (red. anm.).

I dette nummeret er det gitt noen eksempler på statistiske metoder innen epidemiologi. Området er stort og mange viktige temaer er ikke berørt. Statistikere har bidratt og bidrar fortsatt vesentlig til utviklingen av epidemiologisk forskning. Vi trenger både teoretisk orienterte statistikere og statistikere som arbeider med konkrete anvendelser i epidemiologi. Ny programvare har bidratt til at avanserte analyser er lettere å gjennomføre i dag enn enklere analyser var for ti år siden. Men det er en utfordring å forstå hvilke analyser man gjør, hvilke forutsetninger og modeller som ligger til grunn, hvilke skjevheter som kan oppstå og tolke resultatene. Vi trenger derfor fortsatt å rekruttere godt utdannede statistikere og andre med god kompetanse i statistisk metodikk til epidemiologisk forskning.

Til slutt vil vi takke forfatterne for mange flotte bidrag. Vår takk går også til referee-ene for meget grundige tilbakemeldinger. Uten deres positive innstilling og gode kommentarer ville det neppe vært mulig å utgi dette temanummeret. Endelig vil vi takke for utmerket samarbeid med redaktør Trond Peder Flaten.

## Gjesteredaktører

### Randi Selmer

Divisjon for epidemiologi  
Nasjonalt folkehelseinstitutt  
randi.selmer@fhi.no

### Tormod Brenn

Institutt for samfunnsmedisin  
Universitetet i Tromsø  
tormod.brenn@ism.uit.no