

# Noen statistiske modeller innen romlig epidemiologi

Geir Aamodt

*Biostatistikk, helsefaglig støtte, 0027 Rikshospitalet*

Telefon: 23071187 Telefaks: 23071178 E-post: geir.aamodt@rikshospitalet.no

## SAMMENDRAG

I denne artikkelen presenteres noen statistiske modeller som kan brukes innen romlig epidemiologi. Modellene er basert på klassiske modeller fra romlig statistikk, men med utvidelser for å ta høyde for epidemiologiske anvendelser. Spesielt ser vi på modeller for småskalafenomener slik som forekomst av sykdom knyttet til et utslippssted og modeller for å beskrive storskalafenomener som å sammenlikne geografiske områder som kommuner. For å beskrive romlig fordeling av sykdom, har romlig epidemiologi vokst seg fram som et selvstendig fag. Faget tilbyr i hovedsak metoder for å lage kart og analysere disse kartene, men det finnes også statistiske tester for å vurdere om forekomst av sykdom er betydelig øket i avgrensede sentra eller om det er romlige trender for ulike sykdommer. Resultatene brukes av både politikere og epidemiologer for å vurdere om man skal gi ressurser til utsatte områder med øket risiko for sykdom, eller om man skal starte mer fokuserte studier.

Aamodt G. **Statistical models in spatial epidemiology.** *Nor J Epidemiol* 2003; 13 (2): 223-228.

## ENGLISH SUMMARY

In this article, statistical models are presented which are used for epidemiological studies where spatial or spatial-temporal components are present. The models are based on classical models from spatial statistics, but with modifications to accommodate epidemiological applications. In particular we will study small scale phenomena such as disease clustering and large scale phenomena such as geographical variation between different areas. Spatial epidemiology has evolved into an independent topic and offers methods both to produce maps of different kinds and to analyse them. Statistical tests have also been developed to test if there are significant increases in incidences in limited areas or if there are spatial trends in incidence rates. The results can be used both by politicians and epidemiologists to decide whether resources should be reallocated or if more focused studies should be initiated.

## INNLEDNING

Epidemiologifaget defineres som læren om sykdommers forekomst og indre og ytre faktorer. Romlig epidemiologi er en naturlig utvidelse av epidemiologifaget, der man også tar hensyn til den romlige komponenten i sykdommens utbredelse og forholdet mellom romlig fordeling av sykdommer og ytre faktorer. Romlig epidemiologi har utviklet seg til et eget fagfelt innen medisinsk statistikk. Faget har fokus rettet mot 1) beskrivelser og karakteriseringer av ulike sykdommers utbredelse i rom og tid, 2) studier av sammenheng mellom utbredelse av sykdommers romlige fordeling og ytre faktorer og 3) finne og definere klynger eller opphopning.

Romlig fordeling av sykdommer visualiseres ofte ved hjelp av tematiske kart. Disse kartene viser forekomsten av sykdom, gjerne som relativ risiko. Kartene framkommer ved å plote forholdet mellom forventet og observert mortalitet (eller morbiditet) for en sykdom for avgrensede områder slik som kommuner. Slike kart over standardisert mortalitetsratio (SMR)

brukes både av politikere, som et verktøy til økonomiske prioriteringer, og av epidemiologer til å vurdere videre studier av sykdom. Det er også av interesse å studere om det har vært opphopninger av sykdommer enten knyttet til spesifikke utslippsområder, eller mer generelt der opphopningen ikke har noen romlig forankring. I medisin som i økologi studeres også sammenhengen mellom forekomst av sykdom og ytre variabler.

Romlig epidemiologi er et metodefag som tilbyr statistiske metoder og modeller for å beregne tematiske kart over SMR og metoder og modeller for å studere opphopning av sykdom. De statistiske metodene og modellene bygger på teori fra romlig statistikk. De modellene vi skal diskutere er ofte kompliserte og lar seg ikke alltid estimere ved hjelp av standard estimeringsmetoder. Moderne estimeringsteknikker basert på simuleringer, brukes derfor for å tilpasse disse statistiske modellene.

I denne artikkelen vil vi først beskrive hva slags datatyper som er vanlig i dette feltet. Videre vil vi presentere statistiske modeller for opphopning av sykdom-

mer i bestemte områder, og modeller for å lage tematiske kart over relativ risiko for sykdommer. Vi vil diskutere hvordan modellene kan utvides til også å favne fordeling av sykdom i rom og tid. Formlene som presenteres i artikkelen er skrevet slik at den matematikkinteresserte leser ser hvordan modellen uttrykkes matematisk.

I de siste årene er det blitt publisert mange artikler i dette fagfeltet. Det har også kommet ut en lærebok (Lawson 2001) og artikkelsamlinger (Halloran et al. 1999, Lawson et al. 1999) om emnet. For teori om romlig statistikk vil vi referere til Cressie (1993).

## ROMLIGE DATA

I romlig statistikk er data stedfestede. Observasjoner er knyttet til lokaliteter, som er gitt ved romlige koordinater – lengdegrad og breddegrad. Slike data kan vise en romlig avhengighet som man må ta hensyn til når data analyseres. Den romlige avhengigheten uttrykkes ved at to observasjoner nær hverandre antas å være høyere korrelerte enn to observasjoner som er lengre fra hverandre.

I romlig epidemiologi skiller man ofte mellom stedfestede og områdefestede data. Stedfestede data eller hendelser (case event data) er registreringer knyttet til en bestemt lokalitet. Lokaliteten skal være spesifisert med entydige koordinater, men kan også være husnummer eller gatenummer. Et eksempel kan være registreringer av lungekreft for ulike hustander i en kommune. Den andre typen data er aggregerte verdier fra enkeltobservasjoner til en verdi som gjelder for et bestemt område. Den romlige komponenten er fortsatt ivarett ved lokaliseringen av området slik som tyngdepunktet til en kommune, et fylke eller et annet geografisk område. Et eksempel er her antall tilfeller av lungekreft pr år for ulike kommuner.

Det er fordeler og ulemper ved begge disse datatypene. Har man informasjon om lokalitet (bosted) for de som har blitt diagnostisert for ulike sykdommer, introduseres skjevhet hvis eksponering er knyttet til arbeidssted eller skole. Flytting vil også vanskeliggjøre en god analyse av årsak til død og lokalitet til dødsattest når eksponering som forårsaket død er annet sted enn der personen dør.

Nivået som data er registrert på, setter også føringer for hvilke slutninger man kan trekke. Tall fra ulike offentlige registre vil være aggregert til kommunenivå og man har derfor ingen mulighet til å studere forekomster av klynger innen kommunen. Verktøy som geografiske informasjonssystemer (GIS), der koordinater til f.eks. veier, kraftledninger og ulike typer bygninger er registrert, muliggjør analyse på detaljert nivå. Et annet problem er fletting av flere datakilder. Ytre variable (miljøvariable) kan være registrert på et for grovt nivå til at man kan bruke disse til å se på finere variasjon.

Romlig epidemiologi favner også statistiske modeller for analyse av utbredelse av sykdommer i rom

og tid. Observasjonene som skal analyseres er derfor både indeksert i rom og tid. Rom og tid data kan både være stedfestede og områdefestede.

## STATISTISKE MODELLER

### *Modeller innen romlig statistikk*

Cressie (1993) skiller i hovedsak mellom tre typer modeller innen romlig statistikk. Den første gruppen favner modeller for romlig fordeling av punkter slik som forekomster av sykdommer i et område (punktprosesser). I tillegg til punktmønstre finnes også modeller for å se på egenskapene til punktene som størrelse og form, i tillegg til deres romlige fordeling (merkede punktprosesser). I den andre gruppen finner vi de klassiske modeller innen geostatistikk som simple Kriging og universal Kriging. Disse modellene brukes for å predikere verdier for en kontinuerlig variabel i en ny lokalitet basert på en mengde observerte verdier av variabelen ved kjente lokaliteter. Den tredje gruppen er modeller for rutedata, der data registreres for hvert pixel i et grid. Slike modeller brukes blant annet innen medisinsk billedbehandling.

For både stedfestede og områdefestede data er det forekomst av hendelser som skal modelleres og man tar derfor utgangspunkt i teori for punktprosesser. Som en spesialmodell finner vi Poissonprosessen. En Poissonprosess kjennetegnes ved uavhengige hendelser i rommet, der sannsynligheten for en hendelse spesifiseres ved en intensitet. Denne intensiteten kan karakteriseres ved hjelp av en romlig intensitetsfunksjon som er avhengig av de romlige koordinatene – lengdegrad og breddegrad. I figur 1 vises til venstre en homogen Poissonprosess der intensiteten ikke er avhengig av romlige koordinater. Til høyre vises en ikke-homogen Poissonprosess der intensiteten er avhengig av de romlige koordinatene. Vi vil nå se på modeller for sykdomsrate eller intensitet for de to datatypene:

### *Beregning av sykdomsrate for stedfestede data*

For stedfestede hendelser antar vi at raten eller intensiteten beskriver et forventet antall forekomster av en sykdom pr flateenhet. Den enkle homogene Poissonprosessen utvides for å ta høyde for problemstillinger innen romlig epidemiologi. Forekomster av sykdom (hendelser) forekommer i en heterogen populasjon både når det gjelder størrelse og sammensetning. Den delen av populasjonen som er under risiko for sykdom vil derfor variere fra sted til sted. Videre kan det forekomme miljøgradienter som fanger opp en romlig trend. Tilslutt vil en hendelse være avhengig av forhold knyttet til lokaliteten og individet, altså av både romlige og ikke-romlige forklaringsvariable.

Vi kan nå sette opp et uttrykk for intensiteten  $\lambda$  ved lokalitet  $x$  som en funksjon av populasjonen, romlige og ikke-romlige forklaringsvariable. De ulike hendelsene i studieområdet benevnes  $i$  slik at lokalitetene til

de ulike hendelsene er  $\{x_i\}$ . Populasjonen under risiko ved lokalitet  $x_i$  er  $g(x_i)$ , mens  $F(x_i)$  er en matrise av forklaringsvariable – både romlige og ikke-romlige. Vektoren  $\beta$  er regresjonskoeffisienter som skal estimeres. Vi får da et uttrykk for intensiteten ved lokalitet  $x_i$ :

$$\lambda(x_i) = g(x_i) m(F(x_i), \beta).$$

Funksjonen  $m$  er en linkfunksjon som knytter risiko-populasjon og forklaringsvariablene sammen. Funksjonen  $m$  tilsvarer en hasardfunksjon slik vi kjenner den fra overlevelsesanalyse. Det finnes flere forslag, slik som en eksponentialfunksjon  $(1 + \exp(F(x) \beta))$  eller en multiplikativ funksjon  $(1 + f(x) \beta)$ .

**Beregning av sykdomsrate for områdefestede data**

Modeller for områdetellinger bygger på den modellen vi har diskutert for stedfestede data. Det forventede antall hendelser innenfor området vil være en funksjon av raten i området  $T$ . Mer generelt skriver vi at det forventede antall hendelser er:

$$E(n) = \int_T \lambda(u) du = \int_T g(u) m(F(u), \beta) du.$$

Fordelingen av antall forekomster innen et område vil igjen følge en Poissonfordeling med forventning og varians som gitt over.

De to modellene for stedfestede og områdefestede data brukes som grunnlag for mer konkrete problemstillinger knyttet til klynger og medisinske temakart.

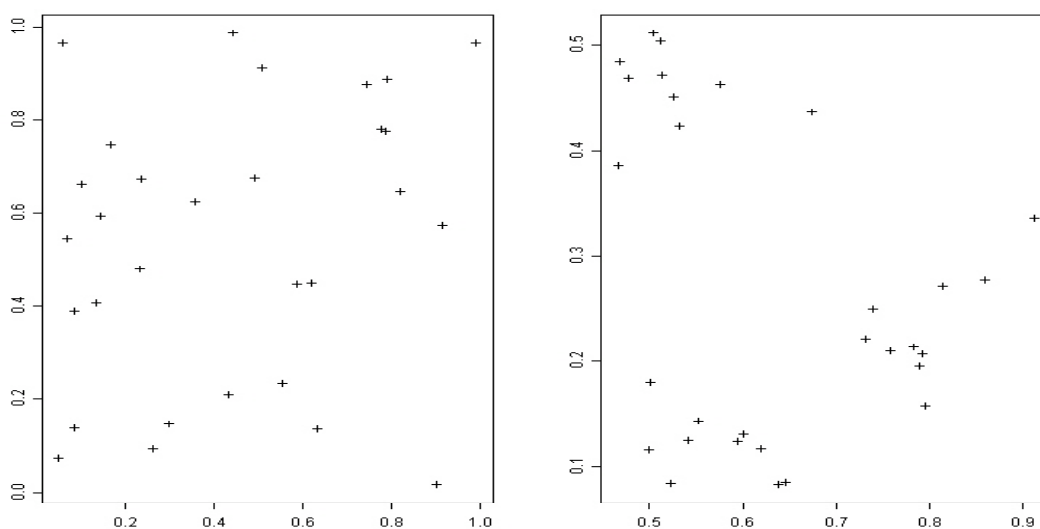
**Medisinske temakart (Disease mapping)**

Bruken av kart for å vise romlig fordeling av forekomst av sykdom er kjent fra lang tid tilbake. I 1854

plottet John Snow gateadressen til ofre for kolera og lokalitet til nærmeste vannkilde på samme kart. Metodene som brukes for å representere forekomst av sykdom har derimot endret seg betydelig de siste årene.

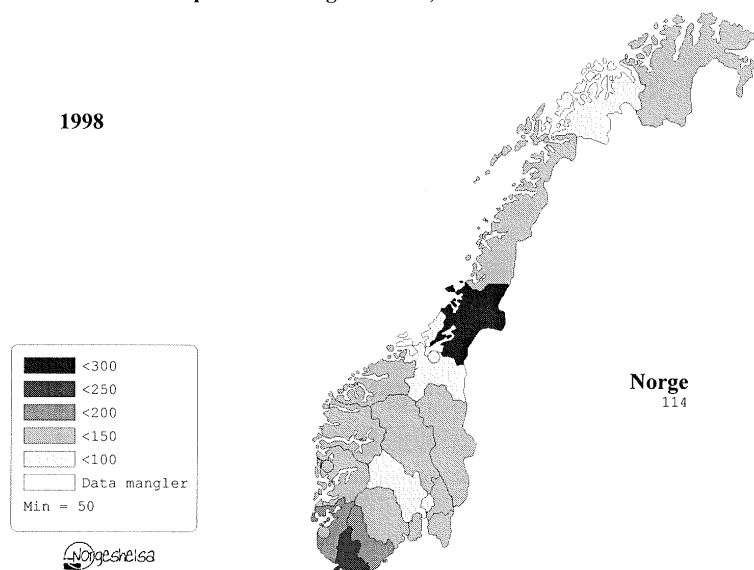
Den enkleste måten å visualisere forekomster av sykdom på er å plote rådata enten som punkter hvis disse er tilgjengelig, eller som antall forekomster for ulike geografiske områder. Folkehelseinstituttet har utviklet programmet Norgeshelsa©. Dette programmet visualiserer en rekke indikatorer om folkehelse på kommunenivå eller fylkesnivå og for ulike tidsperioder ved hjelp av temakart. Programmet lastes ned fra [www.fhi.no](http://www.fhi.no), det er enkelt å installere og bruke. Indikatorene finnes for både befolkningens størrelse, sosiale faktorer, utdanning, et utvalg miljøfaktorer og et utvalg indikatorer om sykkelighet og sykehusopphold. Verdens helseorganisasjon (WHO) har utviklet et tilsvarende program – The HealthMapper – som visualiserer forekomst av et utvalg infeksjonssykdommer i mer enn 60 land. The HealthMapper inneholder også noen demografiske- og miljøvariable.

Det er mer vanlig å visualisere de observerte standardiserte mortalitetsratio eller morbiditetsratio (SMR). Disse kartene tar hensyn til den underliggende populasjonen og gir derfor et bedre bilde av romlig variasjon av relativ risiko. Kartene er basert på en sammenlikning av de observerte verdiene ( $n_i$ ) og de forventede verdier ( $e_i$ ). De forventede verdiene beregnes ved hjelp av den statistiske modellen der man antar ingen endring av risiko i rommet. For stedfestede data plottes estimert intensitet over observert verdi, mens for aggregerte data som f.eks. kommuner, plottes forholdet mellom det observerte antallet  $n$  og det forventede antallet  $e$ .



**Figur 1.** Tilfeldig fordeling eller klynge? Forekomst av sykdom modelleres ofte som punktprosesser. For å studere om forekomster av sykdom opptrer i klynger vil en nullhypotese være at det er ingen samling av forekomster innen studieområdet. I figuren til venstre vises en realisasjon fra en homogen Poissonprosess der forekomster er uavhengig av romkoordinatene. I figurene til høyre vises en realisasjon fra en ikke-homogen Poissonprosess der forekomstene er avhengige av de romlige koordinatene.

Antall barn født pr 1000 vestlige kvinner, 20-34 år



Figur 2. Eksempler på romlig fordeling av helserelevante data.

En naturlig fortsettelse av å kartlegge sykdom er å modellere forholdet mellom romlig fordeling av sykdom og forklaringsvariable. Økologer har utviklet modeller for å studere hvordan sammensetningen av en rekke arter endres som en funksjon av et sett ytre faktorer. Se Biggieri et al. (1999) der metoder fra økologisk analyse drøftes og Bølviken (1998) der fagfeltet geomedisin presenteres med eksempler fra Norge. I tillegg til disse metodene kan klassiske metoder fra geostatistikk (universal kriging) også brukes her.

Flere land har lagt ressurser i produksjon av medisinske temakart. I land som Canada er for eksempel medisinske temakart for romlig fordeling av spedbarnsdødelighet brukt som en indikator for sosioøkonomisk status. Dette kan igjen være grunnlag for ressursfordeling til ulike områder (MacNab et al. 2002).

### Klynger

En klynge eller opphopning definerer vi som et område innen vårt studieområde med betydelig økt risiko. Problemstillingen er aktuell innen offentlig helse for å se på sammenhengen mellom sykdom og forurensningskilder. Diggle et al. (1997) har f.eks. studert opphopning av magekreft rundt et utslippsområde. Vi kan skille mellom ikke-spesifikke og spesifikke klynger. Med ikke-spesifikke klynger mener vi opphopning innen studieområdet uten at vi er interessert i den geografiske beliggenheten til klyngen. For spesifikke klynger derimot er fokus rettet også mot klyngens lokalitet og dens form. Begge typer klynger kan være knyttet til områdefestede eller stedfestede data.

### Risiko knyttet til ikke-spesifikke punktkilder

For å beskrive og videre å teste forekomsten av klynger bygger man på modellene for intensitet som er

beskrevet over. Vi benevner nå koordinatene til totalt  $k$  klynger for  $y_j$ . En enkel modell for intensiteten blir nå:

$$\lambda(x_i) = g(x_i) m \left( \prod_{j=1}^k h(x_i \mid y_j) \right).$$

Her vil funksjonen  $h$  fange opp fordeling av forekomster knyttet til klyngene. Videre er funksjonen  $g(x)$  bakgrunnsraten (populasjonen under risiko) og funksjonen  $m$  er en linkfunksjon. I figur 1 til høyre vises en realisasjon av 30 punkter (hendelser) knyttet til tre klynger. I dette eksempelet er funksjonen  $h$  Gaussisk. Sentra til klyngene er ikke vist.

Klynger kan ha bestemte former som vil være karakteristiske for type utslipp eller forurensningskilder. Funksjonen  $h$  kan derfor spesifiseres mer detaljert for å ta høyde for ulike former for opphopning. For å modellere intensiteten av sykdom knyttet til større veier er man interessert i å modellere klynger knyttet til en linje og ikke et punkt. Lawson et al. (2000) har utviklet en modell for punkter og linjer:

$$\lambda(x_i \mid y, l) = g(x_i) \left[ \prod_j h_p(x_i \mid y_j) + \prod_k h_l(x_i \mid l_k) \right]$$

Her vil funksjonene  $h_p$  og  $h_l$  fange opp intensiteten knyttet til sentra  $\{y_j\}$  og linjer  $\{l_k\}$ .

Det er også vanlig å studere klynger ved hjelp av K-funksjoner. K-funksjonene representerer et mål på variasjon i en Poissonprosess fordi den viser fordeling av avstand mellom nabopunkter. Hvis punktene er spredt tilfeldig vil K-funksjonen følge en rett linje, mens for punkter med klyngeform vil K-funksjonen vise mer variasjon avhengig av avstanden til nabopunktene i klyngen. Chetwynd et al. (2001) studerer klynger for matchede case-control design, der metodene blir illustrert ved diabetes hos barn i Yorkshire, England. K-funksjoner er diskutert i Cressie (1993).

Kulldorff (1995) har utviklet et program StatScan™, som beregner forekomst av klynger i et studieområde. Programmet ”skanner” studieområde ved å systematisk endre midtpunkt til klynger og radius til klyngene. For å unngå problemet med multiple hypotesetester bestemmes antall klynger a priori. Programmet er blant annet brukt for å studere klynger av leukemi blant barn i Sverige (Hjalmer et al. 1996) og amyotrofisk lateralsklerose i Finland (Sabel et al. 2003).

En viktig oppgave er å finne tester for å vurdere om økningen i risiko er betydelig eller ikke innenfor ett eller flere områder  $\{y_{jj}\}$ . Det er utviklet tester for både stedfestede og områdespesifikke data og også for spesifikke eller ikke-spesifikke klynger. For alle testene er nullhypotesen en antagelse om ingen økning av risiko innen de aktuelle sentrene  $\{y_{jj}\}$ . Anderson og Titterington (1997) foreslår en test der man integrerer opp kvadratavviket mellom observert intensitet  $\hat{\lambda}$  og forventet intensitet  $g$  under en nullhypotese om ingen opphoping:

$$T = \int \{\hat{\lambda}(u) - g(u)\}^2 du.$$

Det må også nevnes at for en punktprosess, som er vårt utgangspunkt, vil det alltid opptre opphopninger uten noen spesielle ”kilder” eller ”årsaker”.

### Risiko knyttet til spesifikke punktkilder

En viktig og ofte anvendt problemstilling er å beregne risiko for sykdom knyttet til en bestemt forurensningskilde (spesifikk klynge). For å studere slike problemstillinger knytter det seg ofte metodiske problemstillinger til design. Det anvendes som oftest retrospektive studier og et hovedproblem er kvaliteten på historiske registreringer av utslipp. Videre må nivået på registreringer av utslipp og sykdom (oppløsning) være like noe som også ikke alltid vil være tilfelle. Offentlige registre kan være aggregert til et nivå på grunnlag av konfidensialitet, mens utslipp er registrert ved flere målepunkter innen de aggregerte områdene. For flere sykdommer der sykdomsutbruddet følger år eller tiår etter eksponeringen, vil flytting vanskeliggjøre studien av effekt.

I motsetning til ikke-spesifikke klynger er vi her interessert i å karakterisere formen på klyngen. Intensitetsfunksjonen  $\lambda$  modelleres slik at vi kan ta høyde for ikke bare romlige forklaringsvariable, men også geometriske mål som karakteriserer formen til opphopning av sykdom. Dette kan være radius, avstand eller sektor.

### Rom og tid modeller

Vi har til nå sett på romlige problemstillinger, men tidsdimensjonen er også viktig for å forstå og forklare utbredelser av sykdom. For både medisinske temakart og studier av klynger er denne problemstillingen aktuell. Rom og tid modeller er viktige ikke bare for å beskrive et fenomen, men også for å framskrive nye

hendelser i rom og tid. Dette kan stille strenge krav til modellene og til data som modellene estimeres på bakgrunn av. Modellene må kunne fange opp ofte raske endringer i rom og i tid. Dette kan gå på bekostning av f.eks. tolkbarhet til parametre i modellen.

Flere epidemiologiske studier av sykdom i rom og tid er rent beskrivende og inkluderer ikke statistiske romlige modeller. I f.eks. Scotet et al. (2002) beskrives romlig fordeling av cystisk fibrose i Britannia, Frankrike i løpet av en 40 års periode.

For stedfestede hendelser vil raten kunne formuleres for rom-tid data:

$$\lambda(x, t) = g(x, t) m_x(F_x) m_t(F_t) m_{xt}(F_{xt}).$$

De ulike leddene i likningen over inneholder elementer for henholdsvis rom, tid og interaksjon mellom rom og tid. Ofte spesifiseres denne raten på logaritmisk form for å få fram effektene som additive. Det finnes en rekke spesifiseringer av denne modellen.

Bernardinelli et al. (1995) introduserte modellen  $\log(\lambda_{ij}) = \lambda + \lambda_i + \lambda_j + \lambda_{ij}$ . Her vil risiko  $\lambda_{ij}$  splittes opp i et konstantledd ( $\lambda$ ), et romlig ledd ( $\lambda_i$ ), et temporalt ledd ( $\lambda_j$ ) og interaksjonsledd ( $\lambda_{ij}$ ). I MacNab og Dean (2002) tar man ikke hensyn til rom og tid interaksjon. Her tilpasses en glatt tidskomponent som skal fange opp endring over tid. I denne modellen ligger derfor en antagelse om at det ikke finner sted noen endring av den romlige fordelingen over tid, bare en endring i total antall forekomster over tid. I Aamodt (1996) benyttes redusert rang regresjon for å fange opp romlige endringer over tid. En delmengde av produkter av funksjoner i rom og tid adderes til å gi en så god tilpasning som mulig ( $\lambda_r, h_r, k_r$ ).

Det er også utviklet modeller for klynger i rom og tid og faktisk er noen av de eldste arbeidene innen romlig statistikk knyttet til denne problemstillingen (Mantel 1967). Videre nytter Diggle et al. (1995) K-funksjoner for å modellere økning i risiko i rom og tid.

### Estimering

For å estimere parametre og deres standardfeil i mange av modellene som er diskutert, kan man ofte bruke vanlig sannsynlighetsmaksimering (ML). Som et eksempel kan vi se på estimering av relativ risiko i SMR. Denne parameteren estimeres basert på en Poissonmodell. Mange av modellene er derimot kjennetegnet ved høy grad av kompleksitet og det finnes ikke tilsvarende teori for ML estimering. Det kan derfor være en fordel å bruke metoder basert på simuleringer slik som Markov Chain Monte Carlo metoder (MCMC). Disse metodene kan anvendes innen en klassisk frekventistisk ramme (empirisk Bayes) eller en fullskala Bayesianisk ramme. Disse estimeringsmetodene er for øvrig standard innen romlig statistikk. Bayesianiske metoder får spesiell betydning når tilfeldige effekter (random effects) inkluderes i modellen. I Lawson (2001) diskuteres ulike former for tilfeldige effekter, mens i Militino et al. (2001) drøftes ulike me-

toder for å estimere modeller med tilfeldige effekter. Det finnes ikke enkle og brukervennlige programvare for å tilpasse de modellene vi har diskutert.

## KONKLUSJON

Vi har i denne artikkelen presentert noen smakebiter av statistiske modeller innen romlig epidemiologi. Fagfeltet er fortsatt ungt og det er mye ugjort innen utvikling av modeller for medisinske temakart, klynger og økologiske analyser.

De metodene som vi har sett på er alle grove, men kan være et viktig beslutningsverktøy for videre analyser. I tillegg til problemstillinger fra forurensing, som ofte kommer i offentlig debatt, etableres i dag

biobanker som vil være viktige for å forstå fordeling av sykdom og årsaker til sykdom. Det vil være mange romlige problemstillinger her, og det finnes metoder for å studere disse.

Det er lite trolig at man vil ende opp med enkle statistiske modeller som f.eks. logistisk regresjon er i tradisjonell epidemiologi. Det vil derfor være behov for tverrfaglig kompetanse der metodefolk og medisinske fagfolk arbeider sammen. I Norge har vi, takket være petroleumsvirksomheten, et sterkt miljø innen romlig statistikk.

Takk til mine kollegaer Kathrine Frey Frøslie og Thore Egeland for kommentarer og hjelp underveis og en anonym referee for utfyllende og verdifulle kommentarer.

## REFERANSER

- Aamodt G. Trend-analysis and reduced rank regression for spatio-temporal data. Dr.polit. avhandling, UiO, 1996.
- Anderson NH, Titterton DM. Some methods for investigating spatial clustering, with epidemiological applications. *Journal of the Royal Statistical Society Series A* 1997; **160**: 87-105.
- Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine* 1995; **14**: 2433-2443.
- Biggeri A, Divino F, Frigessi A, Lawson AB, Böhningen D, Lesaffre E, Viel J-F. Introduction to spatial models in ecological analysis. In: Lawson AB, Böhningen D, Lesaffre E, Briggeri A, Viel J-F, Bertollini R (eds.). *Disease Mapping and Risk Assessment for Public Health*. Wiley, New York, 1999.
- Bølviken B. Geomedisin. *Norsk Epidemiologi* 1998; **8** (1): 7-17.
- Chetwynd AG, Diggle PJ, Marshall A, Parslow R. Investigation of spatial clustering from individually matched case-control studies. *Biostatistics* 2001; **2**: 277-293.
- Cressie N. *Statistics for spatial data*. Wiley, New York, 1993.
- Diggle PJ, Chetwynd AG, Haggkvist R, Morris SE. Second-order analysis of space-time clustering. *Statistical Methods in Medical Research*. 1995; **4**: 124-136.
- Diggle PJ, Morris S, Elliot P, Shaddick G. Regression modelling of disease risk in relation to point sources. *Journal of the Royal Statistical Society Series A* 1997; **160**: 491-505.
- Halloran ME, Berry D (eds.). *Statistic and Epidemiology. Environment and Clinical Trials*. Springer, New York, 1999.
- HealthMapper, World Health Organisation.
- Hjalmer U, Kulldorff M, Gustafsson G. Childhood leukemia in Sweden: using GIS and spatial scan-statistics for cluster detection. *Statistics in Medicine* 1996; **15**: 707-715.
- Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Statistics in Medicine* 1995; **14**: 799-810.
- Lawson AB, Böhningen D, Lesaffre E, Briggeri A, Viel J-F, Bertollini R (eds.). *Disease Mapping and Risk Assessment for Public Health*. 1999, Wiley, New York, 1999.
- Lawson AB, Williams F. Spatial competing risk in disease mapping. *Statistics in Medicine* 2000; **19**: 2451-2468.
- Lawson AB. *Statistical Methods in Spatial Epidemiology*. Wiley, New York, 2001.
- MacNab YC, Dean CB. Spatio-temporal modelling of rates for the construction of disease maps. *Statistics in Medicine* 2002; **21**: 347-358.
- Mantel N. The detection of disease clustering and a generalised regression approach. *Cancer Research* 1967; **27**: 283-294.
- Militino AF, Ugarte MD, Dean CB. The use of mixture models for identifying high risks in disease mapping. *Statistics in Medicine* 2001; **20**: 2035-2049.
- Norgeshelsa ©, Folkehelseinstituttet, Nasjonalt Helseindikatorsystem.
- Sabel CE, Boyle PJ, Loytonen M, Catrell AC, Jokelainen M, Flowerdew R, Maasliita P. Spatial clustering of amyotrophic lateral sclerosis in Finland at place of birth and place of death. *American Journal of Epidemiology* 2003; **157** (10): 898-905.
- Scotet V, Gillet D, Dugueperoux I, Audrezet MP, Bellis G, Garnier B, Roussey M, Rault G, Parent P, de Braekeleer M, Ferec C. Spatial and temporal distribution of cystic fibrosis and its mutations in Brittany, France: a retrospective study from 1960. *Human Genetics* 2002; **111**: 247-254.