

# ”PAPPERSPATIENTER” TESTAR LÄKARENS DIAGNOSTIK

## Möjligheter och problem vid klinisk bedömningsanalys

**Kliniska situationer är ofta mycket komplexa och svåra att överblicka, och de bedömningar vi gör som läkare i vår medicinska vardag får konsekvenser i fråga om både sjukdom, hälsa och ekonomi. Våra bedömningar baserar sig på sannolikheter, vilket gör att de blir osäkra.**

**Det är därför viktigt, men också svårt, att studera de bedömningar som görs inom sjukvården. Klinisk bedömningsanalys är en metod som har använts för att studera läkares bedömningar och beslut inom olika områden.**

Medicinen är inte bara ett kunskapsområde, utan också en handlingsinriktad verksamhet, där vi styrs av föreställningar om vad vi vill uppnå. Det kliniska beslutet knyter samman vår biomedicinska kunskap med vårt praktiska handlande och är därför centralt inom medicinsk praktik.

Beslutsteori är ett tvärvetenskapligt ämne som har vuxit fram under de senaste decennierna, och medicinen har varit ett av de tillämpningsområden som man intresserat sig för.

Man brukar skilja på två huvudriktningar inom beslutsteori – en normativ, som handlar om hur vi bör fatta beslut, och en deskriptiv, som handlar om hur vi faktiskt gör när vi fattar beslut. Klinisk bedömningsanalys (Clini-

cal judgment analysis, CJA) är en deskriptiv metod.

Metoden innebär i korthet att en läkare bedömer en serie patientfall, där klinisk information av olika slag (kliniska variabler, s k cues) varierats. Med ledning av den givna informationen gör läkaren en bedömning av varje fall, genom att ta ställning till diagnos eller behandling. Utifrån denna serie av bedömningar kan man sedan med hjälp av multivariata statistiska metoder studera läkarnas bedömningsstrategi.

### »Papperspatienter»

De stiliserade fallbeskrivningarna presenteras ofta i form av en uppsättning »papperspatienter» – varje patient presenteras med hjälp av en kort inledning som är gemensam för alla patienterna, följt av ett värde för var och en av de kliniska variabler som man valt att använda (t ex ålder, blodtrycksvärden, röntgenfynd). En av fördelarna med att använda papperspatienter är att man kan låta flera läkare bedöma samma uppsättning patientfall och använda sig av samma medicinska information. En annan fördel är att man har möjlighet att studera och jämföra sammanvägningen av klinisk information fristående från insamlandet av informationen.

En vanlig invändning mot metoden är att patientfallen blir abstrakta och verklighetsfrämmande, och att de bedömningar som görs därför inte motsvarar verkliga bedömningar. Detta har testats i olika studier genom att man använt papperspatienter baserade på verkliga patientfall, och man har då funnit god överensstämmelse mellan hur läkare har bedömt de verkliga patienterna och hur de bedömt papperspatienterna [1-3].

### Linsmodellen

Metoden har rötter i perceptionspsykologin [4, 5] och bygger på tanken att bedömningar och beslut, i likhet med perceptionen, kan uppfattas som ett sätt för organismen att handskas med en omgivning som styrs av sannolikhetslagar. Teorin bakom klinisk bedömningsanalys brukar sammanfattas i linsmodellen (Figur 1).

Med sjukdom och diagnos som ex-

empel kan linsmodellen kortfattat beskrivas: Ett antal kliniska variabler (cues) förhåller sig med en viss grad av sannolikhet till en sjukdom (S) och har därvid olika stor betydelse eller viktning ( $W_1-W_n$ ). För den som gör en bedömning har de olika variablerna också olika stor betydelse, eller vikt, för diagnosen (D). Dessa »vikter» ( $W_1-W_n$ ) utgör tillsammans bedömarens strategi. Om ( $W_1-W_n$ ) har en fördelning som starkt avviker från ( $W_1-W_n$ ), så kan det leda till bedömningar som skiljer sig mycket från det verkliga förhållandet, dvs (D) skiljer sig kraftigt från (S). Då talar man om dålig kalibrering – bedömarens strategi är inte anpassad till de uppgifter hon skall lösa.

Dålig kalibrering kan bl a uppstå när man befinner sig i en omgivning som skiljer sig från den man har erfarenhet av – ett medicinskt exempel är svårigheten att göra korrekta bedömningar i en miljö där sjukdomar och symtom har andra prevalenser än dem man är van vid.

### Exempel från litteraturen

En serie studier handlar om reumatologers bedömningar av graden av sjukdomsaktivitet hos patienter med reumatoid artrit [3, 6-8]. I en av delstudierna fick två reumatologer bedöma sjukdomsaktiviteten på en visuell analog skala (VAS-skala), först på en serie riktiga patienter och några veckor senare på papperspatienter utformade efter dessa patienter [6]. Med hjälp av regressionsanalys tog man fram deras bedömningspolicy och fann att för olika läkare var olika variabler avgörande när de gjorde sina bedömningar. När man tittade på läkarnas egen uppfattning om vilka variabler de tog hänsyn till, så skilde sig däremot inte deras uppfattningar nämnvärt.

Genom att på det här sättet kunna synliggöra skillnader i de kliniska bedömningarna, anser man att man har större möjligheter att åstadkomma en gemensam policy, inom en mottagning eller mellan olika mottagningar.

Ett annat exempel är en studie där man studerade läkares beslut att förskriva eller inte förskriva hormonsubstitution till kvinnor i klimakteriet [9]. De

### Författare

YLVA SKÅNÉR  
allmänläkare, Hälsocentralen Södra station, Stockholm

LARS-ERIK STRENDER  
docent, överläkare, Landstingets enhet för allmänmedicin, Huddinge

JOHAN BRING  
fil dr, statistiker, Regionalt onkologiskt centrum, Uppsala.

kliniska variabler och nivåer man använde i den här studien var:

- risk att utveckla endometrie-cancer (standard, måttlig, hög),
- vasomotoriska symtom (standard-symtom, kraftiga symtom),
- risk för osteoporos med frakturer (standard, hög).

Dessa möjligheter varierades systematiskt och alla de tolv möjliga kombinationerna användes i studien, vilket underlättade analysen. Men det innebar samtidigt att en tredjedel av papperspatienterna kom att utgöras av kvinnor med hög risk att utveckla cancer. Därmed kom dessa att bli överrepresenterade i studien jämfört med den patientpopulation som läkarna möter i verkligheten. Som norm, »gyllene standard», för hormonförskrivning hade man räknat fram en »ideal» beslutsstrategi, och i jämförelse med denna ideala strategi hade läkarna i undersökningen en mycket för låg förskrivning av östrogen.

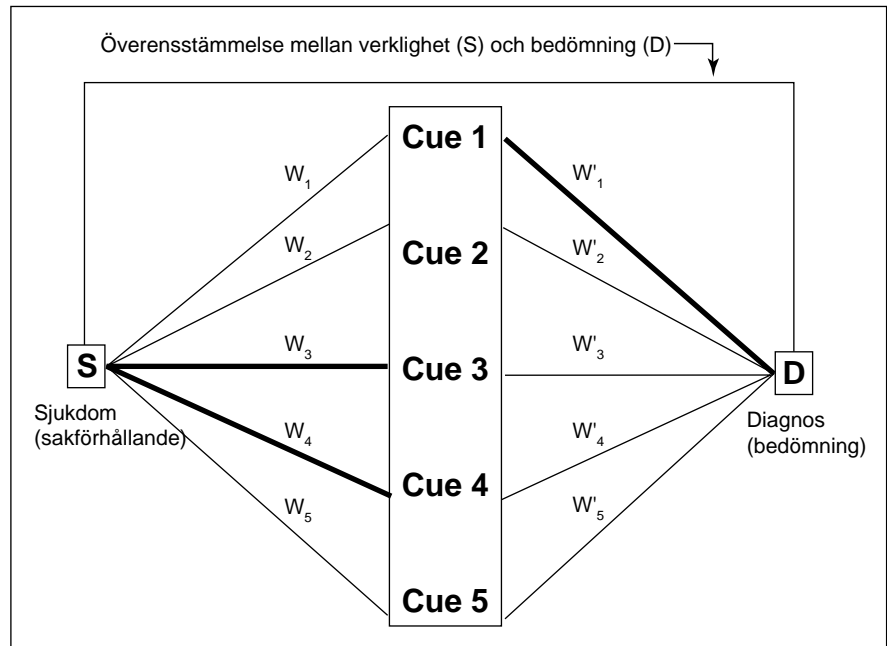
Klinisk bedömningsanalys har också använts i samband med kvalitetsutvecklingsarbete, bl a har man använt den som hjälpmedel för att hjälpa allmänläkare att studera hur de förskriver antibiotika vid mediaotit hos barn [2, 10]. Metoden ger möjlighet att se om t ex skillnader i förskrivningsmönster kan förklaras av att man gör olika bedömningar av den kliniska informationen. I Sverige och i några andra europeiska länder pågår för närvarande ett projekt som handlar om rationell läkemedelsanvändning vid astma och urinvägsinfektioner, där klinisk bedömningsanalys med återkoppling används som pedagogisk interventionsmetod. (Ansvariga för det svenska projektet är IHCAR, avdelningen för internationell hälso- och sjukvårdsforskning vid Karolinska institutet.)

### Vilken klinisk information skall presenteras?

Om man frågar läkare vilken information de använder sig av när de gör en viss medicinsk bedömning så brukar svaret ofta bli en lång lista med olika variabler. När man undersöker detta i en studie så ser man dock ofta att ett fåtal av variablerna räcker för att förklara bedömningarna, men vilka dessa är varierar ofta mellan de olika läkarna.

Att välja ut den information som skall användas vid konstruktionen av papperspatienterna innebär alltså en balansgång mellan att försöka skapa en realistisk situation – som innebär att man erbjuder en stor mängd överskottsinformation – och att ställa rimliga krav på läkarna när det gäller det antal papperspatienter som de skall bedöma.

För att skapa papperspatienter måste man alltså på något sätt göra ett urval av



**Figur 1.** En sjukdom uttrycker sig i olika kliniska variabler (cue 1–5), var och en med olika grad av sannolikhet (vikterna  $W_1$ – $W_5$ ). En bedömare viktar i sin tur dessa variabler, och dessa vikter (vikterna  $W'_1$ – $W'_5$ ) avspeglar bedömarens diagnostiska strategi. I det här exemplet har cue 3 och cue 4 tydliggen ett starkt samband med sjukdomen (S) medan bedömaren har valt att låta cue 1 vara den kliniska variabel som betyder mest för diagnosen (D).

klinisk information, och man kan då gå till väga på olika sätt.

Man kan fråga de läkare som skall ingå i studien vilken information de tror att de använder och sedan konstruera papperspatienterna utifrån detta. Man kan också göra journalstudier eller prospektiva studier för att ta reda på vilken information som används av bedömarna.

Man kan ha speciella skäl att vilja testa vissa variabler. Om exempelvis ett kliniskt syndrom definieras genom vissa kriterier kan man välja att använda just dessa kriterier i sin studie.

### Variabelnivåer

Presentationen av variabelvärdenas olika nivåer kan göras på olika sätt. Variablerna kan varieras kontinuerligt (t ex ålder eller blodtrycksvärden), enligt någon form av ordinalskala (inga/lätta/måttliga/svåra förändringar) eller dikotomt (patienten har eller har inte en viss variabel). Man kan välja om man vill ge siffervärden (hjärtfrekvens 84), beskrivningar i ord (lätta ankelödem) eller om man vill presentera variabelvärdena i form av histogram, där stapelhöjden antas representera ett symtoms svårighetsgrad [11].

### Representativ eller ortogonal studie?

Den uppsättning patienter som används i en studie kan ses som ett urval av patientfall taget från en tänkt population.

Om man konstruerar sina papperspatienter utifrån verkliga patienter så talar man om en representativ uppläggning. Fördelen med detta är att man får en patientuppsättning som är representativ för de patienter läkarna möter i sin var-

dag. Ett problem med det här förfaringsättet kan vara att de flesta papperspatienterna kommer att vara ganska lika varandra, »genomsnittsfall», vilket gör den statistiska analysen svagare. Interkorrelationen mellan variablerna kan komma att vara hög, vilket i så fall försvårar analysen.

Om man istället varierar variabelvärdena – slumpmässigt eller med hjälp av något systematiskt förfarande – så att alla nivåer har samma möjlighet att bli representerade, så brukar man tala om en ortogonal utformning. Man får då en maximal spridning av kombinationer och ingen eller låg interkorrelation mellan variablerna, vilket är till fördel vid den statistiska analysen.

Problemet blir då att papperspatienterna kan skilja sig från dem som läkarna möter i sin kliniska vardag. Dels kan extremfallen få en mer dominerande roll, som i östrogensubstitutions-exemplet ovan, dels kan interkorrelationsmönstret bli orealistiskt. Om alla verkliga patienter som har en hög nivå av en viss variabel samtidigt har en hög nivå av en annan variabel, men denna koppling saknas hos papperspatienterna, så kan fallen te sig orealistiska och validiteten av bedömningarna bli tveksam.

För att kunna dra slutsatser när det gäller reproducerbarheten av de enskil-

da läkarnas bedömningar, använder man ofta dubblettfall. Om läkarna har en bedömningsstrategi som de tillämpar på ett konsekvent sätt, så bör identiska fallbeskrivningar ge upphov till identiska bedömningar. Eftersom det totala antalet papperspatienter måste begränsas sker användningen av dubletter alltid på bekostnad av nya och annorlunda fall. Avvägningen mellan antalet dubletter och nya fall är beroende av vilken frågeställning man har i sin studie.

### Vilka bedömningar vill vi studera?

För varje papperspatient som presenteras ber man läkaren göra en eller flera bedömningar. Det kan handla om att ställa en diagnos, att bedöma svårighetsgraden av en sjukdom, att besluta sig för om patienten skall behandlas, att välja ett läkemedel eller att välja kompletterande utredning.

Frågor som handlar om att man har en åsikt om något – diagnos, depressionsgrad eller liknande – kan lämpa sig för att besvaras med hjälp av en visuell analog skala (VAS).

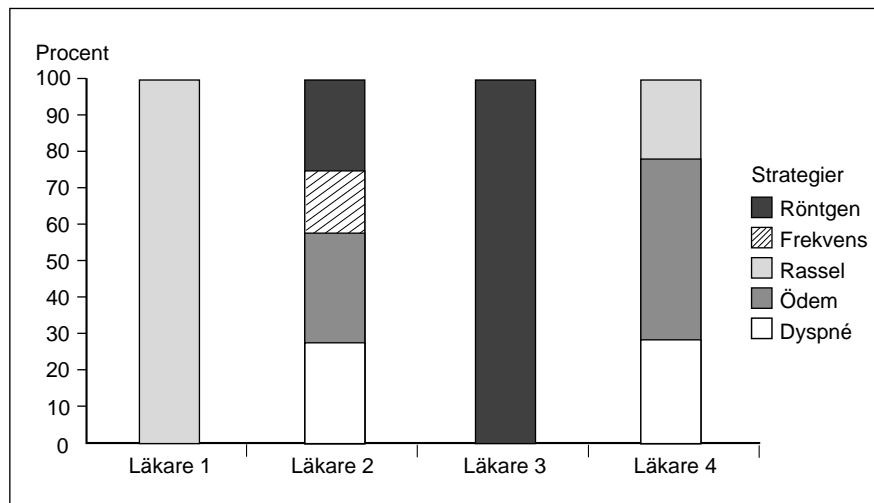
När det gäller frågor som handlar om att man skall göra någonting – t ex välja behandling – är det mer realistiskt att avkräva läkaren ett bestämt ställningstagande genom ett ja- eller nej-svar. Något som är svårt att fånga upp med klinisk bedömningsanalys är det successiva beslutsfattande som ofta används, särskilt när det gäller kroniska sjukdomar – hela spektrum av exspektans, provbehandling, kompletterande utredning är svårt att fånga.

### Analys och resultat

Resultatet analyseras med hjälp av olika former av multipel regressionsanalys. De oberoende eller förklarande (x-)variablerna utgörs av den kliniska information som presenterats (cues), och de beroende (y-)variablerna utgörs av läkarnas bedömningar. Regressionsmodellen används för att beskriva sambanden mellan förändringar i den kliniska informationen och förändringar i läkarnas bedömningar – den gör inte anspråk på att vara en modell för hur det går till när läkarna fattar beslut.

De individuella strategierna kan skilja sig mycket från varandra (Figur 2). Att påvisa dessa skillnader och återföra den kunskapen till de deltagande läkarna för reflektion är ofta en av huvudpoängerna i studier där man använder sig av klinisk bedömningsanalys.

Men man kan också göra jämförelser mellan själva bedömningarna. Kommer enskilda läkare eller grupper av läkare t ex fram till samma bedömningar? De papperspatienter som gett upphov till störst bedömningsskillnader kan vara



**Figur 2.** Strategier vid hjärtsviktsdiagnostik för några av läkarna i en pilotstudie vi genomfört. Dr 1 och dr 3 låter förekomst av auskultatoriska rassel respektive röntgenfynd vara avgörande för diagnosen, medan dr 2 väger samman ödem, dyspné, röntgensvar och hjärtfrekvens och dr 4 ödem, dyspné och rassel.

de som det är mest givande att lyfta fram och diskutera i samband med kvalitetsutvecklingsarbete.

Att läkarna kommer fram till samma bedömningar kan bero på att de har en likartad policy, men det kan också bero på slumpfaktorer, dvs de bedömningar de gör kanske bara till en mindre del kan förklaras av en konsekvent bedömningspolicy. Men ibland kan de komma fram till samma bedömningar med hjälp av olika strategier [12]. Vissa kliniska variabler kanske i verkligheten nästan alltid samvarierar, och vilken av dem som betyder mest för bedömningarna spelar då i praktiken mindre roll. Interkorrelationen mellan olika slag av klinisk information i den patientuppsättning man presenterar är därför betydelsefull. Det här är ett viktigt argument för att använda sig av en representativ utformning – har man använt sig av ett autentiskt patientmaterial så vet man att interkorrelationsstrukturen motsvarar den som läkarna har erfarenhet av.

I vilken utsträckning ställer läkarna i studien korrekta diagnoser?

För att besvara den frågan krävs det att vi kan jämföra bedömningarna av papperspatienterna med uppgifter om »hur det är i verkligheten» – alltså med det vi ser på linsmodellens vänstra sida (Figur 1). Detta förutsätter att vi har en »gyllene standard» för vad som är en korrekt diagnos, vilket ju inte alltid är fallet. Istället kan man göra jämförelser med expertbedömningar (som dock sinesemellan brukar uppvisa stor variation).

Intressantare är det kanske att studera behandlingsstrategier, särskilt i den mån man kan göra jämförelser med behandlingsrekommendationer av något slag – som lokala vårdprogram, SBU-rekommendationer eller principer som tagits fram av nationella konsensusgrupper.

Hjärtsvikt är ett kliniskt syndrom där de diagnostiska kriterierna varierar, inte bara i klinisk praxis utan också mellan

olika vetenskapliga studier [13]. Ett förslag till kriterier är de s k Bostonkriterierna, som använts i flera studier, men vars värde är omdiskuterat [14-17]. Behandlingsrekommendationerna när det gäller hjärtsvikt har ändrats, i och med att ACE-hämmare lyfts fram som basbehandling, men följsamheten till de nya rekommendationerna anses vara ganska låg [16, 18].

De flesta patienter med hjärtsvikt bedöms och tas om hand inom primärvården. Vi har därför påbörjat en studie av allmänläkares hjärtsviktsbedömningar, där vi använder oss av klinisk bedömningsanalys, och planerar en studie där vi skall jämföra allmänläkares och kardiologers bedömningar.

### Referenser

1. Braspenning J, Sergeant J. General practitioners' decision making for mental health problems: Outcomes and ecological validity. *J Clin Epidemiol* 1994; 47: 1365-72.
2. Chaput de Saintonge M, Hattersley LA. Antibiotics for otitis media: Can we help doctors agree? *Fam Pract* 1985; 2: 205-12.
3. Kirwan JR, Chaput de Saintonge M, Joyce CRB, Currey HLF. Clinical judgment in rheumatoid arthritis. I. Rheumatologists' opinions and the development of »paper patients». *Ann Rheum Dis* 1983; 42: 644-7.
4. Leary DE. From act psychology to probabilistic functionalism: The place of Egon Brunswik in the history of psychology. In: Ash MG, Woodward WR, eds. *Psychology in twentieth century thought and society*. Cambridge: Cambridge University Press, 1987.
5. Brehmer B, Joyce CRB, eds. *Human judgment, the SJT view*. Amsterdam: Elsevier Science Publishers BV, North-Holland, 1988.
6. Kirwan JR, Chaput de Saintonge M, Joyce

- CRB, Currey HLF. Clinical judgment in rheumatoid arthritis. II. Judging »current disease activity» in clinical practice. *Ann Rheum Dis* 1983; 42: 648-51.
7. Kirwan JR, Chaput de Saintonge DM, Joyce CRB, Currey HLF. Clinical judgment in rheumatoid arthritis. III. British rheumatologists' judgments of »change in response to therapy». *Ann Rheum Dis* 1984; 43: 686-94.
8. Kirwan JR, Currey HLF. Clinical judgment in rheumatoid arthritis. IV. Rheumatologists' assessments of disease remain stable over long periods. *Ann Rheum Dis* 1984; 43: 695-7.
9. Elstein AS, Holzman GB, Ravitch MM, Metheny WA, Holmes M, Hoppe R et al. Comparison of physicians' decisions regarding estrogen replacement therapy for menopausal women and decisions derived from a decision analytic model. *Am J Med* 1986; 80: 246-58.
10. Chaput de Saintonge DM, Hathaway NR. Antibiotic use in otitis media: patient simulations as an aid to audit. *BMJ* 1981; 283: 883-4.
11. Fish HU, Hammond KR, Joyce CRB, O'Reilly M. An experimental study of the clinical judgment of general physicians in evaluating and prescribing for depression. *Br J Psychiatry* 1981; 138: 100-9.
12. Brehmer B. The psychology of linear judgement models. *Acta Psychol* 1994; 87: 137-54.
13. Marantz PR, Alderman MH, Tobin JN. Diagnostic heterogeneity in clinical trials for congestive heart failure. *Ann Intern Med* 1988; 109: 55-61.
14. Marantz PR, Tobin JN, Wasserteil-Smoller S, Steingart RM, Wexler JP, Budner N. The relationship between left ventricular function and congestive failure diagnosed by clinical criteria. *Circulation* 1988; 77: 607-12.
15. Remes J, Miettinen H, Reunanen AS, Pyörälä K. Validity of clinical diagnosis of heart failure in primary health care. *Eur Heart J* 1991; 12: 315-21.
16. Dahlström U, Boman K, Edvardsson N, Pehrsson K, Persson S. Hjärtsvikt – en svårställd diagnos. Kan poängsystem identifiera patienterna? *Läkartidningen* 1995; 92: 1360-3.
17. Carlson K, Lee DCS, Goroll AH, Leahy M, Johnson RA. An analysis of physicians' reasons for prescribing long-term digitalis therapy in outpatients. *J Chron Dis* 1985; 38(9): 733-9.
18. Osignerad. Behandling av akut och kronisk hjärtsvikt. Information från Läkemedelsverket 1992; 3: 221-39.

# PULSFÄLTGEL-ELEKTROFORES

## Ny metod att spåra bakteriers smittvägar

**Pulsfältgelelektrofores, PFGE, är en ny metod som möjliggör klonal identifiering av bakteriestammar. Den har visat sig vara ett värdefullt epidemiologiskt verktyg vid kartläggning av meticillinresistenta stafylokockers smittvägar. Dessa bakterier utgör ett stort nosokomialt problem i stora delar av västvärlden. PFGE-metoden, som bygger på restriktionsenzymanalys av bakteriernas kromosomala arvs massa, har prövats med framgång i Umeå.**

Antibiotika med förmåga att motstå bakteriernas nedbrytande enzymer (beta-laktamaser) introducerades vid mitten av 1960-talet. Kort därefter påvisades stafylokocker med resistens mot samtliga antibiotika i betalaktamgruppen, t ex penicilliner och cefalosporiner (s k meticillinresistenta stafylokocker, MRSA). Stafylokocker har i varierande grad även blivit resistenta mot andra, icke-betalaktamantibiotika såsom aminoglykosider, tetracykliner, erytromycin, tienamyciner och kinoloner.

Det första större utbrottet av nosokomiala infektioner orsakade av MRSA rapporterades från Barcelona, Spanien 1989 [1]. Därefter har MRSA spritt sig och förekommer numera endemiskt vid många sjukhus i stora delar av Västeuropa och USA, där de utgör ett betydande nosokomialt problem [2-4].

I Sverige har hitintills de flesta MRSA-stammarna isolerats från patienter eller personal som på något sätt haft kontakt med sjukhus eller sjukvård utomlands, där de smittats av dessa bakterier. Vid några tillfällen har smittörföringen till sjukvårdspersonal och medpatienter kunnat påvisas, vilket krävt betydande och kostsamma sjukhushygieniska insatser för att kunna eliminera fortsatt smittspridning.

### Kartläggning av smittvägen

Eftersom kliniska infektioner med dessa multiresistenta bakterier är svårbehandlade försöker man på olika sätt

att identifiera och isolera asymtomiska smittbärare liksom patienter med kliniska infektioner. För att förebygga ytterligare nosokomial spridning kartläggs spridningsvägar både inom och utom sjukhusen.

För detta ändamål krävs möjligheter att typbestämma olika isolat av MRSA och därigenom säkerställa bakteriernas ursprung och eventuella släktskap med andra isolat. Vid typning av MRSA-bakterier har man hittills fått förlita sig på metoder med låg specificitet och diskriminationsförmåga, t ex biokemisk typning, fagtypning, antibiotikaresistensmönster och serologisk typning.

På senare år har metoder som analyserar bakteriernas DNA-innehåll såsom bestämning av plasmidmönster introducerats i detta syfte. Tyvärr har även den senare metoden visat sig ha låg specificitet för att påvisa bakteriestammar med identiskt genetiskt ursprung, s k klonal specificitet. Avsaknaden av typningsmetoder som med stor säkerhet identifierar en klonal bakteriespridning i populationen har försvårat möjligheterna till säkra epidemiologiska utredningar av spridningsmönster för MRSA.

### Ny lovande metod

En ny metod, pulsfältgelelektrofores (PFGE), har visat sig vara en lovande metod för identifiering av bakteriestammar. Med denna metod kan DNA-fragment upp till 50Mb (50 000 000 baspar) separeras. Äldre gelanalytmetoder av restriktionsenzymbehandlad arvs massa

### Författare

TOR MONSEN

ST-läkare, kliniska bakteriologiska laboratoriet

JOHAN WISTRÖM

överläkare, infektionskliniken

ROLF LUNDHOLM

överläkare, sjukhushygieniska avdelningen

STIG HOLM

professor, kliniska bakteriologiska laboratoriet, samtliga vid Norrlands Universitetssjukhus, Umeå.