

Kvalitetsindikatorer i sjukvården 1: Flera metodfrågor obesvarade

I en optimistisk artikel om kvalitetsindikatorer för övervakning av internmedicinska enheter presenteras i Läkartidningen 3/99 ett axplock goda exempel på vad som kan uppnås. Dessa exempel kan emellertid även inspirera en mindre optimistisk statistiker till några kritiska reflexioner och frågor om indikatorernas användbarhet. En metoddebatt om kvalitetsindikatorer i sjukvården förefaller vara högeligen önskvärd.

Det är just nu »inne» att i många olika medicinska sammanhang beräkna statistiska indikatorer som skall spegla kvaliteten i en viss verksamhet. Det föreligger uppenbarligen ett starkt tryck på diverse sjukvårdsmyndigheter att tillhandahålla sådana indikatorer för att olika verksamheter skall kunna bedömas. Indikatorerna antas dessutom fungera som ett system för tidig varning om någonting skulle gå snett. I en trosviss artikel i Läkartidningen 3/99 presenterar Britton, Karlander och Andersson exempel på några kvalitetsindikatorer inom internmedicinen. De hävdar – utan särskilt övertygande bevis – att indikatorerna är känsliga och stimulerar till kvalitetsförbättringar.

Det finns i artikeln inte några djupare funderingar om metodologiska problem förknippade med kvalitetsindikatorerna. Det kan därför vara på sin plats att redovisa några hitills inte besvarade frågor av metodologisk karaktär, inspirerade av de exempel som givits i artikeln.

Man kan ana att filosofin bakom utnyttjandet av indikatorer har inspirerats av industriella tillämpningar, där man

tex har en stor maskin som med bestämd precision skall spotta fram maskin av exemplar av en produkt av en given dimension. Denna situation finns beskriven i åtskilliga statistiska läroböcker.

Den gängse ansatsen är då att betrakta det hela som en statistisk process, där varje enhet har sitt eget värde med avseende på den aktuella dimensionen, och det tas ett slumpmässigt stickprov om enheter tex varje timme för kontroll. De utvalda enheterna mäts och medelvärde markeras successivt på ett sk kvalitetskontrollkort. Om medelvärdet skulle hamna utanför de på förhand beräknade kontrollgränserna stängs maskinen av för närmare genomgång.

Riskerna för felslut i olika riktningar, dvs för att stänga av maskinen i onödan (trots att »processen är under kontroll»), liksom att underlåta att stänga av den när så borde ha gjorts (när »processen inte är under kontroll»), låter sig beräknas. Om det dessutom finns uppgifter om kostnaderna för respektive möjligt felslut är grunden lagd för en rationell strategi. Detta är alltså en situation där man kvalitetskontrollerar en produktion av en mängd exemplar (items) av samma slag.

Den industriella analogin vilsedande i vården

Arbetet vid en speciell klinik eller ett visst sjukhus kan även det beskrivas som en process, vilken kan vara under kontroll eller ej. Men den enkla analogin till kvalitetskontrollkortet fungerar dessvärre inte så bra – en klinik är inte en maskin, och patienterna är inte likartade »items» som skall vara så identiska som möjligt när de passerat »maskinen».

Dessutom tillkommer den aspekten att det inte enbart gäller att studera indikatorernas variation över tiden vid en och samma klinik eller vid ett och samma sjukhus – man vill dessutom jämföra resultaten vid olika kliniker, som under ideala omständigheter skall förväntas hålla exakt samma höga kvalitet på vården. (Jämför situationen när tex MacDonalds vill kontrollera att hamburgarna smakar exakt likadant i alla butiker.)

Det nu så moderna arbetet med kvalitetsindikatorer inom sjukvården aktualiserar därför flera grundläggande metodfrågor, som inte låter sig besvaras med hjälp av den metodologi som redan finns väl utarbetad för kvalitetskontroll i industriella sammanhang. Här kommer några sådana frågor att presenteras.

Fråga 1: Var går gränsen för det oacceptabla?

I den aktuella artikeln redovisas åldersstandardiserad mortalitet vid akut hjärtinfarkt under 1996 vid 32 sjukhus. Mortaliteten är här alltså att betrakta som ett slags utfallsmått vid beskrivningen av kvaliteten vid de olika inrättningarna. Om mortaliteten i princip är densamma inom alla kliniker skall endast slumpmässiga avvikelser förekomma mellan klinikerna.

Det påpekas i artikeln att man »måste dock ta hänsyn till slumpmässiga variationer. Enbart skillnader som är statistiskt signifikanta bör därför beaktas. Dock uppkommer ju även några sådana slumpvis då uppåt ett 50-tal kliniker undersöks».

(Naturligtvis behöver man inte undersöka 50 kliniker för att en sk slumpsignifikans skall uppkomma, det må räcka med en enda klinik. Men chansen att råka få med minst en slumpsignifikans ökar givetvis med antalet undersökta kliniker.)

Slumpen kan väntas »pricka» en klinik

Som framgår av Figur 1 i Brittons och medarbetares artikel (återges även här) är det vid fyra kliniker som mortaliteten skiljer sig signifikant från genomsnittet; vid tre av dessa är mortaliteten lägre. Man kan emellertid förmoda att intresset inte just är centrerat kring dessa tre utan kring den stackars klinik som råkade få en signifikant högre mortalitet än genomsnittet.

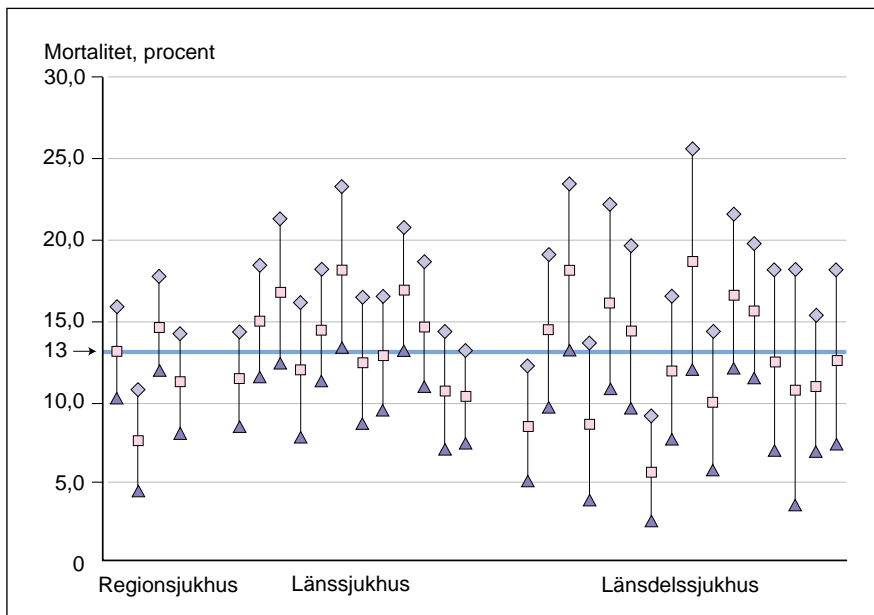
Vid de dubbelsidiga konfidensintervall som beräknats är chansen att »av en slump» hamna signifikant över genomsnittet 2,5 procent och förväntat antal slumpsignifikanser i denna riktning $32 \times 0,025$, dvs 0,8. Således förväntar man sig att ungefär att en klinik skall avvika signifikant med högre mortalitet

Författare

ADAM TAUBE

professor emeritus, statistik, Institutionen för informationsvetenskap, Uppsala universitet.

E-post: Adam.Taube@dis.uu.se



Figur 1. Åldersstandardiserad mortalitet vid akut hjärtinfarkt under 1996. Figuren visar klinikvisa frekvenser med 95 procents konfidensintervall samt medelfrekvens för hela patientmaterialet. Statistiskt signifikant skillnad från medelvärdet föreligger då ingen del av konfidensintervallet berör medelvärdelinjen.

trots att det egentligen inte finns några systematiska skillnader mellan kliniker.

Slutsatsen här blir att det i Figur 1 inte föreligger fler signifikanta avvikelser från genomsnittsmortaliteten än vad som kan förväntas av en ren slump i en situation då alla kliniker har exakt samma kvalitet på sitt arbete.

Nu är det emellertid så att kliniker inte alls står för några »slumpmässiga» stickprov av patienter: Var och en av dessa kliniker har sin egen upptagningspopulation med speciell yrkesstruktur m m. Därför skall man i princip inte förvänta sig ens att det förekommer enbart slumpmässiga avvikelser, även om vårdens kvalitet vore exakt densamma vid alla kliniker. Man kan faktiskt förvänta sig vissa (signifikanta) skillnader även mellan kliniker med precis samma kvalitet på vården. Hur stora skillnader som kan förväntas på grund av strukturvariationer m m tycks dock inte finnas någon objektiv metod för att beräkna.

Det är uppenbart inte meningsfullt att beakta avvikelser som ej är statistiskt signifikanta, men alla statistiskt signifikanta avvikelser är dessvärre inte heller intressanta. Var gränsen skall dras för en mortalitet som avviker oacceptabelt mycket från genomsnittet synes inte vara möjligt att ange.

Fråga 2: Existerar endast en enda optimal nivå?

I ett annat indikator exempel redovisar Britton och medarbetare i sin Figur 2 den åldersstandardiserade trombolysfrekvensen vid akut hjärtinfarkt vid åtta kliniker under tre år. Det är här alltså fråga om ett s k processmått, som skall ange om man vid de olika kliniker följer vedertagna behandlingslinjer. Jag citerar: »Trombolysbehandling redovi-

sas under en treårsperiod regelbundet av åtta kliniker. De flesta ligger relativt stabilt över åren, men skiljer sig åt i nivå från 25 till 45 procents trombolysfrekvens. Båda nivåerna kan inte vara optimala.»

Är det verkligen så att det existerar endast en möjlig optimal nivå som gäller för samtliga kliniker? Eller har varje kliniker sin egen optimala nivå? Eftersom »patientmixen» i princip inte är densamma vid de olika kliniker kan det ju faktiskt även vara så att det är optimalt att inte ha exakt samma trombolysfrekvens överallt. Måhända bör den av flera skäl inte vara exakt densamma i Tornedalen och i Storstockholm...

Fråga 3: Hur stor variation mellan sjukhus kan förväntas?

Om det nu är så att en viss variation i trombolysfrekvens kan vara betingad av olika sammansättningar på patientpopulationen eller andra omständigheter så inställer sig följdfrågan hur stor variation mellan kliniker som i någon mening är naturlig. Det är inte lätt att få svar på den frågan. Det har publicerats åtskilliga artiklar om »stora skillnader mellan olika sjukhus», men sällan ser man vilken variation som borde kunna förväntas under likartade, normala betingelser.

I artikeln påpekar författarna vidare: »Vid ett av sjukhusen sjönk plötsligt trombolysfrekvensen från en under fyra år konstant hög nivå till en betydligt läg-

re (se Figur 2). Närmare analys visade att underbehandling skett i flera fall. Den troligaste anledningen bedömdes vara att hjärtavdelningen det året genomgått en genomgripande omorganisation med personalförändringar och dessutom en förflyttning.»

För läsaren är det knappast möjligt att få klarhet om vilken sänkning man här har ansett vara av anmärkningsvärd storlek, eftersom man inte vet vilka kriterier som kan ha tillämpats vid bedömningen av variationen från ett år till ett annat. Från den aktuella figuren är det inte lätt att gissa vilken kliniker det är fråga om, måhända är det kliniker 4 eller kliniker 7. Man har emellertid trots sig finna en plausibel förklaring till nedgången och drar därmed även slutsatsen att indikatorn »verkligen avspeglar betydelsefulla kvalitetsförändringar och att riskerna kan öka vid stora omorganiseringar».

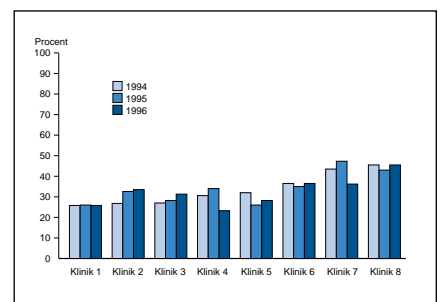
De presenterade indikatorernas stora sensitivitet motiveras troligtvis, men dessvärre förblir frågan om hur stor »naturlig» variation som kan förväntas helt obesvarad.

Metoddiskussion önskvärd

Det är uppenbart från Brittons och medarbetarens artikel att tolkningen av de indikatorer som nu tillämpas inte alls är färdigpenetrerad, liksom att det återstår olösta statistiska metodologiska problem. Följaktligen är – som tidigare påpekats av Nilsson och Bring i Läkartidningen 5/99 – en metodologisk debatt inom den medicinska professionen och med deltagande av statistisk expertis synnerligen angelägen.

PS

Efter det att detta manus färdigställts har Birgitta Stegmayr och medarbetare (Läkartidningen 22/99) publicerat resultat grundade på data från Riks-Stroke. Även i detta arbete påstås att det föreligger stora skillnader mellan kliniker och mellan regioner. Det görs emellertid – som vanligt – inte något som helst försök att utreda hur stor (naturlig) variation som är att förvänta.



Figur 2. Åldersstandardiserad trombolysfrekvens vid akut hjärtinfarkt vid åtta kliniker som redovisat årliga uppgifter under tiden 1994–1996.