

Rankningslistor är inte att lita på

Andra sätt att presentera data kan behövas



ANDERS ANELL, professor, Ekonomihögskolan, Lunds universitet anders.anell@fek.lu.se



OSKAR HAGBERG, statistiker, Regionalt cancercentrum syd, Lund oskar.hagberg@skane.se

Uppföljning av indikatorer på kvalitet i vården har sedan länge använts för klinisk forskning och internt förbättringsarbete. På senare tid, och i ökande grad, har samma information använts för externt ansvarsutkrävande [1, 2]. Exempel är att indikatorer på kvalitet kopplas samman med ekonomisk ersättning, s.k. målrelaterad ersättning, eller att vårdgivares kvalitet rangordnas som underlag för politikernas beslut eller vårdkonsumenters val av vårdgivare. Kan man då lita på rankningslistor?

Ett sätt att belysa frågan är att jämföra utfallet av olika rankningslistor för att se om resultaten är konsekventa. I en aktuell studie publicerad i den ledande amerikanska tidskriften Health Affairs jämförs fyra olika nationella rankningssystem för sjukhus i USA [3].

Artikeln huvudsakliga budskap är att de olika rankningssystemen producerar väldigt olika resultat, trots att syftet är att vägleda vårdkonsumenter till de bästa alternativen. Inget sjukhus rankades som högrepresterare av alla fyra systemen. Bara 10 procent av de 844 sjukhus som rankades som högrepresterare av ett system rankades som högrepresterare av något av de andra systemen. Flera sjukhus som rankades högt av något av systemen rankades tvärtom lågt av något av de andra systemen.

Olika utgångspunkt ger olika resultat

Artikeln belyser flera bidragande förklaringar till olika bedömningar. En första skillnad är vilka kvalitetsdimensioner som fokuseras. Två av systemen (Leapfrog och Consumer Reports) fokuserar på patientsäkerhet, dock med olika definitioner av begreppet. Det tredje systemet (Healthgrades) fokuserar på kvalitet utifrån risk för komplikationer och mortalitet. Det fjärde systemet (U.S. News) fokuserar på bästa

kvalitet för patienter med de mest komplicerade tillstånden.

Med dessa olika utgångspunkter är det inte så konstigt att de olika rankningssystemen omfattar olika val av indikatorer. I ett av systemen (Healthgrades) beaktas dessutom enbart indikatorer på resultat för patienterna, medan de andra systemen också omfattar struktur- och processindikatorer.

Ytterligare skillnader är inklusions- och exklusionskriterier för sjukhus och hantering av saknade data för de sjukhus som omfattas. Även riskjusteringen av utfallsindikatorer varierar; för två av systemen (U.S. News och Healthgrades) är inte riskjusteringen redovisad, eftersom denna utgör en affärshemlighet.

Systemen skiljer sig även avseende hur de kommunicerar resultaten. Ett system (Leapfrog) ger varje sjukhus ett bokstavsbyte (A–F), två system (Consumer Reports och U.S. News) bedömer sjukhusen med en 100-gradig skala och ett system (Healthgrades) anger de bästa 50 respektive 100 sjukhusen.

Även finansieringen skiljer sig åt för de olika rankingarna. Tre av systemen tillåter att berörda sjukhus, mot en avgift, använder sin ranking i annonser och marknadsföring, medan ett system (Consumer Reports) inte tillåter detta utan finansierar sin verksamhet genom avgifter från prenumeranter i form av konsumenter.

Kan motverka syftet

Till en del kan skillnader i olika rankningssystem förstås vara motiverade. Den snabba tillväxten i antalet olika indikatorer som används i allt fler och olika uppföljningar kan emellertid också motverka syftet att förbättra verksamhetens kvalitet. Andra studier från amerikansk vård visar att det finns liten överensstämmelse mellan de indikatorer som används i olika uppföljningar från finansierares perspektiv [4]. En genomgång av 48 uppföljningssystem på delstatsnivå och regional nivå identifierade fler än 500 indikatorer, men bara 20 procent av dem användes av mer än ett program [5].

En annan studie av 29 privata försäkringsplaner identifierade 550 indikatorer, med liten överlappning med de indikatorer som används av offentliga program [6]. För enskilda sjukhus innebär

skillnaderna att man får ägna betydande resurser åt att rapportera in de olika data som begärs.

För patienter kan det också vara svårt att dra några bestämda slutsatser från de olika uppföljningarna, eftersom de pekar i olika riktningar. Det kan vara en bidragande förklaring till att patienter i USA, trots 20 års erfarenhet av öppna jämförelser av sjukhusens kvalitet, i liten utsträckning beaktar informationen vid sina val av vårdgivare [7].

Annat resultat i Öppna jämförelser

I ett svenskt perspektiv framstår pluralismen när det gäller olika indikatorer och uppföljningssystem i amerikansk sjukvård som avskräckande och något som vi lyckats undvika. Samtidigt framstår de olikheter som faktiskt finns i Sverige, med knappt 10 miljoner invånare, som märkliga i jämförelse med amerikanska hälso- och sjukvårdsföretag av samma storlek, med helt enhetliga system för uppföljning.

Grunden för uppföljningen i svensk vård är till stor del densamma, genom nationella kvalitetsregister och nationella system för enkäter till patienter och befolkning, men tillämpningarna varierar mellan landstingen. Exempelvis varierar uppföljningssystemen i primärvården och även hur indikatorer används inom ramen för t.ex. målrelaterade ersättningssystem.

Jämförelsen i Health Affairs gör oss uppmärksamma på att resultaten av jämförelser beror på vilka dimensioner och indikatorer som valts ut. Med ett annat system för ranking mellan landsting än det som används i Öppna jämförelser skulle ordningen kanske sett annorlunda ut.

Systematiska skillnader eller slump

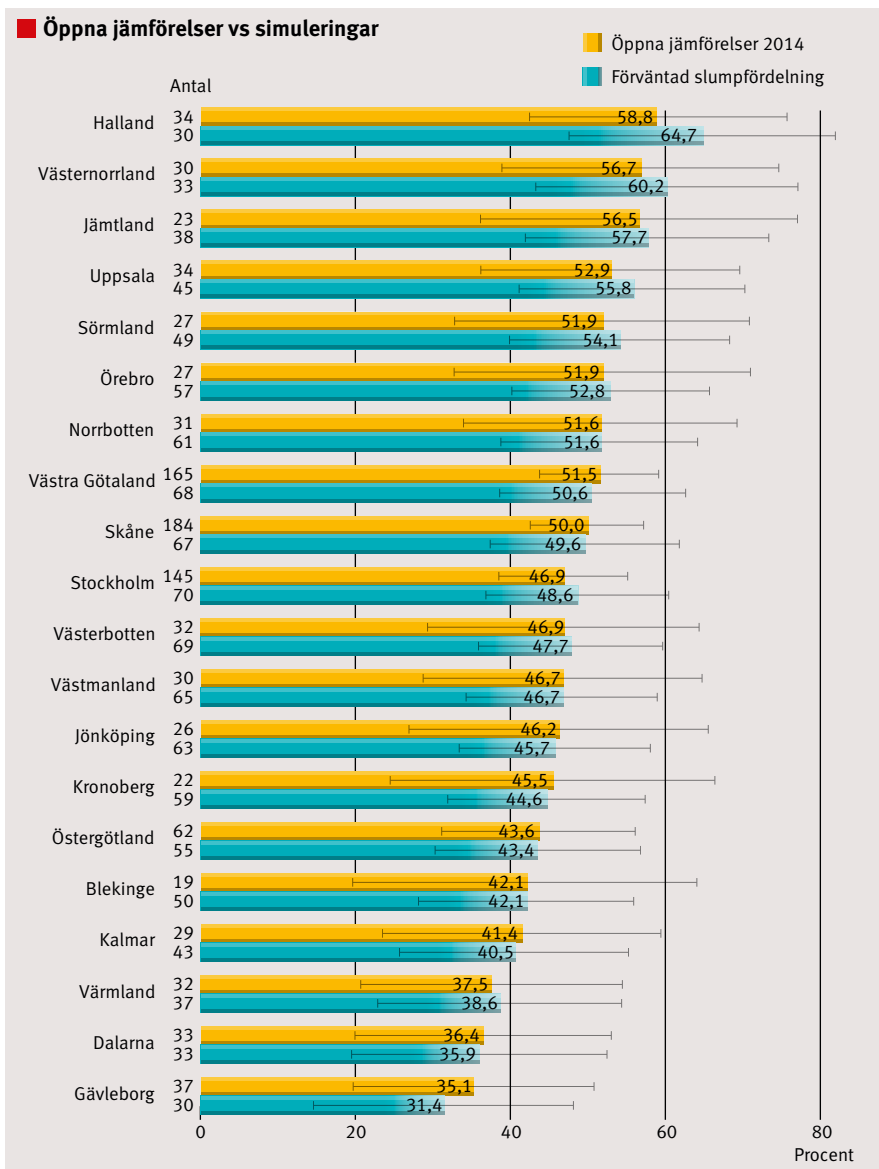
Det finns också ett annat osäkerhetsmoment vid ranking av olika vårdgivare i form av slumpmässig variation. I Socialstyrelsens »Öppna jämförelser av cancervården år 2014« sägs t.ex. följande

■ SAMMANFATTAT

I ökande grad används uppföljning av vårdgivares kvalitet som underlag för styrning och öppna jämförelser.

En jämförelse av fyra rankningssystem för sjukhus i USA visar att de bedömer sjukhus olika. Flera faktorer bidrar till skillnaderna.

Slumpmässig variation kan leda till att rankningslistor blir falskt vägledande. Alternativa sätt att presentera data bör då övervägas.



Figur 1. Andel patienter med urinblåsecancer i tumörstadium T2–T4 som genomgår kurativt syftande behandling. Öppna jämförelser 2014 jämfört med medelvärde av 1000 simuleringar.

om vården av urinblåsecancer i Sverige [8]: »Det är stora skillnader mellan landstingen när det gäller andelen som får en kurativt syftande behandling, 35–59 procent. Jämförelsen omfattar 1030 patienter som under åren 2012–2013 hade en tumör i stadium T2–T4.«

Att skriva så kan betraktas som vilseledande. Patienter får en behandling därför att ansvariga läkare gjort bedömningen att den är motiverad. Där finns ingen slump. Däremot kan sjukhus inte styra över vilka patienter som söker vård. Detta ger effekter som från vårdens perspektiv är slumpmässiga. Den korrekta frågan är därför: »Finns

det systematiska skillnader som är större än de som kan förklaras av slump?»

I Figur 1 har jämförelsen i Öppna jämförelser 2014 ställts emot medelvärdet av 1000 datorsimuleringar under samma förutsättningar. Patienter har antagits vara oberoende av varandra och ha samma sannolikhet (48,1 procent) att få en kurativt syftande behandling. Resultaten har ordnats från det »sämsta« till det »bästa«.

Simuleringen visar att en variation mellan landsting på 35–59 procent inte alls är förvånande. I själva verket är den variation man kan förvänta sig av ren slump något större: 31–65 procent.

Precis som man kan förvänta sig kännetecknas de som rankas högt och lågt av färre observationer. Tolkningen av data bör alltså snarast vara att det inte finns några belegg för systematiska variationer mellan landsting.

Samma kritik av rankingslistor utifrån att bättre och sämre sjukhus ofta inte skiljer sig signifikant från genomsnittliga sjukhus har redovisats i en aktuell studie från USA, baserad på data från Medicare Hospital Compare [9].

Anledning att fundera strategiskt

I takt med att uppföljning av indikatorer på kvalitet i vården får allt större spridning i svensk vård finns anledning att fundera strategiskt på hur uppföljningen ska användas [1]. Vilka former för ansvarutkrävande finns, och hur påverkar de professionens motivation till förbättringsarbete? Hur bör data presenteras mot bakgrund av valt syfte, och vilka egenskaper hos indikatorer blir då väsentliga?

I denna kommentar har främst osäkerheten i rankingslistor lyfts fram. Rankingslistor baseras implicit på önskemålen att göra summativa bedömningar av vårdgivare: vilka som har hög respektive låg kvalitet. Det ställer höga krav på valda indikatorer, att data riskjusteras utifrån olikheter mellan patienter och att osäkerhet till följd av slump hanteras.

Om syftet med uppföljningen är att identifiera systematiska avvikelser bör presentationen av data ge stöd för en sådan användning. Det innebär i princip att rankingslistor bör undvikas. Även om rankingslistor kombineras med konfidensintervall, ger de sken av att det är möjligt att på ett enkelt sätt skilja mellan olika vårdgivares kvalitet. Rankingslistor blir då falskt vägledande.

Med ett diagnostiskt syfte i uppföljningen kan i stället alternativa tabeller i form av exempelvis trattendigram tillämpas. De vårdgivare som tycks avvika kan bli föremål för närmare analys där ytterligare information bedöms. En närmare bedömning utifrån ytterligare insamlade data får sedan avgöra om avvikelserna kan fastställas och vilka åtgärder som i så fall kan anses lämpliga.

■ *Potentiella bindningar eller jävsförhållanden: Inga uppgivna.*

REFERENSER

1. Anell A. Styrning i vården genom uppföljning av indikatorer på kvalitet och effektivitet. KEFU Skriftserie 2014:4. Lund: Institutet för ekonomisk forskning vid Lunds universitet; 2014.
2. Braspenning J, Hermens R, Calsbeek H. Quality and safety of care: the role of indicators. In: Grol R, Wensing M, Eccles M, et al (eds). Improving patient care: the implementation of change in health care. Second edition. Hoboken, NJ: John Wiley & Sons; 2013.
3. Austin MJ, Jha AK, Romano PS, et al. National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Affairs*. 2015;34(3):423-30.
4. Cassel CK, Conway PH, Delbanco SF, et al. Getting more performance from performance measurement. *N Engl J Med*. 2014;371(23):2145-7.
5. Bazinsky K, Bailit M. The significant lack of alignment across state and regional health measure sets. *Health care performance measurement activity: an analysis of 48 state and regional measure sets*. Bailit Health Purchasing. 10 sep 2013. <http://www.buyingvalue.org/resources>
6. Higgins A, Veselovskiy G, McKnown L. Provider performance measures in private and public programs: achieving meaningful alignment with flexibility to innovate. *Health Affairs*. 2013;32(8):1453-61.
7. Hussey PS, Luft HS, McNamara P. Public reporting of provider performance at a crossroads in the United States: summary of current barriers and recommendations on how to move forward. *Med Care Res Rev*. 2015;71(5):5S-16S.
8. Öppna jämförelser 2014. *Cancersjukvård. Jämförelse mellan landsting*. p. 66. Stockholm: Socialstyrelsen, Sveriges Kommuner och landsting; 2014. Artikelnr 2014-12-2.
9. Paddock SM, Adams JL, Hoces de la Guardia F. Better-than-average and worse-than-average hospitals may not significantly differ from average hospitals: an analysis of Medicare Hospital Compare ratings. *BMJ Qual Saf*. 2015;24:128-34.