

Surrogatmått – en ersättning för det man egentligen vill mäta

Lättmätta men irrelevanta?

■ Ett rimligt krav på en metod som skall användas i klinisk praxis är att det finns dokumentation som visar att användning av metoden för patienten gör större nytta än skada, eller, med andra ord, att balansen mellan nytta och risk är gynnsam. Detta gäller inte bara behandling (t ex med läkemedel), kirurgiska ingrepp, sjukgymnastik, psykoterapi och metoder inom alternativmedicinen, utan också diagnostik (t ex screening för tidig upptäckt av bröst- eller prostatacancer).

När det gäller läkemedel anger Läkemedelslagen att det åligger tillverkaren att visa »ändamålsenlighet« för att erhålla godkännande för marknadsföring: »Ett läkemedel är ändamålsenligt om det är verksamt för sitt ändamål och vid normal användning inte har skadeverkningar som står i missförhållande till den avsedda effekten.« Det som avgör ett läkemedels ändamålsenlighet är således balansen mellan nytta och risk.

För andra behandlingsmetoder och för diagnostiska åtgärder är kraven inte lika väldefinierade; det kan räcka med att några inom vården tycker att det förefaller spännande och rimligt att pröva en ny metod, t ex laparoskopisk kirurgi, för att metoden skall kunna komma till användning i klinisk praxis.

De flesta exempel som diskuteras i föreliggande artikel avser behandling med läkemedel, men samma principer kan tillämpas när det gäller andra interventioner.

Vad är ett surrogatmått?

Det engelska uttrycket för surrogatmått är »surrogate endpoint« eller »surrogate marker«. Begreppet har av Robert Temple vid amerikanska Food and Drug Administration (FDA) definierats på följande sätt: »A surrogate endpoint of a clinical trial is a laboratory measurement or a physical sign used as a substitute for a clinically useful endpoint that measures directly how a patient feels, functions or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint« [1].

Den ordlista som sammanställts av Statens beredning för medicinsk utvärdering (SBU) ger följande definition:

»Surrogatmått är ett effektmått som har ett samband med den effekt som är det egentliga slutmålet för en åtgärd och som i en klinisk prövning används som ersättare för detta.«

Vad är det man egentligen vill mäta?

Det vanligaste sättet på vilket en behandling kan göra nytta är att åtgärden får patienten att må bättre, t ex genom lindring av

SAMMANFATTAT

Värdet av en behandlingsmetod bestäms av balansen mellan de positiva och de negativa effekterna. Att dokumentera nytta kräver ofta att stora grupper av patienter följs över lång tid, vilket är dyrt och ofta förenat med etiska, praktiska och metodologiska problem.

Att indikera nytta genom att i stället mäta något som är enkelt att mäta kan förefalla tilltalande. Ett surrogatmått är ett laboratorievärde eller ett fysikaliskt tecken som används som ersättning för det man egentligen vill mäta. Förhoppningen är att förändringar i surrogatmättet skall avspeglar förändringar i det man egentligen ville mäta.

Den stora fördelen med ett väl fungerande surrogatmått är att man med färre patienter och på betydligt kortare tid kan dokumentera nyttan av en behandling.

Dokumentationen är dock bristfällig för att förändringar av surrogatmättet verkligen avspeglar något kliniskt betydelsefullt.

Evidensbaserad medicin

symtom som smärta, feber, ångest, andnöd eller nästäppa. »Må bättre« kan också innebära ett ökat allmänt välbefinnande eller en förbättrad hälsorelaterad livskvalitet.

En behandling kan också göra nytta genom att minska risken såväl för framtida icke-dödliga komplikationer till sjukdomen som för progress av funktionsnedsättning. Det kan t ex handla om att reducera risken för exacerbationer vid astma, för synnedsättning vid diabetes, för kotfraktur vid osteoporos eller för icke-letal hjärtinfarkt vid hyperkolesterolemi. Progress av funktionsnedsättning, t ex lägre expiratoriskt flöde vid astma, får anses vara ett surrogatmått så länge patienten inte märker av nedsättningen [2]. Detsamma gäller progress

av njurfunktionsinskränkning vid hypertoni. Först när patienten kan notera den försämrade funktionen kan man tala om ett egentligt patientrelaterat effektmått.

Ett tredje sätt på vilket en behandling kan göra nytta är genom att förlänga patientens liv, t ex vid cancer, aids eller efter hjärtinfarkt.

Varför mäter man inte alltid det man egentligen vill mäta?

Det kan tyckas rimligt att den som vill introducera en ny behandlingsmetod åläggs att dokumentera nyttan av densamma i form av gynnsam effekt enligt något av ovanstående kriterier. Problemet är att det ofta finns metodologiska svårigheter, tar lång tid, kräver ett stort antal patienter i kliniska prövningar, är dyrt eller är tvivelaktigt ur etisk synvinkel. Det kan därför vara lockande att i stället mäta ett surrogatmått, eftersom detta oftast är enklare, inte tar lika lång tid, kräver färre patienter, kostar avsevärt mindre och kan förefalla etiskt oantastligt.

Det sistnämnda kan dock ifrågasättas, eftersom Helsingforsdeklarationen kräver att klinisk forskning både skall vara meningsfull och ge kunskap till nytta för patienter med den sjukdom som studeras.

Svårigheter vid dokumentation av symtomlindring

Vid icke livshotande tillstånd är den vanligaste orsaken till behandling att man vill att patienten skall må bättre, med lindrigare symtom eller ökad hälsorelaterad livskvalitet, helst bådadera.

Mätning av subjektiva symtom vid t ex angina pectoris eller refluxsjukdom medför flera metodologiska problem. Patienter med angina pectoris kan beskriva sina symtom på olika sätt: som »smärta«, »sveda«, »trånghets känsla« eller »tyngdkänsla«. Refluxpatienten kan vara mest besvärad av »halsbränna« eller »sveda bakom bröstbenet«. Man kan således stöta på problem redan vid benämning av de viktigaste symtomen.

Ett annat problem vid mätning av symtom är att patienter med samma diagnos inte alltid har identiska problem. Man räknar med att ca 40 procent av astmapatienter får nattsömnen störd av sin lungsjukdom. För denna grupp är minskad nattlig astma en betydande behandlingsvinst. För de 60 procent som inte lider av nattliga symtom finns inte denna behandlingsvinst att hämta. Om man i en klinisk prövning blandar patienter

med och utan nattlig astma, och har lindring av nattlig astma som ett viktigt mått för nytta, är risken stor att positiva resultat för en patientgrupp späds ut av det faktum att en större grupp inte har något att vinna när det gäller detta symtom.

Ytterligare ett problem vid mätning av subjektiva symtom är mätskalan. Ofta används en sk ordinalskala (t ex ingen-lätt-måttlig-svår-mycket svår), där man vet ordningsföljden på skalstegen men inte om skalstegen är lika höga, vilket medför betydande begränsningar i den statistiska bearbetningen. Vid t ex smärta används inte sällan en visuell, analog skala (VAS), en 100 mm lång linje. Svårigheterna med denna är dels att veta hur kliniskt relevant en skillnad på t ex 5 mm är, dels att det kan ifrågasättas om en förändring på 1 mm i mitten av skalan representerar en lika stor förändring i smärta som 1 mm i den nedre eller den övre delen av skalan.

Sammantaget kan man konstatera att det finns flera metodologiska problem när man vill dokumentera att en behandling leder till symtomlindring. Det kan därför vara lockande att i stället använda något instrument för att få fram ett mätvärde som lämpar sig väl för statistisk bearbetning. Vid astma kan man välja att i stället för symtom mäta den forcerade exspiratoriska volymen under en sekund (FEV_{1s}), eller under 10 millisekunder, »peak expiratory flow« (PEF). Vid refluxsjukdom förlitar man sig ofta mer på gastroenterologens subjektiva bedömning av slemhinnerodnad än på patientens bedömning av sina symtom.

Hur dokumentera ökad hälsorelaterad livskvalitet?

Hälsorelaterad livskvalitet är ett informativt effektmått i kliniska studier [3]. Inom flera terapiområden finns sjukdomspecifika livskvalitetsmått med känd reliabilitet, validitet och känslighet för kliniskt relevanta förändringar. Viktiga fördelar med livskvalitet som effektmått är att mätningen är patientbaserad, mångdimensionell och ger en bild av hur patienten mår och fungerar i vardagen.

Kliniska prövningar är en blandning av klinik och statistik. Metodiken vid kliniska prövningar bygger på att testa i förväg uppställda hypoteser. Det innebär att man i prövningsprotokollet skall definiera en eller ett fåtal frågeställningar, vilken mätmetod som skall användas och när mätning skall ske. Formulär för livskvalitetsmätning innehåller ofta ett stort antal frågor, vilket är väl motiverat för att möjliggöra en allsidig beskrivning av hur patienten mår. Problemet är att om man vill signifikantesta för varje enskild fråga är risken för sk slumpsignifikanser stor. Om man i stället slår ihop närliggande frågor riskerar man att förlora viktig information. Ytterligare ett alternativ kan vara att i förväg välja ut en av frågorna som primär frågeställning. Då kommer man ifrån problemet med multipla signifikantest, men riskerar i stället de problem som ovan beskrivits vid mätning av enskilda symtom.

Mätning av livskvalitet i kliniska prövningar tillför utan tvivel värdefull information om patienternas allmänna välbefinnande. Ändå ser man relativt sällan att mätning av livskvalitet är primär frågeställning. Den främsta orsaken till detta är den inbyggda konflikten mellan antalet frågor i livskvalitetsformulär och de statistiska metoderna för hypotesprövning, vilka utgör basen för utvärdering av resultat av kliniska prövningar.

Hur mäta risk för komplikationer och funktionsnedsättning?

Många behandlingar ges inte primärt för att öka patientens aktuella välbefinnande utan för att förebygga framtida komplikationer, recidiv eller progress av sjukdomen. Flera olika behandlingar har visat sig minska risken för icke dödliga komplikationer under de första åren efter hjärtinfarkt. Många behandlingar inom onkologin ökar signifikant tiden till recidiv. Vid astma ger behandling med inhalationssteroider långsam-

mare progress av lungfunktionsnedsättning. Gemensamt för dessa exempel är att komplikationsincidensen är relativt hög inom något eller några år. Det är därför möjligt att visa nytta genom kliniska prövningar som pågår under ett fåtal år och omfattar upp till några tusen patienter.

Svårigheterna är betydligt större om man vill visa att blodtryckssänkande behandling vid mild-måttlig hypertoni hos medelålders individer minskar risken för slaganfall, hjärtinfarkt eller hjärtsvikt. En klinisk prövning för att fastställa detta skulle kräva en mycket stor, långvarig och dyr prövning. Man kan förstå de företag som hellre väljer att i en tolvveckorsstudie visa att blodtrycket går ner. Vid typ II-diabetes är det betydligt enklare att dokumentera korttidseffekter på blodsocker eller HbA_{1c} än att visa att tiden till synnedsättning, neuropati eller njursvikt förlängs.

Hur dokumentera förlängd överlevnad?

Nytan kan knappast ifrågasättas av en behandling som visats ge förlängd överlevnad. Den kliniska relevansen är ju odiskutabel, och mätmetoden ger inte mycket utrymme för olika tolkningar. Detta gäller studier med signifikant effekt på totalmortalitet, men knappast prövningar som visar t ex reducerad kardiovaskulär mortalitet men ingen skillnad i total dödlighet.

Svårigheterna förknippade med att dokumentera att en diagnostisk åtgärd eller behandling leder till förlängd överlevnad hänger förstas samman med prognosen vid den studerade sjukdomen. För sjukdomar som subaraknoidalblödning eller pankreascancer kan det vara möjligt att med ett rimligt patientantal och på relativt kort tid visa att en ny behandlingsmetod kan ge förlängd överlevnad. För andra sjukdomar är det svårt att med kliniska prövningar visa att en viss åtgärd förlänger livet. Det gäller även för sjukdomar som är vanliga och som är förenade med hög mortalitet.

Ett aktuellt och omdiskuterat exempel är mammografiscreening och dess effekt på totalmortaliteten. Vid astma finns det indikationer från observationella studier att behandling med inhalationssteroider förbättrar överlevnaden, men detta går knappast att studera i en prospektiv klinisk prövning.

Vilka är fördelarna med surrogatmått?

Ett surrogatmått som predicerar nytta kan bidra till att nya effektiva behandlingar blir tillgängliga för sjukvården flera år innan dokumentationen för verklig klinisk nytta blir tillgänglig. Ett exempel på detta är simvastatin, som godkändes för marknadsföring på basis av mätning av lipidnivåer och där senare kliniska prövningar har påvisat nytta vad gäller kardiovaskulär morbiditet och mortalitet.

Kostnaden för kliniska prövningar med effekt på ett surrogatmått som primär frågeställning är avsevärt lägre, vilket förhoppningsvis kan ge utslag vid prissättningen av läkemedlet.

... och nackdelarna?

Effekt på ett surrogatmått har i sig inte något värde för patienten. Ofta väljs surrogatmått baserat på logiskt tänkande. En riskfaktor, t ex blodtryck eller serumkolesterol, samvarierar med morbiditet och mortalitet. En intervention reducerar riskfaktornivån, och därför är det rimligt att förmoda att interventionen minskar risken för morbiditet och mortalitet.

En anledning till att en viss reduktion av en riskfaktor inte behöver ge motsvarande reduktion i morbiditet och mortalitet är att riskfaktorn kan vara en varningssignal för andra rubbningar i kroppen. Det förhöjda blodtrycket kan vara en del av ett metabolt syndrom med störningar i t ex insulinresistens och lipidmetabolism. Ett blodtryckssänkande preparat kan då »skruva ur varningslampan«, dvs sänka blodtrycket men en-

dast obetydligt minska risken för morbiditet och mortalitet genom att andra riskfaktorer förbises.

En annan anledning till att en gynnsam effekt på ett surrogatmått inte behöver vara gynnsam för patienten är att alla läkemedel även har negativa effekter som kan påverka balansen mellan nytta och risk. Surrogatmålet ger ingen information om riskerna. Ett exempel är Atromidol (klofibrat), under 1960- och 1970-talen en storsäljare vid hyperkolesterolemi. Klofibrat har en gynnsam effekt på lipidnivåerna och borde logiskt sett kunna ge minskad morbiditet och mortalitet. En stor europeisk placebokontrollerad klinisk prövning under 1970-talet visade överraskande en signifikant ökad mortalitet i klofibratgruppen. Ökningen gällde även det som behandlingen skulle minska risken för, nämligen kardiovaskulär mortalitet [4]. Gynnsam effekt på ett surrogatmått behöver alltså inte betyda minskad risk för patienten.

Ytterligare en orsak till varför det finns brister i sambandet mellan effekter på surrogatmått och effekter på morbiditet/mortalitet är att läkemedel kan ha flera positiva effekter. Detta verkar vara fallet för vissa av statinerna. Utöver gynnsamma effekter på lipidmönstret har de visats ha flera andra biologiska effekter, bl a antiinflammatoriska och plackstabiliserande. Graden av dessa effekter tycks inte vara kopplad till graden av sänkning av LDL-kolesterol. I den nyligen rapporterade Heart Protection Study hade simvastatin en gynnsam effekt på morbiditet och mortalitet oberoende av om kolesterolnivån före studien var förhöjd, normal eller låg under gängse behandlingsmål [5]. Det förefaller rimligt att behandla patienter med ökad kardiovaskulär risk oavsett lipidnivåer, och i stället för att behandla laboratorievärden välja preparat och dos på grundval av vad som i kliniska prövningar visat sig påverka morbiditet och mortalitet.

Vilka förändringar avspeglar surrogatmålet?

Ett mycket stort antal kliniska prövningar med olika typer av läkemedel har visat att sänkning av blodtrycket vid hypertoni minskar risken för slaganfall. Detta har medfört att det råder enighet om att sänkning av ett förhöjt blodtryck är en tillförlitlig markör för minskad risk för slaganfall. Detta betyder inte att det råder enighet om blodtryckssänkning som surrogatmått för nytta vad gäller andra komplikationer, t ex hjärtinfarkt och hjärtsvikt. I den pågående ALLHAT-studien avbröts doxazosinarmen i förtid av etiska skäl, då antalet kardiovaskulära komplikationer i denna grupp var 25 procent större än i klortalidongruppen, trots likvärdig blodtryckssänkning [6].

Signifikant blodtryckssänkning i en tolv veckor lång klinisk prövning av ett nytt blodtryckssänkande läkemedel indikerar nytta i form av reducerad risk för slaganfall. Prövningen ger dock föga vägledning om balansen mellan nytta och risk. Dels ger den inte information om andra tänkbara positiva effekter i form av reducerad risk för hjärtinfarkt, hjärtsvikt eller njurskada, dels ger den knappast någon vägledning för bedömning av riskerna vid långtidsbehandling.

I kliniska prövningar vid aids användes tidigt antalet CD4-celler som surrogatmått för nytta, definierat som förlängd tid till insjuknande i aids för HIV-positiva eller förlängd överlevnad för aidssjuka. I en översikt 1994 [7] visade det sig att antalet CD4-celler ökade signifikant i sju av åtta studier där behandlingen förbättrade prognosen. Problemet var att CD4-cellerna ökade signifikant även i sex av åtta studier där behandlingen inte förbättrade prognosen.

Senare har man valt att använda virusförekomst i blod, »viral load«, som surrogat för nytta. På basis av detta effektmått kom ett flertal HIV-medel snabbt ut på marknaden, och för några av dessa har klinisk nytta senare visats i flera studier. Virusförekomst kan dock inte i alla lägen förutsäga klinisk ef-

fekt. Goda kliniska resultat kan ses även om man inte helt lyckas pressa ned virusreplikationen [8].

Det finns misslyckanden

Kartan överensstämmer dock inte alltid med verkligheten. Listan över misslyckanden med surrogatmått kan dessvärre göras lång. Här är två exempel:

- *Patienter med kombinationen koronarsjukdom–ventrikulära extrasystolier (ES) utgör en högriskgrupp för plötslig död.* För drygt tio år sedan introducerades i USA två nya antiarytmika, som mycket effektivt reducerade antalet ES enligt långtidsregistrering av EKG. Rent logiskt borde dessa läkemedel kunna minska risken för plötslig död. Logiken ledde till att mer än 200 000 patienter per år behandlades med dessa läkemedel i USA.

Efter ingående diskussioner om det etiska i att genomföra en prövning för att studera mortalitet accepterades en placebokontrollerad studie med plötslig död som primär frågeställning. Studien avbröts i förtid av säkerhetskommittén sedan det visat sig att 33 fall av plötslig död inträffat med aktiv behandling, mot nio fall med placebo. Totalt avled 56 patienter med aktiv behandling, mot 22 i placebogruppen [9]. Man räknar med att fler än 50 000 amerikaner dog till följd av behandling med enkainid eller flekainid, vilket är ungefär lika många som dog i Vietnamkriget. Detta är ett drastiskt exempel på riskerna förknippade med att behandla patienter på basis av »mekanismer« eller på vad som förefaller vara logiskt rimligt.

- *Minskad bentäthet är väl korrelerad till risk för frakturer.* Det är därför logiskt att anta att läkemedel som ökar bentätheten minskar frakturrisken. En studie vid Mayo-kliniken visade att natriumfluorid ger en signifikant ökning av bentätheten. Uppmuntrade av dessa resultat startade forskarna en klinisk prövning avseende postmenopausala kvinnor med nedsatt bentäthet. Den nya studien bekräftade att bentätheten ökade, men visade överraskande en signifikant ökad risk för frakturer [10]. Natriumfluorid gav uppbyggnad av ett sköra skelett, som lättare frakturerade.

Konklusion

En fördel med tillförlitliga surrogatmått är att de kan bidra till att effektiva behandlingar snabbare och till lägre kostnader kan komma patienterna till nytta. Exempel på detta finns framför allt inom HIV-området. Vid utveckling av nya behandlingar kan surrogatmått i tidiga studier ge vägledning om huruvida det är värt att gå vidare eller inte. För att verkligen dokumentera nytta krävs dock att patientrelaterade effektmått studeras senare under produktutvecklingen. Det finns anledning att vara skeptisk till läkemedel som långt efter introduktionen på marknaden inte har någon dokumenterad klinisk nytta utan bara har visats påverka laboratorievärden eller fysikaliska tecken.

Ett av problemen med surrogatmått är att de flesta inte är validerade, dvs det är inte visat att en förändring av markören förutsäger en viss klinisk effekt. Ett större problem är att surrogatmättet inte heller ger information om eventuella risker med en behandling. Surrogatmättet kan därför bara ge en del av den information som behövs för att bedöma balansen mellan nytta och risk med en intervention.

Referenser

1. Temple RJ. A regulatory authority's opinion about surrogate endpoints. In: Nimmo WS, Tucker GT, editors. Clinical measurement in drug evaluation. New York: J Wiley; 1995.
2. Behandling av astma och KOL. En systematisk kunskapssamman-

ställning. Stockholm: SBU, Statens beredning för medicinsk utvärdering; 2000. Rapport nr 151.

3. Sullivan M, Tunsäter A. Hälsorelaterad livskvalitet informativt effektmått i kliniska prövningar. Tillämpningsexempel: astma/kroniskt obstruktiv lungsjukdom. Läkartidningen 2001;98:4428-33.
4. WHO cooperative trial on primary prevention of ischemic heart disease using clofibrate to lower serum cholesterol: mortality follow-up. Report of the Committee of Principal Investigators. Lancet 1980; ii:379-85.
5. Collins R. Heart Protection Study. Föredrag vid 74th annual scientific sessions of the American Heart Association, Anaheim, California November 13 2001.
6. The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major cardiovascular events in hypertensive patients randomised to doxazosin vs chlorthalidone. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. JAMA 2000;283:1967-75.
7. Fleming TR. Surrogate markers in aids and cancer trials. Stat Med 1994;13:1423-35.
8. Carosi G, Castelli F, Suter F, Maggiolo F, Orani AM, Pan A, et al. Antiviral potency of HAART regimens and clinical success are not strictly coupled in real life conditions: Evidence from the MASTER-1 study. HIV Clinical Trials. 2001;2:399-407.
9. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, et al. Mortality and morbidity in patients receiving encainide, flecainide or placebo. N Engl J Med 1991;324:781-8.
10. Riggs BL, Hodgson SF, O'Fallon WM, Chao EY, Wahner HW, Muhs JM, et al. Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. N Engl J Med 1990;322:802-9.

SUMMARY

Surrogate markers –
easy to measure but irrelevant?

Bengt Furberg

Läkartidningen 2002;99:1672-5

Many treatments aim at making a patient feel better due to symptom relief or improved quality of life. Other treatments are given in order to prevent future complications, non-fatal or fatal. There are methodological difficulties in assessing subjective symptoms and quality of life. To document reduced risk of complications takes large, long-term, costly clinical trials. Thus, the use of a surrogate marker to document benefit may seem appealing. A well-validated surrogate marker can demonstrate benefit of a given treatment with fewer patients in a shorter time and at considerably lower cost. It is much easier to show in a 12-week study that a new compound lowers blood pressure than to show that it lowers the risk of stroke, myocardial infarction or renal impairment. One problem with surrogate markers is that very few are well-validated, another is that even those which are cannot be expected to reflect adverse effects, expected or unexpected, and are thus of limited value in the global assessment of benefit vs. risk. It is recommended that results obtained with surrogate markers be regarded as preliminary until results of large trials with clinically relevant outcomes are available.

Correspondence: Bengt Furberg, Krassevågen 7, SE-434 46 Kungälv, Sweden (bengt.furberg@artimplant.se)