

OmniStat-OmniRisk

Ett par enkla datorprogram för att lära sig grunderna i att

1. utvärdera diagnostiska metoders prestanda och testresultat

OCH

2. jämföra risker/händelser/komplikationer mellan grupper

Utarbetat av

Ulf Nyman, docent,

ulf.nyman@skane.se

Mattias Kristiansson, underläkare

mattias.kristiansson@gmail.com

Röntgenavdelningen, Lasarettet Trelleborg

OmniStat

För utvärdering av diagnostiska metoders prestanda och testresultat

Avsikten med detta program är att ge den verksamme läkaren ett enkelt instrument att snabbt kunna värdera sannolikheten för att en patient har/inte har en misstänkt sjukdom vid ett positivt respektive negativt utfall av en diagnostisk metod. Denna sannolikhet baseras på metodens *sensitivitet* och *specificitet* SAMT skattning av den *kliniska sannolikheten (pre-test eller prior probability)* för sjukdom hos den enskilde patienten eller *prevalens* av sjukdomen i den aktuella populationen, vilka tillsammans utgör det s.k. *testresultatet (post-test eller revised probability)*. Dyliga diagnostiska sannolikhetsanalyser brukar kallas Bayes-analys efter den engelske prästen och matematikern Thomas Baye's (död 1761) texter om sannolikheter, "Essays towards solving a probability in the doctrine of chances". Praktiska exempel i denna information är hämtade från diagnostik av venös tromboembolism (VTE).

Instruktioner

I övre halvan av programmet lägger man in värden från en aktuell studie i de fyra fälten motsvarande a, b, c och d i Tabell 1. Programmet räknar sedan ut prevalens, sensitivitet, specificitet, positivt/negativt negativt prediktionsvärde (PPV/NPV), likelihood ratio (LR) och nedre/övre gräns för ett 95 % konfidensintervall (KI). Under "frågetecknen" finns förklaringar till de olika begreppen.

I den undre halvan kan man själv skriva in valfria värden för prevalens, sensitivitet och specificitet varefter programmet räknar ut PPV, NPV och LR.

Sensitivitet och specificitet

En diagnostisk metods prestanda anges med dess *sensitivitet* och *specificitet*, dvs. metodens känslighet att diagnostisera en specifik sjukdom hos dem som har sjukdomen respektive utesluta sjukdomen hos friska individer. Sensitivitet påverkas av sjukdomens allvarlighetsgrad eller utbredning medan specificiteten endast påverkas av egenskaper hos de friska individerna. Sensitiviteten är t.ex. högre för ultraljud vid diagnostik av utbredda symtomatiska tromboser jämfört med små asymtomatiska. Sensitivitet och specificitet är däremot oberoende av sjukdomens prevalens. Uträkning av

sensitivitet/specifitet förutsätter att man jämfört metodens resultat med någon standard eller referensmetod. Resultatet av en sådan jämförelse brukar sättas upp i en fyrfältstabell enligt följande:

Tabell 1

		Referensmetod	
		Sjuk	Frisk
Diagnostisk metod	Positiv	a (n=255) sant positiva	b (n=35) falskt positiva
	Negativ	c (n=45) falskt negativa	d (n=665) sant negativa
		a/a+c	d/b+d
		Sensitivitet	Specifitet

Sensitiviteten anger sannolikheten för ett positivt utfall av den diagnostiska metoden hos dem som har sjukdomen, dvs. $a/a+c$; $255/300 = 0,85$ (85 %) i Tabell 1.

Specifiteten är sannolikheten för ett negativt utfall om man är frisk, dvs. $d/b+d$; $665/700 = 0,95$ (95 %) i Tabell 1.

Prevalens

Prevalensen beskriver förekomst/frekvens av sjukdomen i en viss population. Blanda inte ihop prevalens med incidens, som anger antalet individer som insjuknar i en sjukdom i relation till antalet invånare och år. När man skattar den kliniska sannolikheten för sjukdom hos en enskild patient, brukar begreppet prevalens ersättas med de engelska begreppen "pre-test" eller "prior probability".

Tabell 2

		Referensmetod		
		Sjuk	Frisk	
Diagnostisk metod	Positiv	a (n=255) sant positiva	b (n=35) falskt positiva	a+b (n=290)
	Negativ	c (n=45) falskt negativa	d (n=665) sant negativa	c+d (n=710)
		a+c (n=300)	b+d (n=700)	a+c/a+b+c+d Prevalens

Prevalensen i Tabell 2 är antalet patienter med sjukdom dividerat med totala antalet patienter som ingick i studien, dvs. $a+c/a+b+c+d$; $300/1000 = 0,3$ (30 %).

I vetenskapliga studier har man funnit, att när den kliniska sannolikheten för lungemboli (LE) eller djup ventrombos (DVT) skattats som hög baserad på antingen empiri (symtom, status, riskfaktorer, etc.) eller poängbaserade system, är sannolikheten (pre-test probability) för att blodpropp föreligger 60-80 % medan om skattningen är låg reduceras sannolikheten till cirka 10 %. Dessa variationer i klinisk sannolikhet/prevalens påverkar i hög grad utfallet av den diagnostiska metodens prediktionsvärden.

Testresultat och prediktionsvärden

I den kliniska situationen är vi framför allt intresserade av att veta sannolikheten för att sjukdom föreligger när en diagnostisk metod är positiv, s.k. *positivt prediktionsvärde (PPV)*, eller för att vara frisk om utfallet är negativt, s.k. *negativt prediktionsvärde (NPV)*. Dessa prediktionsvärden är beroende av metodens sensitivitet, specificitet och sannolikheten för sjukdomen (pre-test probability). Dessa tre parametrar ger det s.k. testresultatet = engelskans ”post-test” eller ”posterior probability”, dvs. det sammanlagda resultatet av klinisk sannolikhetsbedömning och metodens prestanda.

Tabell 3

		Referensmetod		
		Sjuk	Frisk	
Diagnostisk metod	Positiv	a (n=255) sant positiva	b (n=35) falskt positiva	a/a+b PPV
	Negativ	c (n=45) falskt negativa	d (n=665) sant negativa	d/c+d NPV

PPV är sannolikheten för att vara sjuk om den diagnostiska metoden utfallet positivt, dvs. $a/a+b$; $255/290 = 0,88$ (88 %).

NPV är sannolikheten för att vara frisk om metoden utfallet negativt, dvs. $d/c+d$; $665/710 = 0,94$ (94 %).

Om den kliniska sannolikheten/prevalensen är 10 % (Tabell 4) respektive 70 % (Tabell 5) blir prediktionsvärdena helt förändrade jfr med Tabell 3. Notera hur värdena i de olika fälten ändras jfr med Tabell 3 beroende av ändrad klinisk sannolikhet medan sensitivitet och specificitet är oförändrade.

Tabell 4; prevalens 10 %

		Referensmetod		
		Sjuk	Frisk	
Diagnostisk metod	Positiv	85	45	85/(85+45) PPV 65%
	Negativ	15	855	855/(15+855) NPV 98%
	Summa	100	900	1000
		85/100	855/900	100/1000
		Sensitivitet	Specificitet	Prevalens
		85%	95%	10%

Tabell 5; prevalens 70 %

		Referensmetod		
		Sjuk	Frisk	
Diagnostisk metod	Positiv	595	15	595/(595+15) PPV 98%
	Negativ	105	285	285/(105+285) NPV 73%
	Summa	700	300	1000
		595/700	285/300	700/1000
		Sensitivitet	Specificitet	Prevalens
		85%	95%	70%

Exemplen illustrerar att för samma siffror för sensitivitet och specificitet vid t.ex. datortomografisk diagnostik av lungembolism (LE) leder situationen i Tabell 4 till att vid låg klinisk sannolikhet är risken för överdiagnostik överhängande (PPV endast 65 %). Diagnosen måste alltså grundas på otvetydiga tecken på LE eller så måste utredning kompletteras med andra metoder för att säkert påvisa/utesluta LE. Vid hög klinisk sannolikhet (Tabell 5) kan man inte med säkerhet utesluta LE vid en negativ DTLA (NPV endast 73 %). I denna situation kan det verka logiskt att gå vidare med ytterligare utredning för att påvisa/utesluta LE, t.ex. bilateral undersökning av nedre extremiteternas vener eller skintigrafi. Å andra sidan kan man anta att de missade embolierna är små och kanske betydelselösa. För att testa den sistnämnda tesen kan man utföra s.k. ”managementstudier” där man lämnar patienter med negativt testutfall utan behandling under en viss uppföljningstid, t.ex. tre månader, för att studera frekvensen av tromboemboliska episoder under denna period. En stor sådan uppföljningsstudie är gjord i Holland och visade att patienter med klinisk sannolik LE eller positiv D-

dimer (30% prevalens för LE i denna grupp) men negativ DTLA löpte endast 1,5 % risk att drabbas av tromboembolism under tre månaders uppföljning om de lämnades utan antikoagulantia. Det är alltså viktigt att kunna värdera/veta den kliniska betydelsen av ett falskt negativt test utfall, så att man inte lägger på ytterligare diagnostiska undersökningar i onödan.

Att räkna med odds

Odds (oddstal) innebär att man jämför sannolikheten för ett utfall med ett annat genom att bilda en kvot mellan sannolikheterna (antalet fall av en "händelse" dividerat med antalet fall av en "icke-händelse"). Sannolikheten (p) för hjärterkort (H) i en kortlek är $1/4$ (0,25) och för övriga kort $3/4$ (0,75). Oddset (chansen) för att *inte* få ett hjärterkort vid utdelning är då sannolikheten för att *inte* få ett hjärterkort (0,75) dividerat med sannolikheten att få ett hjärterkort (0,25 eller $1-0,75$), vilket blir $0,75/0,25$, dvs. 3 mot 1. Chansen att få ett icke-hjärterkort är alltså 3 gånger så stor som att få ett hjärterkort.

Likelihood ratio

I engelsk litteratur förekommer begrepp "likelihood ratio" (LR), som inte har fått något riktigt bra svenskt namn (sannolikhetskvot?). LR är ett oddsbegrepp som beskriver chansen att vara sjuk istället för frisk vid ett positivt respektive negativt utfall av en diagnostisk metod. Den ger ett enkelt och snabbt kvantitativt mått på hur värdefull metoden är relativt den kliniska sannolikheten. För att illustrera LR använder vi oss av Tabell 3.

Positiv likelihood ratio (LR+) erhålls genom att dividera sannolikheten för ett positivt testutfall om man har sjukdomen ($a/a+c$; $255/300 = 0,85 = \text{sensitivitet}$) med sannolikheten för ett positivt utfall om man inte har sjukdomen ($b/b+d$; $35/700 = 0,05 = 1-\text{specificitet} = 1-0,95$). LR+ ($\text{sens}/1-\text{spec}$) blir alltså $0,85/0,05 = 17$. Chansen för ett positivt test är alltså 17 gånger så hög om man är sjuk jämfört om man är frisk.

Negativ likelihood ratio (LR-) erhålls genom att dividera sannolikheten för ett negativt testutfall om man är sjuk ($c/a+c$; $45/300 = 0,15 = 1-\text{sensitivitet} = 1-0,85$) med sannolikheten för ett negativt utfall om man är frisk ($d/b+d$; $665/700 = 0,95 = \text{specificitet}$). LR- ($1-\text{sens}/\text{spec}$) blir alltså 0,16. Chansen för ett negativt test för den som bär sjukdomen är alltså drygt 6 gånger ($1/0,16$) mindre jämfört med om man är frisk.

Storleken på LR talar om i vilken grad den diagnostiska metoden kommer att öka (vid LR+) eller minska (vid LR-) den kliniska sannolikheten av den misstänkta sjukdomen enligt följande:

Hög grad:	LR+ >10;	LR- <0,1
Måttlig grad:	LR+ >5-10;	LR- 0,1<0,2
Ringa grad:	LR+ >2-5;	LR- 0,2<0,5
Obetydligt:	LR+ >1-2;	LR- 0,5<1
Ingen påverkan	LR+ 1;	LR- 1

Diagnostisk odds ratio (DOR; oddskvot) kan bildas genom LR+/LR- ($a \times d / c \times b$) varvid man får ett enda värde för bedöma den diagnostiska metodens prestanda.

Konfidensintervall

Ett intervall inom vilket en populationsparameter (t.ex. ett medelvärde eller en proportion) ligger med en viss sannolikhet. Sannolikheten anges ofta med 95 % säkerhet och är beroende av antalet individer i populationen. Ju fler individer desto snävare blir intervallet.

Exempel: 292 patienter med en negativ datortomografi på misstanke lungembolism följs under 3 månader utan antikoagulantia. Man studerar frekvensen tromboemboliska episoder under denna tid som ett mått på att datortomografen faktiskt var falskt negativ vid undersökningstillfället. Under denna tid diagnostiserade en episod av venös tromboembolism hos 5 patienter, dvs. 1,7 %. Med 95 % säkerhet ligger denna proportion inom intervallet 0,7-3,9 %. Sannolikheten är alltså mindre än 2,5 % att fler än 3,9 % av de undersökta patienterna hade en missad lungemboli vid datortomografiundersökningen. Oftast anges 4 % som en acceptabel övre gräns för det 95 %-iga konfidensintervallet i diagnostiska uppföljningsstudier avseende venös tromboembolism.

Beräkning av konfidensintervallet i OmniStat bygger på att binomialfördelningen approximeras med normalfördelningen, vilket innebär att man måste ha minst 5 individer i varje inmatningsruta för att konfidensintervallet skall vara giltigt. Har man färre individer så får man osäkra värden varför det kan krävas beräkning med mer exakta och komplexa metoder.

Metodologiska kvalitetskriterier enligt QUADA

Vid läsning av vetenskapliga studier där diagnostiska metoder utvärderats måste man vara medveten om de metodologiska problem som kan föreligga för att kunna avgöra om studiens resultat är relevanta för den egna situationen. Här följer exempel på några viktiga kvalitetskriterier som en vetenskaplig studie bör uppfylla för att man skall kunna "lita" på resultaten:

- 1) Studien bör vara prospektiv med ett rimligt spektrum av konsekutiva patienter som misstänks ha sjukdomen. Demografi avseende ålder, kön, symtomatologi och komorbiditet bör vara angiven både för inkluderade och exkluderade patienter eftersom testets utfall kan variera med populationens egenskaper
- 2) Definierade inklusions- och exklusionskriterier.
- 3) Den diagnostiska metod som skall utvärderas måste jämföras med en acceptabel referensmetod som t.ex. obduktion, kirurgi, angiografi, klinisk uppföljning ("management-studie") eller en kombination av metoder.
- 4) Tidsintervallet mellan test- och referensmetodens utförande måste vara kort nog så att patientens tillstånd inte hinner ändra sig mellan undersökningarna.
- 5) Test och referensmetod skall vara utförda oberoende av varandra, dvs. utfallet av den ena metoden får inte påverka om den andra metoden skall utföras eller inte, t.ex. lungangiografi (facit) utförs ej om lungskintigrafin (testmetod) är normal ("partial verification bias").
- 6) Samma referensmetod bör användas oberoende av testresultatet ("differential verification bias")
- 7) Testmetoden bör inte vara del av referensmetoden ("incorporation bias").
- 8) Adekvat beskrivning av hur test- och referensmetod utfördes liksom diagnostiska kriterier så att det kan översättas till den egna verksamheten.
- 9) Bedömningen av test- och referensmetod måste utföras "blindat" så att bedömningen av testmetoden inte påverkas om man känner resultatet av referensmetoden eller tvärtom.
- 10) Samma kliniska data bör finnas tillgängliga då testmetoden utförs, som skulle finnas tillgängliga när testmetoden används i sitt kliniska sammanhang.
- 11) Inkonklusiva/intermediära resultat bör vara rapporterade.
- 12) Sensitivitet och specificitet skall vara angiven eller kunna räknas ut.
- 13) Överrensstämmelse mellan olika bedömare ("interobserver agreement") och av samma bedömare vid olika tillfällen ("intraobserver agreement") bör också vara utvärderad liksom testmetodens utfall om den upprepas. Det sistnämnda kravet kan vara omöjligt att uppfylla för radiologiska metoder.

OmniRisk

Att jämföra risker/händelser/komplikationer mellan grupper

Riskbedömning

Risk anger sannolikheten för viss händelse av negativ karaktär. I en medicinsk studie är risktalet antalet personer som drabbades av händelsen dividerat med totala antalet personer i gruppen. Termen risk kan ersättas av incidens som har fördelen att vara neutralt, dvs. kunna användas om både "positiva" och "negativa" händelser. En jämförelse av risker/komplikationer kan t.ex. göras mellan två behandlingsmetoder eller diagnostiska undersökningar. Beräkning av olika risktal kommer att ske genom illustration av risken för nefrotoxisk biverkan av två olika kontrastmedel i samband med coronarangiografi (Aspelin et al.). Sjutton av 65 patienter som erhöll kontrastmedlet Omnipaque utvecklade en njurpåverkan, men enbart 2 av 64 patienter som injicerades med Visipaque. Patienterna som erhöll Omnipaque bildar kontrollgruppen och de övriga testgruppen.

Risk för nefropati av Omnipaque blir 26 % (17/65) och av Visipaque 3 % (2/64).

Relativ risk (=riskkvot eller risk ratio) utgör kvoten mellan risktalen hos två undersökta grupper och blir i detta fall 0,12, dvs. risken för nefropati av Visipaque är endast 12 % av risken för nefropati av Omnipaque (Riskkvoten är inte identisk med oddskvot men de båda kvoterna skiljer sig inte mycket om riskerna är låga).

Odds kvot (=odds ratio) anger kvoten mellan två odds (sannolikheten av en "händelse" dividerat med sannolikheten av en "icke-händelse"), dvs. oddset av en "händelse" i testgruppen (Visipaque: 2 med nefropati dividerat med 62 (64 minus 2) med icke-nefropati = $2/62 = 0,032$) dividerat med motsvarande odds i kontrollgruppen (Omnipaque: 17 nefropati dividerat med 48 (65 minus 17) icke-nefropati = $17/48 = 0,35$) och blir detta fall 0,09 (= $0,032/0,35$), dvs. chansen för nefropati med Omnipaque är 11 (= $1/0,09$) gånger högre jfr med Visipaque. Oddskvot kan användas som (ungefärlig) ersättning för relativ risk.

Relativ riskförändring är lika med riskdifferensen i absoluta tal (utan tecken) mellan test- och kontrollgrupp som andel av risktalet i kontrollgruppen dvs. $(26 \% - 3 \%) / 26 \% = 0,88$.

Absolut riskförändring är den differensen i absoluta tal (utan tecken) mellan risktalen i test- och kontrollgrupp, i detta exempel $26 \% - 3 \% = 23 \%$.

”Number needed to treat” är antal personer som behöver behandlas för att en av dem sannolikt ska dra nytta av den gynnsammare effekten i test- eller kontrollgrupp. ”Number needed to treat” är det inverterade värdet av den absoluta riskförändringen, i detta exempel $1/0,23 = 4,3$; dvs. minst var 5:e patient som erhåller Visipaque enligt denna studie kommer att dra nytta av detta kontrastmedelsval.

Referenser

- 1) Taube A, Malmquist J. Räkna med vad du tror. Bayes' sats i diagnostiken. *Läkartidningen* 2001;98:2910-2913.
- 2) SBU:s ordlista. Förklaringar av termer för utvärdering av medicinska metoder. www.sbu.se.
- 3) Statens beredning för medicinsk utvärdering. Blodpropp – förebyggande, diagnostik och behandling av venös trombembolism. Utvärdering av diagnostiska metoder. SBU-rapport 158, kapitel 3.1. www.sbu.se, <http://www.arwen.se/radiologi/vte.html>.
- 4) van Belle and Christopher Study Investigators. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. *JAMA* 2006;295:172-179.
- 5) Ejlertsson G. Statistik för hälsovetenskaparna. Studentlitteratur, Lund, 2003.
- 6) Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III How to use an article about a diagnostic test. A. Are the results valid. *JAMA* 1994;271:389-391.
- 7) Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients. *JAMA* 1994;271:703-707.
- 8) Bland JM, Altman DG. Statistics notes. The odds ratio. *BMJ* 2000;320:1468.
- 9) Weinstein S, Obuchowski NA, Sieber ML. Fundamentals of clinical research for radiologists. Clinical evaluation of a diagnostic test. *AJR* 2005;184:14-19.
- 10) Whiting et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*. 2003;25:1-13.
- 11) Whiting et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health technology Assessment* 2004, vol.8, No 25.
<http://www.nchta.org/project/htapubs.asp>
<http://www.nchta.org/fullmono/mon825.pdf>
- 12) Bossuyt et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Clinical Chemistry* 2003;49:1-6.
- 13) Bossuyt et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and Elaboration. *Clinical Chemistry* 2003;49:7-18.
- 14) Bossuyt et al. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *AJR Am J Roentgenol*. 2003;181:51-56.
- 15) Aspelin et al. Nephrotoxic effects in high-risk patients undergoing angiography. *NEJM* 2003;348:491-499.