

## **Índices de eficacia de un tratamiento. NNT**

Joaquín Primo  
Unidad de Digestivo, Hospital de Sagunto

Ponencia presentada en el Symposium "Gestión del conocimiento y su aplicación en la Enfermedad Inflamatoria Crónica Intestinal", organizado por GETECCU (Grupo Español de Trabajo en Enfermedad de Crohn y Colitis Ulcerosa). Valencia, 24 de enero de 2003.

El efecto de un tratamiento, representado por una variable binaria y evaluado mediante un ensayo clínico aleatorizado, puede expresarse mediante distintos índices (1-3). Aunque todos ellos comparan la probabilidad, o riesgo, de un determinado evento en el grupo experimental con el riesgo en el grupo control, la forma de expresar la eficacia de un tratamiento puede influir en el grado de aceptación por parte de los médicos y los pacientes (4-8). Utilizaremos como ejemplo para calcular los distintos índices el estudio de Present *et al.* (9), en el que se consiguió el cierre de la mitad o más de las fístulas en el 68% (21/31) de los pacientes con enfermedad de Crohn tratados con infliximab (5 mg/kg iv 0-2-6 semanas) y en el 26% (8/31) a los que se les administró placebo.

### Riesgo relativo (RR)

Es el cociente entre el riesgo del grupo experimental y el riesgo del grupo control. En el ejemplo,  $0,68/0,26 = 2,62$ . Indica que el infliximab es casi tres veces más eficaz que el placebo.

Cuando el riesgo en ambos grupos es similar, el RR es 1. Si el riesgo del grupo experimental es superior al del grupo control, el RR es  $>1$  (si se evalúa un evento deseado significa que el tratamiento es mejor que el control, y si el evento es algo no deseado significa que el tratamiento es peor que el control). Por el contrario, si el riesgo del grupo experimental es inferior al del grupo control, el RR es  $<1$  (si se evalúa un evento deseado significa que el tratamiento es peor que el control, y si el evento es algo no deseado significa que el tratamiento es mejor que el control).

### Reducción relativa del riesgo (RRR)

Es el cociente entre la diferencia de riesgo entre el grupo experimental y el control y el riesgo del grupo control. En el ejemplo,  $(0,68-0,26)/0,26 = 1,62$  ó 162%. Puede calcularse también como  $RR-1$ ,  $2,62-1 = 1,62$  ó 162%. Indica que el infliximab aumentó un 162% la respuesta obtenida con el placebo.

En este ejemplo, en el que el tratamiento experimental aumenta la probabilidad de un evento deseado, podría hablarse de incremento relativo del beneficio. Por ello, en términos más amplios, a la RRR puede denominarse también diferencia relativa del riesgo (DRR), pues el tratamiento experimental puede tanto reducir como aumentar la tasa de un evento deseado (reducción relativa del beneficio e incremento relativo del beneficio, respectivamente) o no deseado (reducción relativa del riesgo e incremento relativo del riesgo, respectivamente).

La principal desventaja del RR y de la RRR es que no reflejan la magnitud del efecto del tratamiento en una escala absoluta y no permiten discriminar entre los efectos importantes del tratamiento de los que no lo son (3,10). Así, si la respuesta con infliximab hubiera sido del

6,8% y con placebo 2,6%, el RR y la RRR tendrían los mismos valores que en el ejemplo anterior.  $RR = 0,068/0,026 = 2,62$ , y  $RRR = (0,068-0,026)/0,026 = 1,62$  ó 162%.

### Odds ratio (OR)

El término *odds*, para el que no hay una traducción española comúnmente aceptada, es el cociente entre la probabilidad de que un evento ocurra y la probabilidad de que ese mismo evento no ocurra. Habitualmente se estima por el cociente entre el número de veces que ocurre el evento y el número de veces que no ocurre. En el ejemplo, la *odds* del grupo experimental es  $21/10$  (ó  $0,68/0,32$ ) = 2,1, y la *odds* del grupo control es  $8/23$  (ó  $0,26/0,74$ ) = 0,35. La OR es el cociente entre la *odds* del grupo experimental y la *odds* del grupo control,  $2,1/0,35 = 6$ .

Cuando la probabilidad de que ocurra un evento es similar en ambos grupos, la OR es 1. Si la probabilidad es mayor en el grupo experimental, la OR es  $>1$  (si se evalúa un evento deseado significa que el tratamiento es mejor que el control, y si el evento es algo no deseado significa que el tratamiento es peor que el control), y si la probabilidad es mayor en el grupo control, la OR es  $<1$  (si se evalúa un evento deseado significa que el tratamiento es peor que el control, y si el evento es algo no deseado significa que el tratamiento es mejor que el control).

Aunque poco intuitivo, la OR es uno de los índices más utilizados debido a que puede estimarse también en los estudios de casos y controles, en los que no puede calcularse el RR (por el tipo de diseño de este tipo de estudios no es posible calcular la incidencia entre los expuestos y los no expuestos), y porque mediante regresión logística permite ajustar la relación en estudio por el efecto de otras variables (2).

### Reducción absoluta del riesgo (RAR)

Es la diferencia entre el riesgo del grupo experimental y el riesgo del grupo control. En el ejemplo,  $0,68-0,26 = 0,42$  ó 42%.

En este ejemplo, en el que el tratamiento experimental aumenta la probabilidad de un evento deseado, podría hablarse de incremento absoluto del beneficio. Por ello, en términos más amplios, a la RAR puede denominarse también diferencia absoluta del riesgo (DAR), pues el tratamiento experimental puede tanto reducir como aumentar la tasa de un evento deseado (reducción absoluta del beneficio e incremento absoluto del beneficio, respectivamente) o no deseado (reducción absoluta del riesgo e incremento absoluto del riesgo, respectivamente).

Aunque la RAR informe menos que el RR y la RRR sobre los efectos proporcionales, diferencia claramente entre las probabilidades en ambos grupos e informa mejor de su posible

significación clínica (3,10). Así, por ejemplo, si la respuesta con infliximab hubiera sido del 6,8% y del placebo 2,6%, la RAR sería  $0,068 - 0,026 = 0,042$  ó 4,2%.

#### Número de pacientes necesarios para tratar (NNT)

El término fue introducido por Laupacis *et al.* en 1988 (11), e indica el número de pacientes que se necesitarían tratar con el tratamiento experimental para conseguir un evento adicional a los que se conseguirían con el tratamiento control. Se calcula fácilmente, ya que es el inverso de la RAR. En el ejemplo,  $1/0,42$  ó  $100/42 = 2,38 \approx 3$  (por convenio, habitualmente se redondea al número entero superior).

Es importante subrayar la importancia de la palabra adicional en la definición. El NNT no indica la probabilidad absoluta de que se produzca un evento. Si fuera entendido así, un NNT de 3 sería erróneamente interpretado como que respondería un paciente de cada 3 tratados con infliximab (un 33%, ¡bastante inferior al 68% de respuesta obtenida en el estudio!). En el ejemplo, la RAR es del 42%, y significa que por cada 100 pacientes que sean tratados con infliximab se conseguirá el cierre de las fístulas en 42 pacientes más que con placebo. Por tanto, mediante una simple regla de tres se puede comprobar que para conseguir que respondiera un paciente más con infliximab que con placebo (un evento adicional) se necesitarán tratar a  $(1 \times 100)/42 = 2,38$  pacientes o, de forma equivalente,  $NNT = 1/0,42 = 2,38 \approx 3$  pacientes.

Cuanto mayor sea la diferencia de efecto entre el tratamiento y el control, menor será el NNT, y al contrario. Así, en caso de que la RAR hubiera sido del 4,2%, el NNT aumentaría a  $1/0,042 = 23,81 \approx 24$  pacientes. Cuando el efecto del tratamiento experimental es similar al del tratamiento control, la RAR es 0 y, en consecuencia, el valor del NNT es infinito ( $\infty$ ). El NNT expresa el efecto de un tratamiento en unos términos muy intuitivos, permitiendo fácilmente comparar sus ventajas con sus inconvenientes (efectos adversos, costes) e indicando el esfuerzo que se requiere para conseguir un determinado efecto terapéutico, de forma muy cercana a la toma de decisiones clínicas (2,3,10-12).

Una ventaja importante del NNT es que permite estimar el beneficio de un tratamiento en un paciente individual (3,12). Dado que  $RRR = RAR/RGC$ ,  $RAR = RRR \times RGC$  y, por tanto,  $NNT = 1/(RRR \times RGC)$ , siendo RGC el riesgo en el grupo control. Generalizando, el RGC se corresponde con la probabilidad esperada de que se produzca el evento en un paciente si no fuera tratado (PEEP). Por tanto, la fórmula puede expresarse como  $NNT = 1/(RRR \times PEEP)$ . El beneficio esperado al aplicar un tratamiento con una determinada RRR a un paciente concreto va a depender del riesgo basal del mismo (PEEP). Si consideramos que nuestro paciente es como el paciente "medio" del estudio en que se valora el efecto del tratamiento, la PEEP puede considerarse igual que la tasa del evento en el grupo control del estudio. Si en el estudio hay

algún subgrupo de pacientes con características similares al nuestro, la PEEP puede ser igual que la tasa del evento en ese subgrupo. Por ejemplo, si nuestro paciente fuera similar a los incluidos en el grupo control del estudio de Present *et al.* (9), PEEP = 26%. Así,  $NNT = 1/(1,62 \times 0,26) = 2,38 \approx 3$ . Sin embargo, si consideráramos que la probabilidad de respuesta sin tratamiento de nuestro paciente es menor, por ejemplo del 10%,  $NNT = 1/(1,62 \times 0,10) = 6,17 \approx 7$ . El nomograma de Chatellier *et al.* (13) facilita estos cálculos (fig. 1). De forma más sencilla, el NNT individualizado también puede estimarse mediante la fórmula  $NNT = NNT \text{ del estudio} / RC$ , siendo RC el riesgo comparado de nuestro paciente con el de los pacientes incluidos en el estudio (3,12). Así, si el riesgo de nuestro paciente se considera similar al del estudio ( $RC = 1$ ),  $NNT = 3/1 = 3$ ; si nuestro paciente tuviera el doble de riesgo ( $RC = 2$ ),  $NNT = 3/2 \approx 2$ ; y si tuviera la mitad de riesgo ( $R = 0,5$ ),  $NNT = 3/0,5 = 6$ . Estos cálculos asumen que el efecto del tratamiento es independiente del riesgo basal de los pacientes, hecho que no siempre es cierto.

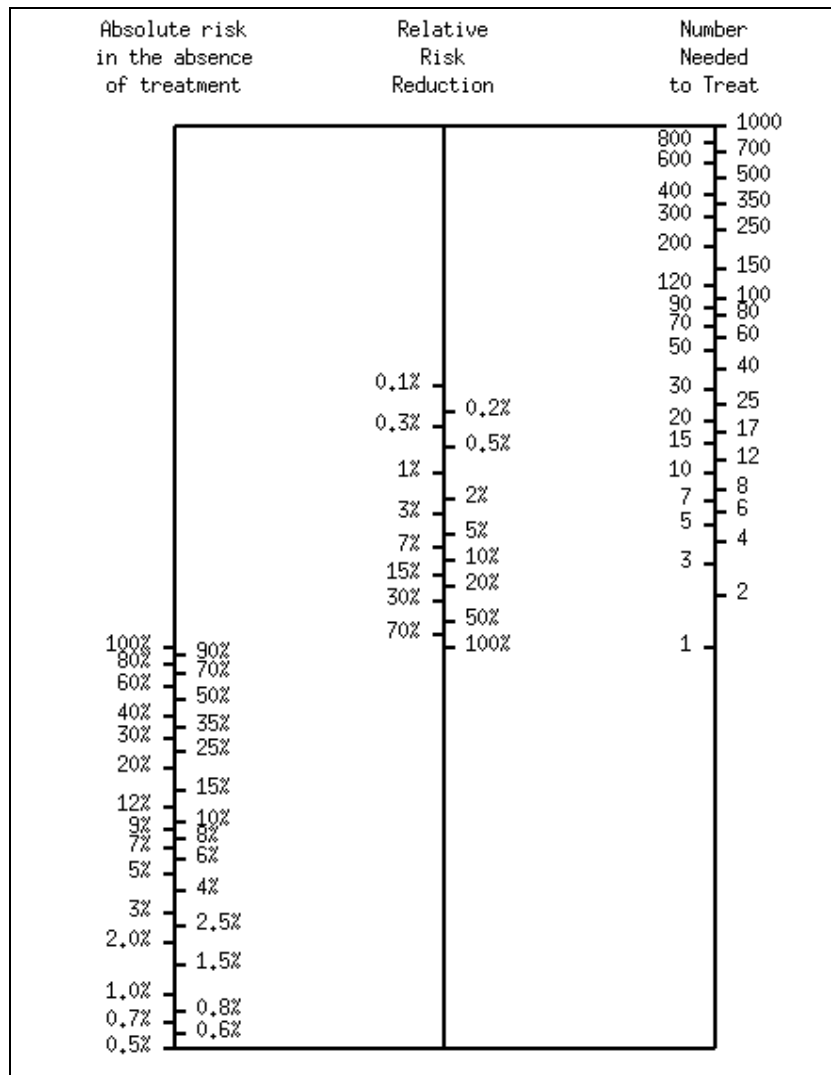


Figura 1. Nomograma de Chatellier *et al.* para el cálculo del NNT (13)

Una característica a tener en cuenta de los NNT es que van asociados al período de seguimiento del ensayo clínico del que derivan. Por ejemplo, en el estudio de Candy *et al.* (14) se mantuvieron en remisión durante un año el 56% de los pacientes con enfermedad de Crohn tratados con azatioprina a dosis de 2,5 mg/kg y el 10% del grupo placebo. En este caso, deberíamos tratar a 3 pacientes (IC 95%: 2 a 5) con azatioprina para mantener una remisión adicional durante un año. En cambio, en el estudio de Hawkey *et al.* (15) el 60% de los pacientes con colitis ulcerosa se mantuvieron en remisión tratados con mesalazina 400 mg/6 h durante 6 meses y el 41% de los que recibieron placebo. En este caso, deberíamos tratar a 6 pacientes (IC 95%: 4 a 18) con mesalazina para mantener una remisión adicional durante seis meses. Una consecuencia de esta asociación es que para comparar los NNT para diferentes períodos de seguimiento es necesario ajustar uno de ellos. Para entender cómo se realiza este ajuste seguiremos el ejemplo expuesto en el libro de Sackett *et al.* (3). El NNT para prevenir un accidente cerebrovascular, un infarto de miocardio o un fallecimiento en hipertensos graves es de 13 a los 1,5 años, mientras que en hipertensos leves es de 128 a los 5,5 años. Asumiendo que la RRR se mantiene constante a lo largo del tiempo, para estimar el NNT a los 1,5 años en hipertensos leves aplicaríamos la siguiente fórmula:  $NNT_{hipotético} = NNT_{observado} \times (\text{Tiempo}_{observado} / \text{Tiempo}_{hipotético})$ . Así,  $NNT_{1,5 años} = 128(5,5/1,5) = 470$ . De este modo puede apreciarse la gran diferencia en el rendimiento de los esfuerzos clínicos para tratar a los hipertensos leves frente a los graves: necesitamos tratar 470 hipertensos leves, pero sólo 13 graves, para prevenir un resultado nocivo adicional en 1,5 años.

Como cualquier otro estimador, el NNT debe expresarse siempre con su intervalo de confianza (IC), habitualmente del 95%, que se calcula fácilmente obteniendo los valores inversos de los límites del IC de la RAR (fig. 2) (3). Cuanto más pequeño sea el número de pacientes incluidos en el estudio que generó el NNT, más amplio será su IC.

$$IC_{95\%} = RAR \pm 1,96 ee \quad ee = \sqrt{\frac{p1(1-p1)}{n1} + \frac{p2(1-p2)}{n2}}$$

ee: error estandar de la RAR  
 p1: probabilidad del evento en el grupo control  
 n1: número total de pacientes en el grupo control  
 p2: probabilidad del evento en el grupo experimental  
 n2: número total de pacientes en el grupo experimental

Figura 2. Cálculo del intervalo de confianza de la RAR (3)

Por ejemplo, en el reciente estudio de Ghosh *et al.* (16), se afirma que el natalizumab (3 mg/kg iv 0-4 semanas) es significativamente más eficaz que el placebo para conseguir la remisión en la semana 6 en pacientes con enfermedad de Crohn moderada a severa: 44% (29/66) vs 27% (17/63) ( $p=0,030$ ). En este caso, la RAR es del 17% (IC 95%: 0,7 a 33,2). Aunque el NNT estimado es de 6, su IC 95% es de 4 a 143, lo que puede contribuir claramente a matizar la “superioridad” del natalizumab frente a placebo.

Cuando la RAR no es significativa (su IC incluye el 0), uno de los límites del IC del NNT es negativo, lo que puede dificultar su interpretación. Por ejemplo, en el trabajo de Present *et al.* (9) se consiguió el cierre de la mitad o más de las fístulas en el 68% (21/31) de los pacientes con enfermedad de Crohn tratados con infliximab a dosis de 5 mg/kg (grupo control) y en el 56% (18/32) de los tratados con infliximab 10 mg/kg (grupo experimental). En este caso, la RAR,  $56-68 = -12\%$ , no es significativa ya que su IC (-35 a 12) incluye al 0. El NNT sería igual a  $1/-0,12 = -8,33 \approx -9$ , y los límites de su IC serían  $1/-0,35 = -2,86 \approx -3$  y  $1/0,12 = 8,33 \approx 9$ . Estos resultados presentan dos dificultades: un NNT negativo (-9) y un IC (-3 a 9) que en apariencia no incluye al estimado. Un NNT negativo indica que el tratamiento tiene un efecto perjudicial (en este ejemplo el tratamiento experimental consigue menos beneficio que el control), y ha sido denominado número necesario para dañar (NND) o para perjudicar (*harm*) (NNH). En el ejemplo, deberían tratarse 9 pacientes con infliximab 10 mg/kg para conseguir una respuesta menos que con infliximab 5 mg/kg. Según la propuesta de Altman (17), la relación entre RAR y NNT puede representarse mediante un gráfico en el que la RAR se sitúa en el eje Y1 con valores desde -100 a 100, y el NNT en el eje Y2 con valores desde -1 a 1. Una RAR de 0 genera un NNT  $\infty$ , mientras que los valores negativos de RAR (desde -100% a 0) generan valores negativos de NNT (desde -1 a  $\infty$ , respectivamente) que significarían pacientes que se perjudican con el tratamiento (NNTH), y los valores positivos de RAR (desde 100% a 0) generan valores positivos de NNT (desde 1 a  $\infty$ , respectivamente) que significarían pacientes que se benefician con el tratamiento (NNTB). Así, en el ejemplo anterior, el NNT estimado de -9 significa NNTH = 9, y los límites de su IC (-3 y 9) significan NNTH = 3 y NNTB = 9, estando incluido entre ellos el  $\infty$  (que corresponde a una RAR de 0). Ello puede expresarse como NNTH = 9 (NNTH 3 a  $\infty$  a NNTB 9) (fig. 3).

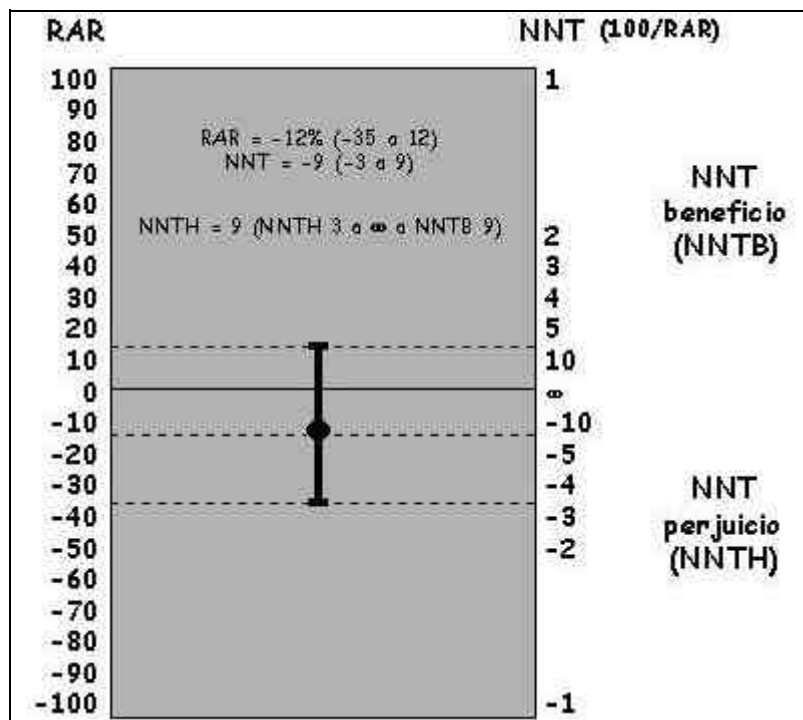


Figura 3. Intervalo de confianza del NNT (17)

Cuando el evento evaluado sucede en un período relativamente corto de tiempo y todos los pacientes tienen el mismo seguimiento, resulta correcto estimar el NNT a partir de una tabla de 2x2. Sin embargo, cuando el período de seguimiento es prolongado, además de conocer la tasa total de eventos interesa precisar el tiempo en que se producen (análisis de supervivencia). En estos casos, el NNT puede calcularse para cualquier tiempo concreto desde el inicio del tratamiento mediante fórmulas específicas (18).

Por último, al igual que en un ensayo clínico individual, también puede calcularse el NNT en un metaanálisis. Sin embargo, es necesario tener en cuenta que este parámetro varía con el riesgo basal de los pacientes y que este riesgo puede variar ampliamente entre los estudios incluidos. Por ello, el NNT derivado de un metaanálisis es preferible calcularlo a partir del modelo de efectos aleatorios (19) y debe de ser siempre interpretado con cuidado (20,21). En la fig. 4 se exponen las fórmulas utilizadas para el cálculo del NNT de un metaanálisis (3). En la revisión sistemática realizada por Sutherland *et al.* (22), el tratamiento con 5-ASA fue superior al placebo en mantener en remisión a pacientes con colitis ulcerosa (53% vs 37%). El metaanálisis de seis estudios muestra una OR de 2,14 (IC 95%: 1,62 a 2,84), sin heterogeneidad. Aplicando la correspondiente fórmula, obtenemos un NNT de 6 (IC 95%: 4 a 9).

	Para RR<1	Para OR<1	
$NNT = \frac{1}{DR}$	$NNT = \frac{1}{PEEP(1-RR)}$	$NNT = \frac{1-PEEP(1-OR)}{PEEP(1-PEEP)(1-OR)}$	
	Para RR>1	Para OR>1	
	$NNT = \frac{1}{PEEP(RR-1)}$	$NNT = \frac{1+PEEP(OR-1)}{PEEP(1-PEEP)(OR-1)}$	
DR: diferencia de riesgo. PEEP: probabilidad esperada de evento sin tratamiento			

Figura 4. Cálculo del NNT de un metaanálisis (3)

Pese a la indudable utilidad del NNT (23), su comunicación en los artículos publicados sobre ensayos clínicos sigue siendo escasa (24).

La validez de los estudios es la condición indispensable para que el cálculo del NNT tenga sentido. Además, es necesario tener en cuenta que asumir que el efecto del tratamiento es constante a lo largo del tiempo e independiente del riesgo basal de los pacientes no siempre es cierto.

En la web de CASPe (*Critical Appraisal Skills Programme*, Programa de habilidades en lectura crítica España) ([www.redcaspe.org](http://www.redcaspe.org)) hay disponibles unas calculadoras realizadas por

nosotros que intentan facilitar todos estos cálculos. Además, también pueden conseguirse las preguntas que debemos formularnos para la valoración crítica de ensayos clínicos y revisiones sistemáticas (23,25-29).

## **BIBLIOGRAFÍA**

1. Abaira V. Medidas del efecto de un tratamiento (I): reducción absoluta del riesgo, reducción relativa del riesgo y riesgo relativo. *SEMERGEN* 2000;26:535-536.
2. Abaira V. Medidas del efecto de un tratamiento (II): *odds ratio* y número necesario para tratar. *SEMERGEN* 2001;27:418-420.
3. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Medicina basada en la evidencia: cómo practicar y enseñar la MBE, 2ª edición. Madrid: Harcourt, 2001.
4. Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med* 1992;92:121-124.
5. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med* 1992;117:916-921.
6. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994;343:1209-1211.
7. Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *BMJ* 1994;309:761-764.
8. Hux JE, Naylor CD. Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? *Med Decis Making* 1995;15:152-157.
9. Present DH, Rutgeerts P, Targan S, Hanauer SB, Mayer L, van Hogezaand RA, et al. Infliximab for the treatment of fistulas in patients with Crohn's disease. *N Engl J Med* 1999;340:1398-1405.
10. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. *Ann Intern Med* 1997;126:712-720.
11. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988;318:1728-1733.
12. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995;310:452-454.
13. Chatellier G, Zapletal E, Lemaitre D, Menard J, Degoulet P. The number needed to treat: a clinically useful nomogram in its proper context. *BMJ* 1996;312:426-429.
14. Candy S, Wright JP, Gerber M, Adams G, Gerig M, Goodman R. A double blind controlled study of azathioprine in the treatment and maintenance of remission in Crohn's disease. *Gut* 1995;37:674-678.
15. Hawkey CJ, Dube LM, Rountree LV, Linnen PJ, Lancaster JF, and the European Zileuton Study Group for Ulcerative Colitis. A trial of zileuton versus mesalazine or placebo in the maintenance of remission of ulcerative colitis. *Gastroenterology* 1997;112:718-724.
16. Ghosh S, Goldin E, Gordon FH, Malchow HA, Rask-Madsen J, Rutgeerts P, et al. Natalizumab for active Crohn's disease. *N Engl J Med* 2003;348:24-32.
17. Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998;317:1309-1312.
18. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999;319:1492-1495.
19. Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *Int J Epidemiol* 2002 31:72-76.
20. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses: sometimes informative, usually misleading. *BMJ* 1999;318:1548-1551.
21. McAlister FA. Relative treatment effects are consistent across the spectrum of underlying risks... usually. *Int J Epidemiol* 2002 31:76-77.

22. Sutherland L, Roth D, Beck P, May G, Makiyama K. Oral 5-aminosalicylic acid for maintenance of remission in ulcerative colitis (Cochrane Review). In: The Cochrane Library, Issue 3, 2002. Oxford: Update Software.
23. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-694.
24. Nuovo J, Melnikow J, Chang D. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. *JAMA* 2002;287:2813-2814.
25. Guyatt GH, Sackett DL, Cook DJ. Users' Guides to the Medical Literature: II. How to use an article about therapy or prevention: A. Are the results of the study valid? *JAMA* 1993;270:2598-2601.
26. Guyatt GH, Sackett DL, Cook DJ. Users' Guides to the Medical Literature: II. How to use an article about therapy or prevention: B. What were the results and will they help me in caring for my patients? *JAMA* 1994;271:59-63.
27. Oxman AD, Cook DJ, Guyatt GH. Users' Guides to the Medical Literature: VI. How to use an overview. *JAMA* 1994;272:1367-1371.
28. Dans AL, Dans LF, Guyatt GH, Richardson S. Users' Guides to the Medical Literature: XIV. How to decide on the applicability of clinical trial results to your patient. *JAMA* 1998;279:545-549.
29. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, for the QUOROM Group. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999;354:1896-1900.