

# COMMENT J'EXPLORE ...

## Une différence de risque de survenue d'un événement dans les études cliniques

A.J. SCHEEN (1), P. ERNEST (2), B. JANDRAIN (3)

**RÉSUMÉ :** La médecine factuelle fait souvent appel à la comparaison de deux interventions thérapeutiques dans des essais cliniques contrôlés avec la démonstration d'une supériorité (par rapport à un placebo ou un comparateur actif) ou au moins d'une non-infériorité (par rapport à un comparateur actif) en ce qui concerne un critère de jugement primaire décidé *a priori* (survenue d'un événement clinique majeur, par exemple). La différence dans la survenue d'un événement entre les deux interventions thérapeutiques testées peut être analysée, sur le plan statistique, par la réduction du risque absolu, la réduction du risque relatif, le hazard ratio ou encore l'odds ratio (rapport de cotes). Cet article discute les nuances, parfois importantes, concernant la signification de ces différents indices et analyse les précautions à prendre et les pièges à éviter dans leur interprétation et leur utilisation. Le clinicien est, en effet, de plus en plus souvent confronté aux résultats d'études cliniques, mais est généralement peu informé quant aux subtilités des analyses présentées.

**MOTS-CLÉS :** *Analyse statistique - Critères de jugement - Efficacité - Essais cliniques contrôlés - Médecine factuelle - Médicament*

### INTRODUCTION

Les essais cliniques contrôlés sont le fondement de la médecine factuelle ou «Evidence-Based-Medicine» (EBM) (1). Ceux-ci visent à démontrer la supériorité d'une approche thérapeutique par rapport à un comparateur, que ce soit un placebo ou un traitement actif de référence (2). Pour ce faire, il convient, dans la plupart des cas, de démontrer une différence statistiquement significative dans la survenue d'un événement (utilisé comme critère de jugement), par exemple un infarctus du myocarde ou un décès dans les essais de prévention cardio-vasculaire (3-5).

Plusieurs indices sont utilisés couramment dans la littérature scientifique pour quantifier l'importance de la différence observée, à savoir la réduction absolue du risque (RAR), la réduction relative du risque (RRR), le hazard ratio

HOW I EXPLORE .... A RISK DIFFERENCE IN THE OCCURRENCE OF AN EVENT IN CLINICAL TRIALS

**SUMMARY :** Evidence-based medicine often requires the comparison of two therapeutic interventions in controlled clinical trials with the demonstration of a superiority (*versus* a placebo or an active comparator) or at least a non-inferiority (*versus* an active reference) concerning a primary endpoint that has been defined *a priori* (occurrence of a major clinical event, for instance). The difference in the occurrence of such an event between two treatments may be statistically analyzed by absolute risk reduction, relative risk reduction, hazard ratio or odds ratio. The present article discusses the nuances, sometimes of importance, concerning the significance of these various indices and analyses the cautions to be taken and the pitfalls to be avoided in their interpretation and use in practice. The clinician is, indeed, increasingly confronted to results of clinical trials, but is generally poorly informed regarding the nuances of these various statistical analyses.

**KEYWORDS :** *Drug - Efficacy - Evidence-based medicine - Outcomes - Randomised clinical trials - Statistical analysis*

(HR) et l'odds ratio (OR) (rapport de cotes en français) (6, 7). Devant cette diversité, le praticien non averti éprouve souvent des difficultés à discerner les nuances entre ces différents indices. Par ailleurs, la différence (ou l'absence de différence) doit parfois être interprétée avec circonspection, en évitant de tomber dans quelques pièges inhérents à la méthodologie des essais cliniques (8). Le but de cet article est de familiariser le lecteur avec les différences sémantiques relatives aux indices utilisés ainsi que de le mettre en garde en mentionnant quelques précautions élémentaires à prendre pour une analyse critique des résultats publiés dans la littérature scientifique (9).

### MODES D'EXPRESSION UTILISÉS DANS LA LITTÉRATURE

Divers indices peuvent être utilisés pour évaluer la différence d'efficacité entre deux traitements et notamment pour estimer, quantitativement, la différence de risque de survenue d'un événement dans un groupe (traité par un nouveau médicament sensé exercer un effet protecteur) par rapport à un autre groupe de référence (par exemple recevant un placebo).

#### 1) RÉDUCTION ABSOLUE DU RISQUE (RAR)

Le risque absolu (RA) peut être défini comme le risque de survenue d'un événement (décès,

(1) Professeur ordinaire, Université de Liège, Chef de Service, Service de Diabétologie, Nutrition et Maladies métaboliques et Unité de Pharmacologie clinique, Département de Médecine, CHU de Liège.

(2) Clinicien investigateur, Unité de Pharmacologie clinique, CHU de Liège.

(3) Interniste endocrino-diabétologue, Clinicien investigateur, Unité de Pharmacologie clinique, CHU de Liège.

maladie) dans un intervalle de temps donné. La différence des risques (DR), appelée aussi différence absolue ou bénéfice absolu, est égale à la différence entre le RA sous traitement ( $r_1$ ) et le RA sans traitement ( $r_0$ ). Elle se calcule par :  $DR = r_1 - r_0$ . La différence des risques donne la taille de l'effet, non ajustée sur la valeur initiale. Il est assez habituel de rapporter la différence des risques sous forme de réduction absolue des risques (RAR ou «absolute reduction in risk») qui est égale à  $RAR = (1 - DR) \times 100$  et est exprimée en pour cent (tableau I).

### 2) RÉDUCTION RELATIVE DU RISQUE (RRR)

Le risque relatif (RR) est le rapport du risque de morbidité ou mortalité lié à un facteur donné sur celui existant en l'absence de ce facteur (10). Le RR se calcule donc par l'équation  $RR = r_1/r_0$ , dans laquelle  $r_1$  est le RA sous traitement et  $r_0$  est le RA sans traitement. En épidémiologie, le facteur pronostique considéré peut être l'âge, le sexe, le diabète, le taux de cholestérol, ... Dans les essais cliniques, le facteur considéré est le traitement testé par rapport au comparateur (placebo ou référence active). La RRR se calcule par l'équation  $RRR = (1 - RR) \times 100$  et est également exprimée en pour cent (tableau I).

La RRR est souvent utilisée par l'industrie pharmaceutique dans les documents promotionnels distribués aux médecins prescripteurs potentiels. Il faut cependant savoir que cet indice a tendance à présenter le bénéfice escompté de l'intervention thérapeutique de façon trop optimiste. Par exemple, une RRR de 50 %, particulièrement impressionnante s'il s'agit de mortalité peut contraster avec la réduction du risque absolu (RAA), généralement plus modeste, ou avec le nombre de sujets à traiter pour éviter un événement (relativement élevé), comme nous le verrons plus loin (tableau II) (11).

### 3) HASARD RATIO (HR)

Le hasard ratio (HR) correspond au rapport des pentes de survenue d'un événement (morbidité ou mortel) en fonction du temps, selon la présence ou non d'un facteur donné (dans un essai d'intervention pharmacologique, un médicament donné). Cette mesure est proche de celle du RR, mais, contrairement à ce dernier, elle tient compte d'un facteur supplémentaire, le facteur temps, c'est-à-dire le moment de survenue de l'événement (par exemple, un

infarctus du myocarde) par rapport à l'initiation de l'intervention. Le HR peut donc être considéré comme un rapport des taux d'incidence et est l'indice recommandé pour comparer des courbes actuarielles (12).

Les courbes actuarielles ont d'abord été popularisées dans les études oncologiques pour analyser la survie des patients en fonction du temps. Par extension, les survies sans progression de la maladie («*progression-free survival*») ou sans rechute («*recurrence-free survival*», «*relapse-free survival*») sont couramment utilisées en cancérologie. La taille de l'effet traitement est alors mesurée à l'aide du rapport des risques instantanés («*hazard ratio*»). Le HR est le rapport du risque instantané dans le groupe traité ( $h_1$ ) divisé par le risque dans le groupe contrôle ( $h_0$ ). En pratique, à partir des données de survie, le HR est estimé soit à partir de la méthode statistique du logrank, soit en utilisant le modèle de Cox (13).

Dans les publications, le HR est parfois présenté comme un RR dont il est souvent proche numériquement. Par exemple, dans un essai de mortalité *versus* placebo, un HR de 0,5 signifie que, sous traitement, le risque instantané de décès est seulement la moitié du risque (instantané) sans traitement.

### 4) ODDS RATIO (OR)

L'odd ou la cote est le rapport entre le pourcentage de sujets qui présentent un événement (morbidité ou fatal) et le pourcentage de ceux qui ne le présentent pas. L'odd est donc égal à  $p / (1-p)$ , où  $p$  est la proportion des sujets présentant l'événement. L'odds ratio (OR, dont une traduction littérale en français peut être «rapport des cotes») est le rapport entre deux odds, à savoir l'odds de l'événement (sa cote) dans le groupe traité divisé par l'odds de l'événement dans le groupe contrôle. Un OR de 1 correspond à l'absence d'effet. En cas d'effet bénéfique, l'OR est inférieur à 1 et il est supérieur à 1 en cas d'effet délétère. Plus l'OR est éloigné de 1, plus l'effet est important (10).

L'OR s'interprète de façon similaire au RR dont il est d'ailleurs fort proche. L'OR est d'autant plus proche du RR que le risque de base dans le groupe contrôle est faible. Par contre, l'OR aura tendance à surestimer l'effet du traitement quand le risque de base est élevé. Dans ce cas, les résultats présentés avec un OR seront plus favorables au traitement que ceux basés

TABLEAU I. COMPARAISON DES DIFFÉRENTS INDICES UTILISÉS POUR QUANTIFIER L'EFFICACITÉ DANS LES ESSAIS CLINIQUES CONTRÔLÉS.

Abréviation	Indice	Explication	Mode de calcul
<b>RAR</b>	Réduction Absolue du Risque	Différence de risque (DR), soit taux d'événements dans le groupe témoin ( $r_0$ ) - taux d'événements dans le groupe expérimental ( $r_1$ )	$RAR = (1-DR) \times 100 = [1 - (r_0 - r_1)] \times 100$
<b>RRR</b>	Réduction Relative du Risque	Taux d'événements dans le groupe expérimental ( $r_1$ ) divisé par le taux d'événements dans le groupe témoin ( $r_0$ )	$RRR = (1-RR) \times 100 = [1 - (r_1/r_0)] \times 100$
<b>HR</b>	Hasard Ratio	Rapport des pentes de survenue d'un premier événement en fonction du temps selon la présence ou non du traitement testé Rapport du risque instantané dans le groupe traité ( $h_1$ ) divisé par le risque instantané dans le groupe contrôle ( $h_0$ )	$HR = h_1/h_0$
<b>OR</b>	Odds Ratio ou Rapport de cotes	Rapport entre l'odd (cote) de l'événement dans le groupe traité (1) divisé par l'odd de l'événement dans le groupe contrôle (0) (où p représente la probabilité de survenue de l'événement dans le groupe considéré)	$OR = [p_1/(1-p_1)] / [p_0/(1-p_0)]$
<b>NNT</b>	Nombre de sujets Nécessaire de Traiter	Nombre moyen de sujets qu'il est nécessaire de traiter pour éviter l'événement sur une période donnée	$NNT = 1 / RAR = 1 / [(1-DR) \times 100]$

TABLEAU II. COMPARAISON DES VALEURS DES DIFFÉRENTS INDICES DÉCRITS DANS LE TABLEAU I CALCULÉ DANS UN ESSAI CLINIQUE DE PRÉVENTION CARDIO-VASCULAIRE D'UNE DURÉE DE 5 ANS COMPARANT UN MÉDICAMENT ACTIF À UN PLACEBO AVEC DEUX SCÉNARIOS QUI DIFFÈRENT PAR UN RISQUE DE BASE FAIBLE (SCÉNARIO A) OU ÉLEVÉ (SCÉNARIO B). DANS LES DEUX CAS DE FIGURE, LA RRR EST COMPARABLE (50%, CORRESPONDANT À UNE RÉDUCTION DE MOITIÉ DE LA MORTALITÉ), ALORS QUE LES VALEURS DE RRA ET DE NNT VARIENT D'UN FACTEUR 20 ENTRE LES DEUX GROUPES. PAR AILLEURS, L'OR S'ABAISSÉ PAR RAPPORT À RRR AVEC L'ACCROISSEMENT DU RISQUE DE BASE (SCÉNARIO B PAR RAPPORT AU SCÉNARIO A).

Indice	Scénario A : risque faible	Scénario B : risque élevé
	Placebo : 20 décès / 1.000 patients / 5 ans Actif : 10 décès / 1.000 patients / 5 ans	Placebo : 400 décès / 1.000 patients / 5 ans Actif : 200 décès / 1.000 patients / 5 ans
RRR = Réduction Relative du Risque	50 %	50 %
OR = Odds Ratio	0,44	0,375
RAR = Réduction Absolue du Risque	1 %	20%
NNT = Nombre de sujets Nécessaire de Traiter	100	5

sur le RR et un OR ne peut plus être interprété comme un RR (voir plus loin : tableau II).

##### 5) NOMBRE DE SUJETS NÉCESSAIRE DE TRAITER (NNT)

Cet indice, nommé par l'abréviation NNT pour Nombre de sujets Nécessaire de Traiter (ou en anglais «*Number Needed to Treat*»), correspond au nombre moyen de sujets qu'il est nécessaire de traiter pour éviter 1 événement sur une période donnée. Il est égal à l'inverse de la différence des RA :  $NNT = 1/DR = 1/(r_1 -$

$r_0)$ . Le NNT doit toujours comprendre un cadre temporel (3 ans, 5 ans, etc.).

Depuis son introduction par Sackett et ses collaborateurs (3, 14), cet indice a obtenu un grand succès, sans doute en raison du fait qu'il apparaît très «parlant» pour les praticiens. Il a cependant tendance à faire croire qu'il existe réellement un individu qui ne présentera pas l'événement avec le traitement alors qu'il l'aurait présenté sans traitement. Il crée donc une individualisation du bénéfice, mais cette interprétation est évidemment fautive. La vraie

interprétation est que tous les patients bénéficient un peu du traitement (il y a réduction de leur probabilité de faire un événement) : cette réduction, en termes de NNT, se traduit en moyenne par une différence de 1 du nombre moyen d'événements attendus avec et sans traitement.

La notion de NNT peut être étendue aux événements délétères. Il s'agit alors du «*Number Needed to Harm (NNH)*» qui quantifie le nombre de patients qu'il faut exposer au traitement pour observer, en moyenne, un effet indésirable du traitement (15).

#### 6) DIFFÉRENCES D'INTERPRÉTATION SELON LES INDICES

De façon schématique, on peut donc séparer les indices en deux grandes catégories, ceux qui prennent en compte le bénéfice absolu (RAR, NNT) et ceux qui testent le bénéfice relatif (RRR, OR). Une différence essentielle entre les deux catégories réside dans le fait que les indices décrivant le bénéfice absolu dépendent non seulement de l'efficacité intrinsèque du traitement, mais sont aussi influencés par le risque de base de la population testée (qui peut varier sensiblement d'un essai à l'autre). Par contre, les indices prenant en compte le bénéfice relatif ne dépendent pas (pour RRR) ou dans une moindre mesure (dans le cas d'OR) de ce facteur confondant. En cas de risque de base faible, une RRR importante pourra se traduire par une RRA faible et donc un NNT élevé, ce qui limite l'intérêt clinique de l'intervention thérapeutique (tableau II).

Compte tenu de cette influence majeure du risque de base, les NNT ne doivent pas être utilisés pour comparer deux essais, encore moins pour comparer deux traitements concurrents pour une même pathologie. Des différences de NNT peuvent témoigner aussi bien d'une différence d'efficacité des traitements considérés que de différences de risques de base des patients entre les essais (15, 16). Pour que la comparaison soit pertinente et fiable, les traitements doivent avoir été mis à l'essai dans des échantillons de population similaires, ayant la même affection mais aussi les mêmes caractéristiques, en utilisant le même comparateur, la même période de temps de suivi et les mêmes critères de jugement. En pratique, toutes ces conditions ne peuvent être garanties que dans des études de comparaison directe («*head-to-head trials*») : voir plus loin)

En ce qui concerne l'évaluation du bénéfice relatif, deux indices peuvent être utilisés,

la réduction du risque relatif (RRR) et l'odds ratio (OR). Cependant, ces deux indices ne sont pas parfaitement identiques. Contrairement à la RRR qui est indépendante du risque de base dans la population étudiée, l'OR varie en fonction de celui-ci et est d'autant plus bas que le risque de base est élevé. Dès lors, d'une façon générale, l'OR surestime la RRR. C'est seulement lorsque l'événement est relativement rare, c'est-à-dire survenant dans moins de 10% des cas dans la population étudiée, que les valeurs des OR correspondent aux valeurs des RRR. Pour un  $RR < 1$  (ce qui est habituellement le cas dans les essais cliniques testant l'efficacité d'un nouveau médicament), l'OR sera plus proche de 0 que le RR et ce, d'autant plus que l'incidence des événements est spontanément élevée. En termes d'efficacité d'un traitement, l'OR aura tendance à surestimer l'effet du traitement quand le risque de base est élevé (au dessus de 25% environ). Dans ce cas, les résultats présentés avec un OR seront plus favorables au traitement que ceux basés sur le risque relatif et un OR ne peut plus être interprété comme un risque relatif. C'est ce que l'on appelle le «*biais intrinsèque*» de l'OR : l'OR est égal au RR multiplié par le biais de l'OR. A titre d'exemple, si le critère d'évaluation est le décès, l'OR est un RR amplifié par le rapport de survie du groupe de référence sur celle du groupe traité.

En clinique, la RRR est généralement préférée à l'OR car elle est d'interprétation plus intuitive et il s'agit d'un indice qui vient naturellement à l'esprit pour mesurer l'effet d'un traitement. Par contre, les OR sont habituellement utilisés dans les méta-analyses regroupant divers essais dont les populations peuvent avoir, par définition, un risque de base et une durée de suivi variables d'une étude à l'autre.

#### PRÉCAUTIONS À PRENDRE DANS LES INTERPRÉTATIONS

Après avoir fait le discernement entre les différents indices susceptibles d'être utilisés, il convient d'attirer l'attention sur certains pièges à identifier et éviter.

##### 1) INTERVALLE DE CONFIANCE (IC)

Quel que soit l'indice pris en compte, il est essentiel de tenir compte de l'intervalle de confiance (IC) de part et d'autre de la valeur obtenue (17). Cet IC est un indicateur de la signification statistique de la différence obtenue.

nue entre les deux interventions thérapeutiques. Ainsi, un IC large (dont les limites franchissent la zone de neutralité correspondant généralement à 1 pour RR, HR et OR) signifie que la différence observée, même importante, n'est pas «statistiquement significative». Dès lors, il existe une telle incertitude autour du résultat obtenu que le chercheur ou le clinicien ne peuvent être certains que l'effet observé ne soit pas dû au hasard et donc difficile (voire impossible) à reproduire. Au contraire, un IC étroit (ne franchissant pas la barre de 1) plaide pour un effet qui, toute chose étant égale par ailleurs, devrait pouvoir être reproduit. Le risque d'erreur d'alpha utilisé le plus souvent en statistique est fixé à 5 % (ce qui correspond au classique  $p < 0,05$ ).

## 2) COMPARAISON INDIRECTE

La tentation est grande de comparer des différences de risque de survenue d'un événement observées dans des études différentes (comparaison indirecte). C'est notamment le cas lorsque l'industrie pharmaceutique veut comparer l'efficacité d'une molécule d'une classe pharmacologique avec une autre molécule de la même classe pour vanter les mérites de «sa» molécule par rapport à certaines de ses concurrents. Il convient de se méfier de toutes les comparaisons indirectes. En effet, une éventuelle différence mise en exergue peut, de fait, être attribuée à une plus grande efficacité de la molécule mise en avant. Cependant, elle peut aussi résulter de différences concernant les caractéristiques des populations étudiées ou encore le protocole d'étude utilisé (critère de jugement pris en compte, durée de l'étude, ...). Il existe donc une série de facteurs confondants potentiels que seule une étude comparative directe («head-to-head trial») est susceptible de gommer. Comme déjà mentionné, le calcul d'un NNT est influencé, bien entendu, par l'efficacité de la stratégie thérapeutique testée, mais aussi par le risque de base de la population étudiée (qui, par définition, peut varier d'une étude à l'autre, même si l'on croit s'adresser à des populations *a priori* comparables) (Tableau II).

## 3) ANALYSE SUR DES CRITÈRES SECONDAIRES OU À VISÉE EXPLORATOIRE

L'analyse statistique qui doit primer est celle qui concerne le critère d'évaluation principal («primary endpoint»). La tentation est parfois grande de mettre en avant des différences de

risque concernant des critères secondaires ou même des critères utilisés uniquement à titre exploratoire. C'est le cas lorsque le critère d'évaluation primaire n'atteint pas la signification statistique alors que l'un ou l'autre des autres critères, plus accessoires, est statistiquement significatif, précisément celui qui plaide en faveur de la molécule étudiée. Selon les règles relatives aux essais cliniques, si le critère d'évaluation primaire n'atteint pas le seuil de signification statistique, la conclusion générale doit être qu'il s'agit d'un «essai négatif» et on ne peut donc conclure à une quelconque supériorité de la molécule mise en avant (5). La différence observée dans le critère secondaire doit être considérée comme génératrice d'hypothèse et non comme preuve en termes de médecine factuelle.

## 4) ANALYSE POST HOC OU A POSTERIORI

Les seules analyses qui ont force de preuve sont celles qui concernent des analyses prévues *a priori*, au moment de la conception de l'étude. Toutes les analyses réalisées au moment de l'analyse finale des résultats, non prévues *a priori*, n'ont qu'une valeur exploratoire et aucunement susceptible d'emporter la conviction (18). Pourtant, de nombreuses analyses de ce type sont réalisées et insérées dans les résultats d'articles scientifiques relatifs à des essais cliniques, en particulier des analyses sur des sous-groupes de patients non prévues au moment de la conception de l'étude. Il faut savoir que la définition de sous-groupes *a priori* impose d'augmenter le nombre de patients à recruter dans l'étude de façon à garantir une puissance statistique suffisante, permettant une conclusion avec un risque d'erreur minimal (par exemple, inférieur à 5%). Dès lors, la tentation est grande de réaliser ces analyses par sous-groupes *a posteriori*.

## 5) MANQUE DE PUISSANCE STATISTIQUE

Plutôt que de démontrer une supériorité, l'objectif de certaines études est de démontrer une non-infériorité (voire une équivalence) d'une molécule avec un médicament de référence, considéré comme étant le «gold standard» dans sa classe. Si le nombre de patients inclus dans l'essai n'est pas suffisamment élevé, l'étude ne pourra pas démontrer une différence significative entre les deux interventions, uniquement par manque de puissance statistique (19). Il s'agit d'une erreur de type B (ou de deuxième espèce), à savoir le fait de conclure qu'il n'y a

pas de différence alors qu'il y en a une (mais qui n'a pu être statistiquement démontrée par un manque de puissance). Il est donc indispensable de calculer, dès le moment de la conception de l'étude, le nombre de sujets à inclure dans l'étude pour être en mesure de conclure à propos de l'objectif fixé, avec un niveau de risque raisonnable (20). Ce calcul de puissance est maintenant exigé dans le descriptif de la méthodologie de l'essai et doit, idéalement, être décrit dans toute publication d'une étude clinique.

#### 6) RECOURS AUX MÉTA-ANALYSES

Un moyen d'augmenter la puissance statistique est de procéder à une méta-analyse de différents essais testant une hypothèse commune. En effet, rassembler les résultats de plusieurs essais, testant une même hypothèse selon un protocole similaire, permet d'augmenter le nombre de sujets finalement analysés et, en l'occurrence, le nombre d'événements d'intérêt dans le groupe expérimental (médicament testé) et dans le groupe témoin, ce qui permet de conclure avec un plus grand degré de confiance. Cependant, la qualité des méta-analyses dépend, de façon cruciale, à la fois de l'exhaustivité et de la fiabilité des essais pris en compte. Nous avons décrit les pré-requis et les principaux pièges à éviter dans un article précédent (21).

#### CONCLUSION

La littérature médicale rapporte, de plus en plus, les résultats d'études cliniques à visée thérapeutique, typiquement des essais qui comparent un médicament actif à un placebo ou une nouvelle molécule à un médicament de référence pris comme comparateur. Le clinicien est ainsi confronté, de façon croissante, à ce type de publication sans qu'il soit nécessairement préparé à pouvoir bien discerner toutes les subtilités inhérentes à une analyse critique des résultats présentés. Cet article est sensé contribuer à une meilleure compréhension des nuances entre les indices statistiques exprimant l'efficacité d'un traitement (un nouveau médicament, par exemple) et des pièges à éviter susceptibles d'aboutir à des conclusions erronées.

#### BIBLIOGRAPHIE

1. Scheen AJ.— Evidence-based medicine. Apport des essais cliniques contrôlés. *Rev Med Liège*, 2000, **55**, 216-219.

2. Scheen AJ.— La mesure de l'efficacité thérapeutique, étape essentielle dans la médecine factuelle. *Rev Med Liège*, 2000, **55**, 206-210.
3. Laupacis A, Sackett DL, Roberts RS.— An assessment of clinically useful measures of the consequences of treatment. *N Eng J Med*, 1988, **318**, 1728-1733.
4. Collins R, MacMahon.— Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet*, 2001, **357**, 373-380.
5. Scheen AJ.— Comment j'explore ... les critères de jugement dans les essais cliniques : réflexion à propos d'études récentes de prévention cardio-vasculaire. *Rev Med Liège*, 2006, **61**, 260-266.
6. Piérard GE, Scheen AJ.— Innover par l'EBM. Glossaire initiatique aux recherches cliniques. *Rev Med Liège*, 2005, **60**, 190-194.
7. Replogle WH, Johnson WD.— Interpretation of absolute measures of disease risk in comparative research. *Fam Med*, 2007, **39**, 432-435.
8. Rodondi N, Cornuz J.— Lecture d'articles médicaux: quelques pièges à éviter. *Rev Med Suisse*, 2007, **3**, 2718-2723.
9. Bergmann JF, Chassany O.— Lire un compte rendu d'essai clinique. *Rev Prat*, 2000, **50**, 838-845.
10. Morabia A.— Risque relatif et odds ratio. *Rev Mal Resp*, 2003, **20**, 757-759.
11. 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.
12. Spruance SL, Reid JE, Grace M, Samore M.— Hazard ratio in clinical trials. *Antimicrob Agents Chemother*, 2004, **48**, 2787-2792.
13. Lee ET, Go OT.— Survival analysis in public health research. *Annu Rev Public Health*, 1997, **18**, 105-134.
14. Cook RJ, Sackett DL.— The number needed to treat: a clinically useful measure of treatment effect. *BMJ*, 1995, **310**, 452-454. Erratum in : *BMJ*, 1995, **310**, 1056.
15. Hutton JL.— Number needed to treat and number needed to harm are not the best way to report and assess the results of randomised clinical trials. *Br J Haematol*, 2009, **146**, 27-30.
16. Stang A, Poole C, Bender R.— Common problems related to the use of number needed to treat. *J Clin Epidemiol*, 2010, **63**, 820-825.
17. Norman GR, Streiner DL.— Do CIs give you confidence? *Chest*, 2012, **141**, 17-19.
18. Rothwell PM.— Treating individuals 2. Subgroup analysis in randomised controlled trials : importance, indications, and interpretation. *Lancet*, 2005, **365**, 176-186.
19. Detsky AS, Sackett DL.— When was a "negative" clinical trial big enough? How many patients you needed depends on what you found. *Arch Intern Med*, 1985, **145**, 709-712.
20. Mélot C.— Que signifie puissance d'une étude ? Comment la calculer ? Comment calculer le nombre de sujets nécessaires ? *Rev Mal Respir*, 2003, **20**, 602-603.
21. Scheen AJ.— Comment j'explore ... les secrets d'une méta-analyse. *Rev Med Liège*, 2003, **58**, 41-46.

Les demandes de tirés à part sont à adresser au Pr A.J. Scheen, Département de Médecine Générale, CHU de Liège, Belgique.