

# COMMENT J'EXPLORE ...

## les différentes facettes statistiques des tables 2x2 en médecine

A. ALBERT (1, 2)

**RÉSUMÉ :** Les tables de comptage 2x2 figurent parmi les représentations statistiques de données binaires les plus utilisées dans la littérature scientifique médicale. Elles sont associées à la définition de concepts épidémiologiques et cliniques fondamentaux, utilisés universellement. Enseignées dans le cours de biostatistique au début du cursus des études de médecine, leur interprétation par les étudiants prête souvent à confusion. Nous avons répertorié au moins huit facettes différentes de ces tables : test d'homogénéité, test d'indépendance, test de McNemar, kappa de Cohen, incidence et risque relatif, prévalence et risque relatif, odds ratio et valeur diagnostique d'un test clinique. Chaque facette est illustrée par un problème clinique exemplatif. Cet article devrait permettre au lecteur de mieux comprendre à quelle table il a réellement affaire.

**MOTS-CLÉS :** Variable binaire - Comparaison de proportions - Coefficient d'accord - Etude prospective - Régression logistique

### INTRODUCTION

En statistique, les tables de contingence 2x2 permettent de représenter la distribution de  $n$  sujets ou objets en fonction de deux variables (critères) binaires. En effet, le croisement de 2 lignes et de 2 colonnes forme une table à 4 cellules contenant chacune un nombre entier appelé comptage. Ces nombres sont souvent exprimés en pourcentages des totaux «marginiaux» des lignes ou des colonnes ou du total général  $n$ . Les tables de comptage 2x2 font partie de tout cours de statistique de base (1). Dans la littérature médicale, elles occupent une place importante car elles se présentent sous diverses facettes et ce, en fonction de la manière dont elles ont été obtenues (méthode d'échantillonnage) ou de la nature du problème envisagé. Elles ont conduit à la définition de concepts épidémiologiques et cliniques utilisés internationalement. En fait, nous avons répertorié au moins huit façons de présenter et d'interpréter les tables de contingence 2x2 et

### HOW I EXPLORE ... THE VARIOUS STATISTICAL FACETS OF 2x2 TABLES IN MEDICINE

**SUMMARY :** In the medical scientific literature, 2x2 contingency tables are widely used to display the distribution of binary data. They are associated with the definition of basic concepts in medicine and epidemiology that are universally utilized. They are usually taught in any basic course in biostatistics early in the medical curriculum; their interpretation however is not always straightforward. We have discerned at least eight different facets of such fourfold tables: homogeneity test, independence test, McNemar test, Cohen kappa coefficient, incidence relative risk, prevalence relative risk, odds ratio, diagnostic value of a clinical test. Each facet is illustrated by a clinical problem used as an example. This article should help the reader to understand which table he/she is actually dealing with.

**KEYWORDS :** Binary variable - Comparison of proportions - Agreement coefficient - Prospective study - Logistic regression

il n'est pas toujours aisé pour le médecin, le chercheur ou l'étudiant de savoir à quelle table il a réellement affaire. Les exemples cliniques de cet article, choisis à dessein à titre illustratif, devraient permettre de les aider dans cet exercice *a priori* difficile.

### DÉFINITION

Dans cet article, nous désignons par  $X$  la variable ligne et par  $Y$  la variable colonne qui définissent la table 2x2. Chacune de ces variables binaires prend la valeur 0 ou 1. Dans les modèles de régression,  $Y$  est la variable dépendante et  $X$  la variable indépendante. En pratique, on représente les comptages de la table 2x2 par les lettres  $a, b, c, d$  dont la somme vaut  $n$  (tableau I).

Dès lors, les totaux «marginiaux» des lignes valent respectivement  $(a + b)$  et  $(c + d)$  et ceux des colonnes  $(a + c)$  et  $(b + d)$ .

TABLEAU I. REPRÉSENTATION GÉNÉRALE D'UNE TABLE DE CONTINGENCE 2 x 2

Variable X	Variable Y		Total
	0	1	
0	a	b	a+b
1	c	d	c+d
Total	a+c	b+d	n

(1) Professeur ordinaire, Université de Liège, Chef de Service, Informatique médicale et Biostatistique.

(2) Note de la Rédaction : Le Comité de Gestion de la Revue Médicale de Liège est heureux de présenter à ses lecteurs la leçon d'adieu donnée le 26 septembre 2013 par le Professeur Adelin Albert à l'occasion de son accession à l'éméritat. Il félicite le Professeur Albert pour sa brillante carrière de biostatisticien et le remercie pour son total dévouement au service de l'institution.

En fait, il n'y a que trois manières d'obtenir une table de contingence 2x2 : (i) soit on fixe les totaux des lignes a+b et c+d, auquel cas les éléments de la table et les totaux des colonnes sont observés et seuls les pourcentages « horizontaux » peuvent être calculés; (ii) soit on fixe les totaux des colonnes a+c et b+d, auquel cas les éléments de la table et les totaux des lignes sont observés et seuls les pourcentages « verticaux » peuvent être calculés; (iii) soit on fixe le total général n, auquel cas les éléments de la table et tous les totaux marginaux a+b, c+d, a+c et b+d sont observés. On peut alors calculer à sa guise les pourcentages horizontaux, verticaux et totaux. Au niveau de l'échantillonnage des sujets, on dit que celui-ci s'effectue, respectivement, de Y conditionnel à X, de X conditionnel à Y, et du mélange (X, Y).

## REPRÉSENTATIONS DES TABLES 2x2

Nous décrivons à présent huit facettes ou façons de présenter et d'interpréter les tables 2x2 : test d'homogénéité, test d'indépendance, test de McNemar, coefficient kappa de Cohen, incidence et risque relatif, prévalence et risque relatif, odds ratio, et valeur diagnostique d'un test.

### FACETTE 1. LE TEST D'HOMOGENÉITÉ

Le test le plus classique associé aux tables 2x2 est, sans nul doute, celui de la comparaison de deux proportions indépendantes, c'est-à-dire issues de deux populations distinctes. Ici Y représente les deux populations et X le critère binaire auquel on s'intéresse. Dans ce cas, la table 2x2 est obtenue en fixant les totaux des colonnes (a + c) et (b + d). Ainsi, le tableau II a été obtenu lors d'une étude sur le tabagisme chez des adolescents en Province de Luxembourg (données non publiées). Celle-ci a été menée auprès de 1.262 garçons et 1.132 filles. Ce n'est qu'au terme de l'étude que la table 2x2 a pu être établie et qu'on a pu connaître les 4 comptages a, b, c et d ainsi que les totaux des lignes. Comme indiqué précédemment, ces nombres ont été observés. La proportion de fumeurs est égale à 30,8% (389/1.262) chez les garçons et 32,9% (372/1.132) chez les filles.

L'hypothèse nulle testée est celle de l'égalité des deux proportions, en l'occurrence ici l'homogénéité des deux groupes par rapport au tabagisme. A cet effet, on calcule le critère :

$$\chi^2_{(1)} = \frac{(ad - bc)^2 n}{(a + b)(c + d)(a + c)(b + d)}$$

TABLEAU II. RÉPARTITION DU TABAGISME CHEZ DES ADOLESCENTS SELON LE SEXE EN PROVINCE DE LUXEMBOURG

Tabagisme	Sexe		Total
	Garçon	Fille	
Non-fumeur	873	730	1633
Fumeur	389	372	761
Total	1262	1132	2394

distribué selon une loi chi-carré à 1 degré de liberté. Il s'agit du célèbre «test du chi-carré» décrit dans tous les ouvrages de statistique. Pour rappel, si la valeur  $\chi^2_{(1)}$  dépasse 3,84, le seuil critique à 5%, on rejette l'hypothèse nulle. On applique parfois la correction de continuité de Yates (2) en remplaçant dans la formule précédente, le facteur

$$(ad - bc)^2 \quad \text{par} \quad (|ad - bc| - \frac{n}{2})^2$$

Notons que dans le cas de petits effectifs attendus (<5), il est recommandé d'utiliser le test exact de Fisher (3) qu'il est plus difficile de calculer manuellement. Il faut donc avoir recours à un logiciel statistique.

L'application du test du chi-carré aux données du tableau II montre qu'il n'y a pas de différence statistiquement significative entre le taux de tabagisme des garçons et des filles. En effet,

$$\chi^2_{(1)} = \frac{(873 \times 372 - 730 \times 389)^2 \times 2394}{1633 \times 761 \times 1262 \times 1132} = 1,14$$

(p=0,29). Si on applique la correction de continuité de Yates, on obtient 1,05 (p=0,31) et la même conclusion.

### FACETTE 2. LE TEST D'INDÉPENDANCE

Celui-ci est souvent confondu avec le test d'homogénéité. Ici, on teste l'hypothèse d'indépendance (absence d'association) entre les deux variables binaires X et Y. Celles-ci sont observées simultanément sur les individus d'une même population. La table 2x2 est donc obtenue en échantillonnant du mélange (X,Y) et seul le total général n de la table 2x2 est fixé. A cet effet, on calcule exactement le même critère que celui utilisé pour le test d'homogénéité (avec ou sans la correction de continuité de Yates), d'où la confusion habituelle entre les deux approches. L'hypothèse nulle revient aussi à dire que la corrélation r entre X et Y est nulle.

Notons que le coefficient de corrélation entre les deux variables binaires X et Y, appelé «coefficient de corrélation de point», est obtenu directement à partir de la formule

$$r = \sqrt{\chi^2_{(1)}/n}$$

Dans une étude sur les nausées et vomissements post-opératoires chez 671 patients hospitalisés en chirurgie, Boogaerts et al (4) se sont intéressés à l'association entre les deux symptômes (voir tableau III).

L'application du test chi-carré aux données du tableau III

$$\chi^2_{(1)} = \frac{(532 \times 53 - 13 \times 73)^2 \times 671}{545 \times 126 \times 605 \times 66} = 181,7$$

montre une association hautement significative ( $p < 0,0001$ ) entre les deux symptômes. L'hypothèse d'indépendance est donc rejetée. Notons que le coefficient de corrélation de point entre nausées et vomissements vaut

$$r = \sqrt{\chi^2_{(1)}/n} = \sqrt{181,7/671} = 0,52$$

### FACETTE 3. LE TEST DE McNEMAR

Contrairement au test d'homogénéité, le test de McNemar (5) permet de comparer deux proportions appariées, c'est-à-dire obtenues sur les mêmes sujets ou sur des sujets appariés. En réalité, on n'a qu'une seule population (ou deux populations appariées) et une seule variable binaire observée dans deux conditions différentes. On peut cependant définir par X la variable relative à la première condition (ligne) et par Y celle relative à la seconde condition (colonne). Dès lors, l'échantillonnage s'effectue en fixant le total général n. L'hypothèse nulle postule l'égalité des deux proportions. A cet effet, on a recours au test de McNemar qui s'écrit :

$$\chi^2_{(1)} = \frac{(b - c)^2}{b + c}$$

distribué comme un chi-carré à 1 degré de liberté. Une correction de continuité peut également être appliquée en remplaçant le numérateur par  $(|b - c| - 1)^2$

Le tableau IV donne la distance parcourue avant et après implantation percutanée d'un dispositif permettant d'élargir le canal lombaire chez 156 patients souffrant de sténose lombaire dégénérative (SLD) accompagnée de

TABEAU III. DISTRIBUTION DES NAUSÉES ET VOMISSEMENTS POST-OPÉRATOIRES CHEZ 671 SUJETS OPÉRÉS

Nausée	Vomissement		Total
	Non	Oui	
Non	532	13	545
Oui	73	53	126
Total	605	66	671

TABEAU IV. CLASSEMENT DE 156 PATIENTS ATTEINTS DE STÉNOSE LOMBAIRE DÉGÉNÉRATIVE AVEC CLAUDICATION INTERMITTENTE NEUROGÈNE EN FONCTION DE LA DISTANCE PARCOURUE AVANT ET APRÈS IMPLANTATION D'UN DISPOSITIF D'ÉLARGISSEMENT DU CANAL LOMBAIRE

Avant intervention	Après intervention		Total
	≤ 500 m	> 500 m	
≤ 500 m	56	37	93
> 500 m	20	43	63
Total	76	80	156

claudication intermittente d'origine neurogène (données non publiées). Le seuil de 500 m a été fixé pour évaluer la condition des patients. Notons qu'avant intervention, 63 (40,4%) patients avaient pu parcourir plus de 500 m, alors qu'après intervention chirurgicale ce nombre s'élevait à 80 (51,3%). Ces deux proportions sont-elles statistiquement différentes ?

L'application du test de McNemar montre un effet significatif de l'intervention chirurgicale sur la distance parcourue par le patient. En effet,

$$\chi^2_{(1)} = \frac{(37 - 20)^2}{37 + 20} = 5,07 \quad (p = 0,024)$$

Si on applique la correction de continuité, on obtient  $\chi^2_{(1)} = 4,49$  ( $p = 0,034$ ) et la même conclusion.

### FACETTE 4. LE COEFFICIENT KAPPA DE COHEN

Le coefficient kappa ( $\kappa$ ) de Cohen permet de mesurer le degré d'accord entre deux observateurs/équipements sur une échelle binaire (6). Considérons l'évaluation de n sujets/objets par deux observateurs sur une échelle binaire. Notons X l'évaluation (binaire) du premier observateur (critère ligne) et Y l'évaluation du second observateur (critère colonne); on établit

TABLEAU V. DIAGNOSTIC DE 187 MASSES MAMMAIRES SUSPECTES PAR MAMMOGRAPHIE (2D) ET TOMOSYNTHESE (3D)

Mammographie	Tomosynthèse		Total
	Bénin	Malin	
Bénin	54	68	122
Malin	14	51	65
Total	68	119	187

ainsi une table 2x2. L'échantillonnage s'effectue donc du mélange car le total n est fixé. L'hypothèse nulle stipule qu'il n'y a pas d'accord entre les deux observateurs ( $\kappa = 0$ ).

A titre d'illustration, le diagnostic (bénin ou malin) de 187 masses mammaires suspectes a été obtenu par mammographie (2D) et par tomosynthèse (3D). Les résultats sont présentés au tableau V (données non publiées – service de sénologie du CHU). Peut-on conclure à un accord entre les deux techniques diagnostiques ?

Pour estimer  $\kappa$ , on calcule d'abord la proportion d'accords observés entre les deux lecteurs, soit  $p_o = (a+d)/n$ . On calcule ensuite la proportion d'accords dus au hasard, c'est-à-dire celle qu'on obtiendrait si les deux lecteurs décidaient sur base du seul hasard; cette proportion vaut  $p_e = [(a+b).(a+c) + (c+d).(b+d)]/n^2$ . Le coefficient kappa de Cohen s'écrit alors

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

Plus kappa est proche de 1, meilleur est l'accord entre les deux observateurs. En général, on associe à  $\kappa$  son intervalle de confiance à 95% (IC95%). Si celui-ci contient 0, il n'y a pas d'accord entre les deux séries de mesures. Il existe une relation inverse entre le test de McNemar et le  $\kappa$  de Cohen. Plus la valeur du kappa de Cohen est faible, plus la valeur du chi-carré de McNemar est élevée et vice versa.

L'application de cette méthode aux données du tableau V donne les résultats suivants. La proportion d'accords observés vaut  $p_o = (54+51)/187 = 0,561$  tandis que la proportion d'accords dus au hasard s'écrit  $p_e = [122 \times 68 + 65 \times 119]/(187)^2 = 0,458$ .

Le kappa de Cohen vaut donc

$$\kappa = \frac{0,561 - 0,458}{1 - 0,458} = 0,19$$

L'IC 95% associé à cette valeur vaut (0,08 – 0,31). Le degré d'accord entre les deux approches diagnostiques est statistiquement différent de 0 mais l'intervalle de confiance montre qu'il est en réalité très faible (puisque la borne inférieure de l'IC 95%, égale à 0,08, est proche de zéro). Ceci est confirmé par la valeur très élevée du test de McNemar  $\chi^2_{(1)} = 35,6$   $p < 0,0001$ ) qui compare les proportions de diagnostics malins obtenus par mammographie (65/187, 34,8%) et par tomosynthèse (119/187, 63,6%) sur les 187 masses mammaires suspectes. La différence entre ces proportions illustre bien la divergence (et donc l'absence de concordance) entre les deux techniques diagnostiques.

#### FACETTE 5. INCIDENCE ET RISQUE RELATIF

En épidémiologie (7), on recherche l'association entre une maladie (présente ou absente) et un facteur de risque ou d'exposition (exposé ou non exposé). Dans ce cas, on note Y la maladie (critère binaire colonne) et X le facteur d'exposition (critère binaire ligne). On distingue trois types d'études épidémiologiques en fonction du mode d'échantillonnage décrit précédemment : (i) dans les études prospectives («exposure-based studies»), les totaux des lignes (facteur de risque) sont fixés et on observe la maladie; (ii) dans les études transversales («population-based studies»), le total général est fixé, on observe à la fois la maladie et le facteur de risque; enfin (iii) dans les études rétrospectives («disease-based studies»), les totaux des colonnes (maladie) sont fixés et on observe le facteur de risque. Les trois cas de figure sont analysés, successivement, dans les facettes 5, 6 et 7.

Dans une étude prospective, comme indiqué, (a + b) et (c + d) sont fixés tandis que (a + c) et (b + d) sont observés. On note  $I_1 = a/(a+b)$  le taux d'incidence (c'est-à-dire la proportion

TABLEAU VI. INCIDENCE À 10 ANS DE MALADIES CORONARIENNES CHEZ DES SUJETS MASCULINS (30-59 ANS) EN FONCTION DU TAUX DE CHOLESTÉROL

Cholestérol	Maladie coronarienne sur 10 ans		Total
	Oui	Non	
Elevé	85	462	547
Bas	28	516	544
Total	113	978	1091

de nouveaux cas) de la maladie dans le groupe «exposé» et  $I_0 = c/(c+d)$  le taux d'incidence dans le groupe «non exposé». Dans ces conditions, le risque relatif (RR) est le rapport entre les deux taux d'incidence donné par la formule  $RR = I_1/I_0$ . Lorsque  $RR > 1$ , le facteur d'exposition est un facteur de risque et lorsque  $RR < 1$ , il est un facteur de protection. L'absence d'association se traduit par  $RR = 1$ . En général, on associe à RR son intervalle de confiance à 95% (8).

Jewell (9) rapporte l'incidence à 10 ans de maladies coronariennes chez des hommes de 30 – 59 ans avec cholestérol élevé et avec cholestérol bas dans l'étude de Framingham (tableau VI).

Le taux d'incidence dans le groupe exposé (cholestérol élevé) est  $I_1 = a/(a+b) = 85/547 = 0,155$  (15,5%). Dans le groupe non exposé (cholestérol bas), il s'élève à  $I_0 = c/(c+d) = 28/544 = 0,0515$  (5,15%). Le risque relatif vaut donc  $RR = 0,155/0,0515 = 3,02$  (IC95% : 2,0 – 4,55). Il existe donc une association significative entre le cholestérol et la maladie coronarienne puisque l'IC95% ne recouvre pas la valeur 1. Un cholestérol élevé augmente d'un facteur 3 le risque de développer une atteinte coronarienne. Ceci est confirmé par le test du chi-carré d'homogénéité ( $p < 0,0001$ ).

#### FACETTE 6. PRÉVALENCE ET RISQUE RELATIF

Dans une étude transversale, l'effectif global  $n$  est fixé, dès lors les totaux marginaux ( $a + b$ ), ( $c + d$ ), ( $a + c$ ) et ( $b + d$ ) sont observés. La maladie est observée en même temps que le facteur d'exposition. On note  $P_1 = a/(a+b)$ , le taux de prévalence (c'est-à-dire la proportion de cas existants) de la maladie dans le groupe «exposé», et  $P_0 = c/(c+d)$ , le taux de prévalence dans le groupe «non exposé». Dans ces conditions, le risque relatif (de prévalence) est

donné par la formule  $RR = P_1/P_0$ . Lorsque  $RR > 1$ , le facteur d'exposition est associé à une augmentation du risque et lorsque  $RR < 1$ , il est associé à une diminution du risque. L'absence d'association se traduit par  $RR = 1$ . Comme précédemment, on associe à RR son IC95%.

A titre d'exemple, le même auteur (8) rapporte le niveau de cholestérol et la présence de maladies coronariennes observés chez les 789 sujets toujours dans l'étude mais au moment de sa clôture à 10 ans (voir tableau VII).

La prévalence de la maladie dans le groupe exposé (cholestérol élevé) vaut  $P_1 = a/(a+b) = 38/409 = 0,0929$  (9,29%) et celle dans le groupe non exposé (cholestérol bas),  $P_0 = c/(c+d) = 33/380 = 0,0868$  (8,68%). Le risque relatif (de prévalence) vaut dès lors donc  $RR = 0,0929/0,0868 = 1,07$  (IC95% : 0,69 – 1,67). Il n'y a donc plus d'association significative entre les deux facteurs comme confirmé par le test du chi-carré d'indépendance ( $p = 0,77$ ).

#### FACETTE 7. ODDS RATIO

Dans une étude rétrospective, appelée aussi étude «cas-contrôles», ( $a + c$ ) et ( $b + d$ ) sont fixés alors que ( $a + b$ ) et ( $c + d$ ) sont observés. On ne peut dès lors pas calculer le risque relatif car la division en ligne n'est plus permise ! Théoriquement, on définit la notion de «odds» (ou «cote» en français) qui est le rapport entre la probabilité d'avoir la maladie sur celle de ne pas avoir la maladie. L'odds peut alors être défini dans le groupe exposé, «odds (exposé)», et dans le groupe non exposé, «odds (non exposé)». Le rapport de ces deux quantités, appelé «Odds Ratio (OR)», est une mesure d'association entre la maladie et le facteur d'exposition. En pratique, l'odds ratio, appelé en français «rapport de cotes ou rapport croisé», se calcule à partir de la formule

$$OR = \frac{ad}{bc}$$

que l'on retrouve dans tous les ouvrages d'épidémiologie (7-8). Lorsque  $OR > 1$ , on parle d'association positive (défavorable) entre le facteur d'exposition et la maladie et lorsque  $OR < 1$ , l'association est négative (favorable). L'absence d'association se traduit par  $OR = 1$ . Le calcul de l'IC95% permet de voir si l'intervalle contient ou non la valeur  $OR=1$ . L'odds ratio est devenu tellement populaire qu'il se calcule

TABLEAU VII. PRÉVALENCE DE MALADIES CORONARIENNES EN FONCTION DU CHOLESTÉROL CHEZ LES SUJETS MASCULINS AU TERME DE L'ÉTUDE

Cholestérol	Maladie coronarienne		Total
	Oui	Non	
Elevé	38	371	409
Bas	33	347	380
Total	71	718	789

TABLEAU VIII. ASSOCIATION ENTRE LE DIABÈTE ET LA CATARACTE CHEZ DES SUJETS ÂGÉS DE 50 A 69 ANS

Diabète	Cataracte		Total
	Oui	Non	
Oui	55	84	139
Non	552	1927	2479
Total	607	2011	2618

aussi pour les études prospectives et transversales. Il est aisé de montrer que si  $RR > 1$ , alors  $OR > RR$ ; de même si  $RR < 1$  alors  $OR < RR$ . En d'autres termes, l'OR surpasse toujours un peu la mesure d'association donnée par RR (8).

Hiller et Kahn (10) ont recherché les cas de diabète chez 607 sujets atteints de cataracte et chez 2.011 sujets non atteints de cette affection (voir tableau VIII).

Le calcul de l'odds ratio conduit à la valeur

$$OR = \frac{55 \times 1927}{84 \times 552} = 2,29.$$

suivante :

L'IC 95% correspondant vaut [1,6 – 3,3] et ne recouvre pas la valeur  $RR=1$ . On conclut donc à une association statistiquement significative entre les deux atteintes. Celle-ci est confirmée par le test de chi-carré d'homogénéité ( $p < 0,0001$ ). Le risque de cataracte est donc au moins 2 fois plus élevé chez les personnes diabétiques que chez les non-diabétiques.

#### FACETTE 8. VALEUR DIAGNOSTIQUE D'UN TEST

Lorsqu'on considère un test clinique ou de dépistage, il est utile de le caractériser par sa «valeur diagnostique», c'est-à-dire sa capacité à diagnostiquer ou dépister la maladie à laquelle on s'intéresse (11-12). Comme en épidémiologie, on note Y la maladie en question (absente ou présente), par contre X désigne, à présent, un test ou examen (clinique ou de laboratoire) qui est, soit «négatif», soit «positif». Donc, X est le critère binaire ligne et Y le critère binaire colonne, établissant ainsi un autre type de table 2x2. En général, comme dans les études rétrospectives, les totaux des colonnes (a + c) et (b + d) sont fixés et les totaux marginaux des lignes observés. On échantillonne de X conditionnellement à Y, et seuls les pourcentages verticaux peuvent être calculés. Plusieurs notions fonda-

mentales sont définies à partir de cette table : (i) la spécificité; (ii) la sensibilité; (iii) l'efficacité; (iv) la valeur prédictive positive; et (v) la valeur prédictive négative du test.

La spécificité (SP) du test est sa capacité à être «négatif» chez les sujets «non malades». On a  $SP = a/(a+c)$ . La sensibilité (SE) du test est sa capacité à être «positif» chez les sujets «malades». On a  $SE = d/(b+d)$ . Ces quantités sont souvent exprimées en pourcent. Plus SP et SE sont proches de 100%, meilleur est le test. L'efficacité (EFF) du test est donnée par la formule  $EFF = (SP+SE)/2$ . Celle-ci correspond à l'aire sous la courbe ROC associée au test. La courbe «Receiver Operating Characteristics (ROC)» est obtenue en reportant en ordonnée la valeur SE et en abscisse la valeur  $1 - SP$  du test (13). Dans le cas présent (table 2x2), on obtient un simple quadrilatère dont l'aire, appelée aire sous la courbe (AUC), est égale à  $(SP + SE)/2$ , c'est-à-dire l'efficacité (voir illustration plus loin dans la figure 1).

Afin d'évaluer l'efficacité diagnostique de la mesure de la glycémie (à l'époque par une méthode colorimétrique appelée «test de Folin-Wu») dans le dépistage du diabète, Remein et Wilkerson (14) ont appliqué ce test chez 510 sujets normaux et chez 70 sujets atteints de diabète. Les données sont reprises dans le tableau IX.

La spécificité du test de Folin-Wu vaut  $SP = a/(a+c) = 461/510 = 0,904$  (90,4%) et la sensibilité s'élève à  $SE = d/(b+d) = 56/70 = 0,80$  (80%). Le taux de faux positifs vaut dès lors  $FP = 1-SP = 0,096$  (9,6%) et le taux de faux négatifs  $FN = 1-SE = 0,20$  (20%). L'efficacité globale du test de Folin-Wu est donnée par la formule

$$EFF = (SP+SE)/2 = (0,904+0,80)/2 = 0,90$$
 (90%).

On introduit également le concept de rapport de vraisemblance, en anglais «Likelihood

TABLEAU IX. EVALUATION DIAGNOSTIQUE DU TEST DE FOLIN-WU DANS LE DÉPISTAGE DU DIABÈTE

Test de Folin-Wu	Diabète		Total
	Non	Oui	
Négatif	461	14	475
Positif	49	56	105
Total	510	70	580

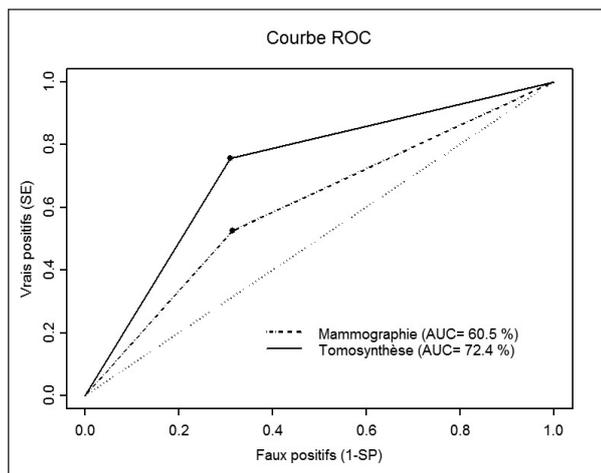


Figure 1. Courbe ROC de la mammographie (2D) et de la tomosynthèse (3D) dans le diagnostic de 187 masses mammaires suspectes

Ratio (LR)», d'un test positif et d'un test négatif. Par définition,  $LR(Positif) = SE/(1-SP)$  et  $LR(Négatif) = (1-SE)/SP$ . Si on désigne par  $\pi$  la «prévalence» (ou probabilité *a priori*) de la maladie dans la population générale, on peut définir la Valeur Prédictive Positive (VPP) du test, c'est-à-dire la probabilité d'être atteint de la maladie Y lorsque le test X est «positif». L'application du théorème de Bayes montre que

$$VPP = \frac{\pi SE}{\pi SE + (1-\pi)(1-SP)}$$

La VPP est encore appelée probabilité *a posteriori* de la maladie pour un test positif. On définit de même la Valeur Prédictive Négative (VPN) du test, c'est-à-dire la probabilité de ne pas être malade en présence d'un test «négatif». On montre par application du théorème de Bayes que

$$VPN = \frac{(1-\pi)SP}{(1-\pi)SP + \pi(1-SE)}$$

Il est important d'insister sur le fait que les quantités VPP et VPN dépendent de la prévalence de la maladie et varient dès lors en fonction de celle-ci. Il est aisé de voir que la VPP augmente avec la prévalence alors que la VPN diminue.

Pour le dépistage du diabète par la méthode colorimétrique de la glycémie selon Folin-Wu, le rapport de vraisemblance d'un test positif est égal à

$$LR(Positif) = SE/(1-SP) = 0,800/0,096 = 8,3.$$

Ceci signifie qu'un test positif est environ 8 fois plus fréquent chez les sujets diabétiques que chez les autres. Le rapport de vraisemblance d'un test négatif vaut

$$LR(Négatif) = (1-SE)/SP = 0,200/0,904 = 0,22.$$

Un test négatif est donc environ 5 fois moins fréquent chez les sujets diabétiques que chez les sujets non diabétiques.

La valeur prédictive positive qui donne la probabilité d'être diabétique en présence d'un test positif vaut, en supposant une prévalence du diabète  $\pi=0,06$  (6%),

$$VPP = \frac{0,06 \times 0,80}{0,06 \times 0,80 + (1-0,06) \times (1-0,904)} = 0,347$$

soit 34,7%. De même, on obtient la valeur prédictive négative en calculant l'expression

$$VPN = \frac{(1-0,06) \times 0,904}{(1-0,06) \times 0,904 + 0,06 \times (1-0,80)} = 0,948$$

soit 94,8%. En conséquence, un test négatif exclut quasiment la maladie.

Dans le diagnostic des masses mammaires suspectes, décrit précédemment, outre la mammographie (2D) et la tomosynthèse (3D), on disposait aussi des résultats de l'anatomopathologie. La spécificité et la sensibilité valaient, respectivement, 70% et 52% pour la mammographie (2D) et 70% et 78% pour la tomosynthèse (3D). Cette dernière possède dès lors aussi une meilleure efficacité diagnostique (AUC) ainsi que le montrent les courbes ROC reprises à la figure I.

## RÉGRESSION LOGISTIQUE ET TABLES 2x2

Aujourd'hui, il est fréquemment fait appel au modèle de régression logistique (RL) pour analyser une table 2x2 (15). Ce modèle repose sur l'équation suivante

$$P(Y = 1|X) = \frac{e^{b_0 + b_1 X}}{1 + e^{b_0 + b_1 X}}$$

où Y et X sont les variables binaires définies précédemment. En introduisant les n données individuelles (Y, X) dans un logiciel statistique, celui-ci fournit directement l'intercept  $b_0$  et le coefficient de régression  $b_1$  ainsi qu'une «p-value».

### RELATION AVEC LE TEST D'HOMOGÉNÉITÉ

Si on utilise la méthode de régression logistique (RL) pour comparer deux proportions indépendantes, on obtient un test chi-carré qui diffère légèrement de celui donné par le test d'homogénéité (facette 1). L'application du modèle logistique aux données du tableau II (tabagisme chez les adolescents) donne un test

dit du «Likelihood Ratio»  $\chi^2_{(1)} = 0,1426$  au lieu de 1,1432 pour le test du  $\chi^2_{(1)}$  d'homogénéité. La p-value vaut 0,2851 au lieu de 0,2580. La différence est minime !

#### RELATION AVEC L'ODDS RATIO

Si on a affaire à une étude rétrospective, l'application de la méthode RL permet d'obtenir immédiatement l'odds ratio (voir facette 7). En effet, on montre aisément que  $OR = \exp(b_1)$ . Notons qu'il n'est pas possible d'obtenir RR. En appliquant la méthode de RL aux données du tableau VIII (cataracte et diabète), on obtient l'intercept  $b_0 = -1,25$  et le coefficient de régression  $b_1 = 0,8267$ . En calculant  $\exp(0,8267) = 2,29$ , on obtient directement l'odds ratio.

#### RELATION AVEC LA VPP

L'application de la méthode de régression logistique aux données du tableau IX (test de dépistage du diabète) donne pour l'intercept et le coefficient de régression, respectivement,  $b_0 = -3,49$  et  $b_1 = 3,63$ . Il est alors aisé de calculer la VPP pour autant que l'on apporte une correction au terme  $b_0$  en fonction de la prévalence  $\pi$  (16). En effet, en supposant comme précédemment  $\pi = 0,06$ , on calcule l'intercept corrigé  $b_0^*$ , soit

$$\begin{aligned} b_0^* &= b_0 - \ln \frac{b+d}{a+c} - \ln \frac{\pi}{1-\pi} \\ &= -3,49 - \ln \frac{70}{510} - \ln \frac{0,06}{0,94} = -4,26 \end{aligned}$$

On obtient alors directement la VPP en introduisant  $b_0^*$  dans l'équation RL et en posant  $X=1$  (test positif) :

$$VPP = P(Y=1|X=1) = \frac{\exp(-4,26 + 3,63 \times 1)}{1 + \exp(-4,26 + 3,63 \times 1)} = 0,35$$

## DISCUSSION

Les tables de contingence 2x2 font partie de la vie quotidienne des médecins et chercheurs. On les retrouve à tous les niveaux de l'analyse statistique d'une base de données. Bien que d'apparence simple, leur interprétation n'est pas toujours aisée. Lorsqu'on exprime les résultats en pourcentages, il est très important de se rappeler que si les totaux des lignes ont été fixés, on ne peut pas calculer les pourcentages en colonne et vice versa. Par contre, lorsque seul le total général  $n$  est fixé, aucune restriction ne s'impose dans le calcul des pourcentages.

Huit facettes différentes ont été décrites des tables 2x2 selon le contexte envisagé (tableau X). La manière dont l'échantillonnage a été réalisé, c'est-à-dire la façon dont la table a été établie, est primordiale : totaux des colonnes fixés (facettes 1, 7 et 8) ou totaux des lignes fixés (facette 5) ou total général fixé (facettes 2, 3, 4 et 6). Il n'y a pas d'autre manière de procéder ! Pour les étudiants, chercheurs, médecins et professionnels de la santé, ces différentes vues sur les tables 2x2 sont essentielles car elles permettent de discerner le problème et la table à laquelle on a réellement affaire.

Il n'est pas inutile de rappeler certains dangers liés à l'interprétation des tables 2x2. S'il est permis d'appliquer le test d'homogénéité (facette 1) lorsque l'échantillon a été obtenu du mélange ( $n$  fixé, facette 2), il est, à l'inverse, proscrit d'appliquer le test d'indépendance lorsque les totaux des colonnes (ou des lignes) sont fixés ! En d'autres termes, il n'est permis de tester l'indépendance de deux critères binaires que s'ils ont été observés en même temps sur les mêmes sujets. Ainsi, dans le tableau II, on ne peut conclure à l'indépendance entre le sexe et le tabagisme, voire calculer le coefficient de corrélation de point. Par contre, dans le tableau III, on peut affirmer que les cas de vomissements sont plus fréquents chez les sujets présentant des nausées (53/126, soit 42,1%) que chez ceux n'en présentant pas (13/545, soit 2,4%).

La distinction entre la comparaison de proportions indépendantes (facette 1) et celle de proportions appariées (facette 3) pose moins de problème. Toutefois, appliquer un test d'homogénéité

TABLEAU X. RÉPARTITION DES TABLES 2x2 EN FONCTION DU TYPE D'ÉCHANTILLONNAGE

Echantillonnage	Totaux fixés	Facettes
X conditionnel à Y	Colonnes	1. Homogénéité
		7. Odds ratio
		8. Test diagnostique
Y conditionnel à X	Lignes	5. RR incidence
Mélange (X,Y)	Total général (n)	2. Indépendance
		3. McNemar
		4. Kappa Cohen
		6. RR prévalence
X=variable ligne, Y=variable colonne		

généité là où le test de McNemar s'impose, ou vice versa, peut conduire à des conclusions aberrantes car ce ne sont pas les mêmes proportions que l'on compare. Ainsi l'application du test d'homogénéité (facette 1) aux données du tableau IV donne  $\chi^2_{(1)} = 12,2$  ( $p=0,0005$ ) mais signifie que la proportion de sujets parcourant > 500 m après intervention est plus élevée chez ceux qui parcouraient déjà > 500 m (43/63, soit 68,3%) que chez ceux qui parcouraient  $\leq$  500 m avant intervention (37/93, soit 39,8%); cette comparaison concerne moins l'efficacité de l'intervention chirurgicale effectuée.

Une confusion peut aussi exister dans le calcul du risque relatif entre les notions d'incidence (facette 5) et de prévalence (facette 6). Pour rappel (9), la prévalence d'une maladie est la proportion de sujets d'une population à risque potentiel de cette maladie qui sont effectivement atteints de la maladie à un moment donné (cas existants). Par contre, l'incidence d'une maladie est la proportion de sujets d'une population à risque potentiel de cette maladie qui sont non atteints de la maladie au début d'une période de temps donnée, mais qui la développent avant la fin de cette période de temps (cas nouveaux). Dans les études prospectives (facette 5), on exclut tous les cas de prévalence au début de l'étude et on inclut tous les nouveaux cas durant la période de suivi. Dans le cas d'études transversales (facette 6), on inclut tous les cas de prévalence, c'est-à-dire atteints de la maladie au moment de l'étude. Notons que dans les études rétrospectives (facette 7), ces notions disparaissent puisqu'on sélectionne soi-même les sujets malades et non malades.

En conclusion, lorsqu'on parcourt la littérature scientifique, il n'est pas inutile de s'arrêter devant une table 2x2 et de se demander à laquelle des 8 facettes celle-ci correspond. D'ailleurs, il est toujours facile de vérifier les conclusions des auteurs puisque l'ensemble des calculs décrits dans cet article peut se faire à la main avec une simple calculatrice manuelle.

## BIBLIOGRAPHIE

1. Albert A.— *Biostatistique*. Presses Universitaires de Liège, Liège 2005.
2. Yates F.— Contingency tables involving small numbers and the  $\chi^2$  test. *J Roy Stat Soc Suppl*, 1934, **1**, 217-235.
3. Fisher RA.— On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P. *J Roy Stat Soc*, 1922, **85**, 87-94.
4. Boogaerts JG, Bardiau FM, Seidel L, et al.— Tropisetron in the prevention of postoperative nausea and vomiting. *J Clin Anesthesia*, 2000, **12**, 402-408.
5. McNemar Q.— Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 1947, **12**, 153-157.
6. Cohen J.— A coefficient of agreement for nominal scales. *Educ Psychol Meas*, 1960, **20**, 37-46.
7. Kahn HA, Sempos CT.— *Statistical methods in Epidemiology*. Oxford University Press, Oxford, 1989
8. Scheen AJ, Ernest P, Jandrain B.— Comment j'explore ... une différence de risque de survenue d'un événement dans les études cliniques. *Rev Med Liège*, 2012, **67**, 597-602.
9. Jewell NP.— *Statistics for epidemiology*. Chapman & Hall/CRC Press, New York, 2004.
10. Hiller R, Kahn HA.— Senile cataract extraction and diabetes. *Br J Ophthalmol*, 1976, **60**, 283-286.
11. Galen RS, Gambino SR.— *Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses*. Wiley, New York, 1975.
12. Albert A, Harris EK.— *Multivariate interpretation of clinical laboratory data*. Dekker, New York, 1987.
13. Swets JA.— ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiol*, 1979, **14**, 109-121.
14. Remein QR, Wilkerson HLC.— The efficiency of screening tests for diabetes. *J Chron Dis*, 1961, **13**, 6-21.
15. Hosmer DW, Lemeshow S.— *Applied logistic regression*. Wiley, New York, 2000.
16. Albert A.— On the use and computation of likelihood ratios in clinical chemistry. *Clin Chem*, 1982, **5**, 1113-1119.

Les demandes de tirés à part sont à adresser au Pr A. Albert, Service d'Informatique médicale et Biostatistique, CHU de Liège, Belgique.  
Email : aalbert@ulg.ac.be