Les Cahiers de l'Analyse des Données
Vol. XXI - 1996 - n°1 - pp.. 45-52.

SUR LES TRAITS ASSOCIÉS PAR PAIRES: MALFORMATIONS CARDIAQUES CONGÉNITALES CHEZ DES ENFANTS AYANT MÊMES PARENTS

[PAIRES MALFORMATIONS]

*J. de TIBEIRO**

1 Les données cardiologiques

1.1 Position du problème

Afin de cerner des relations étiologiques entre les diverses catégories de malformations cardiaques congénitales, F. CLARKE FRASER et ALASDAIR D. W. HUNTER, (1975), ont considéré la distribution de ces malformations chez des sujets consanguins, essentiellement des enfants ayant les mêmes parents (ou un seul parent commun).

On ne sera pas surpris d'apprendre que, chez les frères ou sœurs d'enfants atteints, la fréquence globale des malformations est plus forte que dans la population générale: de façon précise, tandis que dans cette population, la fréquence des malformations s'établit autour de 1%, (pour des enfants observés pendant plusieurs années depuis leur naissance,) le taux est estimé entre 3% et 4% pour les frères ou sœurs d'un patient. On admettra, de même, que la malformation est la même plus souvent que si l'atteinte pathologique avait frappé les deux sujets indépendamment l'un de l'autre: en fait, il s'agit, une fois sur deux de la même malformation. Reste le cas de deux malformations différentes: cette éventualité, considérée à elle seule, pour le frère ou la sœur d'un patient, ayant une fréquence de 1,5% à 2%, dépasse encore la fréquence générale qui est de 1%. La question se pose donc d'apprécier dans quelle mesure, même si les deux malformations sont distinctes, il y a entre elles une certaine similitude.

Les auteurs cités ont fait eux-mêmes, à ce propos, d'importantes observations sur les données collationnées par eux. Mais, comme il s'agit d'une distribution statistique complexe, un autre auteur, Brenda MacGIBBON, (1981), a repris l'analyse de ces données. Et nous croyons pouvoir y revenir ici, du point de vue de l'analyse des correspondances.

(*) Professeur à l'Université de Moncton à Shippagan; Secteur des Sciences;
C.P. 2000 ; Shippagan, Nouveau-Brunswick; CANADA

1.2 Le tableau disponible

Le tableau distingue 13 malformations, que nous énumérons en suivant une terminologie latine ou française dont l'usage nous paraît bien établi.

TdF : Tétralogie de FALLOT
VSD : Ventricularis S pti Defectus
PS : Pulmonalis Stenosis
TGV : Transposition des Gros Vaisseaux
PDA : Patens Ductus Arteriosus
AS : Aortica Stenosis
ASD : Auricularis S pti Defectus
TRU : Truncus Arteriosus
TA : Tricuspidis Atresia
COA : Coarctatio Aortica
DEX : Dextroposition de l'origine de l'Aorte
PTR : Pseudo Truncus Arteriosus
AV : Canal Atrio-Ventriculaire

Après les auteurs cités, nous devons signaler une difficulté qui, dans l'élaboration statistique, n'est pas explicitement prise en compte par nous: un même sujet peut être atteint d'une malformation multiple. En particulier, le nom même de Tétralogie de FALLOT, signale la conjonction de quatre signes, dont trois sont aussi comptés séparément ici: rétrécissement (sténose: PS) de la voie artérielle pulmonaire, communication intraventriculaire (cf. VSD), dextroposition aortique (DEX); le dernier signe, hypertrophie ventriculaire droite (non recensée sur le tableau), n'étant que la conséquence des autres (principalement du premier).

Dans la publication de 1975, les deux auteurs cités précisent comment sont dénombrés les faits dans le tableau repris ici: nous y reviendrons au §4, en conclusion.

coexistence de malformations cardiaques

13	TdF	VSD	PS	TGV	PDA	AS	ASD	TRU	TA	COA	DEX	PTR	AV
TdF	0	13	19	10	4	1	1	0	1	0	1	2	0
VSD	13	0	3	5	3	3	6	1	0	0	2	1	0
PS	19	3	0	2	0	1	1	3	1	0	0	0	0
TGV	10	5	2	0	4	1	2	1	0	1	0	0	0
PDA	4	3	0	4	0	2	0	1	2	0	0	0	1
AS	1	3	1	1	2	0	2	0	1	3	2	0	0
ASD	1	6	1	2	0	2	0	0	1	1	0	0	1
TRU	0	1	3	1	1	0	0	0	0	0	0	1	0
TA	1	0	1	0	2	1	1	0	0	0	0	0	0
COA	0	0	0	1	0	3	1	0	0	0	0	0	0
DEX	1	2	0	0	0	2	0	0	0	0	0	0	0
PTR	2	1	0	0	0	0	0	1	0	0	0	0	0
AV	0	0	0	0	1	0	1	0	0	0	0	0	0

Puisque k(i, i') est le nombre de cas concernant deux sujets consanguins dont l'un est atteint de i, l'autre de i', le tableau est symétrique. La diagonale est nulle, parce qu'excluant les cas de deux sujets présentant la même malformation, on dénombre seulement les cas où il s'agit de malformations distinctes. Par le fait, tel quel, le tableau se prête mal l'analyse. C'est ce qui nous a fait penser à construire une diagonale, en nous inspirant d'une méthode qui a déjà fait l'objet de quelques publications (in *CAD*, 1992).

2 Méthode statistique: la reconstitution d'une diagonale manquante

Rappelons qu'on a un corpus de paires (i, i'), $i \neq i'$; i.e., pour tout (i, i') est donné le nombre de fois que celle-ci est attestée, soit k(i, i'). Le corpus est présenté comme un tableau symétrique $I \times I$, dont la diagonale est nulle; le total des cases non nulles étant alors le double du nombre total des paires du corpus.

On s'interroge sur l'existence d'une association entre i et i'. Pour l'hypothèse opposée de l'indépendance, on prendra pour modèle qu'il existe une loi p_I sur I telle que l'ensemble donné de paires puisse être considéré comme un échantillon issu de la loi définie par les termes extradiagonaux de $p_I \times p_I$.

De façon précise, si cardI = n, l'ensemble C des paires a pour cardinal n(n-1)/2; l'ensemble S(C) des lois p_C sur C et l'ensemble S(I) des lois p_I sur I dépendent, respectivement, de (n(n-1)/2)-1 et de (n-1) paramètres. À toute loi p_I sur I, $p_I \in S(I)$, est associée (par restriction aux termes extradiagonaux) une loi $£(p_I) \in S(C)$; on notera £(C) l'ensemble des lois de cette forme.

Pour la loi p_C sur C associée au corpus donné de paires, on a, dans £(C), une loi $\pi_C = £(\pi_I)$, qui est la plus proche de p_C: e.g. réalise le minimum de Δ, distance du chi^2 calculée avec π_C pour centre. Compte tenu des dimensions, Δ^2 est un chi^2 à ((n(n-1)/2)-1) - (n-1) = n(n-3)/2. Plus exactement, on doit rapporter à la loi de chi^2, la distance carrée Δ^2 entre les lois (considérées sur C) multipliée par l'effectif du corpus de paires.

Pour chercher, approximativement, π_I, on opère sur le tableau k_{II} symétrique à diagonale nulle; et on y crée une diagonale, telle que soit satisfaite la condition suivante: au tableau complété est associée une loi π_{II} sur $I \times I$; et le carré $\pi_I \times \pi_I$ de la loi marginale π_I, de π_{II}, a même diagonale que π_{II}.

Le §2 de [TRAC. MANQ. FLUX] explique les étapes de ce calcul. On retient, pour seule donnée, la loi marginale, k_I, du tableau à diagonale nulle: on caractérise les π_i par la condition que la loi $\{\pi_i-(\pi_i)^2 \mid i \in I\}$ (qui n'est autre que la loi marginale du tableau déduit de $\pi_I \times \pi_I$ en y annulant la diagonale), est proportionnelle à k_I.

Le calcul ne peut, en général, produire une loi πI telle que $£(\pi_I) = p_C$ (i.e. que $\pi_I \times \pi_I$ et k_{II}, restreints au complémentaire de la diagonale, soient proportionnels). Mais en réalisant l'égalité des lois marginales, on a imposé à π_I le plus de contraintes qu'il est possible pour rendre $£(\pi_I)$ proche de p_C.

Du point de vue de la validité statistique, la distance du chi^2 calculée sera supérieure ou égale au minimum: donc, si l'épreuve de validité classique conclut à une différence non significative entre $£(\pi_I)$ et p_C, on saura, *a* *fortiori*, que le corpus étudié ne permet pas d'affirmer une corrélation entre les deux termes d'une paire.

Maintenant, l'analyse de correspondance de k complété (i.e. de π_{II}) affiche (comme de règle), pour trace, la distance (carrée) du chi^2 entre π_{II} et $\pi_I \times \pi_I$, considérées comme lois sur $I \times I$; or la distance du chi^2 entre les deux lois restreintes à C se calcule aisément en fonction de cette trace; en effet, au calcul de celle-ci, les termes diagonaux ne contribuent pas; les autres termes sont ceux qu'il faut pour calculer la distance entre p_C et $\pi_I \times \pi_I$ restreinte à C; à ceci près que dans la trace, figurent des fréquences sous-éstimées dans un rapport égal à celui du total des termes extradiagonaux de k au total des termes de k complété. En multipliant la trace par l'inverse de ce rapport, on a la distance carrée Δ^2 cherchée (en effet, les termes de la distance du chi^2, Δ^2, sont de la forme générale: quotient du carré d'une fréquence par une fréquence). Il reste que ce que nous appelons "distance (carrée) du chi^2" doit être multiplié par le nombre des cas avant de consulter les tables usuelles.

Ces calculs sont faits, au §3, sur l'exemple des données de pathologie cardiaque congénitale. Il faut d'autre part tenir compte de ce que, parmi les facteurs issus de l'analyse du tableau k complété, ne sont suceptibles d'interprétation que les facteurs directs: ce point sera considéré au §3.1, sur l'exemple de nos données.

3 Application de la méthode aux données de cardiologie

Dans le tableau donné à diagonale nulle, les six dernières colonnes ont un poids très faible ce qui compromet la comparaison avec le modèle d'indépendance. L'on considère donc, après le tableau complet, 13×13, le sous-tableau, 7×7, afférent aux malformations les plus fréquentes.

3.1 Recherche de corrélations entre les 13 malformations recensées

13	TdF	VSD	PS	TGV	PDA	AS	ASD	TRU	TA	COA	DEX	PTR	AV
marge	52	37	30	26	17	16	15	7	6	5	5	4	2

loi marginale du tableau 13×13 : total de la marge : 222

Dans le tableau présenté au §1.2, le total des cases est 222 (ce qui, compte tenu de la symétrie, correspond à 111 paires; avec moins de 222 sujets au total, du fait des malformations multiples).

TdF	VSD	PS	TGV	PDA	AS	ASD	TRU	TA	COA	DEX	PTR	AV
19670	7626	4579	3284	1280	1123	978	198	145	100	100	63	16

NB les nombres calculés pour la diagonale sont donnés, ci-dessus, en millèmes.

La méthode considérée au §2 fournit pour termes diagonaux des nombres réels positifs, dont le plus élevé (afférent à la malformation la plus fréquente) est de l'ordre de 20; le total de la diagonale approchant de 40. [Ceci nous rappelle que, selon une estimation citée au §1.1, le nombre des cas de paires de sujets consanguins atteints de la même malformation est quasi égal à celui des paires avec malformations différentes; la diagonale touvée confirme que, sans présumer de corrélations entre malformations distinctes, les cooccurrences de malformations identiques sont plus nombreuses que s'il y avait totale indépendance: 40 << 222].

répartition de 13 malformations cardiaques : tableau avec diagonale ≠ 0 ;

trace : 8.445e-1

rang :	+1	-2	-3	-4	+5	+6	+7	-8	+9	+10	
lambda :	2498	1626	1284	1201	704	352	282	250	125	96	e-4
taux :	2958	1926	1520	1423	833	417	334	296	148	114	e-4
cumul :	2958	4884	6404	7827	8660	9077	9411	9707	9855	9969	e-4

Puisque la matrice analysée est symétrique, chaque facteur est soit direct ($\forall$i : F(i) = G(i); même valeur du facteur pour un élément i, que celui-ci soit représenté par une ligne ou par une colonne); soit inverse ($\forall$i : F(i) = -G(i); valeur opposée selon que l'élément i est considéré comme ligne ou comme colonne).

Dans le cas présent, seul un facteur direct est susceptible d'être interprété: comme exprimant une tendance à ce que les malformations sortant du même côté de l'axe, se trouvent associées chez des sujets apparentés. Avec un facteur inverse, on devrait conclure que la possession par un sujet d'une des malformations sortant d'un côté déterminé de l'axe, constitue, pour les parents de ce sujet une relative portection contre les malformations de ce groupe; ce qui paraît inexplicable.

Ici, le facteur 1 est direct; mais, comme l'indiquent les signes portés sur le tableau des valeurs propres, les facteurs suivants, de rang 2, 3 et 4, sont inverses. Avant tout calcul de critère de validité, on sait donc que seul le facteur F1 est susceptible d'interprétation.

PTR	TRU	PS	TdF	TGV	VSD	PDA	TA	ASD	DEX	AV	AS	COA
-613	-538	-513	-379	-14	+92	+243	+423	+635	+744	+885	+1009	+1490

On trouve, en effet, vers (F1>0), AS et COA; i.e. sténose (ou rétrécissement) de l'aorte, et coarctation de l'aorte: or cette dernière est encore décrite comme un “sténose isthmique de l'aorte”; et, vers (F1<0), TRU et PTR: i.e. absence de cloisonnement entre aorte et artère pulmonaire, un seul vaisseau donnant des branches pour la circulation coronarienne, pulmonaire et systèmique (distribuant le sang oxygéné); ou, tableau circulatoire équivalent, mais dû à l'atrésie (occlusion de l'artère pulmonaire); malformations qu'il n'est pas surprenant de trouver sur l'axe 1, proches de la tétralogie de FALLOT et de la simple sténose pulmonaire.

Reste à effectuer les calculs de distance suggérés au §2.

On a : total des termes sur le carré, k $\approx$ 261 ;

total, extradiagonal, sur nos cas, k' = 222 ;

la distance carrée(ou trace de l'analyse) est à multiplier par 261/222; d'où la nouvelle distance carrée : $\Delta^2 = (261/222) \times 0,8445 = 0.993 \approx 1$.

Il s'agit d'un chi^2 à 78 cases (e.g. sous la diagonale) soit 77 paramètres (degrés de liberté) pour S(C); mais il y a pour la loi, dans S(I), 12 paramètres, donc, pour Δ^2, un nombre de degrés de liberté:

$$77\text{-}12 = 65 = 13 \times (13\text{-}3)/2 \quad ;$$

conformément à la formule générale du §2.

Il y a 111 cas :

donc notre chi^2 (critère) est: $111 \times \Delta^2 \approx 110$.

Compte tenu du nombres de paramètres, est distribué quasi normalement:

$$\sqrt{(2\times \text{chi}^2)} - \sqrt{((2\times 65) - 1)} \approx \sqrt{(220)} - \sqrt{(129)} \approx 14,8 - 11,3 \approx 3,6 \ ;$$

on est au delà du seuil à 1‰. Mais si l'on élimine la première valeur propre, le chi^2 restant est multiplié par $\approx 0,70$ (part des autres v.p. dans la trace); la quantité à considérer devient: $\sqrt{(154)} - \sqrt{(129)} \approx 12,4 - 11,3 \approx 1,1$; valeur qui n'est aucunement significative; ce qui s'accorde avec le fait qu'il n'y a pas lieu d'interpréter les facteurs de rang ≥ 2.

3.2 Recherche de corrélations entre 7 malformations fréquentes

coexistence de malformations cardiaques

7	TdF	VSD	PS	TGV	PDA	AS	ASD
TdF	0	13	19	10	4	1	1
VSD	13	0	3	5	3	3	6
PS	19	3	0	2	0	1	1
TGV	10	5	2	0	4	1	2
PDA	4	3	0	4	0	2	0
AS	1	3	1	1	2	0	2
ASD	1	6	1	2	0	2	0

Les calculs du §3.1 sont repris avec un tableau 7 × 7.

7	TdF	VSD	PS	TGV	PDA	AS	ASD
marge	48	33	26	24	13	10	12

loi marginale du tableau 7×7 : total de la marge : 166

Les termes diagonaux calculés sont du même ordre qu'au §3.1.

7	TdF	VSD	PS	TGV	PDA	AS	ASD
diag	26117	7986	4391	3629	921	527	776

NB les nombres calculés pour la diagonale sont donnés, ci-dessus, en millèmes.

Après élimination des modalités de faible fréquence, on a des valeurs propres moindres qu'au §3.1.

répartition de 7 malformations cardiaques : tableau avec diagonale ≠ 0 ;

trace : 2.931e-1

rang	:	+1	-2	-3	+4	-5	+6	
lambda	:	1449	637	518	269	56	1	e-4
taux	:	4944	2174	1769	918	191	4	e-4
cumul	:	4944	7118	8887	9805	9996	10000	e-4

PS	TdF	TGV	VSD	PDA	ASD	AS
-466	-310	+140	+309	+365	+663	+674

On trouve associées, vers (F1<0), PS, la sténose pulmonaire, et TdF, la tétralogie de FALLOT, dont PS est l'une des composantes; tandis que, du côté (F1>0), sont des malformations dont trois {ASD, PDA, VSD} sont des défauts des barrières - ou septa - séparant les chambres cardiaques en cavités droites et gauches.

Restent les calculs de distance.

On a : total des termes sur le carré, k $\approx$ 210 ;

total, extradiagonal, sur nos cas, k' = 166 ;

la distance carrée(ou trace de l'analyse) est à multiplier par 210/166; d'où la nouvelle distance carrée : $\Delta^2 = (210/166) \times 0{,}2931 = 0.37$.

Il s'agit d'un chi^2 à 21 cases (e.g. sous la diagonale) soit 20 paramètres (degrés de liberté) pour S(C); mais il y a pour la loi, dans S(I), 6 paramètres, donc, pour Δ^2, un nombre de degrés de liberté:

$$20 - 6 = 14 = 7 \times (7 - 3)/2 \quad ;$$

conformément à la formule générale du §2.

Il y a 83 cas :

donc notre chi^2 (critère) est: $83 \times \Delta^2 \approx 30{,}8$.

Un chi2 à 14 degrés de liberté dépasse 30,8 avec une prob <1% .

Mais si l'on élimine la première valeur propre qui apporte la moitié de l'inertie, le chi^2 à diviser par 2: 15,4; valeur très proche de la moyenne pour un chi^2 à 14 degrés de liberté. Ce qui, comme au §3.1, s'accorde avec le fait que seul le facteur 1 est susceptible d'interprétation.

4 Conclusion

Le tableau des associations de malformations différentes a été complété par une diagonale non nulle d'après le critère de la trace minima. Les épreuves de validité statistiques, fondées sur des modèle probabilistes usuels, s'accordent avec l'interprétation de l'analyse de correspondance, pour signaler des associations plausibles.

Mais il faut reconnaître que, dans la mesure même où ces associations sont interprétables du point de vue de la cardiologie, elles peuvent résulter, en partie, de la terminologie suivant laquelle sont décrites les malformations, souvent complexes. En effet, dans le tableau reçu de F. CLARKE FRASER et coll., un sujet atteint de plusieurs lésions est récensé comme n'en ayant qu'une seule, considérée comme "majeure" en ce sens qu'elle peut être à l'origine de la malformation prise dans son ensemble.

Références bibliographiques

K. Ben SALEM: "Le critère de la trace minima pour estimer les données manquantes dans un tableau de correspondance"; [TRAC. MANQ.], in *CAD*, Vol. XVII, n°1, pp. 97-112; (1992);

K. Ben SALEM, M.-M. THOMASSIN: "Estimation de la diagonale d'un tableau carré d'après le critère de la trace minima: application à l'analyse des flux de population à la Martinique entre 1975 et 1982"; [TRAC. MANQ. FLUX],in *CAD*, Vol. XVII, n°3, pp. 261-274; (1992);

F. CLARKE FRASER, ALASDAIR D. W. HUNTER: "Etiologic Relations Among Categories of Congenital Heart Malformations", in *The American Journal of CARDIOLOGY*, Vol. 36, pp. 793-796; (1975);

Brenda MacGIBBON: "A log-linear model of a paired sibling study"; in *Proc. Statistics'81 Canada Conference* (Montréal), pp. 193-197; (1981);