

REVUE DE STATISTIQUE APPLIQUÉE

J. P. NAKACHE

A. GUEGUEN

H. PIERART

Utilisation du modèle logistique dans l'étude de l'influence des variables initiales et du traitement sur l'évolution de l'acouphène

Revue de statistique appliquée, tome 34, n° 2 (1986), p. 5-14

http://www.numdam.org/item?id=RSA_1986__34_2_5_0

© Société française de statistique, 1986, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

UTILISATION DU MODÈLE LOGISTIQUE DANS L'ÉTUDE DE L'INFLUENCE DES VARIABLES INITIALES ET DU TRAITEMENT SUR L'ÉVOLUTION DE L'ACOUPHÈNE

J.P. NAKACHE (1), A. GUEGUEN (1), H. PIERART (2)

(1) *Unité de Recherche INSERM U.88*

(2) *Institut IPSEN*

RÉSUMÉ

Le modèle logistique en pas à pas avec prise en compte d'interactions a été utilisé en vue de déterminer les variables pronostiques d'une amélioration de l'acouphène (acouphène isolé de moins d'un an) et d'apprécier l'effet des traitements : Tanakan, Placebo, produits de comparaison.

Dans ce modèle, la variable à expliquer est l'évolution globale de l'acouphène considérée dans cette étude sous la forme d'une variable à deux modalités — aggravation ou pas de changement, amélioration nette — et les variables explicatives sont des mesures quantitatives et qualitatives relevées sur les sujets au moment de leur inclusion dans l'essai.

Les scores d'amélioration de l'acouphène, indépendamment du traitement, sont fonction de l'ancienneté et des interactions ancienneté-siège et ancienneté-périodicité. Seul le Tanakan augmente le score d'amélioration; et l'évolution globale de l'acouphène est semblable sous Placebo et sous produits de comparaison.

SUMMARY

The stepwise logistic regression with interactions has been used in order to point out the variables responsible of a tinnitus improvement and to appreciate the treatment effect — Tanakan, Placebo, therapeutic equivalencies —

In this model, the dependent variable represents the global tinnitus improvement taken here as a variable with two categories (no change or worsening, improvement) and the independent variables are quantitative and qualitative measurements collected on the subjects when entering in the clinical trial.

Irrespective of the treatment, the tinnitus improvement scores are function of the tinnitus age and the interactions tinnitus age-tinnitus side, tinnitus age-tinnitus periodicity. Only the Tanakan treatment increases the improvement score, and, the global tinnitus evolution is similar under Placebo and under therapeutic equivalencies.

INTRODUCTION

Le but de cet article est de présenter une partie des résultats de l'étude multicentrique comparative du traitement des acouphènes effectuée par l'Institut IPSEN. Cette partie concerne l'utilisation du modèle logistique, avec prise en compte d'interactions, pour étudier l'influence du traitement et des varia-

bles de description de l'acouphène — relevées au moment de l'inclusion des sujets dans l'essai — sur l'évolution globale de l'acouphène.

Très fréquents, les acouphènes s'expriment, de façon très variable, du simple bourdonnement d'oreille perçu uniquement au repos jusqu'au bruit intense insupportable perturbant toute activité. Il s'agit, dans cette étude, de patients suivis en ambulatoire et présentant des acouphènes isolés et récents (d'une durée de moins d'un an).

Cette étude nous a permis de mettre en évidence, d'une part la supériorité du Tanakan sur les produits de comparaison et, d'autre part, de déterminer les variables pronostiques d'une amélioration servant à caractériser différentes classes de sujets d'amélioration décroissante de l'acouphène.

DONNÉES DE L'ÉTUDE

En vue de mettre en évidence les variables pronostiques d'une amélioration de l'acouphène et d'apprécier les effets des traitements, nous avons pris en compte dans nos analyses les données des deux protocoles de l'étude : Tanakan T^(P) comparé au Placebo (P), et Tanakan T^(PC) face aux deux produits de comparaison (P.C. 1, P.C. 2).

La variable privilégiée de l'étude, ou variable à expliquer, est l'évolution globale de l'acouphène mesurée à la dernière consultation, et, considérée dans nos analyses sous la forme d'une variable à deux états : aggravation ou pas de changement et amélioration nette.

La distribution de cette variable suivant les traitements est la suivante :

	T ^(P)	Plac.	T ^(PC)	P.C. 1	P.C. 2	Total
Aggravation ou pas de changement	24	27	44	33	33	161
Amélioration nette	31	18	76	21	14	160
Total	55	45	120	54	47	321

En ce qui concerne les variables explicatives, nous avons utilisé une analyse discriminante barycentrique qui nous a permis de sélectionner, parmi les nombreuses variables de description de l'acouphène relevées au moment de l'inclusion des sujets dans l'essai, 9 variables en raison de leur liaison plus ou moins importante avec l'évolution globale de l'acouphène. Six de ces variables sont binaires (siège, périodicité, composition, rythme, hypertension et travail), deux variables sont qualitatives ordinales (sévérité et intensité de la gêne) et la dernière variable est quantitative (âge). La variable explicative traitement est prise en compte dans le modèle sous la forme de quatre variables binaires T^(P), T^(PC), P.C. 1 et P.C. 2 référencées au Placebo : ainsi, pour un sujet ayant reçu le Placebo les valeurs de ces quatre variables sont nulles.

Le tableau 1 fournit la liste des variables (et leur codage).

TABLEAU 1

Liste des variables de l'étude et leur codage

Variables	Codage des variables
Siège	0 = atteinte unilatérale 1 = atteinte bilatérale
Périodicité	0 = permanent 1 = intermittent
Composition	0 = son unique 1 = sons multiples
Rythme	0 = continu 1 = rythmé
Hypertension	0 = non 1 = oui
Travail	0 = non 1 = oui
Sévérité	1 = légère 2 = modérée 3 = gênante 4 = sévère
Gène (intensité)	(modalités ordonnées de 0 à 3)
Ancienneté	anc. de l'acouphène (en jours)
Tanakan (Plac.)	0 = non 1 = oui
Tanakan (P. comp.)	0 = non 1 = oui
P. comp. 1	0 = non 1 = oui
P. comp. 2	0 = non 1 = oui
Evol. globale	0 = aggravation ou pas de changement 1 = amélioration nette

RÉGRESSION LOGISTIQUE

Le modèle logistique fournit une estimation de la probabilité d'amélioration nette de l'acouphène P (A.N./X) pour un sujet X dont le profil est (X_1, X_2, \dots, X_p) . Cette probabilité est telle que :

$$\log \frac{P}{1-P} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

où β_j ($j = 1, 2, \dots, p$) sont les coefficients de régression du modèle.

P (A.N./X) est d'autant plus grande que la valeur de la combinaison linéaire :

$$S(X) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

est grande.

La régression logistique permet :

- d'estimer les coefficients β_i du modèle;
- de tester la significativité de chaque coefficient ($H_0 : \beta_i = 0$);
- de tester l'égalité de deux coefficients ($H_0 : \beta_i = \beta_j$);
- de comparer deux modèles;
- de vérifier l'adéquation du modèle.

Les différentes régressions logistiques présentées, ont été effectuées en pas à pas décroissant avec élimination successive des variables basée sur le critère de la vraisemblance ($C = -2 \log V$). Le test du rapport de vraisemblance permettant de comparer deux modèles a été utilisé pour définir la règle d'arrêt de la procédure du pas à pas.

1^{ère} régression logistique : à partir des 13 variables explicatives

248 sujets ne présentent pas de données manquantes pour les 13 variables explicatives du tableau 1. Une première régression logistique en pas à pas, effectuée à partir du fichier (248×13), nous a conduit aux résultats fournis dans le tableau 2 suivant :

TABLEAU 2
Régression logistique en pas à pas effectuée à partir du tableau (248×13)

Variables	Test (Student) des coeff. de régression	Significativité du test
Ancienneté	-3,79	$p < 0,001$
Siège	-2,85	$0,01 < p < 0,001$
Tanakan (P.C.)	2,26	$0,05 < p < 0,02$
Périodicité	2,12	$0,05 < p < 0,02$
Composition	-2,11	$0,05 < p < 0,02$
Sévérité	1,71	non signif.
Produit 2	1,47	non signif.
Tanakan (Plac.)	1,41	non signif.
Produit 1	-0,09	non signif.

La valeur du critère de la vraisemblance associé au modèle avec les 13 variables du tableau 1 est $C = -144,46$. Pour le modèle avec les 9 variables du tableau 2, on obtient $C = -145,23$.

Le test du rapport de vraisemblance entre ces deux modèles étant non significatif, on ne conservera dans la suite que ces 9 variables.

L'élimination de 4 variables a permis d'augmenter le nombre de sujets sans données manquantes qui passe à 307. Dans le but d'affiner les résultats de la régression logistique précédente, nous avons effectué une régression logistique en pas à pas à partir du tableau (307×9) conduisant à des résultats semblables aux précédents.

2^e regression logistique

Elle a été effectuée en éclatant les variables :

- sévérité de l'acouphène, en raison de son importance dans l'étude et malgré le test non significatif du coefficient de régression qui lui est associé
- ancienneté de l'acouphène, dans le but de tester la linéarité de son effet.

Le découpage de ces variables et leur transformation en variables binaires est le suivant :

Découpage de l'ancienneté en 4 classes ≤ 30 jours 31-70 jours 31-160 jours > 160 jours	Transformation en 3 variables binaires : ANC1, ANC2, ANC3 ANC1 = 1 si ANC (31-70) 0 sinon ANC2 = 1 si ANC (71-160) 0 sinon ANC3 = 1 si ANC (> 160) 0 sinon
Découpage de la sévérité en 4 classes 1) Légère 2) Modérée 3) Gênante 4) Sévère	Transformation en 3 variables binaires : SEV1, SEV2, SEV3 SEV1 = 1 si Modéré 0 sinon SEV2 = 1 si Gênant 0 sinon SEV3 = 1 si Sévère 0 sinon

Remarque : pour un sujet ayant un acouphène léger < 30 jours
 ANC1 = 0 ANC2 = 0 ANC3 = 0
 SEV1 = 0 SEV2 = 0 SEV3 = 0

Les résultats de cette régression logistique sont présentés dans le tableau 3 suivant :

TABLEAU 3

Régression logistique en pas à pas avec Ancienneté et Sévérité transformées en variables binaires

Variable	t (Student)	Significativité
ANC3	-4,09	p < 0,001
Siège	-3,70	p < 0,001
Périodicité	2,67	p = 0,001
ANC2	-2,66	p = 0,01
Tan. (P.C.)	2,34	p = 0,02
ANC1	-2,25	p = 0,02
Tan. (Plac.)	1,69	non signif.
Composition	-1,66	non signif.
Produit 2	-1,27	non signif.
SEV2	1,00	non signif.
SEV1	1,68	non signif.
Produit 1	-0,22	non signif.
SEV3	0,03	non signif.

L'examen du tableau 3 montre que :

- les coefficients de régression associés aux variables ANC1, ANC2 et ANC3 sont significativement différents de 0,
- le siège et la périodicité influent significativement sur l'évolution globale,
- la sévérité, qui n'est pas significative sous forme semi-quantitative, le demeure quand elle est découpée en classes,

- la composition n’est pas loin de la significativité,
- le Tanakan (face aux produits de comparaison) influe significativement et favorablement sur l’évolution globale de l’acouphène,
- le Tanakan (face au Placebo) est à la limite de la significativité.

Remarques

a) les résultats du tableau 3 concernant l’ancienneté nous ont conduit à tester la linéarité de l’effet de cette variable. Pour cela nous avons effectué les tests des hypothèses nulles : $\beta_{\text{ANC1}} = \beta_{\text{ANC2}}$, $\beta_{\text{ANC2}} = \beta_{\text{ANC3}}$ et $\beta_{\text{ANC1}} = \beta_{\text{ANC3}}$. Ces tests ne sont pas significatifs ce qui indique la non linéarité de l’effet de l’ancienneté sur l’évolution de l’acouphène. En conséquence, nous avons considéré dans la suite de l’étude la variable ancienneté (ANC) sous forme de variable à deux modalités (0 si $\text{ANC} \leq 30$ jours et 1 si $\text{ANC} > 30$ jours).

b) la comparaison des modèles : le modèle avec les variables traitement $T^{(\text{PC})}$, $T^{(\text{P})}$, P.C. 1 et P.C. 2 et le modèle où sont regroupés les deux Tanakan d’une part, et les deux produits de comparaison d’autre part, conduit à une différence non significative ; d’où leur regroupement dans les analyses suivantes en deux variables traitement :

Tanakan : $P^{(\text{P})}$ ou $T^{(\text{PC})}$ 1 – oui 0 – non
 Produits de comparaison : P.C. 1 ou P.C. 2 1 – oui 0 – non

Regression logistique avec interactions

Le modèle logistique utilisé jusqu’ici étant un modèle additif, la liaison mise en évidence entre l’amélioration de l’acouphène et une des trois variables pronostiques (ancienneté, siège et périodicité) est la même quelles que soient les modalités de deux autres. Autrement dit, ce modèle suppose qu’il n’existe pas d’interaction entre ces variables. Or il existe, par exemple, une interaction entre l’ancienneté et le siège comme le montre le tableau 4 suivant :

TABLEAU 4
Interaction entre les variables ancienneté et siège

	Acouphène récent (≤ 30 jours)		Acouphène ancien (> 30 jours)	
	Siège		Siège	
	Unilat.	Bilat.	Unilat.	Bilat.
Non amél. nette	26	13	56	65
Amél. nette	50	19	66	24
% d’amél. nette	(66 %)	(59 %)	(54 %)	(27 %)
Test du χ^2	χ^2 non signif.		χ^2 signif.	

Ainsi dans le but d’affiner les résultats précédents, nous avons été amenés à considérer un modèle avec les variables d’interaction suivantes :

API	si ancien et intermittent	alors	API = 1	API = 0	sinon
AAB	si ancien et bilatérale	alors	AAB = 1	AAB = 0	sinon
ASM	si ancien et sont mult.	alors	ASM = 1	ASM = 0	sinon
IAB	si interm. et bilatér.	alors	IAB = 1	IAB = 0	sinon
ISM	si interm. et sons mult.	alors	ISM = 1	ISM = 0	sinon
BSM	si bilater. et son mult.	alors	BSM = 1	BSM = 0	sinon

TABLEAU 5
Régression logistique en pas à pas avec interactions

Variables	Coeff. de régression	Test de Student	Signif.
	0,31		
AAB (ANC et bilal.)	-1,40	-4,17	p < 0,001
API (ANC et interm.)	1,40	3,63	p < 0,001
ANC (ancienneté)	-0,76	-2,61	p = 0,01
Tanakan	0,83	2,27	p < 0,02
Prod-comparaison	-0,36	-0,91	n.s.

Le tableau 5 suivant fournit les résultats de la régression logistique effectuée en pas à pas à partir des variables : périodicité, siège, ancienneté (≤ 30 jours), composition et interactions API, AAB, ASM, IAB, ISM, BSM.

Le tableau 5 montre que :

- la composition n'est pas significative,
- un acouphène récent (≤ 30 jours) et la prise du Tanakan sont en faveur d'une amélioration,
- la périodicité et le siège n'interviennent significativement dans le modèle que par leurs interactions avec l'ancienneté : ces variables n'ont un effet significatif sur l'évolution globale de l'acouphène que dans le cas où l'acouphène n'est pas récent (> 30 jours).

DÉTERMINATION DES SCORES PRONOSTIQUES D'AMÉLIORATION DE L'ACOUPHÈNE

Pour un sujet X dont le profil est (X_1, X_2, \dots, X_p) , son score pronostique $S(X)$ d'amélioration de l'acouphène est fourni par :

$$S(X) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

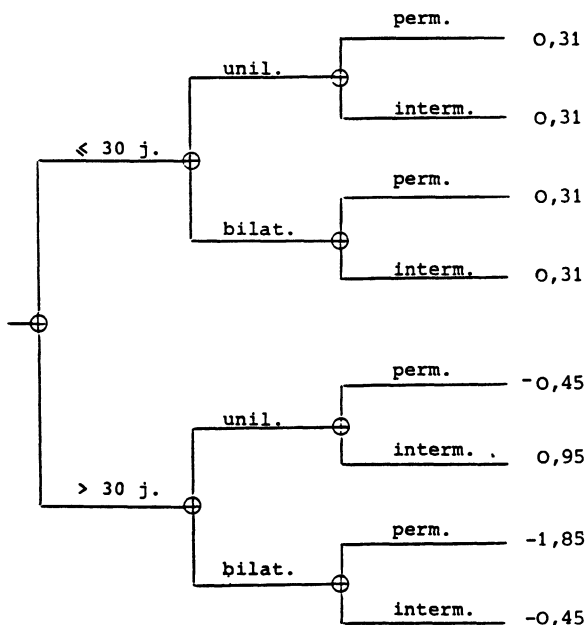
où $\beta_0, \beta_1, \dots, \beta_p$ sont les estimations des coefficients de régression.

Classes de sujets d'amélioration distinctes indépendantes du traitement

Si on ne tient pas compte du traitement, on a à considérer la combinaison linéaire suivante :

$$S_1 = 0,31 - 1,40 (\text{AAB}) + 1,40 (\text{API}) - 0,76 (\text{ANC}).$$

L'arbre suivant fournit les différents profils du triplet (AAB, API, ANC) et leurs scores respectifs.



En ordonnant les différents scores S_1 on obtient les 4 classes de sujets d'amélioration décroissante (indépendantes du traitement) suivantes :

CLASSE 1 : $S_1 = 0,95$

Acouphène ancien (> 30 jours) intermittent avec atteinte unilatérale

CLASSE 2 : $S_1 = 0,31$

Acouphène récent (< 30 jours)

CLASSE 3 : $S_1 = -0,45$

Acouphène ancien intermittent avec atteinte bilatérale ou Acouphène ancien permanent avec atteinte unilatérale

CLASSE 4 : $S_1 = -1,85$

Acouphène ancien permanent avec atteinte bilatérale

Scores d'amélioration de l'acouphène

Pour obtenir ces scores il faut ajouter à la valeur du score S_1 : + 0,83 pour le Tanakan, - 0,36 pour les produits de comparaison et 0 pour le Placebo; ceci montre que, seul le Tanakan, augmente la valeur du score pronostique d'amélioration de l'acouphène.

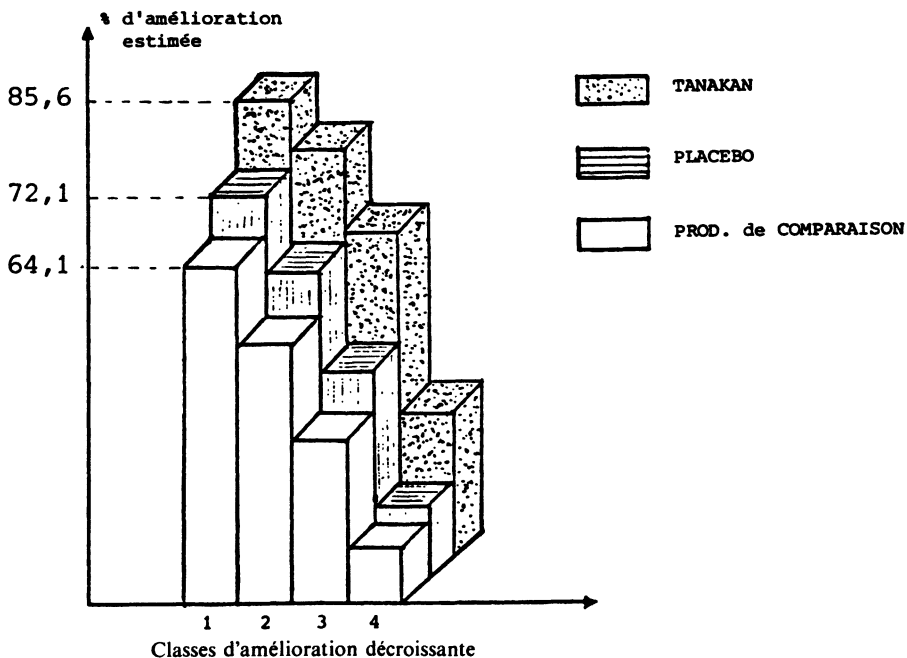
Dans le tableau suivant (tableau 6) les 12 groupes de sujets (classes d'amélioration décroissante par traitement) sont représentés avec par groupe : le nombre de sujets, le pourcentage d'amélioration observé et le pourcentage d'amélioration estimé par le modèle.

Le test d'ajustement du modèle qui conduit à un $\chi^2 = 2,65$ (à 9 ddl) est non significatif, ce qui témoigne d'un bon ajustement.

TABLEAU 6
Classes pronostiques d'amélioration décroissante suivant les traitements

Trait.	Classes	NB. de sujets	Sans amél.	Amél. nette	% Amél. obs.	% Amél. est.
Prod. De comp.	classe 1	5	2	3	60,0	64,4
	2	39	20	19	48,7	48,8
	3	34	23	11	32,4	30,8
	4	23	21	2	8,7	9,9
Placebo	classe 1	1	0	1	100,0	72,1
	2	13	5	8	61,5	57,7
	3	22	14	8	36,4	38,9
	4	9	8	1	11,1	13,5
Tanakan	classe 1	19	1	18	94,7	85,6
	2	56	14	42	75,0	75,8
	3	65	30	35	53,8	59,3
	4	33	22	11	33,3	26,4

Représentation graphique



Remarque

Bien que les pourcentages d'amélioration estimés soient supérieurs dans le groupe Placebo par rapport aux groupes Produits de comparaison, la différence est non significative. En effet la régression logistique en pas à pas avec interactions (tableau 5) a montré que le coefficient de régression associé à la variable P.C. (produits de comparaison) n'est pas significativement différent de 0, ce qui implique que l'évolution globale est semblable sous Placebo et sous PC.

CONCLUSION

Ces analyses ont permis de mettre en évidence, d'une part les variables pronostiques d'une amélioration de l'acouphène, et, d'autre part la supériorité du Tanakan sur les autres produits pour lesquels l'évolution globale n'est pas significativement différente.

La prise en compte d'interaction entre les variables pronostiques, ancienneté, siège et périodicité, a permis de définir avec précision le profil des classes d'amélioration distinctes.