

M. KAMINSKI

J. GOUJARD

C. RUMEAU-ROUQUETTE

Prévision du risque de prématurité à partir de caractères maternels connus en début de grossesse

Journal de la société statistique de Paris, tome 114 (1973), p. 229-236

http://www.numdam.org/item?id=JSFS_1973__114__229_0

© Société de statistique de Paris, 1973, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) 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/>

PRÉVISION DU RISQUE DE PRÉMATURITÉ A PARTIR DE CARACTÈRES MATERNELS CONNUS EN DÉBUT DE GROSSESSE

A survey has been made, between 1963 and 1969, in maternities of parisian hospitals, among pregnant women during the first three months of pregnancy. From collected data, we have established a risk of prematurity function following maternal characters knowed at the beginning of pregnancy (weight, size, tobacco consumption of the mother, socioprofessional class of the husband, data about previous pregnancies).

In den pariser Entbindungsanstalten wurden während sieben Jahren, von 1963 bis 1969 Untersuchungen bei den Frauen gemacht, deren Schwangerschaft nicht älter als drei Monate war. Mit den erhaltenen Ergebnissen hat man versucht die Möglichkeit eines Risikos einer Frühgeburt zu bestimmen, aufgrund der Kenntnisse der mütterlichen Charakteristika zu Beginn der Schwangerschaft. (Gewicht, Grösse und Alter der Mutter, Tabakconsum der Mutter, soziale und berufliche Situation des Ehemannes, Verlauf der vorhergehenden Schwangerschaften.)

Una encuesta ha sido realizada entre 1963 y 1969 en las casas de maternidad de Paris, dirigiendose a mujeres en estado durante el primer trimestre del embarazo. A partir de los resultados obtenidos, se estableció una función de riesgos de prematuridad, segun los caracteres de la madres al principio del embarazo (peso, estatura y edad de la madre, consumación de tabaco por la madre, categoria socio-profesional del marido, informaciones sobre los embarazos anteriores).

Le but de cette étude est la construction d'une fonction qui permette de définir un groupe de femmes présentant, dès le début de la grossesse, un risque élevé d'avoir un enfant prématuré. Pour qu'on puisse mettre en œuvre très tôt dans la grossesse une surveillance particulière de ce groupe à risque, il faut bien sûr que la prévision soit faite à partir de caractères connus eux aussi très tôt : pratiquement nous avons utilisé des caractères connus même avant le début de la grossesse, puisqu'il s'agit de caractères socio-économiques, du poids, de la taille, de l'âge, de la consommation de tabac de la mère, et de l'issue de ses grossesses antérieures.

L'ÉCHANTILLON

Les données proviennent d'une enquête prospective, réalisée entre 1963 et 1969 par la Section Maternité-Pédiatrie et l'Unité de Recherches statistiques de l'Institut national de la Santé et de la Recherche médicale, avec la collaboration de 12 maternités hospitalières parisiennes (le protocole de cette enquête a été décrit à propos d'une étude antérieure [1]). Au cours de cette enquête, près de 20 000 femmes ont été interrogées et examinées au cours du premier trimestre de leur grossesse. Actuellement 9 500 dossiers seulement sont en cours d'analyse.

Mais pour notre étude de prévision nous avons préféré nous limiter à une population plus homogène en ne considérant que les femmes nées en France métropolitaine, et qui ont

accouché à l'hôpital (des recherches sont en cours pour celles qui ont accouché ailleurs, ou bien qui ont avorté). Ont également été exclues les femmes qui ont accouché, pour la grossesse étudiée :

- de jumeaux, car il s'agit d'un phénomène particulier, diagnostiqué en cours de grossesse, et méritant une étude à part;
- de mort-nés, enfants dont le poids et le terme peuvent être faussés lorsqu'il y a rétention *in utero*.

De plus, comme nous voulions utiliser, dans la prévision, des renseignements concernant les grossesses antérieures, dont on connaît la valeur pronostique, nous n'avons considéré dans cette étude que des femmes multigestes (c'est-à-dire, ayant déjà eu au moins une grossesse, terminée par un avortement ou un accouchement).

Le terme d'un enfant à la naissance est calculé en semaines révolues à partir du premier jour des dernières règles. Un enfant est prématuré si son terme est inférieur ou égal à 36 semaines. Les dates des dernières règles sont connues d'après les déclarations faites par les femmes au cours de l'interrogatoire. Mais nous avons constaté parfois des anomalies en comparant :

- le poids de naissance et le terme calculé des enfants (enfants trop « lourds » pour des termes courts);
- les déclarations des femmes et l'examen clinique du troisième mois.

Ces deux types d'anomalies sont particulièrement fréquentes chez les femmes non mariées, le plus souvent liés à des erreurs dans les dates des dernières règles, et nous les avons exclues de cette étude.

Nous disposons donc d'un échantillon de 4 008 femmes métropolitaines, mariées, multigestes, n'ayant eu ni jumeaux, ni mort-né : sur cet échantillon, le taux de prématurité observé est de 8,1 %.

LES VARIABLES

Pour cet échantillon, nous avons retenu 13 variables quantitatives (variables 1 à 3) ou qualitatives à deux ou plusieurs classes (variables 4 à 13), connues dès le début de la grossesse, et qui sont les suivantes :

1. Poids habituel de la mère.
2. Taille de la mère.
3. Age de la mère.
4. Catégorie socio-professionnelle du mari.
5. Travail extérieur de la mère.
6. Consommation de tabac de la mère.
7. Primiparité (si la ou les grossesses précédentes se sont toutes terminées par des avortements).
8. Existence d'au moins 4 grossesses antérieures.
9. Existence d'au moins une fausse-couche dans le passé.
10. Existence d'au moins un mort-né dans le passé.
11. Existence d'au moins un enfant pesant 2 500 g ou moins dans le passé.
12. Existence d'au moins un enfant de terme inférieur ou égal à 8 mois dans le passé.
13. Issue de la grossesse précédente (fausse-couche, mort-né, poids de naissance de l'enfant né vivant).

Notre but est, rappelons-le, de construire une fonction de risque de prématurité. Or, si on considère les variables à deux classes, on constate qu'elles ont la propriété d'apporter une notion de risque. Une femme par exemple qui a eu un prématuré dans le passé a un risque plus élevé d'en avoir un à la grossesse étudiée. Ceci nous a conduit à construire pour les variables quantitatives des classes traduisant cette idée de risque : une femme très jeune par exemple a un risque élevé d'avoir un enfant prématuré... Cette méthode de regroupement en classes est également une solution au problème des variables dont l'effet sur la prématurité n'est pas monotone : c'est ce qu'on observe par exemple pour le poids de la mère, avec un excès d'enfants prématurés chez les femmes de petit poids, et chez les femmes de poids élevé.

Certaines variables peuvent ne pas être connues pour toutes les femmes : on a donc créé pour ces variables une classe supplémentaire « renseignement inconnu », information qui peut avoir, nous allons le voir, une valeur prédictive. Une remarque importante est que, pour chaque variable, chaque femme appartient à une classe et une seule.

LA MÉTHODE D'ANALYSE

Il faut maintenant, à l'aide de ces variables en classes, construire une fonction permettant de classer chaque femme dans un groupe à risque plus ou moins élevé d'accouchement prématuré.

Nous allons utiliser une régression linéaire multiple dans laquelle chacune des variables, dépendante ou indépendante, est une variable binaire. Cette méthode de régression binaire a été présentée par Suits [2] et utilisée en particulier par Feldstein [3] dans la prévision du risque de mortalité périnatale.

Nous allons transformer chacune des variables en classes en autant de variables 0-1 qu'il y a de classes, en procédant de la manière suivante : à chaque classe d'une variable, on associe une variable qui, pour un sujet, vaut 1 si le sujet appartient à cette classe, 0 sinon. En fait, dans l'estimation (par la méthode des moindres carrés) des paramètres de la régression, un caractère à n classes est transformé en $n - 1$ variables binaires ; ceci permet d'avoir à inverser une matrice régulière, et d'ailleurs $n - 1$ variables binaires définissent complètement un caractère à n classes : si une femme appartient à l'une des $n - 1$ premières, elle n'appartient pas à la dernière ; si elle n'appartient pas à l'une des $n - 1$ premières, elle appartient obligatoirement à la $n^{\text{ème}}$. Le modèle pour la prévision du risque de prématurité s'écrit finalement :

$$Y = \beta_0 + \sum_{i=1}^p \sum_{j=1}^{n_i-1} \beta_{ij} X_{ij} + \varepsilon$$

où

p est le nombre de caractères (ici 13),

n_i est le nombre de classes du caractère i ,

$Y = 1$ si l'enfant est prématuré,

$= 0$ si l'enfant est à terme.

$X_{ij} = 1$ si la femme appartient à la classe j du caractère i ,

$= 0$ sinon.

ε est un terme d'erreur aléatoire, de moyenne nulle, non corrélé avec les X_{ij} , non corrélé avec le terme d'erreur aléatoire des autres observations.

Si :

$$b_0, b_{11}, \dots, b_{1n_1-1}, \dots, b_{ij}, \dots, b_{pn_p-1},$$

sont les estimations de :

$$\beta_0, \beta_{11}, \dots, \beta_{1n_1-1}, \dots, \beta_{ij}, \dots, \beta_{pn_p-1}$$

$$\hat{Y} = b_0 + \sum_{i=1}^p \sum_{j=1}^{n_i-1} b_{ij} X_{ij}$$

nous donne une estimation de Y pour chaque femme.

En posant :

$$b_{im_i} = 0, \quad i = 1, \dots, p$$

$$\hat{b}_{ij} = b_{ij} + \frac{b_0}{p},$$

\hat{Y} peut s'écrire sous la forme :

$$\hat{Y} = \sum_{i=1}^p \sum_{j=1}^{n_i} \hat{b}_{ij} X_{ij}$$

LES RÉSULTATS

La valeur de la fonction \hat{Y} pour une femme est la somme des 13 coefficients b_{ij} associés aux classes auxquelles elle appartient, et plus \hat{Y} est grand, plus son risque d'avoir un enfant prématuré est grand. Y valant 0 ou 1, \hat{Y} devrait varier entre ces mêmes bornes, mais pour des raisons pratiques nous avons multiplié tous les coefficients par 10.

Les résultats figurent pour l'ensemble des 13 caractères dans les tableaux 1 à 13 : chaque ligne correspond à une variable X_{ij} , et les tableaux donnent l'effectif de chaque classe, le taux d'enfants prématurés dans la classe, et le coefficient \hat{b}_{ij} , qui mesure le risque de prématurité associé à la classe j du caractère i .

Nous remarquons tout d'abord que les coefficients \hat{b}_{ij} correspondant aux différentes classes d'un même caractère sont d'autant plus grands que les taux de prématurés sont plus élevés, c'est-à-dire que les variables prédictives isolément du risque de prématurité le restent après ajustement sur l'ensemble des variables.

On note également, comme nous l'avons déjà signalé, qu'il y a parfois un excès de prématurés dans les classes extrêmes d'un caractère comme, par exemple, pour le poids de la mère : la méthode de régression binaire permet de traiter ce type de problème, alors qu'une régression linéaire multiple conservant cette variable sous sa forme quantitative suppose au moins un effet monotone de la variable.

Il faut aussi insister sur l'intérêt qu'il y a à utiliser « l'information » apportée par la notion de données manquantes : en effet, les renseignements ne manquent pas « au hasard » comme on le suppose habituellement, quand on cherche à estimer ces données manquantes ; nous voyons ici que « poids de la mère inconnu » par exemple ou « catégorie socio-professionnelle inconnue » sont des facteurs de risque de prématurité.

Si nous regardons l'ensemble des caractères, nous voyons que les femmes à risque élevé d'accouchement prématuré (tel qu'il est envisagé en début de grossesse) sont les femmes très jeunes, celles qui ont un bas niveau socio-économique, celles qui ont plus de quatre grossesses passées, ou bien, à un moindre degré, qui sont primipares, celles qui ont déjà eu

dans le passé un mort-né, un enfant prématuré ou de petit poids. Un poids faible ou élevé ou inconnu, une petite taille, ou une taille inconnue, le fait de fumer, surtout plus de 10 cigarettes par jour, sont aussi des facteurs de risque. Par contre, être de grande taille, avoir plus de 40 ans, être dans la catégorie socio-professionnelle la plus élevée sont plutôt des facteurs associés à un risque faible de prématurité.

Voyons maintenant globalement quelle est la valeur prédictive de la fonction de risque \hat{Y} . Pour cela, nous avons défini à partir de cette fonction trois classes correspondant à un risque croissant de prématurité.

- \hat{Y} inférieur ou égal à 9,0 : risque faible,
- \hat{Y} compris entre 9,1 et 13,0 : risque moyen,
- \hat{Y} supérieur à 13,0 : risque élevé.

Le tableau 14 donne l'effectif de chaque classe, la répartition des mères d'enfants de plus de 36 semaines dans les trois classes, celle des mères de prématurés, et le taux de prématurité dans chaque classe. Rappelons que le taux global de prématurité est de 8,1 %. Il est de 5 % dans le groupe à bas risque (qui comprend les deux tiers de la population) et de 16,8 % dans le groupe à haut risque (qui comprend 14 % de la population).

Mais ce en qui concerne la répartition des mères :

- 68 % des mères d'enfants nés à terme sont bien classées dans le groupe à risque faible, 13 % sont mal classées dans le groupe à risque élevé;
- 30 % des mères de prématurés sont bien classées dans le groupe à haut risque, et 41 % sont mal classées et échappent à la prévision, ce qui est encore beaucoup trop.

En tenant compte de la probabilité *a priori* de la prématurité (8 %), la probabilité moyenne d'une erreur de classement est égale à :

$$r = 0,92 \times 0,13 + 0,08 \times 0,41 \text{ soit } 15 \%$$

Mais bien sûr les deux types d'erreur n'ont pas le même coût.

DISCUSSION

Notre fonction de risque ne fait intervenir que des signes connus « avant » le début de la grossesse. Mais, en toute rigueur, elle n'est applicable qu'à des femmes ayant atteint le troisième mois de leur grossesse. En effet les femmes entraient dans l'enquête au cours du premier trimestre de leur grossesse, et par conséquent y ont échappé les femmes ayant avorté précocement, particulièrement avant le troisième mois. Et nous ne savons pas comment ces femmes se placeraient dans les trois classes de risque. Une autre difficulté vient de ce que nous avons exclu les mort-nés : ce serait pleinement justifié s'il était possible en cours de grossesse de prédire la mortinatalité. Or une étude antérieure [4] a montré que la prévision d'un nombre important de cas était possible, mais pas bien entendu de la totalité.

Malgré tout, la mortinatalité est beaucoup plus fréquente chez les prématurés, et la non-exclusion des mort-nés entraînerait une confusion entre facteurs de risque de mortinatalité et facteurs de risque de prématurité.

Nous avons vu que la fonction de risque laissait échapper un taux important d'enfants prématurés. On peut se demander si la prise en compte d'un nombre plus important de caractères permettrait d'améliorer la prédiction : une étude de ce genre a été faite par

Rantakallio [5] pour la prévision du risque de faible poids de naissance, mais elle donne des résultats assez semblables. Une amélioration est obtenue si la prévision est faite plus tard avec l'introduction de caractéristiques du déroulement de la grossesse [6] —, et la meilleure estimation du risque de prématurité est certainement une estimation faite durant tout le déroulement de la grossesse.

M. KAMINSKI, J. GOUJARD, C. RUMEAU-ROUQUETTE

TABLEAU 1

Poids de la mère	Effectifs	% prématurés	\hat{d}_y
≤ 45 kg	278	11,5	2,8
46-70 kg	3 171	7,4	0,4
> 70 kg	223	9,0	1,8
Inconnu	336	11,0	2,8

TABLEAU 2

Taille de la mère	Effectifs	% prématurés	\hat{d}_y
≤ 150 cm	293	11,3	1,7
151-170 cm	3 188	7,7	0,4
> 170 cm	78	5,1	-1,7
Inconnue	454	8,8	1,9

TABLEAU 3

Age de la mère	Effectifs	% prématurés	\hat{d}_y
≤ 20 ans	272	13,2	6,2
21-39 ans	3 661	7,7	0,4
> 40 ans	72	6,9	-4,9
Inconnu	3	0	-5,1

TABLEAU 4

Catégorie socio professionnelle du père	Effectifs	% prématurés	\hat{d}_y
1. — Ind., cadres sup., prof. lib.	397	2,8	-3,7
2 — Cadres moyens, employés, OQ, OS	3 378	8,0	0,4
3 — Apprentis, manoeuvres, pers. serv.	291	14,3	6,1
Inconnue	66	18,3	9,0

TABLEAU 5

Travail extérieur de la mère	Effectifs	% prématurés	\hat{d}_y
Oui	2 022	7,8	-0,1
Non	1 958	8,4	0,4
Inconnu	28	10,7	-0,1

TABLEAU 6

Consommation de tabac	Effectifs	% prématurés	\hat{d}_y
Non fumeuses	3 414	7,4	0,4
Moyennes fumeuses (1-9 cig./jour) n'inhalant pas	104	8,8	1,5
Moyennes fumeuses inhalant	215	6,7	-1,4
Grandes fumeuses (≥ 10 cig./jour) n'inhalant pas	174	10,0	2,1
Grandes fumeuses inhalant	70	10,8	2,4
Inconnue	23	8,7	0,5

TABLEAU 7

Primipare	Effectifs	% prématurés	\hat{b}_Y
Oui	299	7,7	1,7
Non	3 709	8,1	0,4

TABLEAU 8

Nombre de grossesses antérieures	Effectifs	% prématurés	\hat{b}_Y
≤ 4	559	13,4	5,2
> 4	3 449	7,2	0,4

TABLEAU 9

Fausse couche dans le passé	Effectifs	% prématurés	\hat{b}_Y
Oui	1 197	8,9	0,5
Non	2 811	7,7	0,4

TABLEAU 10

Mort-né dans le passé	Effectifs	% prématurés	\hat{b}_Y
Oui	171	11,1	5,0
Non	3 837	7,9	0,4

TABLEAU 11

Enfant de poids $\leq 2 500$ g dans le passé	Effectifs	% prématurés	\hat{b}_Y
Oui	520	15,2	2,9
Non	3 488	7,0	0,4

TABLEAU 12

Prématuré (terme ≤ 8 mois) dans le passé	Effectifs	% prématurés	\hat{b}_Y
Oui	750	13,9	5,0
Non	3 258	6,7	0,4

TABLEAU 13

Issue de la grossesse précédente	Effectifs	% prématurés	\hat{b}_Y
Fausse couche	730	8,5	0,5
Mort-né	82	8,5	-2,7
Enfant né vivant poids $\leq 2 500$ g	230	17,0	4,3
Enfant né vivant poids 2 501-4 000 g	2 715	7,4	0,4
Enfant né vivant poids $> 4 000$ g	214	4,7	-2,3
Enfant né vivant poids inconnu	51	9,8	-0,1

TABLEAU 14

Prévision de la prématurité à l'aide de \hat{Y}

\hat{Y}	Effectifs	Répartition des mères d'enfants		% de prématurés
		Terme		
		> 36 semaines	≤ 36 semaines	
$\leq 9,0$	2 631	68 %	41 %	5,1 %
9,1-13,0	801	19 %	29 %	11,5 %
$> 13,0$	576	13 %	30 %	16,8 %
Total	4 008	100 %	100 %	

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] RUMEAU-ROUQUETTE C., GOUJARD J. et ÉTIENNE C. — Relation entre les métrorragies du début de la grossesse et les malformations congénitales. Résultats d'une enquête prospective portant sur 9 525 grossesses. *J. Gyn. Obst.*, Paris, 70, 557-562, 1971.
- [2] SUITS D. B. — Use of dummy variables in regression equations. *J. Amer. Statis. Ass.*, 52, 548-551, 1957.
- [3] FELDSTEIN M. S. — A binary variable multiple regression method of analysing factors affecting perinatal mortality and other outcomes of pregnancy. *J. Roy. Statis. Soc.*, 129, 61-74, 1966.
- [4] RUMEAU-ROUQUETTE C., GOUJARD J., KAMINSKI M., et SCHWARTZ D. — Mortalité périnatale en liaison avec les antécédents obstétricaux et l'usage du tabac. *J. Gyn. Obs. Biol.*, Repr. I, 723-729, 1972.
- [5] RANTAKALLIO P. — Groups at risk in low — birth weight infants and perinatal mortality. *Acta Poediatrica Scandinavica*, Suppl. 193, 1969.
- [6] KAMINSKI M. and PAPIERNIK E. — A multifactorial study of the risk of prematurity : a comparison between an empirical prediction and discriminant analysis. A paraître dans *Journal of perinatal medicine*.