BUREAU D'APPLICATION DES METHODES
STATISTIQUES ET INFORMATIQUES

Document de Travail

DT 04/2001

Note sur la mortalité infantile

Samuel AMBAPOUR

BAMSI B.P. 13734 Brazzaville

DT 04/2001

Note sur la mortalité infantile

Samuel AMBAPOUR*

Résumé : La méthode biométrique imaginée par J. Bourgeois-Pichat permet de séparer la mortalité infantile en mortalité endogène et mortalité exogène. Elle est présentée et appliquée ici, du point de vue de la théorie statistique.

Mots clés : Mortalité infantile, Mortalité endogène, Mortalité exogène, Loi multinomiale, Maximum de vraisemblance.

*BAMSI BP 13.734, Brazzaville,
CASP BP 1085, Brazzaville
E-mail : ambapour_samuel@yahoo.fr

Introduction

On classe généralement les décès de moins d'un an en deux catégories ; d'une part, ceux qui proviennent des causes antérieures à la naissance ou à l'accouchement ; d'autre part, ceux liés aux périls extérieurs. Cette séparation des décès respectivement en endogènes et exogènes, suppose en principe l'utilisation des statistiques des décès par cause.
Ces statistiques sont souvent défectueuses voire indisponibles pour certains pays. Pour pallier cet handicap, on a parfois recours à des méthodes approximatives, parmi lesquelles la méthode biométrique de J. Bourgeois-Pichat (1951).
Dans cette note, la méthode biométrique est appliquée du point de vue de la théorie statistique (R. Nadot ; 1971).

1. Les deux composantes de la mortalité infantile

1.1. Définition de la mortalité infantile

De façon classique, on définit la mortalité infantile, comme la proportion des enfants nés vivants qui meurent avant d'atteindre leur premier anniversaire. Même si conventionnellement, cette définition est admise, il faut noter qu'il y a un peu d'arbitraire dans les limites fixées : la naissance et la fin de la première année. En effet, cette définition nous renvoie à un autre concept assez complexe, celui de la naissance. Le dictionnaire de démographie définit la naissance ou de façon précise la naissance vivante, comme une expulsion ou une extraction complète de l'utérus d'un produit de conception, qui, après séparation d'avec le corps de la mère, respire ou donne tout autre signe de vie (Pressat, 1979). Intuitivement, on peut dire que la naissance n'est pas un commencement, mais un événement dans une suite qui a débuté neuf mois plus tôt (Bourgeois-Pichat, 1951).
Ceci dit, quelle différence y a t-il entre un enfant qui meurt dans le sein de sa mère (le jour ou il aurait dû naître) et celui qui meurt le premier jour de sa naissance. Certainement qu'il ne doit pas y avoir de différence biologique majeure. La preuve, dans un cas comme dans l'autre, on utilise le même vocable, celui de la mortalité périnatale.
En effet, sous ce dernier terme, on regroupe l'ensemble de la mortinatalité (extraction de l'utérus d'un produit de conception sans vie, après une certaine durée de

gestation) et de la mortalité néonatale (mortalité durant la première semaine). Comme on peut le constater, dans les deux cas, on met en exergue le caractère endogène de la mortalité. Les décès constatés ici, sont dus à des causes antérieures à la naissance ou résultant de la naissance elle-même. Ils constituent la mortalité infantile endogène.

Les décès endogènes sont à distinguer de ceux qui sont imputables au milieu extérieur qui constituent la mortalité exogène : ''l'enfant meurt parce qu'il rencontre dans le milieu où il vit la cause de sa mort''.

La statistique des causes des décès est souvent insuffisante pour séparer ces deux types de décès. Les différences de législation en la matière compliquent parfois les comparaisons internationales.

Notons toutefois que la distinction entre les deux types de mortalité, perd de sa pertinence au fur et à mesure de la baisse de la mortalité infantile (Leridon et Toulemon, 1997).

1.2. La mortalité exogène

Comme les décès exogènes résultent d'une cause extérieure, ils sont donc par principe tous évitables par la prévention et les soins. Sont en cause ici, les décès à caractère accidentel. On dit que si ''l'enfant meurt c'est parce que la mort lui a été apportée du dehors, soit par des microbes, soit par refroidissement, soit par une alimentation mal adaptée''.

Dans beaucoup de pays, si la mortalité infantile a reculé, cela est dû pour l'essentiel à la baisse de la mortalité exogène qui n'est pas loin d'être nulle. Dans les pays en développement, cette composante de la mortalité est encore très importante.

1.3. La mortalité endogène

Dans ce cas, l'enfant apporte la mort avec lui dès la naissance. Les causes des décès endogènes peuvent être classées en trois groupes :

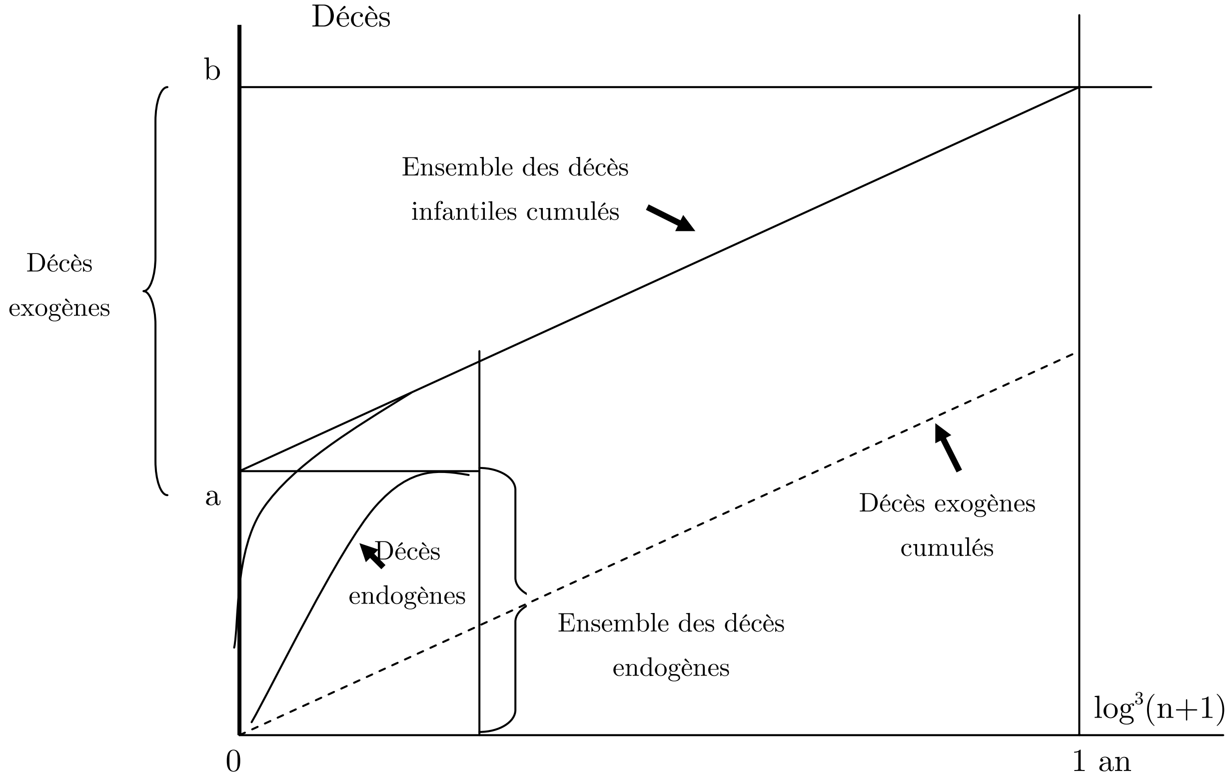
- les maladies héréditaires ; il s'agit des défauts de constitution, difficile à éviter ;
- les malformations congénitales. Si le mécanisme de la grossesse est connu, on peut se débarrasser de ces causes ;
- des causes endogènes acquises au moment de l'accouchement (traumatisme obstétrical) ; ce dernier représente pour l'enfant une aventure où les risques sont loin d'être négligeables.

Contrairement à la mortalité exogène, il est plus difficile de prévenir et de guérir une maladie endogène.

2. Le modèle statistique de la méthode biométrique

2.1. Rappel de la méthode biométrique de J. Bourgeois-Pichat

Fig.1 (Pressat ; 1973)



J. Bourgeois-Pichat a mis au point un procédé qui permet de séparer la mortalité infantile en mortalité endogène et mortalité exogène. Ce procédé exige que l'on connaisse seulement la répartition par âge des décès de moins d'un an. Ayant constaté que les décès exogènes de la première année se répartissaient selon l'âge d'une manière à peu près indépendante du niveau de la mortalité, il a établi ce qui suit (Henry, 1972) :

- les points ayant comme abscisse une fonction bien déterminée de l'âge exact et comme ordonnée les décès cumulés de la naissance jusqu'à cet âge exact sont alignés à partir d'au moins un mois ;
- cette droite prolongée vers la gauche coupe l'axe des ordonnées en un point dont l'ordonnée est égale aux décès endogènes. La différence au total est constituée par les décès exogènes.
La loi ainsi déterminée peut être illustrée par le graphique ci-dessus (Pressat, 1973).

2.2. Le modèle statistique

En suivant R. Nadot, on peut dire que le modèle biométrique de J Bourgeois-Pichat repose sur les trois hypothèses suivantes :
- H_1 : au-delà d'un certain âge A que l'on suppose d'environ un mois (28, 30, 31), tous les décès sont de nature exogène ;
- H_2 : au-delà de A, le total des décès jusqu'à un âge B donné est une fonction de cet âge ;
- H_3 : on peut extrapoler cette fonction, pour les décès exogènes, avant l'âge A.
Si l'on note par P_B le rapport entre le nombre de décès, entre les âges 0 et B et le nombre N de naissances annuelles, on peut écrire P_B sous la forme suivante :

$$P_B = \alpha + \beta \log^3(B+1) \qquad (1)$$

On a pu vérifier que l'expression (1) pour $B > A$, fournissait un bon ajustement dans un bon ajustement dans beaucoup de pays. Le paramètre α dans ces conditions représente le taux de mortalité endogène.
Prenons maintenant la première année de vie de l'enfant et on la découpe en périodes dont les limites sont les âges a_{i-1} et $a_i\left(i=1,k\right)$ avec :
$a_0 = 0$, $a_1 = A$, $a_k = 365$

Et soit le vecteur $y = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_k \end{bmatrix}$

Où les y_i sont les décès observés dans chaque période. $\sum y_i = S_1$ est le nombre de décès dans les douze premiers mois. L'on peut considérer y comme l'observation d'une variable aléatoire Y, qui suit une loi multinomiale (Tassi, 1992) :

$$P[Y = y] = \frac{N!}{y_1!y_2!...y_k!(N - S_1)!} \times p_1^{y_1} \times p_2^{y_2} \times ... p_k^{y_2} \times \pi^{N-S_1} \tag{2}$$

Où $\pi = 1 - \sum p_i$; p_i est la probabilité de décéder entre a_{i-1} et a_i.

La variable aléatoire Y suivant une loi multinomiale, l'espérance mathématique des y_i s'écrit : $E[y_i] = Np_i$; on peut alors formuler l'hypothèse 2 comme suit :

$$\sum E[y_i] = N(\alpha + \beta x_i) \tag{3}$$

Où $x_i = \log^3(a_i + 1)$; α et β sont deux paramètres à estimer et représentent respectivement le taux de mortalité endogène (hypothèse 3) et exogène.

Toujours sous l'hypothèse 2, les probabilités de la loi de Y s'écrivent :

$$p_1 = \alpha + \beta x_1 \tag{4}$$

$$p_i = \beta(x_i - \alpha_{i-1}) \;;\; i = 2, k \tag{5}$$

$$\pi = 1 - \alpha - \beta x_k \tag{6}$$

et la loi :

$$P[Y = y] = (\alpha + \beta x_i)^{y_1} \times \beta^{S_2} \times (1 - \alpha - \beta x_k)^{N-S_1} \tag{7}$$

où : $S_2 = \sum y_i$; $i = 2, K$

2.2. Estimation par la méthode du maximum de vraisemblance

Les paramètres du modèle peuvent être estimés en appliquant la méthode du maximum de vraisemblance. La log-vraisemblance de (7), s'écrit :

$$\zeta = y_1 Log(\alpha + \beta x_i) + S_2 Log\beta + (N - S_1) Log(1 - \alpha - \beta x_k) \tag{8}$$

Où $\zeta = LogP$

Le calcul des dérivées premières par rapport à α et β et l'annulation de ces dérivées conduit à l'obtention des estimateurs correspondants ; on obtient les équations de vraisemblance suivantes :

$$\frac{\partial \zeta}{\partial \alpha} = y_1(\alpha + \beta x_i)^{-1} - (N - S_2)(1 - \alpha - \beta x_k)^{-1} = 0 \tag{9}$$

$$\frac{\partial \zeta}{\partial \beta} = y_1(\alpha + \beta x_i)^{-1} S_2 \beta^{-1} - (N - S_1)(1 - \alpha - \beta x_k)^{-1} xk = 0 \tag{10}$$

Les estimateurs $\hat{\alpha}$ et $\hat{\beta}$ de α et β peuvent être obtenus de façon simple moyennant un peu d'algèbre et quelques manipulations.

2.2.1. Estimateur de β

Considérons l'équation (9) ; elle peut s'écrire :

$$y_1(\alpha+\beta x_i)^{-1}=(N-S_2-y_1)(1-\alpha-\beta x_k)^{-1}$$
$$y_1(\alpha+\beta x_i)^{-1}=(y_1+N-S_2-y_1)(\alpha+\beta x_1+1-\alpha-\beta x_k)^{-1}$$
$$y_1(\alpha+\beta x_i)^{-1}=(N-S_2)\left[1-\beta(x_k-x_1)\right]^{-1} \qquad (11)$$

On peut réécrire (11) de deux façons différentes,

- d'abord comme suit :

$$(\alpha+\beta x_1)(N-S_2)=y_1\left[1-\beta(x_k-x_1)\right],$$

on a alors :

$$\alpha+\beta x_1=(N-S_2)^{-1}=y_1\left[1-\beta(x_k-x_1)\right]$$

- ensuite de la manière suivante :

$$(1-\alpha-\beta x_k)(N-S_2)=(N-S_2-y_1)[1-\beta(x_k-x_1)],$$

on tire :

$$(1-\alpha-\beta x_k)=(N-S_2-y_1)(N-S_2)^{-1}[1-\beta(x_k-x_1)] \qquad (13)$$

On porte (12) et (13) dans (10) et en tenant compte de (11), on obtient :

$$x_1(N-S_2)\left[1-\beta(x_k-x_1)\right]^{-1}+S_2\beta^{-1}-x_k(N-S_2)\left[1-\beta(x_k-x_1)\right]^{-1}=0$$

En réduisant cette expression au même dénominateur et après simplification, on trouve :

$$\beta N x_1+S_2-\beta N x_k=0$$
$$\beta N(x_k-x_1)=S_2$$

d'où :

$$\hat{\beta}=S_2N^{-1}(x_k-x_1)^{-1} \qquad (14)$$

2.2.2. Estimateur de α

Si l'on porte l'expression de $\hat{\beta}$ dans (12), on obtient :

$$\hat{\alpha}+S_2N^{-1}(x_k-x_1)^{-1}x_1=y_1(N-S_2)^{-1}\left[1-\left(S_2N^{-1}(x_k-x_1)^{-1}\right)(x_k-x_1)\right]$$

Soit encore

$$\hat{\alpha}+S_2N^{-1}(x_k-x_1)^{-1}x_1=y_1(N-S_2)^{-1}\left[1-S_2N^{-1}\right]$$

Et finalement :

$$\hat{\alpha}+S_2N^{-1}(x_k-x_1)^{-1}x_1=y_1N^{-1}$$

D'où :

$$\hat{\alpha}=y_1N^{-1}(x_k-x_1)^{-1}x_1 \qquad (15)$$

Si l'on tient compte de (14),

$$\hat{\alpha} = y_1 N^{-1} - \hat{\beta} x_1 \tag{16}$$

Et après réduction au même dénominateur et simplification, on a en définitive :

$$\hat{\alpha} = \left[y_1 x_k - \left(S_2 + y_1 \right) x_1 \right] \left[N \left(x_k - x_1 \right) \right]^{-1} \ ; \tag{17}$$

soit encore :

$$\hat{\alpha} = \left(y_1 x_k - S_1 x_1 \right) \left[N \left(x_k - x_1 \right) \right]^{-1} \tag{18}$$

2.3. Propriétés des estimateurs

On se limite aux propriétés de l'estimateur du taux de mortalité endogène.

2.3.1. Espérance mathématique

L'estimateur du taux de mortalité endogène est sans biais ; en effet :

$$E[\hat{\alpha}] = E\left[\left(y_1 x_k \right) - S_1 x_1 \left[N \left(x_k - x_1 \right) \right]^{-1} \right],$$

$$E[\hat{\alpha}] = N^{-1} \left(x_k - x_1 \right)^{-1} \left[x_k E[y_1] - x_1 E[S_1] \right]$$

En servant de (3), on trouve que :

$$E[S_1] = \sum N p_i = N\left(\alpha + \beta x_1 \right),$$

d'où :

$$E[\hat{\alpha}] = \alpha,$$

2.3.2. Variance

$$Var[\hat{\alpha}] = Cov[\hat{\alpha}, \hat{\alpha}]$$

En remplaçant $\hat{\alpha}$ par sa valeur dans (16), on a :

$$Var[\hat{\alpha}] = Cov\left[\left(y_1 N^{-1} - \hat{\beta} x_1 \right), \left(y_1 N^{-1} - \hat{\beta} x_1 \right) \right].$$

En développant cette expression, on obtient :

$$Var[\hat{\alpha}] = N^{-2} Cov[y_1, y_1] + x_1^2 Cov\left(\hat{\beta}, \hat{\beta} \right) - 2N^{-1} x_1 Cov\left[y1, \hat{\beta} \right].$$

Il nous faut maintenant calculer les différentes covariances de l'expression ci-dessus :

- $Cov[y_1, y_1] = N\left(\alpha + \beta x_1 \right)\left(1 - \alpha - \beta x_1 \right)$, puisque y suit une loi multinomiale de paramètre N et $\alpha + \beta x1$;
- $Cov\left[\hat{\beta}, \hat{\beta} \right] = Cov\left[\left(S_2 N^{-1} \left(x_k - x_1 \right)^{-1} \right), \left(S_2 N^{-1} \left(x_k - x_1 \right)^{-1} \right) \right]$, d'après (14). Ainsi,
- $Cov\left[\hat{\beta}, \hat{\beta} \right] = \left[N^{-2} \left(x_k - x_1 \right)^2 \right]^{-1} Cov[S_2, S_2]$.

Comme S_2 suit une loi multinomiale de paramètre N et $\beta\left(x_k - x_1 \right)$, on aura donc :

$$Cov[S_2, S_2] = N\beta(x_k - x_1)[1 - \beta(x_k - x_1]$$

Par la suite on a :

$Var\left[\hat{\beta}\right] = N^{-1}\left[\beta\left(x_k - x_1\right)^{-1} - \beta^2\right]$;

- $Cov\left[y_1, \hat{\beta}\right] = N^{-1}\left(x_k - x_1\right)^{-1} Cov\left[y_1, S_2\right]$;

or on sait que y_1 et S_2 suivent ensemble une loi multinomiale de paramètre $\alpha + \beta x_1$ et $\beta(x_k - x_1)$, donc :

- $Cov\left[y_1, \hat{\beta}\right] = -\beta\left(\alpha + \beta x_1\right)$

Enfin, on trouve la variance de $\hat{\alpha}$ par :

$$Var\left[\hat{\alpha}\right] = N^{-1}\left[\alpha\left(1-\alpha\right) + \beta x_k x_1\left(x_k - x_1\right)^{-1}\right] \qquad (19)$$

2.4. Intervalle de confiance

Si N est assez grand, on peut supposer que la loi des observations est pratiquement normale. Dans ces conditions $\hat{\alpha}$ suit donc une loi normale de moyenne α et de variance $N^{-1}\left[\alpha\left(1-\alpha\right) + \beta x_k x_1\left(x_k - x_1\right)^{-1}\right]$. Dans ce cas, $u = \left(\hat{\alpha} - \alpha\right)\left[Var\left[\hat{\alpha}\right]\right]^{-1/2}$ suit une loi normale réduite $N\left(0,1\right)$. L'intervalle de confiance à 95% est donc :

$$\left|\hat{\alpha} - \alpha\right| < 2\left[Var\left[\left[\alpha\right]\right]\right]^{1/2} \qquad (20)$$

Dans$\left(19\right)$, on peut constater que α et β sont inconnus ; il n'est donc pas possible de calculer $Var\left[\hat{\alpha}\right]$. R. Nadot, recommande de remplacer cette variance par $\widehat{\mathrm{var}\left[\alpha\right]}$ où les valeurs de $\hat{\alpha}$ et $\hat{\beta}$ remplacent α et β.

3. Application aux données de la ville de Brazzaville

3.1. Les données

Ce sont des données relatives au mouvement naturel de la population observé sur la période 1961-1974 (CNSEE, 1974). Ces données mensuelles ont été utilisées comme telles. D'une part elles n'ont pas été désaisonnalisées, et d'autres part on ne s'est pas fié à leur qualité. Sur ce dernier point, notons par exemple, qu'il existe un écart très important entre ces données et celles issues du recensement de 1974. Ces dernières apparaissent comme sous-évaluées. Ces données d'état civil sont synthétisées dans le tableau 1.

Tableau 1 : **Mortalité infantile et naissances par année**

Années	y_1	S_1	N
1961	357	689	7517
1962	365	679	8913
1963	385	698	9842
1964	372	727	10294
1965	492	830	11171
1966	493	931	11590
1967	545	925	12492
1968	480	820	12652
1969	516	966	13925
1970	501	1049	13906
1971	569	1118	14924
1972	729	1314	15232
1973	573	995	14843
1974	738	1316	16483

y_1 : *Nombre de décès dans le premier mois ;* S_1 *: nombre de décès dans les douze premiers mois ;* N *: le nombre de naissances annuelles.*

3.2. Les résultats

Les paramètres du modèle statistique peuvent maintenant être calculés, en utilisant les données du tableau 1. Pour ce faire, on prend pour x_1 la valeur $x_1 = \log^3(365+1)/12$ et pour x_k la valeur $x_k = \log^3(365+1)$. Les résultats d'application des formules 14 et 15 sont consignés dans le tableau 2 ci-après :

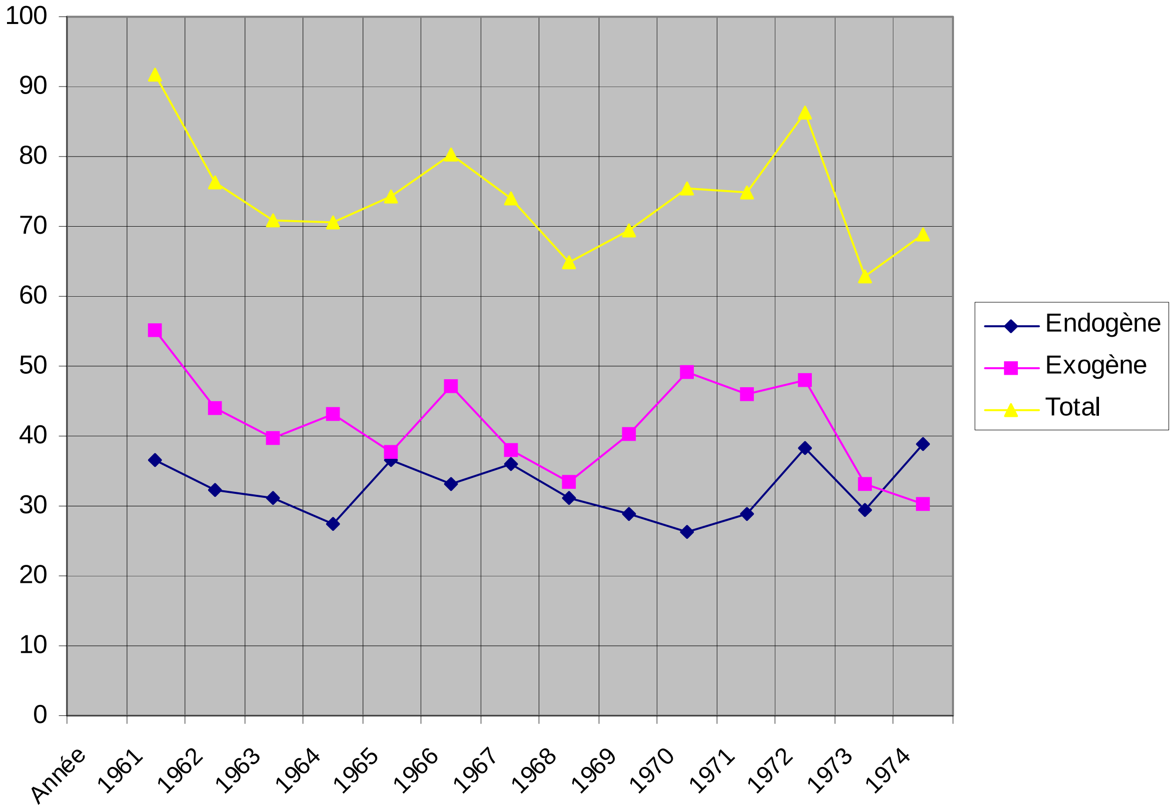
Tableau 2 : **Taux de mortalité infantile (pour 1000 naissances)**

Années	Endogène	Exogène	Total
1961	36,49	55,17	91,66
1962	32,17	44,01	76,18
1963	31,19	39,73	70,92
1964	27,55	43,07	70,62
1965	36,50	37,79	74,29
1966	33,12	47,20	80,32
1967	36,05	38,00	74,05
1968	31,24	33,57	64,81
1969	29,00	40,37	69,37
1970	26,21	49,22	75,43
1971	28,96	45,96	74,91
1972	38,29	47,97	86,26
1973	29,53	33,27	62,80
1974	38,76	30,16	68,92

Comme on l'a déjà indiqué, les résultats du tableau 2 confirment le fait que, la médecine a fait plus de progrès dans la lutte contre la mortalité exogène que dans celle contre la mortalité endogène. La mortalité exogène due à des facteurs sociaux et hygiéniques, peut être combattue en créant des conditions agréables de vie et des structures sanitaires appropriées.

La mortalité endogène baisse lentement. Si l'on prend des chiffres des années extrêmes, on a même l'impression que cette mortalité infantile a augmenté ou est restée sensiblement stationnaire. Et pourtant, on observe une tendance à la baisse de la mortalité infantile totale. En fait, « les maladies endogènes des enfants, comme certaines maladies chroniques des adultes font l'objet d'une mauvaise thérapie et d'une mauvaise prophylaxie ».

Fig.2: Evolution de la mortalité infantile



Conclusion

La méthode présentée dans cette note, malgré sa facilité de mise en œuvre et s'imposant en bien des cas, ne doit pas être appliquée de façon mécanique. Les trois hypothèses énoncées doivent faire l'objet d'une vérification au moins approximativement. Souvent, c'est la fonction utilisée pour l'âge qui n'est pas adaptée. C'est par exemple le cas du Québec étudié par R. Nadot, où les courbes annuelles des décès cumulés en fonction de l'âge montraient une concavité négative. En ce qui concerne la ville de Brazzaville, nous n'avons pas jugé utile de procéder à cette vérification pour des raisons déjà évoquées.

Bibliographie

Bourgeois-Pichat, J., (1951). La mesure de la mortalité infantile. I. Principes et méthodes, *Population*, 2, avril – juin.

Bourgeois-PIchat, J., (1964). Evolution récente de la mortalité infantile en Europe, *Population*, 3, juin juillet.

CNSEE., (1974). Mouvement naturel de la population (1960-1974), Brazzaville, Congo.

Henry, L.,(1972). *Démographie. Analyse et Modèles.* Larousse, Paris.

Leridon, H., Toulemon, L., (1997). Démographie. *Approche statistique et dynamique des populations.* Economica.

Nadot, R., (1971). Mesure de la mortalité infantile. Etude statistique de la méthode biométrique de Mr Jean Bourgeois-Pichat, *Population*, 5

Pressat, R., (1973). *L'Analyse Démographique*, 3ème Edition, PUF.

Pressat, R., (1979). *Dictionnaire de Démographie*, PUF.

Tassi, P., (1992). *Méthodes Statistiques*, 2ème Edition, Economica.

SERIE DES DOCUMENTS DE TRAVAIL DU BAMSI

01/2001 ''STATIS : une méthode d'analyse conjointe de plusieurs tableaux de données''.

Samuel AMBAPOUR

02/2001 ''Estimation des frontières de production et mesures de l'efficacité technique''.

Samuel AMBAPOUR

03/2001 ''Estimation d'un modèle d'emploi de court terme avec ajustement partiel''.

Samuel AMBAPOUR

04/2001 ''Note sur la mortalité infantile''

Samuel AMBAPOUR