

REVUE DE STATISTIQUE APPLIQUÉE

PAUL DAMIANI

Études sur la morbidité

Revue de statistique appliquée, tome 21, n° 1 (1973), p. 33-58

http://www.numdam.org/item?id=RSA_1973__21_1_33_0

© Société française de statistique, 1973, 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/>

ÉTUDES SUR LA MORBIDITÉ

Paul DAMIANI

La morbidité, c'est-à-dire la fréquence des maladies dans la population, est une notion difficile à mesurer de façon directe. Les évaluations chiffrées que l'on peut établir en ce domaine dépendent essentiellement de la définition adoptée pour le mot "malade".

On a analysé dans cette étude, deux aspects différents de la morbidité, l'un à partir de la statistique des causes de décès, l'autre à partir d'une enquête auprès des médecins.

En appliquant des modèles mathématiques aux données disponibles, on a pu évaluer, par sexe et par âge, le nombre annuel de nouveaux malades (taux d'incidence) et le nombre de malades présents à un moment donné (taux de prévalence) pour les principales maladies.

1^{ère} Partie

LA MORBIDITE MESUREE A PARTIR DE LA STATISTIQUE DES CAUSES DE DECES

Dans cette 1^{ère} partie, on a étudié la morbidité qui est impliquée (1) dans la mortalité, c'est-à-dire la morbidité représentée par les malades dont le décès est attribué à la maladie considérée, ou d'une façon générale, dont le certificat médical de décès fait mention de la maladie.

CHAMP DE L'ETUDE ET DONNEES DE BASE

Les renseignements obtenus seront d'autant plus significatifs que l'on considère des maladies qui, par rapport au nombre de malades, sont souvent impliquées dans les déclarations de causes de décès parce qu'elles constituent des états de fond, ou des antécédents notables, dont les médecins ont tendance à faire état dans les déclarations même quand ils ne leur attribuent pas directement le décès. De plus ces déclarations doivent être assez nombreuses.

Parmi les malades répondant à ces critères et à titre d'exemples d'application de la méthode proposée, on a choisi les trois maladies suivantes (les numéros entre parenthèses sont ceux de la nomenclature internationale des maladies, 7^e révision).

1 – Tuberculose de l'appareil respiratoire (n° B1 de la liste abrégée),

2 – Diabète sucré (n° B20 de la liste abrégée),

3 – Alcoolisme y compris la cirrhose du foie (n°s 307, 322, 581 de la liste détaillée). On a inclus la cirrhose parce qu'elle est le plus souvent d'origine alcoolique. On a retenu les déclarations faisant mention de la cirrhose dans la proportion de 80 %, estimation minimale de l'étiologie alcoolique de la cirrhose du point de vue médical (2).

Les données de base (3) sont les répartitions, suivant le sexe et l'âge, des décès attribués à ces causes ou dont la déclaration faisait état de la maladie à titre d'élément associé, au cours de la période 1960-1963. Les données utilisées dans l'étude ont été rapportées à l'année 1962.

CORRECTION DES DONNEES DE BASE

Les statistiques des causes de décès présentant une proportion non négligeable de décès pour lesquels la cause n'est pas spécifiée : cause non déclarée, cause mal définie (mention de manifestations purement symptomatiques le plus souvent), et décès attribué à la sénilité (rubrique B45 de la liste abrégée). Cette proportion atteint 14,1 % des décès en 1962 et varie beaucoup d'un département à l'autre.

Pour corriger les statistiques des causes de décès des causes non spécifiées, on a utilisé la méthode suivante inaugurée par l'INSEE et exposée par Ledermann (4).

Pour une cause de décès k , pour un âge et un sexe donnés, on observe pour le département j les proportions suivantes sur l'ensemble des décès :

z_{kj} proportion observée des décès de cause k ,

z_{oj} proportion des décès de cause non spécifiée.

On suppose que la proportion θ_k des décès de cause k dissimulés dans les décès de cause non spécifiée est constante quel que soit le département et ne dépend que de la cause k , de l'âge et du sexe considérés.

Pour le département j , la proportion réelle des décès de cause k est donc $z_{kj} + \theta_k z_{oj}$ et la proportion réelle des décès de cause non spécifiée qui peuvent être attribués à d'autres causes que la cause k est : $(1 - \theta_k) z_{oj}$. Ces deux quantités sont supposés être sans corrélation.

Cette condition, s'écrit :

$$(Z_k + \theta_k Z_o) Z_o' = 0$$

en utilisant les notations matérielles suivantes :

Z_k matrice (1 × n) des proportions départementales z_{kj} centrées

Z_o matrice (1 × n) des proportions départementales z_{oj} centrées

On en tire :

$$\theta_k = - Z_k Z_o' (Z_o Z_o')^{-1}$$

On remarque que si l'on calcule la régression linéaire existant entre z_k et z_o :

$$z_k = a_k + b_k z_o + \epsilon_k$$

on trouve que le coefficient de régression b_k est donné par la formule

$$b_k = Z_k Z_o' (Z_o Z_o')^{-1}$$

On a donc :

$$b_k = - \theta_k$$

Pour l'ensemble des causes de décès, la somme des coefficients est égale à 1.

En effet :

$$\sum_k \theta_k = - \left(\sum_k Z_k \right) Z_o' (Z_o Z_o')^{-1}$$

Or

$$\sum_k Z_k + Z_o = U$$

où U est la matrice (1 × n) dont tous les éléments sont égaux à 1.

On en tire :

$$\sum_k \theta_k = (Z_o - U) Z_o' (Z_o Z_o')^{-1}$$

Comme $U Z_o' = 0$, il vient :

$$\sum_k \theta_k = 1$$

En revenant aux données de base, on voit que la proportion moyenne observée des décès de cause k est égale à \bar{z}_k et que la proportion moyenne réelle est :

$$\bar{z}_k + \theta_k \bar{z}_o = \bar{z}_k - b_k \bar{z}_o = a_k$$

c'est l'ordonnée à l'origine de la droite de régression de z_k en fonction de z_o .

On appelle *coefficient correctif des causes non spécifiées*, le rapport :

$$\delta_k = \frac{a_k}{\bar{z}_k}$$

Pour chacune des causes de décès étudiées et par sexe, on a calculé les coefficients correctifs par groupe d'âge et pour l'ensemble des âges. On constate que les coefficients correctifs par groupe d'âge ne sont pas significativement différents du coefficient correctif pour l'ensemble des âges. C'est donc ce dernier coefficient qu'on retient pour corriger les décès quel que soit l'âge.

Les calculs ont été réalisés sur les décès pour lesquels la cause k a été déclarée comme cause principale et on a supposé que les coefficients ainsi trouvés s'appliquaient également aux décès pour lesquels la cause k était mentionnée comme cause associée.

Les résultats sont les suivants :

Coefficients correctifs des causes non spécifiées

Cause k	Sexe masculin	Sexe féminin
Tuberculose de l'appareil respiratoire.	1,166	1,242
Diabète sucré.	1,224	1,309
Alcoolisme et cirrhose du foie d'origine alcoolique.	1,262	1,341

Dans la suite de l'étude, quand on parlera de décès dus à la cause k, il s'agira de l'ensemble des décès attribués à la cause k à titre principal ou associé auquel on a appliqué les coefficients correctifs des causes non spécifiées.

PRINCIPE DE LA METHODE

Définition

La méthode s'applique à une population de "malades" déterminée pour chacune des maladies choisies et pour chaque sexe.

Un individu peut entrer dans la population des malades considérée à tout âge et ne peut en sortir que par décès, décès que le médecin attribuera à la maladie considérée à titre principal ou à titre d'élément associé au processus mortel.

Sont donc exclues du champ de l'étude les personnes qui ont été atteintes, au cours de leur vie, de la maladie considérée mais dont le certificat de décès ne fait pas mention de cette maladie, soit parce qu'elles ont été guéries de cette maladie sans séquelle ayant contribué au processus mortel, soit parce que le médecin a ignoré, négligé ou estimé inutile de mentionner cette affection.

Plan

Le but de l'étude est de déterminer la distribution par âge des entrées en maladie de chacune des populations de malades considérées. Cette distribution peut être calculée à partir de la distribution des décès dus à cette maladie si on connaît les quotients de mortalité par âge de la population des malades et si on peut évaluer également la variation de cette mortalité dans le temps.

Ayant déterminé la répartition par âge des entrées en maladie (incidence), on en déduit la répartition par âge des malades présents à un moment donné (prévalence).

Notations

On définit $(n + 1)$ groupes d'âge, tels que le groupe d'âge i est limité par les âges :

$$x_i \text{ et } x_{i+1} \quad (i = 0, 1, \dots, n)$$

Pour une maladie et un sexe donnés, on définit pour le groupe d'âge i les variables suivantes :

D_i nombre annuel de décès dans le groupe d'âge,

E_i nombre annuel d'entrées en maladie dans le groupe d'âge,

Y_i, Y_{i+1} nombre annuel d'entrées en maladie respectivement à l'âge x_i et à l'âge x_{i+1} en supposant que les entrées se font uniquement aux extrémités du groupe d'âge et non dans l'intervalle.

P_i effectif des malades du groupe d'âge.

On a les relations suivantes :

$$E_i = \frac{1}{2}(Y_i + Y_{i+1}) \quad (i = 1, 2, \dots, n)$$

$$E_0 = Y_0 + \frac{1}{2}Y_1$$

$$Y_{n+1} = 0$$

Les données de base ayant aussi à l'étude ont été au préalable corrigées des variations dues aux effets des guerres : pertes militaires et civiles pendant les guerres, baisses de natalité correspondantes. On a admis que les coefficients correctifs, différents suivant le sexe et l'âge, étaient les mêmes que ceux qui ont été calculés pour la population générale, par ajustement graphique de la pyramide des âges en 1962.

Expression des entrées en maladie Y

Pour chaque maladie et par sexe, on définit la table de mortalité à l'instant t, correspondant à l'année 1962, par les éléments suivants :

l_i nombre de survivants à l'âge x_i ($l_{n+1} = 0$),

$d_i = l_i - l_{i+1}$, nombre de décès entre l'âge x_i et l'âge x_{i+1} ,

$q_i = \frac{d_i}{l_i}$ quotient de mortalité à l'âge x_i pour le groupe d'âge

(x_i, x_{i+1})

Dans le cas où la mortalité ne varie pas dans le temps, le nombre de décès D_i s'obtient, à partir des entrées Y_j , par la relation :

$$D_i = \sum_{j=0}^i Y_j \frac{d_i}{l_j}$$

Or, on observe, en réalité, une baisse de mortalité au cours du temps. Par suite de cette baisse, le nombre de décès observés D_i est inférieur au nombre de décès théoriques obtenu en appliquant la table de mortalité calculée à l'instant t.

Si l'on conserve la même table de mortalité, on est amené à modifier la formule précédente en affectant les entrées Y_j d'un coefficient correctif dépendant de l'âge x_i au décès et de l'âge x_j à l'entrée en maladie. On prend un *coefficient correctif* de la forme $\frac{1 + \lambda_j}{1 + \lambda_i}$, avec $1 + \lambda_0 = 1$.

La relation liant les décès aux entrées devient alors :

$$D_i = \sum_{j=0}^i \frac{1 + \lambda_j}{1 + \lambda_i} Y_j \frac{d_i}{l_j}$$

Sous forme matricielle, on peut écrire :

$$D = Y \Lambda M \Lambda^{-1}$$

avec les notations suivantes :

D est la matrice $[1 \times (n + 1)]$ des décès D_i ,

Y est la matrice $[1 \times (n + 1)]$ des entrées en maladie Y_j ,

Λ est la matrice diagonale $[(n + 1) \times (n + 1)]$ d'élément général $(1 + \lambda_i)$

M est la matrice triangulaire stochastique $[(n + 1) \times (n + 1)]$

$$M = \begin{bmatrix} \frac{d_0}{l_0} & \frac{d_1}{l_0} & \frac{d_2}{l_0} & \cdots & \frac{d_n}{l_0} \\ & \frac{d_1}{l_1} & \frac{d_2}{l_1} & \cdots & \frac{d_n}{l_1} \\ & & & \cdots & \\ & & & & \frac{d_n}{l_n} \end{bmatrix}$$

De la relation précédente, on déduit la valeur de Y :

$$Y = D \Lambda M^{-1} \Lambda^{-1}$$

On peut donc calculer les entrées en maladie Y si l'on connaît M et Λ .

Forme des quotients de mortalité

On admet que les quotients de mortalité q_i relatif à une maladie et un sexe donnés, sont égaux aux quotients correspondants à la mortalité générale, pour le même sexe q_{Gi} , multipliés par un coefficient α , sauf pour le dernier groupe d'âge où le quotient de mortalité est égal à 1 :

$$q_i = \alpha q_{Gi} \quad (i = 0, 1, \dots, n - 1)$$

$$q_n = 1$$

Le paramètre α , appelé *coefficient d'aggravation* pour la maladie et le sexe étudiés, est supposé ne pas varier avec l'âge.

C'est la méthode adoptée dans les assurances pour le calcul des risques aggravés.

Forme des coefficients correctifs $(1 + \lambda_i)$

On peut calculer les coefficients correctifs $(1 + \lambda_{G_i})$ correspondant à la population générale. On constate graphiquement que ces coefficients vérifient des relations de la forme :

$$\log (1 + \lambda_{G_i}) = h x_i$$

où h est un coefficient constant quel que soit l'âge.

On admet que les coefficients correctifs relatifs à une population déterminée de malades vérifient des relations de même forme.

On en déduit que les coefficients $(1 + \lambda_i)$ peuvent s'écrire :

$$(1 + \lambda_i) = (1 + \lambda_{G_i})^\gamma$$

où γ est un *coefficient de variation de mortalité* constant quel que soit l'âge mais différent suivant la maladie et le sexe.

Modèle

Avec les hypothèses faites sur q_i et sur $(1 + \lambda_i)$, la relation qui donne les entrées en maladie Y_i s'écrit :

$$Y_i = D_{i-1} \left(\frac{1 + \lambda_{G, i-1}}{1 + \lambda_{G_i}} \right)^\gamma \left(1 - \frac{1}{\alpha q_{G, i-1}} \right) + D_i \frac{1}{\alpha q_{G_i}} \quad (i = 1, 2, \dots, n-1)$$

$$Y_o = \frac{D_o}{\alpha q_{G_o}}$$

$$Y_n = D_{n-1} \left(\frac{1 + \lambda_{G, n-1}}{1 + \lambda_{G_n}} \right)^\gamma \left(1 - \frac{1}{\alpha q_{G, n-1}} \right) + D_n$$

Les valeurs des entrées en maladie Y_i seront connues si l'on détermine les valeurs des coefficients α et γ .

Distance généralisée entre les distributions des entrées en maladie et des décès.

On admet que l'action de défense du corps humain ainsi que l'action des médicaments tendent à reculer le plus possible la date du décès c'est-à-dire à rendre maximum l'intervalle de temps séparant l'entrée en maladie et le décès.

On peut exprimer cette condition en écrivant que la distance généralisée entre la distribution des entrées en maladie et celle des décès est maximum.

La *distance généralisée* entre deux distributions est tirée d'une notion proposée par Mahalanobis. Cette notion intervient en analyse statistique dans le calcul de la fonction discriminante.

Si on appelle \bar{x}_1, \bar{x}_2 les moyennes des deux distributions et V_1, V_2 leurs variances, le carré de la distance généralisée entre ces deux distributions est défini par la relation :

$$\Delta^2 = K \frac{S_e}{S_i}$$

où K est un facteur de proportionnalité,

S_e est la somme des carrés des écarts entre les distributions,

S_i est la somme des carrés des écarts à l'intérieur des distributions.

L'effectif des deux distributions est le même et égal à N .

La moyenne générale \bar{x} s'écrit alors :

$$\bar{x} = \frac{1}{2} (\bar{x}_1 + \bar{x}_2)$$

La valeur de S_e est donnée par :

$$\begin{aligned} S_e &= N(\bar{x}_1 - \bar{x})^2 + N(\bar{x}_2 - \bar{x})^2 \\ &= \frac{N}{2} (\bar{x}_1 - \bar{x}_2)^2 \end{aligned}$$

La valeur de S_i est égale à :

$$S_i = N(V_1 + V_2)$$

On trouve donc, en prenant $K = 2$:

$$\Delta^2 = \frac{(\bar{x}_1 - \bar{x}_2)^2}{V_1 + V_2}$$

On prendra comme valeur de la distance généralisée :

$$\Delta = \frac{\bar{x}_2 - \bar{x}_1}{(V_1 + V_2)^{1/2}}$$

où la distribution (1) est celle des entrées, la distribution (2) celle des décès.

Calcul pratique des coefficients α et γ .

On prend pour α et γ leurs valeurs qui rendent maximale la distance généralisée. C'est un problème d'optimisation.

On a, en effet, à calculer les valeurs que doivent prendre des variables, soumises à des contraintes, de façon à maximiser une certaine fonction.

$$\left\{ \begin{array}{l} \text{variables : } \alpha \text{ et } \gamma \\ \text{contraintes : } Y_i \geq 0, Y_o \geq D_o, Y_n \leq D_n \\ \\ \alpha \geq 1 \\ \\ \gamma \geq 0 \text{ (si on suppose une baisse de mor-} \\ \text{talité)} \\ \\ \text{fonction à maximiser : } \Delta \end{array} \right.$$

Dans le plan (α, γ) les contraintes sont représentées par des courbes dont la courbure est très faible. En calculant les valeurs de Δ pour des couples de valeurs de α et γ , on constate expérimentalement que la solution du problème correspond aux valeurs prises par α et γ à l'un des sommets du polygone des contraintes, comme dans le cas de la programmation linéaire.

Calcul des taux d'incidence et de prévalence

Connaissant α et γ , on détermine Y_i et E_i par les formules données précédemment. On calcule, de plus, la population des malades P_i par la relation :

$$P_i = \left(L_i - \frac{1}{2} D_i \right) a_i$$

où L_i est le nombre de malades vivant à l'âge x_i défini par $L_i = D_i/q_i$ et

$$a_i = x_{i+1} - x_i$$

A partir de ces valeurs, en appelant P_{Gi} la population générale du groupe d'âge, on calcule :

– *le taux d'incidence* ou nombre annuel de nouveaux malades pour 1.000 habitants de la population générale :

$$1.000 \frac{E_i}{P_{Gi}}$$

– *le taux de prévalence* ou proportion des malades pour 1.000 habitants de la population générale :

$$1.000 \frac{P_i}{P_{Gi}}$$

RESULTATS

On a indiqué, dans le tableau 1, le coefficient d'aggravation α et le coefficient de variation de mortalité γ pour chacune des maladies étudiées et par sexe.

Dans les tableaux 2, 3 et 4 figurent les taux d'incidence et de prévalence suivant l'âge pour ces mêmes maladies et par sexe. Ces résultats sont illustrés par les graphiques 1, 2 et 3.

Tableau 1
Coefficients d'aggravation et de variation de mortalité

Maladie	Coefficient d'aggravation de la maladie α		Coefficient de variation de mortalité γ	
	Sexe masculin	Sexe féminin	Sexe masculin	Sexe féminin
Tuberculose de l'appareil respiratoire	1,254	1,052	10,036	9,300
Diabète sucré	1,333	1,751	1,046	1,268
Alcoolisme et cirrhose du foie d'origine alcoolique	1,444	1,506	10,765	14,195

Tableau 2
Tuberculose de l'appareil respiratoire

Groupe d'âge (x_i, x_{i+1})	Taux d'incidence % _c		Taux de prévalence % _c	
	Sexe masculin	Sexe féminin	Sexe masculin	Sexe féminin
Moins d'un an	1,74	2,00	0,85	1,47
1 à 4 ans	0,31	0,43	2,47	2,26
5 à 14 ans	0,31	0,83	2,48	4,16
15 à 24 ans	2,38	3,06	6,72	15,74
25 à 44 ans	1,70	1,37	51,19	63,12
45 à 64 ans	0,63	0,01	40,54	21,21
65 à 74 ans	—	—	19,44	11,50
75 et plus	—	—	12,99	6,17
Ensemble	1,11	1,02	26,48	25,87

Tableau 3

Diabète sucré

Groupe d'âge (x_i, x_{i+1})	Taux d'incidence ‰		Taux de prévalence ‰	
	Sexe masculin	Sexe féminin	Sexe masculin	Sexe féminin
Moins d'un an	1,08	1,27	0,47	0,14
1 à 4 ans	0,28	0,56	1,69	2,41
5 à 14 ans	0,05	0,20	2,74	4,63
15 à 24 ans	0,16	0,08	2,46	5,76
25 à 44 ans	0,21	0,29	6,17	5,86
45 à 64 ans	0,28	0,60	11,18	17,85
65 à 74 ans	0,38	0,63	14,95	23,82
75 ans et plus	—	—	17,71	21,55
Ensemble	0,22	0,35	6,66	10,37

Tableau 4

Alcoolisme et cirrhose du foie d'origine alcoolique

Groupe d'âge (x_i, x_{i+1})	Taux d'incidence ‰		Taux de prévalence ‰	
	Sexe masculin	Sexe féminin	Sexe masculin	Sexe féminin
Moins d'un an	—	—	—	—
1 à 4 ans	—	—	—	—
5 à 14 ans	—	—	—	—
15 à 24 ans	4,22	2,86	4,48	1,48
25 à 44 ans	3,33	2,47	86,37	61,24
45 à 64 ans	1,56	1,15	79,97	51,40
65 à 74 ans	—	—	33,09	17,19
75 ans et plus	—	—	18,88	5,54
Ensemble	1,98	1,35	44,94	29,26

COMPARAISON AVEC LES DONNEES DISPONIBLES

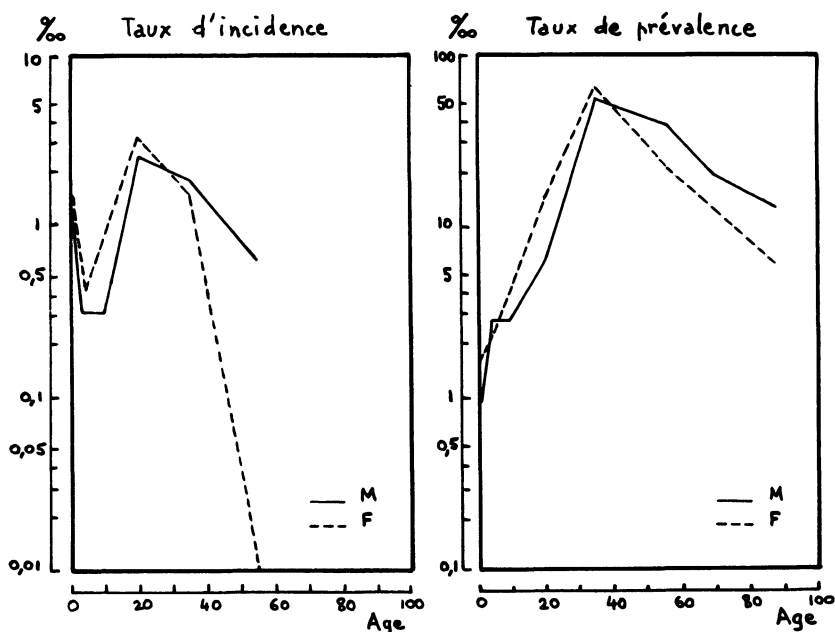
Pour juger de la valeur de la méthode, il convient de comparer les résultats obtenus dans cette étude aux données ou évaluations disponibles par ailleurs sur la morbidité.

a) Tuberculose pulmonaire

D'après la statistique annuelle des nouveaux cas de maladie de longue durée pris en charge par le régime général de la Sécurité Sociale (5), le taux d'incidence de la tuberculose respiratoire pour 1962 est de 1,9 pour 1.000 personnes protégées. Ce taux est supérieur à celui trouvé dans l'étude mais il concerne tous les malades connus de la Sécurité Sociale et non pas seulement ceux pris en compte dans l'étude (ceux dont le certificat de décès fait mention de la tuberculose pulmonaire).

L'Institut National de la Santé et de la Recherche Médicale a évalué à 1,4 ‰ le taux d'incidence relatif à toutes les formes de tuberculose (6). Ce résultat est compatible avec le taux trouvé dans l'étude qui ne concerne que la tuberculose pulmonaire.

Il est intéressant de noter que la structure par sexe et âge de la morbidité, telle qu'elle résulte de l'étude, correspond aux constatations médicales (7).

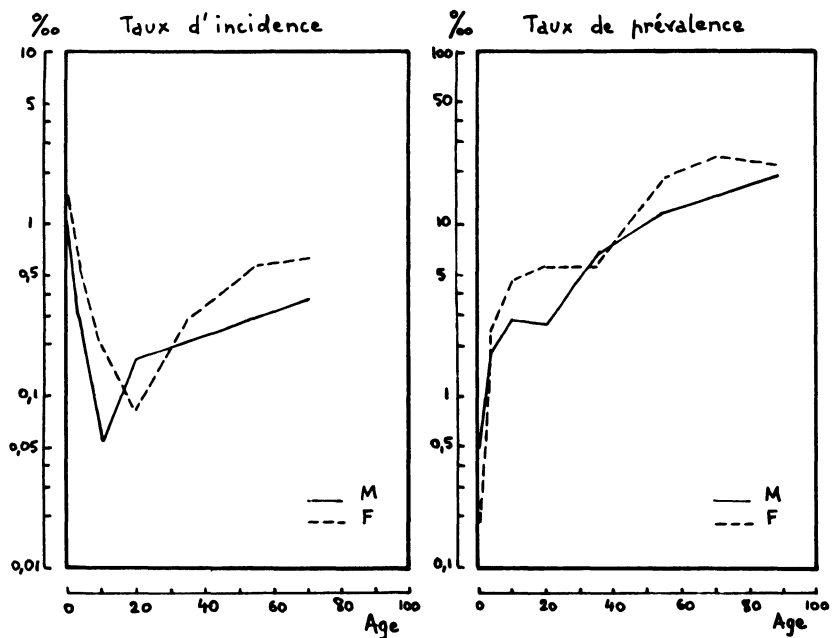


Graphique 1 - Tuberculose pulmonaire

b) Diabète sucré

La statistique des maladies de longue durée du régime général de Sécurité Sociale indiquait en 1964 un taux d'incidence de 0,35 pour 1.000 personnes protégées, ce qui correspond aux résultats de l'étude.

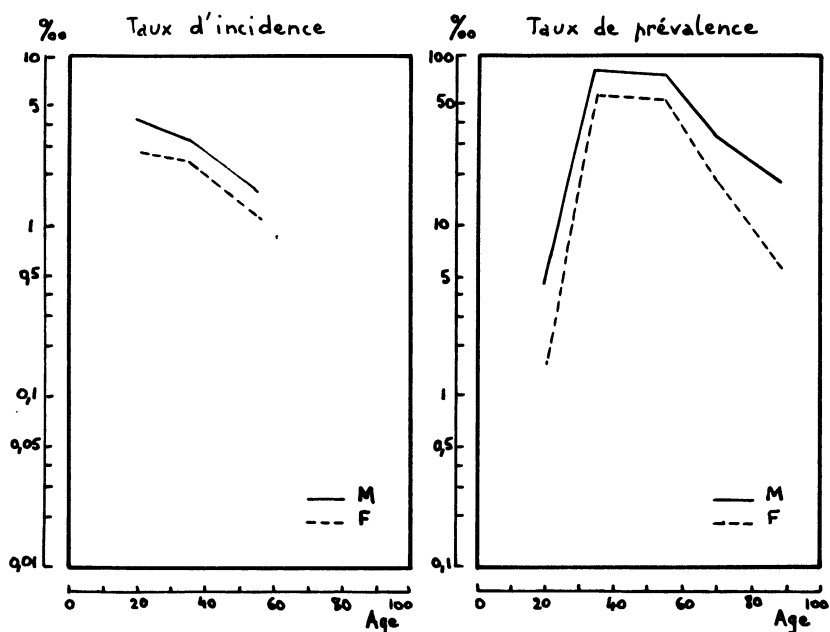
L'enquête américaine du National Health Survey réalisée en 1964-1965 (8) donne des taux de prévalence de 10,5 ‰ et de 13,8 ‰ pour le sexe masculin et le sexe féminin respectivement. Ces taux sont plus élevés que ceux de l'étude, ce qui est normal puisque l'enquête américaine tient compte de tous les cas déclarés par les personnes interrogées. On constate que les variations par sexe et par âge des taux relevés dans l'enquête américaine sont comparables à celles de l'étude.



Graphique 2 - Diabète sucré

c) Alcoolisme et cirrhose du foie d'origine alcoolique

D'après Ledermann (9), le nombre de buveurs excessifs, estimé à l'aide d'une autre méthode, correspondrait, pour la période 1950-1954, à 7 % de la population adulte. Ce taux est supérieur à celui des alcooliques estimé par l'étude mais il est certain que tous les buveurs excessifs ne meurent pas d'alcoolisme ou de cirrhose du foie. Les deux évaluations ne sont pas contradictoires.



Graphique 3 – Alcoolisme et cirrhose du foie d'origine alcoolique

CONCLUSION

La méthode utilisée a permis de déduire certains renseignements sur la morbidité à partir de la statistique des causes de décès : taux de fréquence correspondant aux nouveaux malades dans l'année et effectif des malades à un moment donné.

La morbidité ainsi étudiée n'est qu'une morbidité partielle limitée aux cas de maladies qui seront reconnues, par les déclarations médicales, comme impliquées dans les processus mortels.

Compte tenu de cette réserve et compte tenu également des erreurs d'observation portant sur les données de base, les résultats obtenus sont compatibles avec les données dont on dispose par ailleurs. Sur certains points la méthode apporte même des renseignements nouveaux.

BIBLIOGRAPHIE

- (1) M. AUBENQUE — "La statistique des causes de décès et la notion d'implication morbide". Etudes et Conjoncture, I.N.S.E.E., n° 3 — 1965.

- (2) G. PEQUIGNOT – “Enquête sur les circonstances diététiques de la cirrhose alcoolique en France”, – Bulletin de l’Institut National d’Hygiène – Tome 13, n° 3 – 1958.
- (3) “Statistique des causes de décès” – I.N.S.E.E. – Volumes 1960 – 1961 et 1962 – 1963 – Imprimerie Nationale.
- (4) S. LEDERMANN – “La répartition des décès de cause indéterminée”. Revue de l’Institut International de Statistique – 1956 –I-III, p. 47-57.
- (5) Supplément au Bulletin Mensuel de Statistique Sociales – Ministère de la Santé publique et de la Sécurité Sociale.
- (6) A. LOTTE et A.M. d’ALLEST – Mortalité par tuberculose en France dans la population de tous âges. Situation actuelle et évolution depuis 1950. INSERM – Recherche et information en santé publique – n° 5 – 1968.
- (7) LOTTE, BERNARD et FOUQUET – “La notion d’âge phthisiogène et son support statistique”. Semaine des hôpitaux, 22 janvier 1955.
- (8) “Characteristics of persons with diabetes” – U.S. July 1964 – June 1965 – Vital and Health statistics – Series 10 – n° 40.
- (9) S. LEDERMANN – “Alcool, alcoolisme, alcoolisation” – Travaux et documents, cahiers n° 29 et 41 de l’Institut National d’Etudes Démographiques, INED, 1956 et 1964.

2^{ème} PARTIE

LA MORBIDITE MESUREE A PARTIR D’UNE ENQUETE AUPRES DES MEDECINS.

Dans cette deuxième partie, on a analysé la morbidité telle qu’elle peut être constatée par des médecins praticiens exerçant en clientèle privée.

CHAMP DE L’ETUDE

On a utilisé les résultats, pour 1966 et 1967, d’une enquête annuelle effectuée auprès d’un échantillon de médecins par l’Institut européen de documentation et de recherche sur les maladies (I.D.R.E.M.) (1).

Pour les années considérées, l’échantillon de médecins interrogés est composé de 1600 médecins, répartis dans la France entière, à raison de 400 par trimestre. Il s’agit de médecins exerçant effectivement en clientèle privée : en conséquence, sont éliminés les médecins qui n’exercent qu’en milieu hospitalier

ainsi que les spécialistes des disciplines complémentaires (anesthésistes, radiologues, biologistes, ...) L'enquête ne porte pas sur les soins dentaires à l'exception de ceux qui sont dispensés par les stomatologistes.

L'échantillon de médecins est représentatif de l'ensemble des médecins praticiens exerçant en clientèle privée par rapport aux critères suivants : région géographique, type de résidence, spécialité, âge du médecin.

Ces médecins communiquent le double de leur ordonnance pour chacun des malades vus en consultation (au cabinet médical) ou en visite (au domicile du malade) au cours de deux journées d'exercice. Ces documents anonymes comportent, en particulier, l'âge et le sexe du malade ainsi que la nature du diagnostic. Le médecin doit indiquer également s'il s'agit, pour lui, d'une première consultation ou d'une première visite pour la maladie considérée.

Afin que l'enquête ne reflète que la pratique médicale de clientèle, on ne tient pas compte des actes tels que les examens de contrôle administratif, les examens systématiques de santé, de grossesse, etc. . .

DONNEES DE BASE

Les données de base sont la répartition des diagnostics constatés suivant l'âge et le sexe des malades pour l'ensemble des consultations et visites ainsi que pour les premières consultations de visites.

Les diagnostics ont été regroupés suivant les principales affections en utilisant la nomenclature internationale des maladies, 7^{ème} révision.

Le groupe d'âge i a pour âge central x_i ; sa population pour un sexe donné est P_i ; la population totale par sexe est P .

On a calculé, par groupe d'affections et par sexe, les taux suivants pour le groupe d'âge i :

- C_i nombre annuel de consultations et de visites par habitant,
- C_{ii} nombre annuel de premières consultations et visites par habitant.

Ce dernier taux C_{ii} ne correspond pas au taux d'incidence (nombre annuel de nouveaux malades par habitant) car un malade peut voir plusieurs médecins pour une même maladie.

DEFINITIONS

Dans cette étude, on appelle "malades" la population des personnes qui ont vu un médecin praticien exerçant en clientèle privée, en consultation ou en visite. Sont donc exclues du champ de l'étude les personnes qui ne vont pas voir de médecin praticien, soit qu'elles ne voient aucun médecin, soit qu'elles aillent consulter dans un établissement hospitalier.

Par ailleurs, la "durée de la maladie" telle qu'elle est définie dans cette étude correspond à la période de maladie pendant laquelle le malade a recours à un mé-

decin praticien. C'est ainsi, par exemple, que des périodes d'hospitalisation ou de soins dans un établissement hospitalier ne sont pas prises en considération.

La notion de durée de la maladie utilisée dans l'étude s'apparente à celle obtenue à partir des dossiers de Sécurité Sociale.

PRINCIPE DE LA METHODE

On cherche une relation liant le nombre annuel de premières consultations et visites par habitant c_{11} et le nombre annuel de consultations et visites par habitant c_i .

On admet alors que le nombre annuel de nouveaux malades par habitant (taux d'incidence) et le nombre de malades présents à un moment donné par habitant (taux de prévalence) sont reliés au taux c_i par des relations de forme analogue à celle trouvée ci-dessus.

On a choisi, comme base de ces formules, le taux c_i car ce taux est supposé connu de façon exacte.

1. Liaison entre les taux c_i et c_{11}

On mesure tout d'abord la liaison existant entre les taux c_i et c_{11} . On constate graphiquement que, par groupe d'affections et quel que soit le sexe, on peut ajuster entre ces deux variables une expression de la forme :

$$c_{11} = c_{10} e^{-\alpha x_i} c_i^\beta \quad (1)$$

En utilisant les logarithmes, on obtient un modèle de régression linéaire dans lequel c_{11} est la variable aléatoire. On en déduit les coefficients c_{10} , α et β en appliquant la méthode des moindres carrés.

On peut expliquer la relation (1) ainsi trouvée de la façon suivante.

On suppose que le nombre de premières consultations et visites déclaré pour une maladie donnée est sous-estimé, car le médecin peut oublier de noter une première consultation ou une première visite dans le cas où il s'agit d'un ancien malade soigné pour une nouvelle maladie. Ce risque d'erreur croît avec l'âge. On remplace donc le nombre c_{11} constaté de premières consultations et visites par habitant par le taux corrigé :

$$c'_{11} = c_{11} e^{\alpha x_i}$$

La liaison entre ce taux corrigé et c_i est par suite de la forme :

$$c'_{11} = c_{10} c_i^\beta \quad (2)$$

2. Nombre annuel de nouveaux malades par habitant (taux d'incidence)

On suppose qu'il existe entre le nombre annuel de nouveaux malades par habitant m_{ii} et le taux c_i la relation suivante analogue à la relation (2) :

$$m_{ii} = m_{10} c_i^\gamma \quad (3)$$

La variable m_{ii} est fonction croissante de c_i , de plus, la dérivée dm_{ii}/dc_i est fonction décroissante de c_i . On doit donc avoir la condition :

$$0 \leq \gamma \leq 1$$

Le nombre de premières consultations et visites par habitant c_{ii} doit être supérieur ou égal au nombre de nouveaux malades par habitant m_{ii} , car un malade peut voir plusieurs médecins pour une même maladie. Les inégalités suivantes doivent donc être vérifiées :

$$m_{ii} \leq c_{ii}$$

On admet enfin que les coefficients m_{10} et γ sont tels que le nombre total de nouveaux malades soit maximum, c'est-à-dire que l'on doit avoir la relation suivante pour chaque groupe d'affections et les deux sexes réunis :

$$\sum_i P_i m_{ii} \text{ maximum}$$

La détermination des coefficients m_{10} et γ est alors un problème de recherche opérationnelle. On a à calculer les valeurs de deux variables soumises à des contraintes de façon à maximiser une certaine fonction.

Ce problème se résout de la manière suivante. Dans le plan des variables $\log m_{10}$ et γ , les contraintes sont représentées par des droites constituant un polygone.

Pratiquement, on détermine les coordonnées des sommets de ce polygone. On calcule pour chaque sommet la valeur de la fonction à maximiser. On retient le sommet pour lequel cette valeur est maximum.

Cette détermination se fait par groupe d'affections, les deux sexes réunis.

3. Nombre de malades à un moment donné par habitant (taux de prévalence)

On suppose également que le nombre de malades à un moment donné par habitant m_i est relié à c_i par la relation suivante analogue aux relations (2) et (3)

$$m_i = m_o c_i^\delta \quad (4)$$

Pratiquement, on calcule m_i à partir de la durée de la maladie pour le groupe d'âge i , définie par :

$$l_i = \frac{365 m_i}{m_{1i}}$$

En utilisant les formules (3) et (4), il vient :

$$l_i = l_o c_i^\epsilon$$

avec

$$l_o = \frac{365 m_o}{m_{10}}$$

et

$$\epsilon = \delta - \gamma$$

Calcul de ϵ

On détermine une valeur approchée de ϵ par la méthode suivante :

On connaît, pour un groupe d'affections donné et pour l'ensemble des sexes, la répartition de la population P_i suivant le nombre c_i de consultations par habitant. La répartition cumulée s'écrit :

$$\sum_{j=1}^i \frac{P_j}{P} = F(c_i)$$

On admet que la répartition de la population P_i suivant la durée de la maladie l_i est une exponentielle. On a donc :

$$\sum_{j=1}^i \frac{P_j}{P} = 1 - e^{-\lambda l_i}$$

d'où :

$$z_i = -\text{Log}[1 - F(c_i)] = \lambda l_i = \lambda l_o c_i^\epsilon$$

Pratiquement, on essaie plusieurs valeurs de ϵ . On ajuste par la méthode des moindres carrés la droite de régression entre z_i et la variable c_i^ϵ . On conserve la valeur de ϵ pour laquelle l'ajustement est le meilleur : soit ϵ' cette valeur.

La valeur ϵ' ainsi trouvée est une valeur approchée du coefficient cherché. On remarque qu'il existe une liaison linéaire entre ϵ' et γ . On remplace alors ϵ' par la valeur ϵ ajustée par la méthode des moindres carrés :

$$\epsilon = 3,216 - 1,919\gamma$$

Calcul de l_o

On admet qu'il y a au maximum une consultation ou visite par jour de maladie, c'est-à-dire que l'on a la relation :

$$365 m_i \geq c_i$$

Cette relation s'écrit :

$$l_o \geq \frac{1}{m_{10}} c_i^{1-\delta}$$

Par groupe d'affections et par sexe, on prend pour l_o la valeur correspondant à la valeur minimale de c_i , soit c^* . On a donc :

$$l_o = \frac{1}{m_{10}} (c^*)^{1-\delta}$$

Calcul de l_i et de m_i

Ayant déterminé ϵ et l_o , on calcule l_i par :

$$l_i = l_o c_i^\epsilon$$

On en déduit m_i par la relation :

$$m_i = \frac{1}{365} l_i m_{11}$$

RESULTATS

Le tableau 1 donne le nombre annuel moyen de consultations et de visites par nouveau malade ($n = c/m_1$) et la durée moyenne de la maladie ($l = l_o(\bar{c})^\epsilon$ où \bar{c} est le nombre moyen de consultations et de visites par habitant) suivant les principaux groupes d'affections.

Le tableau 2 fournit le nombre annuel de nouveaux malades (taux d'incidence) et le nombre de malades à un moment donné (taux de prévalence) pour 1.000 habitants suivant les principaux groupes d'affections.

Un graphique représente les variations du taux d'incidence et de prévalence avec l'âge pour l'ensemble des maladies.

Il convient de souligner que la méthode employée fournit des résultats précis pour le taux d'incidence. Elle ne donne, par contre, que des renseignements approchés pour le taux de prévalence par maladie par suite de l'imprécision de la détermination du coefficient l_o .

Il s'ensuit que les durées de maladies données par l'étude doivent également être considérées comme des valeurs approximatives.

Validité des résultats

Il n'y a pas de justification théorique aux hypothèses faites dans l'étude. La valeur du modèle employé ne peut se juger qu'en comparant les résultats obtenus avec les données disponibles par ailleurs à condition que ces données correspondent à une définition analogue de la morbidité.

La seule enquête à caractère général, ayant donné des résultats détaillés comparables à ceux de cette étude au point de vue définition, est l'enquête effectuée par la Fédération nationale des Organismes de Sécurité Sociale (F.N.O.S.S.) sur 400.000 dossiers maladie du régime général au cours de la période 1952–1954 (2). A partir de cette enquête ont été publiés, en particulier, les nombres annuels moyens de nouveaux cas de maladie par personne protégée dans les caisses de province, suivant l'âge et le sexe, ainsi que la durée moyenne de la maladie.

1. Ensemble des maladies

La courbe donnant les variations avec l'âge du taux d'incidence de l'étude a même allure que celle obtenue à partir des résultats de l'enquête de la F.N.O.S.S. Le taux moyen d'incidence de l'étude est cependant supérieur de 50 % environ au taux moyen de l'enquête de la F.N.O.S.S. Cette différence s'explique par l'accroissement de la consommation médicale entre les deux enquêtes et par le fait que l'enquête auprès des médecins couvre l'ensemble des consultations et visites, y compris celles non remboursées par la Sécurité Sociale.

L'enquête "santé" pilote (3) effectuée par l'I.N.S.E.E. et le C.R.E.D.O.C. en 1965-1966 auprès de 400 ménages de la région parisienne donne 1,35 affections incidentes par personne, résultat tout à fait comparable avec ceux de l'étude.

D'après l'enquête de la F.N.O.S.S., la durée moyenne de la maladie serait de 26 jours en province contre 15 jours et 26 jours pour le sexe masculin et le sexe féminin respectivement dans l'étude.

2. Tuberculose de l'appareil respiratoire

Le taux moyen d'incidence de l'étude est inférieur à celui de l'enquête de la F.N.O.S.S. et inférieur également au taux d'incidence de 1,4 % ,concernant toutes les formes de tuberculose estimé par l'Institut national de la Santé et de la Recherche médicale (4).

On peut expliquer cette différence par le fait que l'enquête auprès des médecins couvre principalement les primo-infections et concerne peu les tuberculoses de réinfection.

3. Cancers et leucémies

Le taux d'incidence de l'étude est de 4,5 % pour le sexe masculin et de 4,1 % pour le sexe féminin.

Le taux d'incidence en province est de 1 % environ d'après l'enquête de la F.N.O.S.S. Si on ajoute à ce taux le nombre de nouvelles affections de longue durée du régime général de Sécurité Sociale par personne protégée, on obtient au total un taux d'incidence de 2,3 %.

Le taux d'incidence admis par l'Institut du Cancer est de 3 %.

La durée de maladie de l'étude est comparable avec celle trouvée par l'enquête F.N.O.S.S.

4. Diabète sucré

Le taux d'incidence de l'étude est supérieur au taux obtenu en ajoutant au taux de l'enquête F.N.O.S.S. celui correspondant aux affections de longue durée.

La durée de la maladie de l'étude est comparable à celle trouvée par l'enquête F.N.O.S.S.

Le taux de prévalence de l'étude correspond à celui de l'enquête du National Health Survey américain en 1964-1965 (5).

5. Autres maladies

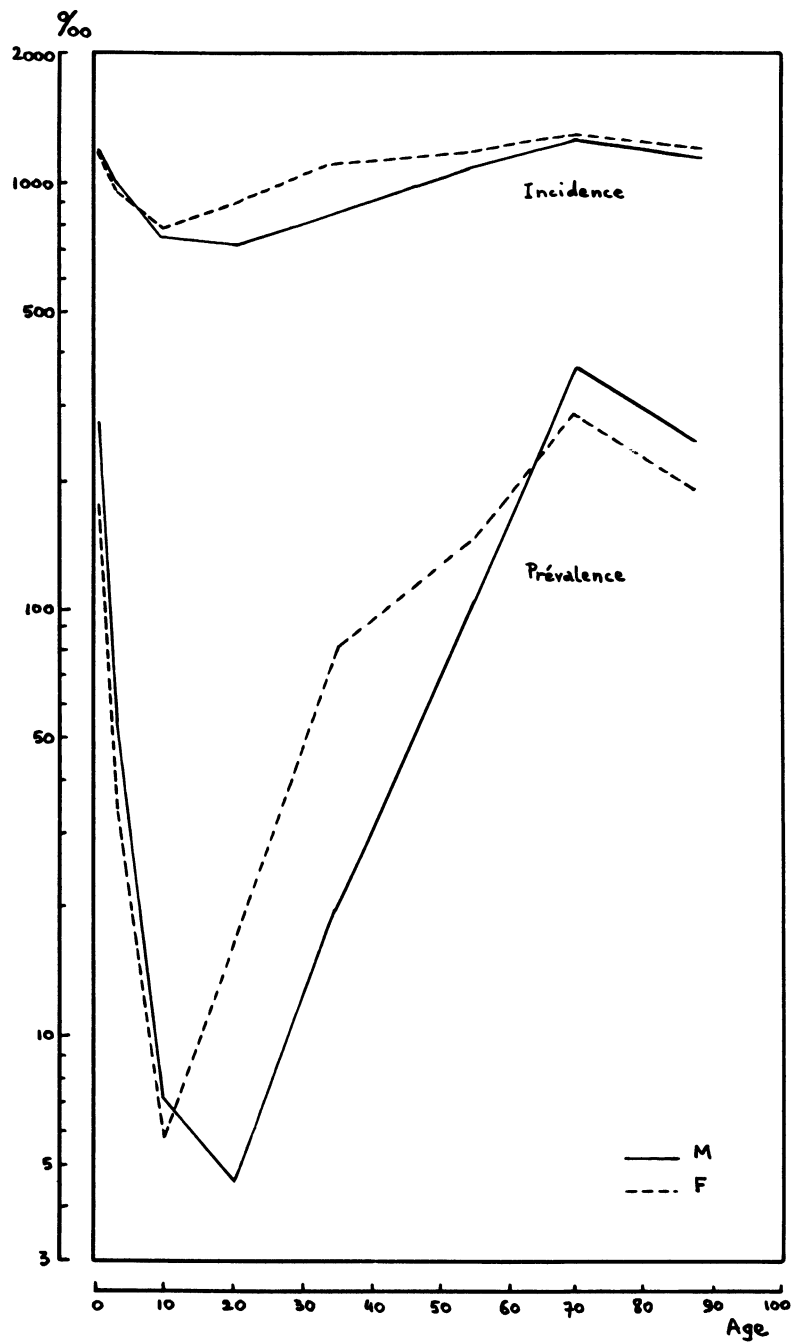
Pour les autres maladies retenues on trouve, dans l'ensemble, des résultats comparables avec ceux de l'enquête de la F.N.O.S.S.

CONCLUSION

Les résultats obtenus dans cette étude sont représentatifs de la morbidité constatée par les médecins praticiens en clientèle privée. Compte tenu de cette définition, ces résultats sont comparables avec les données disponibles par ailleurs, en particulier à partir des dossiers de Sécurité sociale. Ils apportent enfin des renseignements nouveaux sur les variations avec l'âge de l'incidence et de la prévalence des principales maladies.

BIBLIOGRAPHIE

- (1) Renseignements statistiques sur les maladies soignées en clientèle privée — Enquêtes 1966 et 1967. Suppléments B4 et B6 au bulletin mensuel de statistiques sociales du Ministère des Affaires sociales.
- (2) Statistiques médico-sociales. Une enquête sur 400.000 cas de maladie. F.N.O.S.S. — 1962.



Graphique — Taux d'incidence et de prévalence . Ensemble des maladies

- (3) Enquête pilote sur les soins médicaux 1965 – 1966. 3 fascicules – C.R.E.D.O.C. – 1968.
- (4) A. LOTTE et A.M. d'ALLEST. – Mortalité par tuberculose en France dans la population de tous âges. Situation actuelle et évolution depuis 1950. INSERM – Recherche et information en santé publique – n° 5, 1968.
- (5) Characteristics of persons with diabetes. United States – July 1964, June 1965. Vital and Health statistics : series 10, n° 40.

Tableau 1

Nombre de consultations par nouveau malade et durée de la maladie suivant le groupe d'affections

Groupe d'affections	Numéros de la nomenclature (1)	Nombre de consultations par nouveau malade $n = \bar{c}/\bar{m}_1$		Durée de la maladie (en jours) \bar{T}	
		Sexe masculin	Sexe féminin	Sexe masculin	Sexe féminin
Ensemble des maladies dont :	001 à 795	3,2	4,0	15	26
Tuberculose de l'appareil respiratoire	001 à 008	12,4	10,7	60	33
Cancers et leucémies	140 à 205	10,2	10,0	130	165
Diabète sucré	260	16,2	17,1	102	159
Alcoolisme (2)	307 et 322	14,0	6,6	73	62
Lésions vasculaires cérébrales	330 à 334	8,2	8,1	44	48
Maladies du système nerveux	340 à 369	3,9	4,2	52	94
Maladies du coeur (3)	400 à 434	9,5	9,4	51	49
Infections aiguës des voies respiratoires supérieures	470 à 475	2,0	2,0	7	7
Grippe	480 à 483	1,8	1,8	5	4
Bronchites	500 à 502	3,4	3,1	20	33
Maladies du foie et de la vésicule biliaire (4)	580 à 586	4,5	4,9	40	30
Arthrites et rhumatismes (5)	720 à 727	4,6	5,1	37	52

(1) Classification internationale des maladies (7^e révision – 1955)

(2) Non compris la cirrhose du foie

(3) Y compris le rhumatisme articulaire aigu

(4) Y compris la cirrhose du foie

(5) Non compris le rhumatisme articulaire aigu

Tableau 2

Incidence et prévalence suivant le sexe et l'âge, par groupe d'affections

m₁ = taux d'incidence, m = taux de prévalence pour 1000 habitants

Groupe d'affections	Taux et sexe	Age en années								
		Total	Moins d'un an	1-4	5-14	15-24	25-44	45-64	65-74	75 et plus
Ensemble des maladies	m ₁ M	926,1	1219,3	995,2	756,0	733,2	870,5	1092,4	1259,1	1143,4
	F	1057,6	1209,2	984,3	790,4	898,7	1095,2	1174,3	1277,8	1208,4
	m M	74,0	280,0	55,8	7,3	4,6	20,1	108,1	376,8	254,6
	F	103,3	180,2	33,8	5,9	17,4	85,7	147,9	296,5	196,3
dont :										
Tuberculose de l'appareil respiratoire	m ₁ M	3,3	—	1,6	1,5	1,9	3,7	5,6	4,9	2,8
	F	1,7	—	1,3	0,9	1,8	2,4	1,8	1,6	0,9
	m M	0,9	—	ε	ε	0,2	0,3	2,5	1,7	0,3
	F	0,2	—	ε	ε	0,3	0,4	0,2	0,1	ε
Cancers et leucémies	m ₁ M	4,5	—	—	0,1	0,5	1,1	8,9	21,4	17,7
	F	4,1	—	—	ε	0,2	1,7	7,8	11,1	10,8
	m M	9,5	—	—	ε	ε	ε	8,9	76,6	47,6
	F	5,7	—	—	ε	ε	ε	9,8	21,5	20,4
Diabète sucré	m ₁ M	2,8	—	—	—	ε	1,7	6,2	10,1	5,8
	F	2,9	—	—	—	ε	1,3	5,3	10,1	6,8
	m M	5,0	—	—	—	ε	0,1	7,7	40,1	5,9
	F	8,4	—	—	—	ε	ε	13,0	53,0	13,8
Alcoolisme	m ₁ M	2,6	—	—	—	0,8	3,5	5,0	3,9	1,9
	F	1,1	—	—	—	ε	1,5	2,1	2,0	ε
	m M	2,3	—	—	—	ε	0,7	7,5	1,8	ε
	F	0,8	—	—	—	ε	0,3	2,1	1,6	ε
Lésions vasculaires cérébrales	m ₁ M	5,2	—	—	—	—	0,5	6,0	33,8	42,9
	F	4,6	—	—	—	—	0,4	3,9	15,5	32,5
	m M	3,2	—	—	—	—	ε	1,1	26,2	42,0
	F	2,6	—	—	—	—	ε	0,6	6,9	30,3
Maladies du système nerveux	m ₁ M	30,3	5,8	5,4	8,8	16,0	36,7	50,2	51,0	33,1
	F	44,1	4,5	5,6	8,5	28,4	57,0	66,7	65,4	39,9
	m M	7,3	ε	ε	0,2	0,8	7,0	16,7	17,5	5,2
	F	18,1	ε	ε	0,1	3,2	22,6	35,3	32,7	8,1
Maladies du cœur	m ₁ M	21,1	—	—	2,0	2,8	6,5	42,0	99,0	93,7
	F	20,2	—	—	2,0	3,2	8,1	25,1	71,8	80,3
	m M	8,4	—	—	ε	0,1	0,3	10,8	66,6	59,5
	F	6,7	—	—	ε	0,1	0,4	4,3	34,6	43,5
Infections aiguës des voies respiratoires supérieures	m ₁ M	125,5	610,2	410,5	138,3	95,2	91,8	62,1	47,7	35,4
	F	129,3	551,0	389,6	150,1	114,8	113,9	76,0	64,4	33,2
	m M	6,6	117,5	43,8	2,9	1,2	1,1	0,4	0,2	0,1
	F	5,4	90,1	33,5	3,8	1,9	1,7	0,7	0,5	0,1
Grippe	m ₁ M	46,0	51,2	48,1	39,9	45,5	51,8	44,2	41,0	35,5
	F	45,4	33,0	46,5	45,7	43,6	47,2	45,0	44,6	35,6
	m M	0,6	0,9	0,7	0,3	0,6	0,9	0,5	0,4	0,2
	F	0,6	0,1	0,6	0,5	0,5	0,6	0,5	0,5	0,2
Bronchites	m ₁ M	30,8	68,1	44,2	20,4	15,0	20,5	41,2	63,7	66,2
	F	25,6	64,8	45,3	20,7	10,8	18,6	26,3	36,4	37,4
	m M	8,1	31,1	5,6	0,2	0,1	0,3	4,1	20,3	23,3
	F	4,9	76,0	17,3	1,0	0,1	0,7	2,2	7,5	8,3
Maladies du foie et de la vésicule biliaire	m ₁ M	20,1	—	2,9	9,4	9,4	19,2	36,5	36,9	24,7
	F	31,3	—	8,9	10,8	16,2	34,0	50,1	50,6	38,8
	m M	4,3	—	ε	0,3	0,3	1,8	11,7	12,1	3,9
	F	4,0	—	ε	0,1	0,4	3,0	8,7	9,5	4,5
Arthrites et rhumatismes	m ₁ M	33,6	—	—	2,7	14,2	34,9	65,5	79,4	59,5
	F	47,2	—	—	—	12,9	40,4	85,3	110,8	86,0
	m M	10,0	—	—	ε	0,2	3,0	23,3	45,8	17,8
	F	20,7	—	—	—	0,1	3,1	37,4	84,5	37,3

ε = résultat inférieur à 0,05