

# JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

DANIEL SCHWARTZ

## Épidémiologie de la mosaïque du tabac

*Journal de la société statistique de Paris*, tome 89 (1948), p. 475-490

<[http://www.numdam.org/item?id=JSFS\\_1948\\_\\_89\\_\\_475\\_0](http://www.numdam.org/item?id=JSFS_1948__89__475_0)>

© Société de statistique de Paris, 1948, 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/>

## IV

### ÉPIDÉMIOLOGIE DE LA MOSAÏQUE DU TABAC<sup>(1)</sup>

---

Il n'y a plus aucune originalité, à l'heure actuelle, à présenter à notre Société des communications de biologie, et, en particulier, dans le domaine de la Pathologie, plusieurs communications nous ont éclairés, au cours de ces dernières années, sur le rôle joué par la Statistique dans l'étude de certaines maladies infectieuses. Mais, pour des raisons bien compréhensibles, on ne nous a parlé jusqu'ici que des maladies humaines et nous n'avons pas encore pénétré dans le domaine de la pathologie végétale. D'autre part, en ce qui concerne l'agent pathogène, il ne nous a pas été fait mention spéciale des virus, et comme on ne compte plus maintenant les domaines où a pénétré la statistique, mais bien plutôt ceux où elle n'a pas pénétré, M. Penglaou croyait pouvoir dire, dans une communication récente, que le domaine des virus était encore vierge d'investigations statistiques. C'est pour combler cette double lacune que nous avons cru intéressant de vous parler d'épidémiologie de la mosaïque du Tabac.

---

(1) Communication présentée à la Société de Statistique le 17 mai 1948.

Disons d'abord que ce domaine offre, par rapport à l'épidémiologie des maladies humaines, d'incomparables facilités d'études.

Dans le domaine des maladies humaines, en effet, des difficultés graves se présentent à tous les niveaux de l'enquête statistique. Considérons un être humain le jour où il prend conscience de sa maladie (instant déjà bien tardif, hélas, car l'expression des symptômes est un aboutissement et non un début, et de nombreux jours déjà se sont écoulés depuis que le fonctionnement normal de l'individu a subi une déviation vers le fonctionnement pathologique). Ce sujet malade ne consulte pas toujours un médecin, et du fait même que cette éventualité suppose de sa part une décision facultative, certainement non aléatoire, il résulte une première erreur d'échantillonnage.

Lors même que le malade consulte, une deuxième erreur risque de fausser les documents de base : c'est l'erreur de diagnostic. Tout permet de penser que, dans certains cas, cette erreur est liée à des causes systématiques. Le cas du cancer en est un bel exemple, puisque depuis plus d'un demi-siècle la mortalité par cancer, telle qu'elle résulte des statistiques, ne cesse de croître, et qu'on a pu attribuer cette croissance, non pas uniquement au développement de la maladie, mais aussi, en grande partie, au perfectionnement du diagnostic.

Enfin, une troisième erreur importante provient du fait que les cas *examinés*, et *diagnostiqués* ne sont pas nécessairement *signalés*. Les maladies ne sont en effet, signalées avec une certaine rigueur que dans les cas mortels, puisque depuis 1937, le permis d'inhumer ne peut être délivré par le maire que sur production d'un certificat de décès que le médecin doit accompagner d'un certificat des causes du décès. Mais, si les *taux de mortalité* sont bien connus, il n'en est pas de même des *taux de morbidité*. Les seules maladies à déclaration obligatoire sont les maladies contagieuses, qui sont bien les plus intéressantes pour l'épidémiologiste, mais qui sont précisément celles que le malade préfère ne pas déclarer, pour éviter les consignes d'isolement, et qui, de fait, sont bien souvent cachées aux Services compétents. Pour ce qui est des autres maladies, leur importance et leur mode de répartition sont inconnus, sauf cas particuliers d'enquêtes menées pour des catégories limitées : fonctionnaires, clients de certaines compagnies d'assurance, etc... pour lesquelles se pose encore le même problème de la signification de l'échantillonnage.

Ainsi parvenus, incomplets ou faussés, aux Services d'exploitation, les renseignements statistiques sont-ils sans valeur?

Il est, certes, interdit d'être aussi pessimiste. En France notamment, l'Institut National d'Hygiène et la Statistique Générale de la France ont suffisamment prouvé leur compétence en cette matière, et montré qu'à défaut de pouvoir réunir un échantillonnage parfait, ils connaissent au moins les limites de son imperfection. Des recoupements ont, d'autre part, montré que pour certaines maladies (diphthérie, thyphoïde, poliomyélite, méningite) les documents de base sont exacts.

Mais d'autres difficultés, indépendamment de la qualité de l'échantillonnage, sont inhérentes à une hétérogénéité certaine, aussi bien du côté des *sujets malades* que du côté des *agents pathogènes*.

a) Du côté sujet, la population étudiée n'est pas génotypiquement homogène;

en effet, si les individus ont en commun un bon nombre de gènes majeurs, du fait qu'ils appartiennent, par exemple, au type Homo Sapiens et, en général, à ses populations à peu près homogènes, il est certain que des gènes essentiels du point de vue du tableau pathologique sont distribués entre les sujets de façon inconnue, dont on ne peut même pas affirmer qu'elle soit aléatoire : nous n'en voulons pour exemple que le caractère « groupe sanguin » dont l'importance a été signalée, à propos de la poliomyélite, par Kossowitch et Grotten. Si l'on ajoute que le milieu, ou environnement, dans lequel sont plongés ces sujets déjà hétérogènes, est lui-même extrêmement divers, aussi bien par ses caractères généraux (facteurs climatiques), que par des caractères liés plus directement au « fait humain » (conditions sociales, d'hygiène générale et particulière, habitudes personnelles, etc...) il en résulte que les corrélations aperçues par le statisticien sont d'une interprétation causale extrêmement délicate : en ce qui concerne, par exemple, l'incidence du groupe sanguin, nous savons que, d'après Mme Léone Bourdel, les activités professionnelles sont en corrélation avec les groupes sanguins. S'il en est ainsi, la propagation de la poliomyélite est-elle, en définitive, liée aux groupes sanguins (caractères génétiques) ou aux activités professionnelles (qui peuvent agir par des caractères d'environnement)?

b) Du côté agent pathogène, l'hétérogénéité n'est pas moins vraisemblable.

Même si nous admettons que tous les diagnostics retenus dans les listes de base soient exacts, le jugement doit être différé pour tout ce qui concerne l'*univocité* de la maladie : l'agent doit être considéré par principe comme support possible de variation sur le point même de ses propriétés de transmissibilité ; or, dans le cas des épidémies humaines naturelles, l'agent pathogène particulier est connu très imparfaitement, du seul fait qu'il est d'apparition spontanée et non pas expérimentalement inoculé. Par ailleurs, la notion même de spécificité pour l'agent pathogène n'est pas à l'abri de toute critique et on peut penser que l'individu malade héberge une population hétérogène de souches plus ou moins sélectionnées par le sujet lui-même. Cette hypothèse, suggérée notamment dans le cas des bactéries par l'inquiétante augmentation des microbes sulfamidorésistants, vient, dans le cas des virus, d'être confirmée par les récents travaux de Johnson qui, dans le jus infectieux d'une plante, isole des souches de caractères distinctifs et constants.

En regard des considérations précédentes, quelles facilités nous offre notre étude d'épidémiologie végétale !

Précisons d'abord que notre enquête n'a pas été menée à l'échelon national, bien que l'organisation de contrôle de la culture du tabac comporte le droit de regard sur toutes les plantations du pays. Nous avons préféré limiter nos investigations au domaine restreint d'un Institut de recherches, ce qui nous a permis à la fois de recueillir des données statistiques exactes, et de compléter l'observation par l'expérimentation.

Il est à peine besoin de souligner que, pour notre collection d'individus la libre décision de consulter le médecin, et le souci d'échapper à la déclaration des maladies contagieuses, sont des causes d'erreur que nous ne redoutons point. En contre-partie de ces précieuses qualités du client, le médecin des

plantes est contraint à une vertu également rare chez nos praticiens : il lui faut examiner tous les individus, les sains comme les malades, pour distinguer ceux-ci de ceux-là. Des relevés sur champ ont donc été effectués tous les jours, pendant la période d'observation qui s'étend depuis le repiquage des pieds jusqu'à leur cueillette (juin-juillet-août).

On peut ajouter que les erreurs de diagnostic sont tout à tout fait négligeables. Le tableau pathologique de la mosaïque sur le génotype Tabac utilisé nous est parfaitement connu, et le personnel qualifié chargé des relevés ne saurait le confondre avec celui d'une autre maladie.

Ainsi, les trois causes principales d'erreur, qui, dans les populations humaines viennent perturber l'échantillonnage, nous sont ici épargnées de façon totale.

Le problème se présente sous des auspices également favorables quant à l'homogénéité des deux populations en présence, la population-sujet et la population agent-pathogène.

a) Du côté sujet, nous avons affaire à une variété fixée ; l'exploration biométrique de populations numériquement importantes nous a toujours montré la parfaite homogénéité de ces populations du point de vue des caractères morphologiques ; il est vrai que cette régularité générale de comportement ne fournit qu'une présomption d'homogénéité pour les caractères dépendant du domaine pathologique ; mais le seul fait que notre variété est issue d'une série d'autofécondations successives poursuivie depuis près de trente ans, est de nature à nous rassurer sur l'homozygotie quasi totale de nos sujets : les quelque 100.000 graines issues d'un de ces plants donnent ainsi naissance à une population qui, toutes choses égales d'ailleurs, représente pour le phytopathologiste ce que seraient, pour le clinicien humain, 100.000 jumeaux vrais. Ajoutons que, du point de vue *milieu*, ces jumeaux sont mis en culture côte à côte sur une parcelle choisie pour son homogénéité, qu'ils sont élevés de façon synchrone et dans les mêmes conditions d'environnement.

b) Du côté agent pathogène, nous sommes également favorisés. On sait que les virus ne sont plus considérés comme des petits microbes, depuis que Stanley, travaillant précisément sur la mosaïque du tabac, a montré que l'agent infectieux de cette maladie s'identifie avec une nucléo-protéine spécifique. Depuis cette mémorable découverte, le virus de la mosaïque du tabac est une des substances chimiques les plus étudiées et un des agents pathogènes les mieux connus.

Profitant de cet avantage, nous avons pu introduire initialement dans le système, non seulement un virus purifié par voie physico-chimique, mais même une souche unique que des dilutions et inoculations successives nous ont permis d'obtenir, selon toute vraisemblance, à partir d'une seule molécule du liquide infectieux.

Il est clair que les qualités précédentes font déjà, du Tabac comme sujet, et du virus mosaïque comme agent, un matériel de choix. Ajoutons que l'association du sujet et de l'agent fournit au pathologiste une maladie dont l'évolution est particulièrement simple : en effet, un tabac qui est infecté sur champ par la mosaïque n'est destiné d'ordinaire ni à en guérir, ni à en mourir : le taux de morbidité, à un instant donné, n'est donc pas seulement le pourcen-

tage d'individus malades, mais aussi le pourcentage des individus qui ont été malades.

Disons, enfin, qu'à cet ensemble de rares qualités s'ajoute une inestimable vertu : l'immobilité du sujet; ainsi, du fait que les pieds du Tabac restent, leur vie durant, à l'endroit où on les plante, résulte non seulement l'absence de foyers mobiles de contagion, mais surtout la précieuse coïncidence de l'unité pathologique avec l'unité topographique. Quel est, en effet, le premier travail du statisticien qui réunit les données épidémiologiques parvenues de toute la France? Il étudie la répartition des maladies par départements, ou au mieux par villes, ou par quartiers de ville : ses unités pathologiques sont donc toujours des populations qui diffèrent du point de vue pathologique par bien autre chose que leurs coordonnées dans le plan.

Au contraire, la possibilité pour nous d'avoir, régulièrement distribuées en quadrillage d'un plan, des unités pathologiques ne différant que par leurs coordonnées, devait évidemment nous conduire à étudier dans des conditions favorables la répartition topographique des individus malades, et à voir sous un jour intéressant le premier problème rencontré par l'épidémiologiste désireux de définir sa discipline : le problème de l'endémie et de l'épidémie.

#### ENDÉMIE ET ÉPIDÉMIE

La distinction entre ces deux états existe depuis fort longtemps; on la retrouve déjà chez Ambroise Paré. Elle repose sur des notions assez intuitives pour qu'on n'ait pas cru souvent utile de les définir, mais les diverses définitions qui ont été données ne sont pas toujours, à première vue, concordantes.

La définition la plus courante de l'*endémie* est basée sur une certaine constance de la fréquence des cas, fréquence qui est parfois considérée dans le *temps* (le taux de malades dans une aire donnée a, « bon an, mal an », une valeur à peu près constante), parfois dans l'*espace* (pendant une époque donnée, la fréquence des cas est à peu près la même dans les divers départements d'un pays, ou les diverses villes d'un département, etc...); parfois on exige que les deux conditions de constance, dans l'espace et dans le temps, soient réunies.

Si on se reporte à la définition étymologique des deux états, il faut admettre qu'il y a *endémie* ou *épidémie* selon que la maladie est présente dans la population, ou, au contraire, d'origine externe. Cette distinction ne doit pas être mise sur le compte de l'agent pathogène (ce qui conduirait, par exemple, à considérer le rhume comme une maladie endémique par le seul fait que le virus correspondant est présent en permanence chez la majorité des êtres humains). En effet, les mots d'*endémie* et d'*épidémie* ont été conçus bien avant la notion d'agent pathogène. Dans ces conditions, si on conclut à l'*endémicité* d'après la présence, non de l'agent, qui a une valeur causale, mais de la maladie elle-même, qui est un effet, il est difficile de concevoir cette « présence » autrement que sous l'aspect d'une certaine constance du taux de malades, qui nous ramène, en définitive, à la première définition.

Un troisième groupe de définitions fait de l'*épidémie* une manifestation de la contagion. Mais « Toute maladie infectieuse a un certain degré de contagio-

sité », ajoute Charles Nicolle. « Cependant, il est nombre d'infections qui ne se présentent que sous forme de cas isolés, parce que les conditions que la contagion exige ne se produisent pas » et, avec sa foi dans la naissance, la vie et la mort des maladies infectieuses, Charles Nicolle conclut « Sachons, en conservant ces termes commodes : endémicité, épidémicité, qu'il n'y a pas, entre les catégories qu'elles étiquettent, barrière qu'un agent pathogène ne puisse franchir, où derrière laquelle il puisse s'enfermer ».

Ainsi, là encore, la contagion ne peut être prévue d'après les propriétés de l'agent pathogène, et l'existence d'une contagion toujours possible ne saurait être décelée autrement que par ses effets, c'est-à-dire, en définitive, par une propagation de la maladie dans un certain domaine (espace, temps, etc...).

Ainsi, de quelque façon qu'on essaie de définir l'*endémie* et l'*épidémie*, on est ramené aux postulats suivants : la distinction entre ces deux états est basée sur la répartition des cas dans certains domaines; les deux états diffèrent l'un de l'autre par le degré de régularité de cette répartition, la frontière qui les sépare n'ayant d'autre signification que celle d'une commodité.

Ayant ainsi posé le problème, il ne semble pas que le pathologiste puisse mieux faire que d'en confier l'étude au statisticien : qui donc, en effet, mieux que celui-ci, pourrait examiner une loi de distribution, l'ajuster à une loi de probabilité, et choisir, en définitive, entre deux hypothèses, celle de causes aléatoires ou, au contraire, systématiques, en situant la frontière à la limite de signification conventionnellement admise de la probabilité 5 %?

L'épidémiologie de la mosaïque du Tabac, grâce aux nombreux avantages que nous avons exposés, devait permettre une étude fructueuse de cette définition statistique.

Une expérimentation a été entreprise, depuis 1944, à l'Institut des Tabacs de Bergerac, comportant chaque année l'observation d'une ou plusieurs plantations de quelques milliers de pieds chacune. Le virus, sous forme d'une souche parfaitement isolée de la mosaïque *Nicotiana virus 1*, était introduit initialement dans le système, par un moyen indirect qui se rapproche des conditions naturelles de contamination : il a été nettement établi qu'après récolte des plants de tabac dans une culture présentant de nombreux malades, le virus est susceptible de persister dans le sol (sous une forme d'ailleurs non explicitement connue) et d'infecter la culture de l'année suivante. Le sol est alors un réservoir virus.

Nous avons donc constitué de tels réservoirs-virus, en nous attachant à réaliser autant que possible, leur homogénéité sur toute la superficie d'un champ ; à cette fin, pour préparer le terrain de la culture de l'année ( $n + 1$ ) nous avons inoculé, au cours de l'année  $n$ , avec notre suspension de protéine-virus, soit la totalité des plants, soit une très grande proportion de plants convenablement randomisés.

Les plants une fois repiqués sur un tel sol au cours de l'année ( $n + 1$ ), un grand nombre de mosaïques se manifestèrent. L'objet de notre étude était de répondre à la question suivante : la répartition des pieds malades parmi l'ensemble des pieds présents est-elle compatible avec l'hypothèse que

*l'apparition de maladie a une probabilité p constante dans la population totale* (indépendante, notamment, des coordonnées du plant, de l'état de santé ou de maladie de ses voisins, etc...)?

Pour répondre à cette question, il fallait décomposer la surface du champ en tranches et étudier la répartition des taux de malades entre ces tranches. Deux catégories de tranches étaient particulièrement intéressantes, non seulement parce que les plus simples, mais aussi parce que susceptibles de mettre en évidence l'influence de facteurs isolés à dessein : les *lignes*, c'est-à-dire les rangées de plants dirigées selon la longueur du champ, et les *rangs*, dirigés dans le sens perpendiculaire. Les lignes, matérialisées par les sillons, correspondaient, comme dans une culture normale, à la direction des façons culturales (labours, hersages, sarclages, etc...). Les rangs, d'ordinaire sans signification particulière, servaient ici à isoler le facteur *ouvrière* (l'opinion étant couramment répandue que les mauvaises ouvrières, en blessant les racines au moment du repiquage, favorisent l'apparition de la mosaïque). Chaque ouvrière fut donc chargée de complanter un certain nombre de rangs, convenablement randomisés dans l'ensemble de la plantation.

En prenant comme tranches du champ, les *rangs* par exemple, nous avions  $n$  rangs de  $K$  pieds, où la maladie se présentait avec les fréquences  $f_1, f_2, \dots, f_n$ . Pour décider si ces fréquences, étaient compatibles avec l'existence d'une probabilité de maladie, constante dans la population totale, il suffisait, par exemple, de former l'expression :

$$\chi^2 = K \sum \frac{(f_i - p)^2}{p(1-p)}$$

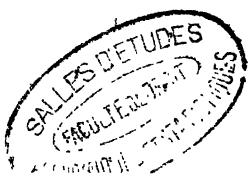
où  $p$  désigne le taux de malades, et de comparer la valeur trouvée à la limite de signification donnée par les tables pour  $(n - 1)$  degrés de liberté. Le même raisonnement s'applique à la distribution par *lignes*.

Les répartitions ainsi examinées, pendant trois ans, sur plusieurs parcelles, donnèrent les résultats suivants : la répartition des lignes a toujours été compatible avec l'existence d'une probabilité de maladie constante; au contraire, pour la répartition par rangs, deux cas se sont présentés :

a) dans le premier cas, la répartition par rangs s'est montrée conforme à l'hypothèse d'une probabilité de maladie constante. Dans ce cas, toute autre décomposition du champ en tranches (par carrés, par exemple, etc...) a donné le même résultat;

b) dans le deuxième cas, au contraire, on constate que l'hypothèse faite est hautement improbable. Ce fait tient à l'existence de rangs à nombre de malades nettement exagéré, en comparaison du taux général de malades. Ces malades ne sont d'ailleurs pas répartis au hasard dans ces rangs spéciaux, ils sont groupés en files ou *séquences*. Des tests statistiques permettent de montrer combien ces séquences sont improbables : en effet, dans l'hypothèse faite d'une probabilité constante  $p$  de maladie, il est facile de voir que la probabilité pour un pied d'appartenir à une séquence de  $n$  malades est (en première approximation)

$$P = N p^n (1-p)^{K-n}.$$



On peut donc ajuster la répartition observée des séquences à la répartition calculée.

Dans le cas, particulièrement marquant, de l'année 1945, nous avons obtenu dans une plantation les répartitions suivantes :

Pieds malades en séquence de . . . . .	1 (malades isolés)	2	3	≥ 4
Nombre calculé C de ces pieds . . . . .	452	163	44	13
Nombre observé O de ces pieds . . . . .	373	131	65	103

$\chi^2 = \Sigma \frac{(O-C)^2}{C}$  = plus de 6.000, alors que la limite de signification donnée par les tables est de 7,8. La valeur élevée de  $\chi^2$  provient des grandes séquences ( $\geq 4$ ), et traduit manifestement l'influence d'un facteur systématique. Ce facteur est certainement lié aux ouvrières planteuses, dont on avait à dessein isolé l'influence en faisant planter selon les rangs; de fait, en étudiant la fréquence des malades dans les lots plantés par les diverses ouvrières, on a pu déceler une nette systématité; ces fréquences sont, pour certaines ouvrières, exagérément grandes, et ceci parce que ces ouvrières sont responsables d'un excès, non pour les malades isolés, mais pour les grandes « séquences » de malades. Ces ouvrières ne sont d'ailleurs pas de « mauvaises ouvrières » comme le veut la légende : les ouvrières qui, au cours de l'été 1945, ont réalisé les « séquences improbables » sont, au contraire, les meilleures ouvrières de l'Institut.

Precisons que le facteur « ouvrière » n'est pas seulement relatif aux manœuvres opératoires des ouvrières planteuses, il contient en même temps les qualités du *plant* distribué à chaque ouvrière, et on peut évidemment expliquer les « séquences » en admettant que certains pieds aient contracté la maladie sur le semis, et que, dans les bottes de plants remis aux ouvrières, ils aient contaminé leurs voisins. Sans insister davantage sur cette hypothèse, qui pour diverses raisons ne nous semble pas tout à fait satisfaisante, ni sur les interprétations possibles du phénomène des « séquences » actuellement encore à l'étude, nous désirons attirer l'attention sur le fait essentiel des deux types de répartitions manifestés dans des conditions d'expérience apparemment identiques; *ces deux types de répartition nous semblent en effet correspondre de façon tout à fait satisfaisante à l'endémicité et à l'épidémicité.*

La statistique se révèle, dans une telle distinction, une auxiliaire indispensable; si, en effet, dans l'exemple cité à propos des « séquences », le nombre et l'importance de celles-ci étaient suffisants pour apparaître sans l'aide d'aucun critérium statistique, précisons qu'il n'en a pas toujours été ainsi; dans ces conditions, si des séquences importantes se présentent, rien ne permet de conclure *a priori* qu'elles correspondent à une influence systématique, plutôt qu'à une distribution aléatoire où la probabilité de maladie est élevée; inversement, l'absence de grandes séquences ne permet pas de conclure *a priori*, à l'existence d'une probabilité de maladie faible, plutôt qu'à l'inexistence de causes systématiques : autrement dit, *seul l'emploi de la statistique*

permet de conclure à une endémie de taux élevé là où les apparences conduiraient à voir une épidémie, et à une épidémie, là où les apparences conduiraient à voir une endémie à taux faible. Le statisticien saura, mieux que quiconque, sur quelles bases conventionnelles repose cette distinction entre l'endémie et l'épidémie; cette distinction, dans le cas présent, nous a révélé la plus ou moins grande probabilité pour que la maladie soit due, indépendamment de la contamination de base à partir du sol, à l'influence d'un facteur systématique, ici : le facteur « ouvrière ». Peut-on demander plus à cette distinction, que de nous orienter dans le labyrinthe de la pathologie causale, et les mots d'endémie et d'épidémie pourraient-ils mieux, dans ce sens, mériter l'épithète que leur donnait Charles Nicolle : « termes commodes »?

La méthode employée est susceptible de généralisation. Si en effet, les cas de maladie n'ont été examinés ci-dessus que du point de vue de leur répartition dans le plan, il est clair qu'on peut étudier cette répartition en fonction de tout autre facteur systématique : le temps, par exemple. Les cas de maladie apparus dans une aire donnée doivent alors être étudiés du point de vue de leur répartition en tranches chronologiques, et d'après la plus ou moins grande régularité de cette répartition, on pourra préciser si le temps est susceptible d'une influence systématique (ce qui se produit, par exemple, quand le taux de malades de l'année  $(n + 1)$  est fonction du taux de l'année  $n$ ).

Souvent cette influence est évidente sans qu'il soit besoin de recourir à la statistique (rougeole, par exemple, dans le cadre des maladies humaines). Dans d'autres cas, au contraire, on ne peut conclure *a priori*. C'est ainsi, par exemple, que pour la poliomylélite, une année à taux de malades élevé peut être encadrée entre deux années normales, et rien, jusqu'ici, n'autorise une prévision. Le tableau ci-dessous le montre bien pour la fameuse épidémie de 1926 en Angleterre :

#### POLIOMYÉLITE (Angleterre et Pays de Galles) (1)

ANNÉE	1913	1914	1915	1916	1917	1918	1919	1920	
Nombre de cas . . . . .	729	509	517	704	357	236	615	329	
ANNÉE	1921	1922	1923	1924	1925	1926	1927	1928	1929
Nombre de cas . . . . .	539	386	644	860	422	1296	895	546	618

Le calcul montre que ces fréquences annuelles ne sont pas compatibles avec l'existence d'une probabilité de maladie constante. Toutefois, du fait que l'aspect épidémique de la maladie nous est décelé par l'étude de la répartition *chronologique*, il ne faut pas conclure à une influence systématique du facteur temps : il peut se faire, par exemple, que les cas de chaque année soient eux-

(1) Rapport d'Hygiène de la Société des Nations, chiffres cités par LEVADITI et LÉPINE, *Les ultraviruses des Maladies humaines*, p. 256.

mêmes à répartition *topographique* épidémique. Si la maladie est contagieuse, il est bien plus intéressant de considérer les nombres de cas en échelle logarithmique (augmentation du taux proportionnelle au taux). Le tableau considéré ci-dessus devient après cette modification :

ANNÉE	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923
Indice de morbidité = log. du nombre des cas.	2,86	2,71	2,71	2,85	2,55	2,37	2,79	2,52	2,73	2,59	2,81
ANNÉE	1924	1925	1926	1927	1928	1929					
Indice de morbidité = log. du nombre des cas . . . . .	2,93	2,62	3,11	2,95	2,74	2,79					

La moyenne des nombres de cas, dans ce nouveau système, étant 2,74, on voit que la fameuse année 1926 n'est pas plus extraordinaire par sa performance élevée 3,11 (écart + 0,37) que l'année 1918 par sa performance faible (écart — 0,37). *Seul le statisticien pourra dire si la répartition des indices est significativement différente de la loi de Gauss, et si, par conséquent, l'accumulation de malades en 1926 n'est pas le seul fait du hasard, ce qui interdirait de conclure formellement à la cause d'après les conditions particulières de l'année 1926.*

La discussion précédente nous a montré que la maladie peut être endémique ou épidémique, selon que la répartition est étudiée topographiquement ou chronologiquement. De même, dans le cas de la mosaïque, et pour la seule répartition topographique, nous avons conclu à une endémie d'après l'examen des lignes, là où l'examen des rangs nous fait conclure à l'épidémie. La relativité de ces termes n'a rien qui doive choquer : il est clair, en effet, que, si l'épidémicité traduit l'existence de facteurs systématiques, ces facteurs passent nécessairement inaperçus dans tout domaine où ils sont convenablement randomisés. Si les ouvrières, au lieu de planter les pieds de tabac par rangs, avaient planté en des emplacements individuels distribués au hasard, la répartition des mosaïques eût semblé endémique dans le sens des rangs aussi bien que dans le sens des lignes. Mais en isolant le facteur ouvrière, nous avons, par l'apparition d'une épidémicité, découvert un maillon de la chaîne causale.

Il est certain qu'une maladie serait toujours épidémique, au sens où nous l'avons défini, pour l'observateur omniscient qui connaîtrait son déterminisme exact : si, par exemple, nous connaissions la répartition exacte des molécules de protéine-virus dans le réservoir sol, nous pourrions séparer les pieds de tabac en deux catégories dont la probabilité de maladie serait différente, selon la présence ou l'absence de l'agent pathogène. Le problème de cette relativité n'est autre, en définitive, que celui des causes énumérées ou au contraire inconnues, et le programme de l'épidémiologiste, soucieux de l'étiologie des maladies, consiste à transformer les causes inconnues en causes énumérées, *c'est-à-dire à faire apparaître l'épidémicité dans le comportement de toute maladie apparemment endémique.*

Cette méthode peut s'intituler : analyse causale du taux d'endémie.

#### ANALYSE CAUSALE DU TAUX D'ENDÉMIE

Dès lors que la répartition des malades est conforme au type endémique, une analyse causale peut être tentée. En effet, le taux d'endémie représente alors une fréquence dont les limites de précision sont parfaitement connues, et la comparaison de deux plantations ne différant que par un facteur systématique, permet d'éprouver l'influence de celui-ci sur le taux d'endémie.

Dans le cas de l'endémie, le nombre  $M$  de malades est assimilable au nombre des réussites dans une série de tirages au sort, où la probabilité de réussite est constante, le nombre des tirages étant le nombre  $N$  de pieds de la plantation. Dès que ce nombre est élevé, la distribution du nombre des réussites est gaussienne, la moyenne et la variance de cette distribution étant données par les formules usuelles de la loi binomiale. On établit donc facilement que deux parcelles de  $N$  pieds ont des taux de malades  $p_1$  et  $p_2$  significativement différents dès que

$$\frac{|p_1 - p_2|}{\sqrt{\frac{p_1 q_1 + p_2 q_2}{N}}} > 2 \quad (\text{avec } q_1 = 1 - p_1 \text{ et } q_2 = 1 - p_2)$$

Par exemple, si  $p_1 = 10\%$  sur 4.000 pieds, un facteur systématique susceptible d'abaisser de 2 % le taux de malades, peut être considéré comme abaissant la probabilité de maladie de façon significative. L'analyse causale du taux d'endémie permet de déceler ainsi avec une grande sensibilité les divers facteurs qui entrent en jeu dans le mécanisme de l'infection.

Ce n'est pas le lieu ici d'énumérer les facteurs envisagés, ni les résultats obtenus. Nous désirons seulement, à propos d'un exemple d'analyse causale du taux d'endémie, montrer comment cette méthode peut apporter une contribution utile au problème des rôles respectifs de l'agent pathogène et du terrain infectieux.

Rappelons que ce problème, lorsque Pasteur eut mis en évidence le rôle des germes pathogènes, fut l'objet d'une querelle mémorable entre les partisans de l'origine exogène de la maladie, ou « contagionnistes » et les partisans de l'origine endogène, ou « localistes ». Du vivant déjà de Pasteur, Claude Bernard, avec le redoutable appui de son autorité et de sa compétence, insistait tout particulièrement sur le rôle du milieu. Pasteur eut même à lutter contre sa grande ombre, puisque Bernard, dans une publication posthume, affirmait : « La levure de bière est la conséquence de la fermentation, et non sa cause... Les organismes élémentaires sont toujours conséquence ». Faut-il croire que Pasteur, ébranlé par ces lignes, ait, sur son lit de mort, comme le veut la légende, prononcé cet aveu : « Claude avait raison. L'agent n'est rien. Le milieu est tout » ?

Quoi qu'il en soit, il n'est pas douteux, à l'heure actuelle, que la présence de l'agent pathogène intervient dans le déterminisme des maladies infectieuses comme une condition nécessaire, mais non suffisante. Certes, au laboratoire, l'inoculation massive de l'agent pathogène spécifique entraîne inévitablement l'apparition de la maladie correspondante. Mais, dans les conditions naturelles, où l'agent pathogène est, en général, beaucoup plus dispersé, la susceptibilité

propre de l'individu reprend toute sa valeur. Cette conception ne saurait mieux s'exprimer qu'en termes de probabilité.

La probabilité de maladie chez un individu apparaît en effet comme le produit de plusieurs probabilités, dont l'une est incontestablement la probabilité de présence de l'agent pathogène, mais dont les autres intéressent le sujet susceptible : telle est, notamment la probabilité de pénétration des éléments infectieux à l'intérieur de l'hôte. En effet, dans le cas, par exemple, du sujet tabac et de l'agent virus, lorsque la probabilité de présence de l'agent pathogène est connue, — ce qui est le cas d'une inoculation expérimentale — et faible, le nombre des individus qui, dans une population importante, reçoivent un élément infectieux au moins, peut être évalué à partir des formules de Poisson : or, l'expérience prouve que le nombre observé diffère systématiquement du nombre calculé par la présence d'un facteur multiplicatif, qu'on peut appeler probabilité de pénétration. Au contraire, si la probabilité de présence de l'agent pathogène est très importante, ce qui est le cas d'une inoculation massive, il y a toujours au contact de l'hôte un nombre suffisant d'éléments infectieux pour que, même avec une faible probabilité de pénétration, l'un d'eux pénètre et détermine la maladie.

Il apparaît ainsi que le chercheur, soucieux de pathologie causale, doit imiter les conditions naturelles en expérimentant dans le cadre des inoculations légères et en appréciant, avec le concours de la statistique, les variations significatives du taux d'endémie qu'entraîne toute modification du sujet susceptible.

Cette méthode peut être utilisée, aussi bien pour les modifications d'ordre génétique, que pour les modifications dues à l'environnement.

a) Le premier problème rencontré par le pathologiste est, en effet, d'ordre génétique, puisque, à part le cas des populations humaines, il a, dans une certaine mesure la faculté de choisir son sujet. Or, il arrive généralement (et c'est le cas, notamment, pour le Tabac) que les individus réfractaires n'appartiennent pas à la même espèce que les individus sensibles, et dans le domaine de la génétique interspécifique, même dans le règne végétal où la stérilité interspécifique peut être conjurée par des moyens chimiques, on n'est jamais certain d'aboutir à une descendance qui allie les qualités intéressantes des deux parents. Par contre, la probabilité de pénétration des éléments infectieux semble avoir des variations notables d'une variété à l'autre de la même espèce, de sorte qu'en abandonnant l'idéal de l'individu réfractaire pour chercher seulement l'individu à faible probabilité de pénétration, on peut obtenir des résultats intéressants, tout en se maintenant dans le cadre relativement simple de la génétique intervariétale.

b) La deuxième voie offerte au pathologiste est celle de la modification des conditions d'environnement. Il a été prouvé que la probabilité de pénétration de l'agent infectieux dans une feuille varie, par exemple, avec la température, le degré hygrométrique de l'air, l'âge de la feuille, sa turgescence, etc... Nous avons constaté nous-mêmes que cette probabilité est notablement inférieure à la normale au moment du repiquage des pieds en raison, sans doute, du flétrissement passager des feuilles.

Dans ces conditions, il devient tentant de chercher des relations entre la probabilité de pénétration et ceux des facteurs d'environnement qui nous

échappent le moins dans les conditions naturelles, c'est-à-dire les conditions de nutrition. Une analyse causale du taux d'endémie fut entreprise dans cette voie en 1946 à l'Institut des Tabacs. Un des facteurs étudiés était la présence d'un engrais azoté, le sulfate d'ammonium, sous forme d'un épandage hivernal. L'expérimentation portant sur un total de 1.300 pieds, dont une moitié seulement avaient reçu l'épandage, donna les résultats suivants :

Taux d'endémie pour les témoins . . . . .	1,2 %
Taux d'endémie après épandage $\text{SO}_4^2 (\text{NH}_4)^2$ . . . . .	9,6 %

A l'aide du critérium indiqué plus haut, on constate aisément que cette différence est hautement significative (probabilité de  $X^2 < 10^{-9}$ ).

Mais l'augmentation du taux de malades après épandage du sulfate d'ammonium pouvait être due à l'influence de cet engrais, soit sur le niveau d'infection du réservoir sol, (action sur la probabilité de présence de l'agent pathogène), soit sur les qualités de sujet susceptible.

Une nouvelle expérimentation fut donc entreprise en 1947, dans laquelle l'agent pathogène était introduit, non par l'intermédiaire du réservoir-sol, mais par inoculation directe le jour du repiquage, dans des conditions de concentration suffisamment faibles pour que tous les pieds ne fussent pas mosaïqués (domaine des inoculations légères où, comme nous l'avons indiqué, peuvent se révéler les qualités de susceptibilités du sujet).

Les résultats suivants furent obtenus, sur 480 pieds en tout :

Taux de malades pour les témoins . . . . .	18 %
Taux de malades après épandage $\text{SO}_4^2 (\text{NH}_4)^2$ . . . . .	35 %

La différence étant, cette fois encore, très significative, on peut conclure que l'épandage de sulfate d'ammonium agit sur la probabilité de maladie par l'intermédiaire de la susceptibilité propre de l'hôte.

Il devenait tentant de mettre cette influence en relation avec la vitesse de croissance du sujet Tabac.

Un système de cotation fut alors mis au point, permettant de chiffrer le développement des pieds après le repiquage. Les « notes » ainsi attribuées, en fonction du nombre de feuilles, n'avaient pas de valeur pour les pieds considérés individuellement, mais seulement pour des collections de pieds suffisamment grandes. Elles étaient objectives, précises et de dispersion connue. Si alors on compare, pour les pieds mosaïqués au cours de l'été, d'une part, et, d'autre part, pour les pieds restés sains, la note qui leur avait été attribuée peu de temps après le repiquage, on trouve une différence nettement significative; par exemple, dans une parcelle observée en 1948, on avait les notes suivantes :

Note moyenne des pieds mosaïqués . . . . .	$1,90 \pm 0,08$
Note moyenne des pieds sains . . . . .	$1,54 \pm 0,08$

Plus précisément d'ailleurs, on a obtenu les taux suivants de malade, parmi les pieds de note donnée :

Note des pieds . . . . .	0	1	2	3	$\geq 4$
% de malades . . . . .	17	27	34	45	67

Des résultats analogues ont été obtenus dans toutes les parcelles observées, que le virus intervint par l'intermédiaire du réservoir sol ou par inoculation directe.

Cette corrélation nous semble tout à fait générale, et l'ensemble des faits signalés nous permet de supposer que la vitesse de croissance agit directement sur la probabilité de maladie (1).

Le sens de cette action rapproche cette conclusion de divers faits constatés aussi bien dans le règne animal que dans le règne végétal. Dans le règne végétal, l'addition d'azote au milieu nutritif accentue les symptômes d'un grand nombre de maladies cryptogamiques, bactériennes ou à virus. D'autre part, dans le règne animal, on a parfois constaté que l'addition de matières azotées au régime alimentaire diminue le pourcentage d'individus réfractaires à certaines maladies infectieuses (par exemple, si on donne à des souris un régime carencé en azote, ces souris sous-alimentées, débiles et misérables, résistent mieux que les témoins à la maladie de Theiler, en ce sens que le nombre des individus réfractaires après inoculation est plus grand, ce qui, par ailleurs, fait justice de l'opinion couramment émise d'après laquelle la sous-alimentation prédispose à toutes les maladies).

L'exemple du tabac, tel que nous l'avons cité, est un peu différent : il s'agit d'individus qui ne sont en aucun cas réfractaires à une inoculation massive, mais chez qui la probabilité de maladie, après inoculation légère, est modifiée par les conditions de croissance.

L'étude de corrélation que nous avons faite à ce propos constitue typiquement une analyse causale du taux d'endémie. Dans une population de plants ou la vitesse de croissance est distribuée de façon aléatoire, la répartition topographique des malades nous fait diagnostiquer une endémie. Mais dès que la répartition des malades est étudiée en fonction d'un autre facteur que les coordonnées du plant, tel que le facteur *vitesse de croissance*, nous ne concluons plus à l'hypothèse d'une probabilité de maladie constante : en même temps que nous apercevons une *épidémie de mosaïque parmi les pieds à forte vitesse de croissance*, une des causes, jusqu'alors non énumérée, de la maladie, est promue au rang de facteur systématique.

#### CONCLUSION

C'est un fait remarquable que, dès la plus haute antiquité, les savants, par leurs observations et leur intuition, ont attribué à des causes naturelles la plupart des maladies contagieuses. « Les médecins arabes, aussi bien que Galien et Hippocrate, fait remarquer le Dr Félix, ne croyaient guère à l'intervention des Dieux dans la propagation des fléaux qui désolent l'humanité, et leurs écrits prouvent la confiance qu'ils avaient dans la médecine et l'hygiène pratique pour les éviter et les combattre avec succès ».

Grâce aux observations précises et patientes de ces savants à la recherche

---

(1) L'hypothèse que la corrélation trouvée indique une influence de la maladie sur la vitesse de croissance n'a pas été retenue pour plusieurs raisons : en particulier, nous insistons sur le fait que la cotation de croissance a été effectuée à une date antérieure à l'apparition de toute maladie.

d'une légalité, l'Épidémiologie fut pendant des siècles une des bases fondamentales de la médecine. Mais elle ne pouvait progresser avec succès, parce qu'elle reposait sur la croyance erronée d'une origine endogène de la maladie. Les théories d'Hippocrate, reprises et à peine modifiées par Baillou et Sydenham, contribuèrent à propager pendant près de deux mille ans la conception des maladies naissant dans l'organisme sous l'influence des conditions de climat. Les quatre « constitutions » d'Hippocrate, types idéaux de saisons où sont associées dans leurs quatre combinaisons possibles le chaud ou le froid, la sécheresse ou l'humidité, figuraient de façon symbolique la profonde influence du milieu ambiant. Point de discrimination, d'ailleurs, entre les divers types de maladie : malgré la description précise des 42 « histoires », destinées à montrer le rigoureux déterminisme et la systématique des syndromes pathologiques, la maladie resta, pour les successeurs d'Hippocrate, « la fièvre », susceptible de « gradations, modifications, affinités ».

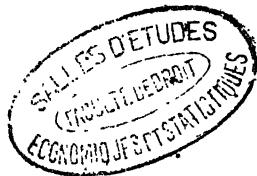
La révolution Pasteurienne changea radicalement l'aspect de la question. La légalité, cherchée pendant des siècles, se réduisit à la présence ou l'absence d'un agent pathogène, cependant que le problème des causes faisait place au problème, jusqu'ici négligé, de la spécificité. Ainsi naissait cette science nouvelle qu'est la Bactériologie, cependant que l'Épidémiologie périclitait rapidement.

Mais on put s'apercevoir bientôt que le succès remporté par les « contagionnistes » sur les « localistes » n'était qu'une victoire à la Pyrrhus, et nous avons montré comment le rôle du milieu, ou terrain infectieux, s'il n'est pas « tout » comme le voulait Claude Bernard, doit, néanmoins, être pris sérieusement en considération dans les conditions naturelles de l'infection.

La synthèse des deux conceptions opposées est possible, et de leur union peut renaître une Épidémiologie exacte et féconde, où la Statistique jouera un rôle de premier plan. Depuis près d'un siècle, les épidémiologistes, soucieux d'établir leur discipline sur des bases solides, ont groupé leurs observations en statistiques, ces signes avant-coureurs de *La Statistique*; il y a 30 ans déjà, le Dr Rist se livrait à d'ingénieux calculs statistiques, dont il vous a montré, en 1936, à quelles découvertes ils ont permis d'aboutir en matière d'épidémiologie typhique et paratyphique; d'autres observations épidémiologiques du Dr Rist, en matière de tuberculose, développées par notre collègue M. Divisia, ont mené à des conceptions intéressantes sur le pouvoir contaminant d'un milieu. De nombreuses études d'Épidémiologie statistique ont été publiées dans les années qui viennent de s'écouler. Il ne fait plus aucun doute, maintenant, que l'Épidémiologie sera d'essence statistique, ou ne sera pas.

Nous avons montré, en effet, comment se présente à l'heure actuelle, le problème des causes en pathologie infectieuse, et indiqué que la synthèse des conceptions « contagioniste » et « localiste » pouvait se réaliser admirablement en termes de probabilité, la probabilité de maladie étant le produit de plusieurs probabilités : probabilité de présence de l'agent pathogène, probabilité de pénétration à l'intérieur de l'hôte, probabilité d'absence des réactions de défense de cet hôte, etc...

Ainsi l'épidémiologie n'échappe pas à cette règle générale des Sciences



modernes qu'est l'envahissement fécond par la notion de probabilité. L'épidémiologiste doit le savoir, qui définira l'endémie et l'épidémie d'après l'ajustement à des lois de probabilité, et cherchera, dans l'endémie, les épidémies masquées qui lui mettront en évidence de nouveaux facteurs systématiques.

Quant au médecin, indépendamment de sa mission thérapeutique, il sait bien lorsqu'il cherche à faire œuvre de prophylaxie, que nous n'exigeons pas de lui le barrage absolu devant toute maladie infectieuse, mais que nous attendons seulement de son action la diminution significative de notre probabilité de maladie.

Le patient, enfin, enveloppé dans un réseau de probabilités, n'aura d'autre ressource que de se répéter un aphorisme célèbre dans sa leçon statistique moderne : l'homme bien portant est un malade probable.

Daniel SCHWARTZ.

#### DISCUSSION

M. LÉVY-BRÜHL demande à M. Schwartz de vouloir bien préciser les précautions qu'il avait prises, lors des expériences si intéressantes qu'il vient d'exposer, pour vérifier l'homogénéité du terrain, comme le font les expérimentateurs anglo-saxons. Il lui paraissait, en particulier, utile de vérifier que les séquences de plants atteints de la maladie selon une certaine direction ne pouvaient être attribuées à des facteurs géologiques ou autres qui en auraient fait une direction privilégiée.

M. SCHWARTZ assure que toutes les précautions nécessaires avaient été prises pour éliminer l'influence du sol. En particulier on observa que les séquences se retrouvaient en considérant les pieds de deux en deux lorsqu'une ouvrière plantait, dans une rangée, les pieds 1, 3, 5, 7... tandis qu'une autre plantait les pieds 2, 4, 6, 8.

---