JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

DANIEL SCHWARTZ

Les essais thérapeutiques et les enquêtes épidémiologiques

Journal de la société française de statistique, tome 140, n° 3 (1999), p. 33-43

<http://www.numdam.org/item?id=JSFS_1999__140_3_33_0>

LES ESSAIS THÉRAPEUTIQUES ET LES ENQUÊTES ÉPIDÉMIOLOGIQUES

Daniel SCHWARTZ *

1. LES ESSAIS THÉRAPEUTIQUES

L'évaluation scientifique de traitements chez l'homme est une pratique relativement récente. Pour la méningite tuberculeuse, un seul succès a pu suffire pour consacrer la streptomycine, parce que la maladie ne guérit jamais spontanément. Mais de telles situations sont exceptionnelles, en règle générale la variabilité qui imprègne les sciences de la vie régit aussi l'évolution des maladies : certains malades guérissent, d'autres non, spontanément ou sous traitement. Ce qu'il faut, c'est savoir si la *probabilité* de guérison est plus forte avec le traitement que sans le traitement, ou avec un nouveau traitement qu'avec l'ancien. Le plus souvent le gain, s'il existe, est modeste. Dans ce domaine des différences fines, la seule démarche rigoureuse est la méthode statistique. Mise au point dans le domaine agronomique, pour comparer des engrais, elle fut utilisée pour la première fois dans le domaine médical, en 1948, par le *Medical Research Council* pour éprouver l'effet de la streptomycine, cette fois dans la tuberculose pulmonaire. De là devait naître, incessamment perfectionnée, une méthodologie conférant à l'évaluation des traitements la rigueur réservée jusqu'alors aux expériences de laboratoire. On peut, en gros, la résumer en quatre principes.

Premier principe : la comparaison

Un groupe de patients recevant le traitement à évaluer, dit groupe **traité**, sera comparé à un groupe **témoin**, soumis à un traitement de référence, en principe le meilleur traitement connu. S'il n'existe pas de traitement actif, et dans ce cas seulement, le groupe témoin peut ne pas recevoir de traitement. Cette nécessité d'une comparaison est souvent oubliée, un médecin qui a obtenu dans un groupe traité 40 % de guérisons s'imagine qu'il a « évalué » le traitement. Comment pourrait-il en être ainsi, alors qu'il ignore quel pourcentage il aurait obtenu sans traitement ? Peut-être 40 % aussi ? Peut-être même plus ? *En bref, qui dit évaluation dit comparaison.*

* INSERM, unité de recherche U 292, Hôpital de Bicêtre, 82, rue du Général Leclerc, 94276 Le Kremlin-Bicêtre Cedex.

Deuxième principe : le test statistique

Pour comparer deux traitements, pour les taux de guérison par exemple, il faudrait connaître ces vrais taux, c'est-à-dire les *probabilités* de guérison. Mais on ne peut pas les connaître, on ne peut que les approcher par des fréquences observées sur des groupes (échantillons) de malades ; valeurs qui diffèrent plus ou moins des vraies valeurs en raison des **fluctuations d'échantillonnage**, ces fluctuations qui font qu'en tirant des échantillons dans une urne contenant 20 % de boules bleues, on observe des pourcentages fluctuant autour de 20 %. Seul le calcul des probabilités permet de dire si la différence observée est ou non raisonnablement imputable aux fluctuations d'échantillonnage. La méthode à utiliser est le **test statistique**.

Le principe en est le suivant : on fait l'hypothèse que la vraie valeur du taux de guéris est la même avec les deux traitements, et on calcule la probabilité d'obtenir sous cette hypothèse une différence au moins aussi grande que celle observée ; si cette probabilité est faible, on rejette l'hypothèse. Ainsi, dans le cas de deux groupes de 200 malades avec 50 % et 30 % de guéris respectivement, le calcul montre que, si les vrais taux de guérison étaient les mêmes avec les deux traitements comparés, on aurait moins de 1 chance sur 10.000 d'observer une différence au moins aussi grande que (50 % − 30 %). L'hypothèse d'égalité des vrais taux de guérison est donc hautement invraisemblable.

Il est usuel de considérer une différence comme **significative** quand la probabilité calculée est inférieure à 5 %. Plus cette probabilité est au-dessous de 5 %, plus le **degré de signification** est élevé. Ainsi, dans l'exemple présenté, où $p = 1$ pour 10.000, il est très peu vraisemblable que la différence provienne des fluctuations d'échantillonnage, très vraisemblable qu'elle traduit une différence réelle : la différence est hautement significative. (On précisera ici significative à 1 pour dix mille.)

Troisième principe : le tirage au sort

Supposons prouvé que le score du groupe traité diffère significativement de celui du groupe témoin. La cause de cette différence est-elle le traitement ?

Un accoucheur, pour sa thèse, compara le taux de complications dans deux séries d'accouchements, dans l'une on avait appelé un accoucheur, et dans l'autre non (l'accouchement avait eu lieu à domicile – l'histoire date d'il y a plusieurs années). Les taux de complications différaient très significativement, mais dans un sens qui le surprit : le taux était plus élevé quand on avait appelé l'accoucheur. C'est alors seulement qu'il réfléchit et comprit ce qui paraît évident : on avait davantage appelé l'accoucheur dans les cas difficiles, les deux groupes n'étaient pas comparables. En fait, dans ce travail, on n'avait pas *constitué* les deux groupes, ils s'étaient formés spontanément, c'était une enquête **d'observation**. Dans une enquête d'observation, les deux groupes se forment à partir d'un ou plusieurs critères, et, comme ceux-ci sont toujours liés à beaucoup d'autres, les deux groupes risquent fort de différer à de multiples égards. Pour obtenir la comparabilité, on ne peut recourir à une enquête d'observation, il faut **expérimenter**.

Encore faut-il, dans l'expérimentation, constituer des groupes comparables. Ce n'est pas simple. On ne peut pas, par exemple considérer comme comparables, les cas de deux époques différentes : les maladies évoluent, et bien plus encore le type de malades, ne fût-ce qu'en raison de l'amélioration des conditions diagnostiques. On ne peut donc utiliser, si attrayante qu'elle soit, la méthode dite « historique » où l'on donne le nouveau traitement dans l'année à venir, le groupe témoin comprenant les malades de l'année passée, qui ont reçu le traitement classique. On ne peut davantage affecter les deux traitements à deux hôpitaux différents car les clientèles de deux hôpitaux ne sont pas comparables. Dans ces deux cas, on a certes constitué les groupes, mais en fonction d'une caractéristique des malades (l'année ou l'hôpital où ils consultent), ici encore, comme toute caractéristique est liée à un écheveau de beaucoup d'autres, les groupes risquent fort d'être différents à de nombreux égards, la situation n'est pas meilleure que dans l'enquête d'observation. La solution consiste donc à affecter un malade à l'un ou l'autre des deux groupes par une décision indépendante de toutes leurs caractéristiques. C'est ce que seul peut faire, par définition pourrait-on dire, le tirage au sort. La technique de tirage au sort est appelée **randomisation**, du mot anglais *random*, emprunté d'ailleurs au mot du vieux français randon (d'où vient randonnée).

On retiendra de cette discussion la difficulté de l'imputation causale : celle-ci n'est possible, ni dans une enquête « d'observation », ni dans une expérimentation où on constitue les groupes à partir d'une de leurs caractéristiques. La seule solution est l'expérimentation avec tirage au sort.

Quatrième principe : les yeux fermés (si possible)

La comparabilité des groupes une fois obtenue au départ, encore faut-il qu'elle soit maintenue tout au long de l'essai. Or la connaissance du traitement par le malade et/ou par le médecin peut retentir différemment dans les deux groupes sur le mode de vie du malade, le comportement du médecin, l'observance des traitements, et même sur le cours de la maladie, qui peuvent être influencés par des phénomènes de suggestion. D'autosuggestion d'abord : des études ont montré qu'un « placebo », traitement inactif imitant dans sa présentation un traitement actif, pouvait, dans de nombreux cas, en reproduire les effets bénéfiques et parfois les inconvénients. D'hétérosuggestion ensuite : on a constaté également que la foi du médecin dans le traitement peut influencer le cours de la maladie. On ne citera ici qu'un seul exemple, particulièrement spectaculaire : quand parut sur le marché un « médicament miracle » contre l'asthme, un médecin décida de soumettre une de ses patientes à un essai rigoureux : il commanda au laboratoire, à des crises successives, tantôt le médicament, tantôt son placebo. Il obtint régulièrement succès dans le premier cas, échec dans l'autre. Lorsqu'il rapporta ces résultats au laboratoire, il s'entendit répondre qu'on lui avait envoyé à chaque fois... du placebo. La foi seule avait agi, illustrant de manière spectaculaire la relation « médecin-malade ».

Pour éliminer cet inconvénient, on a imaginé des essais où les traitements reçus sont ignorés du malade, voire du malade et du médecin : l'essai en **double**

insu ou **double aveugle** est celui où les traitements à comparer sont présentés sous une forme indiscernable, leur identité n'étant connue ni du malade ni du médecin. De telles précautions ne sont pas toujours possibles, ni même nécessaires. Cependant, toutes les fois qu'on évitera la méthode « aveugle » il faudra se souvenir de ce choix dans l'interprétation des résultats.

Signalons, pour terminer sur une note d'humour, que les anglais appellent essai en « triple aveugle » celui où les statisticiens ont égaré le code et où personne, au moment de l'analyse, ne sait qui a été traité par quoi.

Les quatre principes de l'essai contrôlé sont résumés dans l'encadré.

Les quatre principes essentiels de l'essai thérapeutique contrôlé

1. Evaluation = comparaison du groupe traité à un groupe témoin.
2. Comparaison = test statistique
3. Constitution de groupes comparables = tirage au sort (randomisation)
4. Maintien de la comparabilité dans le déroulement de l'étude = procédures à l'aveugle (quand c'est possible)

Le nombre de sujets nécessaire

Un élément essentiel dans un essai en est le nombre de sujets nécessaire, car de lui dépend la faisabilité de l'essai.

Le nombre de sujets nécessaire, très variable d'un essai à l'autre, dépend – par une formule assez simple – des paramètres suivants :

Le premier de ces paramètres concerne le critère d'évaluation. S'il s'agit d'un événement en tout ou rien – guérison ou non, décès ou survie – c'est sa fréquence. Dans l'essai du vaccin Salk, la poliomyélite frappait environ 50 enfants sur 100.000 dans une année. Avec quelques centaines d'enfants, on aurait selon toute vraisemblance observé 0 cas dans les groupes traité et témoin. On a dû monter à près de 200.000 enfants dans chacun des deux groupes. Ce paramètre ne dépend pas de nous.

Les autres paramètres dépendent de la prétention du médecin. En premier lieu, s'il tient à déceler une différence minime, il faudra un grand nombre de sujets. Le nombre de sujets d'un essai est comme la puissance d'une loupe; plus on veut voir un petit détail, plus il faut une loupe puissante.

En second lieu, le test statistique qui couronne l'essai comporte deux risques inévitables. S'agit-il par exemple, de comparer un groupe traité à un groupe non traité, le premier risque est de conclure que le traitement est efficace alors qu'il ne l'est pas (risque de première espèce α) le second est de ne pas conclure à l'efficacité alors que le traitement est efficace, donc de laisser échapper un traitement intéressant (risque de deuxième espèce β). C'est comme, en justice, les risques de condamner un innocent et de relaxer un coupable. On voudrait bien sûr que ces risques, s'ils ne peuvent être nuls, soient minimes. Mais des faibles risques, cela aussi se paie en nombre de sujets.

A titre d'exemple caricatural, voici les nombres de sujets nécessaires si, avec un risque d'erreur de première espèce α, on tient à avoir un risque au plus égal

à β de laisser échapper une différence Δ entre les taux, de guéris par exemple, des deux groupes (en test bilatéral, dans la zone de pourcentages 20 %), pour les choix suivants :

Δ	α	β	n (par groupe)
10 %	5 %	5 %	400
1 %	1 %	1 %	80.000

Le médecin, bien entendu, voudrait être très exigeant, déceler une petite différence, courir de petits risques, mais alors la formule le conduira vite à un nombre de sujets dépassant les possibilités, et il ne pourra pas dire au statisticien : « je ferai sans vous », car s'il opte pour un effectif moindre, la formule étant incontournable, cela veut dire qu'il se contentera, sans le savoir, d'exigences plus faibles que celles qu'il souhaitait. Le nombre de sujets ne peut donc résulter que d'une négociation : le médecin indique le nombre de sujets disponibles, le statisticien lui précise quelles performances l'essai peut alors apporter. Si elles sont raisonnables, l'essai est faisable d'emblée. Si elles ne le sont pas, on peut faire appel à des aménagements de deux sortes :

a) *une augmentation* du nombre de sujets disponibles peut être obtenue par un essai intercentres. De nombreux essais sont à l'heure actuelle conduits à l'échelle nationale ou internationale.

b) *une diminution* du nombre de sujets nécessaire peut être obtenue par **un plan expérimental** ou une **démarche séquentielle**.

Un exemple particulièrement simple de **plan expérimental** est celui où, au lieu de constituer deux groupes indépendants, on utilise le sujet comme son propre témoin. Par exemple, pour comparer deux somnifères, on peut donner à des patients une fois l'un et une fois l'autre (dans un ordre tiré au sort) et comparer ainsi les résultats sur les mêmes sujets. Ce plan demande un nombre de sujets nettement plus faible, d'abord parce que chacun « sert » deux fois, ensuite parce que la comparaison est plus fine, plus « puissante ».

Dans la **démarche séquentielle**, l'objectif est le suivant. Si la différence entre les deux traitements est, soit importante, soit négligeable, il doit être possible de s'en apercevoir avant la fin prévue de l'essai et de l'arrêter plus tôt. On fait donc le point – lorsque c'est possible – au fur et à mesure du déroulement de l'essai (*cf.* fig. 1).

Un cas particulier est l'analyse intermédiaire. On décide à l'avance de faire une première analyse, disons à mi-parcours, et en fonction du résultat on arrête ou on continue. Cette procédure, très fréquemment utilisée actuellement, soulève en fait un difficile problème sur le plan éthique : si en effet on obtient en analyse intermédiaire une différence non significative, mais appréciable, on a du mal à continuer l'essai, ce qui est pourtant méthodologiquement nécessaire. La solution de plus en plus souvent adoptée - mais est-ce une solution ?- est de confier la décision à un comité indépendant de surveillance formé de cliniciens, de statisticiens et d'éthiciens.

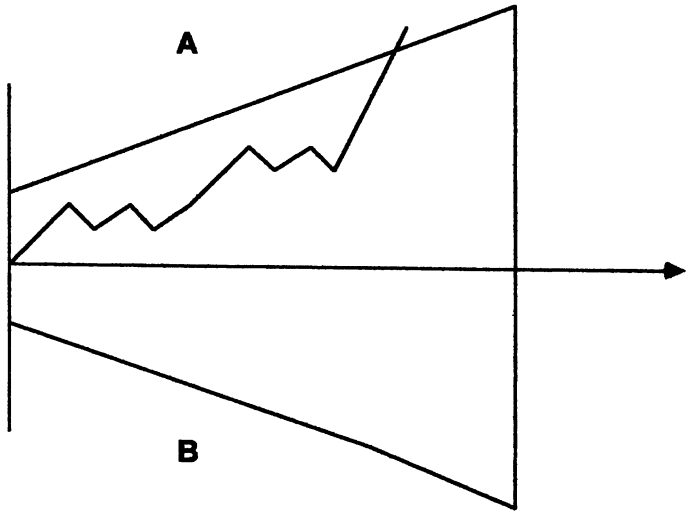

FIG 1. — Un exemple de démarche séquentielle. Les patients sont associés par paires, l'un recevant le traitement A, l'autre le traitement B. On fait le point après chaque paire, et on représente celle-ci par un segment à 45 degrés vers le haut ou vers le bas selon que le résultat avec A est supérieur ou inférieur à celui de B. On adoptera A si le chemin coupe la frontière A (cas de la figure), B s'il coupe la frontière B. S'il coupe la verticale, l'essai est arrêté, la différence n'étant pas significative.

Cette machinerie

La machinerie complexe des essais peut paraître irréaliste, inutile, inéthique.

Irréaliste ? Dans le seul domaine du cancer, on a dénombré, sur le plan international, en dix ans, près de 1.000 essais avec tirage au sort.

Inutile ? Pensons à l'exemple du vaccin Salk contre la poliomyélite. L'essai effectué en 1957 sur près de 400.000 enfants dont la moitié, tirée au sort, reçut un placebo, prouva en 15 mois l'efficacité (et l'innocuité) du vaccin, permettant sa diffusion très rapide. Quel contraste avec les innombrables controverses dont a fait l'objet le vaccin B.C.G., qui n'avait été soumis au départ à aucun essai de ce genre !

Inéthique ? C'est certes l'impression qu'on ne peut manquer de ressentir devant le tirage au sort, l'essai en double insu.

Cependant le Comité National d'Ethique, dans un avis promulgué en 1984, a déclaré que la méthodologie de l'essai thérapeutique était, sous certaines réserves qu'il a énumérées, non seulement acceptable, mais même indispensable.

En fait, le principal problème d'ordre éthique est celui du tirage au sort. Celui-ci n'est évidemment acceptable que dans la « situation d'équivalence » où l'on ne sait absolument pas lequel des deux groupes est avantagé. Quand se profile un nouveau traitement, on espère qu'il est meilleur, mais cela n'est pas prouvé et il y a une part d'inconnu quant aux effets indésirables. On

peut donc très bien admettre pendant un certain temps qu'on ignore s'il vaut mieux être dans le groupe de référence ou dans le groupe recevant le nouveau traitement. Dans cette situation d'équivalence, autant faire la répartition par tirage au sort, parce qu'il y aura au moins l'avantage que le test statistique et son interprétation en termes de causalité seront possibles, ce qui autrement ne serait pas le cas. C'est pour cette raison qu'il est alors éthique de tirer au sort. Quand on n'est pas dans la situation d'équivalence, c'est-à-dire si on a une forte impression qu'un traitement est plus avantageux que l'autre, on ne peut, en principe, pas faire l'essai.

Il y a quelques années, le docteur Maupas avait mis au point un vaccin contre l'hépatite B, et l'avait essayé sur des patients sans groupe témoin. Le jour où, pour le commercialiser, une évaluation rigoureuse s'imposa, près de 3.000 patients avaient été vaccinés avec succès sans aucun inconvénient secondaire. Faire alors un essai correct avec groupe témoin non traité posait donc un bien délicat problème.

En définitive, il est indispensable non seulement de procéder à des essais rigoureux, mais en outre de les mener à un moment opportun, qu'il ne faut pas laisser passer.

2. LES ENQUÊTES ÉPIDÉMIOLOGIQUES

L'épidémiologie qui, dans son acception moderne, concerne les maladies transmissibles ou non, vise à déterminer la fréquence d'une maladie, les facteurs qui la gouvernent (« facteurs de risque ») et les traitements susceptibles de la diminuer ; elle comporte donc trois volets, les épidémiologies **descriptive**, **analytique**, et **expérimentale**. Je me bornerai à décrire le deuxième volet, l'étude des facteurs de risque, et ceci très brièvement, mon seul but étant de montrer le contraste entre ce domaine et celui des essais thérapeutiques sur le plan de l'imputation causale.

Un exemple

Un exemple ancien, mais qui est resté le plus illustre, a été la recherche des causes du cancer du poumon. Le cancer des bronches surgit, autour des années 50, d'une manière explosive et spectaculaire. Une immense floraison de recherches s'ensuivit aussitôt, dans toutes les directions possibles : expériences sur animal, analyse chimique, et surtout enquêtes épidémiologiques.

La plus importante des premières enquêtes fut menée en Grande Bretagne de 1948 à 1952. Doll et Hill y comparaient 1465 malades atteints de cancer des bronches et un nombre égal de témoins indemnes de ce cancer, choisis de telle manière qu'à chaque cancéreux corresponde un témoin de même sexe, de même âge, interrogé dans le même hôpital à la même époque. Le nombre de fumeurs s'avéra plus élevé dans le groupe « cancer des bronches », de manière hautement significative. Le tabac fut aussitôt incriminé. Cependant, les résultats de l'enquête furent critiqués. Un point délicat, entre autres, était que la qualité de la réponse devait différer selon que les patients interrogés

étaient des cancéreux ou des témoins. Ceci pour deux raisons : les enquêteurs avaient peut-être davantage poussé les cancéreux à s'avouer fumeurs, et les malades eux-mêmes avaient pu déformer la réalité, soit consciemment, soit inconsciemment par défaut de mémoire. Dans les enquêtes de ce type, dites **rétrospectives** parce que portant sur les années antérieures à l'interrogatoire, ce « biais d'interrogatoire » est un inconvénient majeur, il entraîne de sérieux doutes.

Pour éviter ce biais, il fut alors décidé de mener des enquêtes **prospectives**, dans lesquelles des sujets « tout venants » seraient interrogés sur leur consommation de tabac ; ils seraient ensuite suivis pendant plusieurs années, et on noterait les décès et leurs causes. On réalise alors la situation dite « aveugle », parce que l'éventuel cancer à venir est ignoré, au moment de l'interrogatoire, par le sujet comme par l'enquêteur. L'inconvénient bien sûr est que, pour avoir un nombre de cancers suffisant, il faut partir d'un échantillon numériquement important. Les deux premières enquêtes furent menées, l'une en Grande Bretagne, l'autre aux U.S.A. L'enquête britannique de Doll et Hill porta sur la totalité du corps médical anglais, soit environ 30.000 personnes. L'enquête américaine, de Hammond et Horn, porta sur près d'un million de sujets, interrogés par 60.000 enquêteurs bénévoles, puis suivis pendant 5 ans. Ces deux enquêtes firent apparaître chez les fumeurs un taux plus élevé, de manière hautement significative, non seulement de cancers bronchiques, mais de toute une série d'autres cancers et d'autres maladies, notamment cardiovasculaires.

Cependant, est-il besoin de le dire, les enquêtes des deux types présentent, l'une comme l'autre, un inconvénient majeur : c'est la difficulté de l'imputation causale.

On s'imagine souvent que l'enquête prospective, parce qu'elle est « aveugle » justifie l'imputation causale. C'est là bien entendu une erreur. Prospectives ou rétrospectives, ces études sont toujours des enquêtes d'observation, les groupes sur lesquels elles portent sont en principe non comparables. C'est ce que confirmaient d'ailleurs les données : les fumeurs différaient des non fumeurs par de nombreux caractères, notamment par une consommation de café et d'alcool en moyenne plus élevée, par leur milieu social, par leur taille également (les fumeurs mesuraient en moyenne... 1 cm de plus). Alors, la cause du cancer est-ce le tabac, l'alcool, le café (...ou le centimètre) ?

C'est ici qu'apparaît le contraste entre essais thérapeutiques et enquêtes étiologiques. Dans les premiers, de nature expérimentale, on peut obtenir par tirage au sort des groupes comparables et conclure à la causalité. Mais dans les enquêtes étiologiques on ne peut bien sûr pas tirer au sort les sujets qui seront exposés au tabagisme ou à tout autre facteur supposé dangereux, on doit se contenter d'enquêtes d'observation qui n'autorisent pas d'emblée l'imputation causale.

Les difficultés de l'imputation causale

Pour accuser le tabagisme à bon escient, il fallut rechercher toute une série d'arguments.

On eut d'abord recours à « l'analyse multivariée » ; c'est là une procédure permettant d'étudier le rôle d'un facteur en neutralisant le rôle d'autres facteurs qui lui sont liés. C'est ainsi que la fréquence du cancer bronchique est augmentée quand les sujets fument mais aussi quand ils boivent de l'alcool. Or les fumeurs boivent davantage que les non fumeurs. On peut donc se demander si le rôle d'un de ces facteurs n'est pas le simple reflet de l'autre. L'analyse multivariée montra que, pour une consommation de tabac donnée, l'alcool ne jouait plus aucun rôle, tandis que pour chaque niveau de consommation d'alcool, la fréquence du cancer bronchique restait plus élevée chez les fumeurs. D'une manière générale, on constata que le rôle du tabac persistait quand on prenait en compte, outre la consommation d'alcool, celle de café, le niveau social, et de nombreux autres facteurs.

On procéda par ailleurs à la comparaison de divers sous-groupes, et on put ainsi montrer, par exemple, que le risque est plus faible chez les petits fumeurs que chez les gros fumeurs, plus faible aussi chez ceux qui ont cessé de fumer.

Aucun de ces résultats n'est à lui seul probant. Ainsi, « l'analyse multivariée » ne prend en compte que les facteurs connus et mesurables, et la comparaison des sous-groupes (petits et gros fumeurs, sujets qui ont ou non cessé de fumer, etc...) ne fait pas disparaître la tare foncière de l'enquête d'observation qu'est la non comparabilité plausible des groupes : ainsi les gros fumeurs sont sans doute différents des petits fumeurs pour diverses caractéristiques, leur excès de cancer n'est donc peut-être pas causé par leur consommation plus élevée de tabac. C'est seulement l'accumulation d'un grand nombre de telles constatations cohérentes, ajoutée à l'expérimentation animale et à l'analyse chimique de la fumée du tabac, qui ont permis d'établir avec certitude la responsabilité du tabagisme.

L'histoire de cet exemple permet de comprendre qu'on puisse, en épidémiologie, distinguer trois types de recherche selon l'objectif affiché en matière d'imputation causale.

Premier objectif

La preuve de la relation causale

L'acquisition de la certitude dans le problème du tabac et du cancer a été un grand moment dans l'histoire de l'épidémiologie, elle apparut comme une éclatante victoire ; mais les moyens mis en oeuvre avaient été considérables.

Des études aussi gigantesques sont parfois menées ; on citera par exemple les enquêtes visant à établir le rôle causal du cholestérol dans l'infarctus du myocarde. Mais de telles études sont rares (sauf dans le domaine des risques professionnels, dont la nature causale est moins difficile à prouver). En fait, les épidémiologistes s'orientent plutôt vers des enquêtes n'ayant pas pour objectif la certitude d'une relation causale.

Deuxième objectif
La présomption de causalité

Le deuxième type consiste à viser seulement une présomption, plus ou moins forte, de relation causale; présomption qui reste très utile, car elle permet soit de guider provisoirement des décisions, soit d'orienter des recherches. Les épidémiologistes ont établi des listes de critères susceptibles d'entraîner cette présomption. Voici, à titre d'exemple, une liste proposée par Bradford Hill.

1. Force de l'association
2. Relation dose-effet
3. Pas d'ambiguïté sur la chronologie
4. Constance des résultats dans diverses études
5. Plausibilité de l'hypothèse
6. Cohérence des résultats
7. Spécificité de l'association

Il s'agit là d'une série d'arguments de bon sens permettant un jugement analogue à celui qui motive l'intime conviction qu'un accusé est coupable ou innocent. Aucun de ces critères, et même leur ensemble, ne permettent une conclusion assurée. A titre d'exemple, rappelons que le tabac, sûrement coupable de causer le cancer bronchique, ne vérifie le critère de spécificité dans aucun des deux sens : il n'est pas la seule cause, et il occasionne d'autres pathologies.

Comme contre exemple, sur un mode humoristique, on évoquera le lever du soleil. La cause est-elle, comme le proclame Chantecler, dans la pièce d'Edmond Rostand, le chant du coq ? La relation vérifie de multiples critères de la liste de Bradford Hill : force de l'association, loi dose-effet (le coq s'époumone – le soleil est éclatant), l'ordre chronologique, si important (le chant précède le lever du soleil), sans parler des autres critères, le dernier de la liste étant le plus frappant : le coq chante-t-il au lever de la lune, les canards chantent-ils au lever du soleil ? Il y a spécificité à double sens...

Troisième objectif
Le renoncement à l'imputation causale

Le pas suivant, devant la difficulté de l'imputation causale, consiste à délaisser la recherche des causes. Ce renoncement fait penser au renard de La Fontaine devant les raisins trop haut perchés. En fait la démarche peut être très rentable. Il est souvent possible de guider l'action à partir de facteurs de risque d'une maladie, sans se préoccuper aucunement de savoir s'ils jouent un rôle causal. Ainsi, pour un sujet, l'existence de parents diabétiques n'est pas un facteur causal, mais un facteur de risque incitant à une surveillance. De même, en périnatologie, on a pu déceler de nombreux facteurs qui, chez la femme enceinte, font présager une mauvaise issue de la grossesse. La démarche statistique permet de ne retenir qu'un tout petit nombre d'entre eux qui, judicieusement combinés, désignent les « grossesses à risque ». Pour ces cas

sera adoptée une conduite ne visant pas à peser sur les facteurs de risque, dont le rôle causal est ignoré : on recommandera seulement aux femmes des visites plus fréquentes. Cette démarche méthodologique ne nous éclaire pas sur le déterminisme du mal à combattre, mais elle guide l'action, illustrant bien le fait qu'il n'est pas toujours nécessaire de comprendre pour agir.

Je voudrais, pour terminer, signaler que l'épidémiologie moderne a été inventée, non en Grande Bretagne comme on l'a longtemps cru, mais en France, par des précurseurs, notamment Pierre Louis, Broca, Gavarret, Villermé ; mais cette invention tomba dans l'oubli pour resurgir plus tard dans les pays anglo-saxons. Il y a à cela des raisons de plusieurs ordres, mais la principale est sans doute que les Français sont rétifs au mode de pensée statistique. C'est là un état de choses grave en santé publique et dans bien d'autres domaines, et qu'il faudrait absolument modifier. Ce mode de pensée devrait être enseigné dès le lycée par un programme adéquat. Si notre Académie pouvait intervenir dans ce domaine[1], elle rendrait au pays un signalé service.

Pour en savoir plus....

SCHWARTZ D. (1994). Le Jeu de la Science et du Hasard. La statistique et le vivant. *Flammarion*, Paris.

BOUVENOT G., ESCHWEGE E. et SCHWARTZ D. (1993). Le Médicament, Naissance, vie et mort d'un produit pas comme les autres *Inserm Nathan*, Paris.

1. (NDLR) C'est aussi le devoir de la SFdS. Notons cependant qu'au moment où cet article est imprimé des avancées encourageantes apparaissent dans les programmes des lycées.