JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

DANIEL SCHWARTZ

Connaissance ou décision. Attitude explicative - Attitude pragmatique

Journal de la société statistique de Paris, tome 135, n° 2 (1994), p. 3-18

<http://www.numdam.org/item?id=JSFS_1994__135_2_3_0>

© Société de statistique de Paris, 1994, 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/

I

COMMUNICATION

CONNAISSANCE OU DÉCISION
Attitude explicative – Attitude pragmatique

Professeur Daniel SCHWARTZ[1]

La distinction entre quête de connaissance et prise de décision est un problème d'ordre très général, mais je l'exposerai principalement dans le domaine de l'épidémiologie et plus particulièrement des essais thérapeutiques, parce que c'est là qu'elle est née et s'est développée avec le plus d'intensité ; on verra ensuite facilement qu'elle s'étend à de très larges secteurs de la recherche scientifique en général.

1. L'essai thérapeutique

Un essai thérapeutique vise à comparer deux traitements. À cet effet un premier groupe reçoit le traitement à évaluer, tandis qu'un groupe témoin reçoit le traitement de référence, en principe le meilleur connu à l'époque, ou aucun traitement si, et seulement si, aucun traitement n'a fait la preuve de son efficacité. Les résultats des deux groupes (taux de guéris par exemple) sont comparés, en général par un test statistique, mais nous reviendrons sur ce point. S'il y a une différence significative, elle peut être attribuée aux traitements si les deux groupes sont, à part les traitements, comparables à tous égards, ce qui nécessite pour commencer de les constituer par tirage au sort. Chacun sait que de tels essais présentent bien des difficultés. Mais bien peu savent que l'une d'entre elles tient à l'ambiguïté du mot « comparer ». Ambiguïté telle qu'une proportion non négligeable d'essais parfois longs, difficiles, coûteux, n'ont pas eu de répercussions en pratique courante, tout simplement parce que le mot « comparer », en fait la formulation même du problème, n'avait pas fait l'objet d'une réflexion préalable suffisante.

Tout essai contrôlé vise à *comparer* les *effets* de *traitements* sur des *patients* ; son protocole doit donc définir avec précision ces 4 mots clefs. Je me propose de montrer que ces définitions peuvent revêtir des aspects radicalement différents selon l'un ou l'autre de deux types de problèmes, orientés, l'un vers la connaissance, l'autre vers la prise de décision.

---

1. Professeur Emérite à la Faculté de Médecine Paris Sud, ancien Président de la SSP. Conférence du 23 février 1994.

Premier type de problème :

Soit à éprouver l'efficacité d'une nouvelle molécule, par exemple d'un hypocholestérolémiant ou d'un hypnotique.

On se rend compte aisément que les définitions logiques des quatre mots clefs seront les suivantes :

a) *Traitements à comparer.* Pour pouvoir attribuer une différence entre les résultats des traitements (problème d'imputation causale), il est essentiel que les deux groupes soient, à part le traitement, toutes choses égales par ailleurs. Sans doute le tirage au sort a-t-il institué la comparabilité au niveau des sujets, mais il faut que cette comparabilité subsiste au cours de l'essai. Ceci exige d'abord que les traitements soient donnés avec le même mode d'administration (voie sanguine, voie orale) et même avec une présentation identique (une pastille rouge n'a pas le même effet qu'une pastille verte...). D'une manière générale, la connaissance du traitement par le malade, ou par le médecin, peut entraîner, par auto- ou hétérosuggestion, des différences, non seulement dans l'appréciation des résultats, mais dans l'évolution même de la maladie. Aussi est-il indiqué, toutes les fois qu'on le peut, de présenter les traitements sous forme indiscernable, leur identité étant ignorée du malade (essai en « aveugle »), voire du malade et du médecin (essai en « double aveugle »). Si le groupe témoin peut ne pas être traité, on lui administrera un *placebo,* substance inerte imitant dans sa présentation le traitement à évaluer.

b) *Critères de jugement :* peu nombreux, un seul si possible, traduisant un mécanisme d'action ou quantifiant l'efficacité : baisse de la cholestérolémie, ou nombre d'heures de sommeil.

c) *Les patients :* on sélectionnera un échantillon selon des critères d'optimisation : homogénéité, sensibilité..., le souci de représentativité étant secondaire ; d'abord, la preuve d'efficacité sur une catégorie quelconque de sujets est déjà en soi un résultat intéressant, mais, en outre, on teste une propriété pharmacologique très générale, vraisemblablement extrapolable d'une population à l'autre, l'essai prolongeant au fond l'expérimentation sur l'animal. Jacques MONOD disait : « Ce qui est vrai pour *Escherichia coli* (c'est une bactérie) est vrai pour l'éléphant ». Sans doute cette extrapolation, valable pour des lois universelles comme le code génétique, ne l'est-elle pas dans tous les domaines. Mais il reste qu'une relation du type causal comme « telle molécule a un effet hypnotique », si elle est valable pour un groupe, l'est sans doute pour d'autres.

d) Enfin la *comparaison* sera effectuée par le test statistique de l'hypothèse nulle.

On reconnaîtra dans les définitions qui précèdent celles de l'essai thérapeutique absolument classique, celui qui est le plus couramment pratiqué. C'est en quelque sorte une évaluation en conditions « de laboratoire ». La démarche est récapitulée, tableau 1, pour les quatre mots clefs.

Tableau 1. Quête de connaissance

Traitements	Conditions égales
Critères	Peu nombreux, explicatifs
Patients	Groupe choisi
Comparaison	Test

Deuxième type de problème :

Choix entre deux traitements, par exemple pour la polyarthrite rhumatoïde, entre un traitement de fond nouveau (N) et un traitement de fond classique (C).

On précisera qu'un tel essai est entrepris pour savoir lequel de ces deux traitements est le plus « intéressant », disons « utile » dans ses conditions habituelles d'emploi, de manière à pouvoir, à la fin de l'essai, opter pour l'un ou pour l'autre. On verra facilement que les quatre mots clefs du protocole doivent être ainsi définis :

a) *Traitements :* le traitement nouveau N pour un groupe, le traitement classique C pour l'autre, chacun avec ses conditions d'administration optimales : la meilleure dose (il n'est pas question de donner des doses égales, cela n'a pas de sens pour des substances différentes !), la meilleure voie d'administration, éventuellement les traitements adjuvants, chacun avec son contexte propre : si le régime sans sel est indiqué pour un des traitements, on ne va pas priver de sel les patients de l'autre groupe si ce n'est pas nécessaire. En bref, on ne se préoccupera aucunement de réaliser des conditions « égales », indispensables dans l'exemple précédent.

En outre, dans un problème de ce type, l'administration des traitements pourra être très souple. On pourra, par exemple, pour tel traitement, prévoir son remplacement par un autre en cas d'intolérance, si cette stratégie est celle qui se présenterait en pratique courante, car c'est elle alors qu'on souhaite évaluer. Mais cette stratégie doit bien sûr être prévue et décrite avec précision dans le protocole.

b) *Critères de jugement :* non seulement les nombreux critères cliniques, et de qualité de vie, permettant d'apprécier l'état d'un malade souffrant de polyarthrite rhumatoïde, mais aussi les inconvénients de ces traitements (coût au sens large). Il faudra ensuite bâtir à partir de ces critères un score exprimant le bilan Bénéfice-Coût.

c) *Patients :* il est cette fois indispensable d'avoir un échantillon de patients aussi représentatif que possible de la population à laquelle on voudra appliquer les conclusions de l'essai : en effet, le résultat N plus (ou moins) « intéressant » que C dans un bilan global n'est pas, comme la propriété pharmacologique testée dans le premier exemple, une conclusion de portée générale vraisemblablement extrapolable d'une population à l'autre. Et c'est pour une population bien déterminée qu'on veut comparer l'intérêt des deux traitements. Il faut donc en avoir un échantillon représentatif.

d) Enfin la *comparaison* est l'élément le plus original de la démarche. Elle ne fera pas appel au test de l'hypothèse nulle, mais à un procédé tout autre, pour au moins deux raisons :

- en premier lieu, s'agissant de comparer des traitements de nature le plus souvent différente, dans un bilan Coût-Avantages, l'hypothèse nulle d'équivalence, outre qu'elle n'a guère de sens, a une probabilité *a priori* nulle ;
- en second lieu, avec un test statistique, la non-conclusion est possible, alors qu'il faut, dans le présent problème, aboutir obligatoirement à un choix : dire lequel des deux traitements est le plus « intéressant ». On doit rechercher des solutions dans la théorie de la décision.

La définition des quatre mots clefs est récapitulée au tableau 2.

Tableau 2. Prise de décision

Traitements	Conditions optimales
Critères	Nombreux (efficacité, coût)
Patients	Groupe représentatif
Comparaison	Méthode décisionnelle

Nous avons, Joseph Lellouch et moi, proposé comme méthode de décision une démarche qui a l'avantage d'être simple. Cette solution consiste à opter, sans test, – je dis bien sans test – pour celui des deux traitements qui a donné, dans l'essai, le meilleur score. Dans ces conditions :

- On conclut toujours : le risque de deuxième espèce β est nul.
- Le risque de première espèce α est, en raison de l'antagonisme des deux risques, de 100 %.
- On encourt alors un risque, non négligeable, qu'on peut appeler risque de troisième espèce γ, de commettre l'erreur de signe, c'est-à-dire de déclarer supérieur à l'autre celui des deux traitements qui est en réalité inférieur à l'autre.
- Les trois risques sont définis au tableau 3. Les valeurs qu'il est souhaitable de leur attribuer vont différer totalement selon qu'il s'agit d'une quête de connaissance ou d'une prise de décision.

Tableau 3. Mode de comparaison de deux traitements et types d'erreurs (Essai thérapeutique)

Conclusion / Réalité	$B-A<0$	$B-A=0$ On ne conclut pas	$B-A>0$
$B-A<0$		Erreur de type II β	Erreur de type III γ
$B-A=0$	Erreur de type I $\frac{\alpha}{2}$		Erreur de type I $\frac{\alpha}{2}$
$B-A>0$	Erreur de type III γ	Erreur de type II β	

La figure 1 concerne la *quête de connaissance*, elle reproduit la démarche classique du test statistique. Soit à comparer, par exemple, deux moyennes. Dans l'hypothèse nulle, la différence est nulle ; la différence des moyennes observées fluctue selon la courbe de gauche. Les verticales L_1 et L_2 définissent l'intervalle de non-conclusion, sortie duquel la différence est significative. Le risque de première espèce est représenté par $\alpha / 2$ de chaque côté. Si l'hypothèse nulle n'est pas vérifiée (différence D non nulle) la fluctuation de la différence est représentée par la courbe de droite, et l'aire entre L_1 et L_2 exprime β. Le troisième risque, γ, dont on ne parle jamais, consiste dans ce cas de figure, où $D > 0$, à observer une différence à gauche de L_1 ; on conclut alors à une différence, mais négative, c'est-à-dire contraire à la réalité. Si l'on ne parle jamais de ce risque, c'est parce que, en règle générale, si on se fixe α et β petits, γ est négligeable : ainsi pour $\alpha = \beta = 5\ \%$, $\gamma \cong 10^{-7}$.

Dans la *prise de décision*, l'absence de test revient à faire se rejoindre (figure 2) les limites L_1 et L_2, confondues sur l'axe de la courbe de gauche. On a bien, dans ces conditions :

risque β = entre L_1 et $L_2 = 0$;

risque α = aire extérieure à $L_1 L_2$ dans la courbe de gauche = 100 % ;

risque γ, aire à gauche de $L_1 L_2$ dans la courbe de droite, non négligeable.

Il convient de s'attarder quelque peu sur les valeurs alors singulières de ces trois risques.

- La valeur $\beta = 0$ correspond bien à notre souhait de décision à tout coup.
- La valeur $\alpha = 100$ % étonne les néophytes. Quoi ? on accepte un risque de 100 % de se tromper ? Cet étonnement appelle deux réponses.

D'abord on ne se tromperait que 100 % des fois où *l'hypothèse nulle est vérifiée* mais, comme on l'a dit, l'hypothèse d'équivalence des deux traitements dans un bilan Coût-Efficacité a une probabilité quasi nulle, elle ne se produirait pratiquement jamais.

Ensuite, à supposer même qu'elle se produise, il arrivera 1 fois sur 2 (figure 2) qu'on observe un meilleur score pour un des deux traitements, mettons A, alors qu'en fait $A = B$. Mais cela se traduira, non par la proclamation, erronée sur le plan scientifique, $A > B$, mais par le choix de A comme traitement à adopter. Qu'importe qu'on ait choisi A plutôt que B puisqu'ils sont, par hypothèse, équivalents ?

Ainsi apparaît-il que le « risque d'erreur » α, souci dominant dans les essais type quête de connaissance est ici sans aucune espèce d'importance.

Reste alors le risque de troisième espèce. C'est là sans aucun doute le plus grave. C'est la rançon de qui ne fait pas de test. Il faut le minimiser, mais ceci peut être fait par un choix judicieux du nombre de sujets nécessaires. Celui-ci, déterminé habituellement en fonction de α et de β, le sera ici en fonction de γ. Son calcul est évident. On voit bien, d'après la figure 2, que, pour avoir un risque au plus égal à γ d'opter pour un traitement inférieur à l'autre d'une quantité Δ, il faut un nombre de sujets

$$n = (\varepsilon_{2\gamma})^2 \frac{2\sigma^2}{\Delta^2} \qquad (1)$$

Figure 1

Hypothèse nulle $\mu_B - \mu_A = 0$		Hypothèse alternative $\mu_B - \mu_A \neq 0$

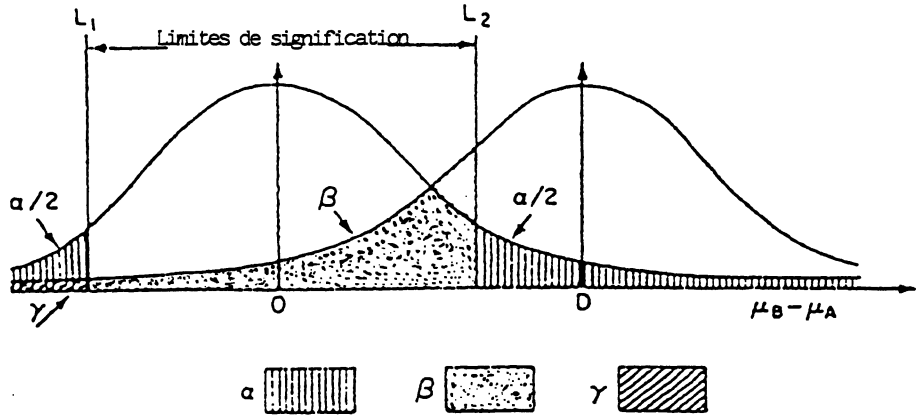
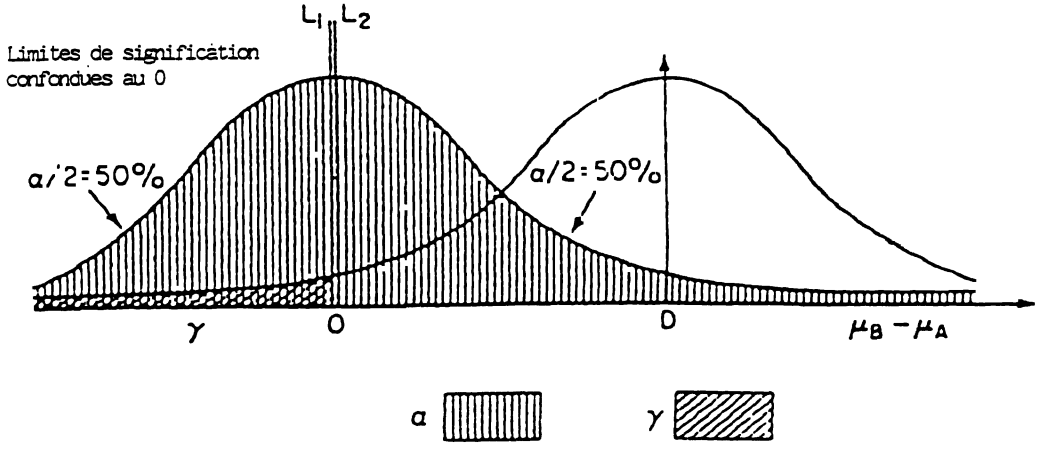


Figure 2

Hypothèse nulle $\mu_B - \mu_A = 0$		Hypothèse alternative $\mu_B - \mu_A \neq 0$



avec σ^2 = variance du score dans chacun des deux groupes et $\varepsilon_{2\gamma}$ = l'écart réduit correspondant au risque 2γ.

On voit la différence, mais aussi la ressemblance entre la formule (1) et celle, classique, donnant le nombre de sujets dans le test d'hypothèse

$$n = (\varepsilon_\alpha + \varepsilon_{2\beta})^2 \times \frac{2\sigma^2}{\Delta^2} \qquad (2)$$

La formule (1) suppose que « l'intérêt » d'un traitement soit mesuré par un indice unique, un « score » ; comme cet intérêt fait intervenir, en réalité, de multiples paramètres d'efficacité et de coût, il faut imaginer des méthodes permettant de ramener le bilan à une variable finale unique. Nous y reviendrons.

Le tableau 4 récapitule la définition des quatre mots clefs du protocole dans les deux types de problèmes envisagés. Ces définitions sont, comme on le voit, quasi opposées. Elles illustrent la distinction entre deux démarches qu'on peut appeler :

Quête de connaissance	ou	attitude explicative
Prise de décision	ou	attitude pragmatique

La définition des quatre mots clefs est si différente d'une attitude à l'autre qu'on ne peut se passer, dans la formulation du problème, de faire un choix.

Tableau 4

Traitement	Conditions égales	Conditions optimales
Critères	Peu nombreux (explicatifs)	Nombreux (efficacité, coût...)
Patients	Groupe sélectionné	Groupe représentatif
Comparaison	Test statistique	Pas de test
	↓	↓
	Attitude explicative	Attitude pragmatique

Le choix d'attitude

Trois exemples montreront les conséquences d'un mauvais choix. Le premier concerne les traitements, le second les critères et le troisième les patients.

Premier exemple : Dans un essai visant à choisir entre deux traitements, administrés l'un par voie orale, l'autre par voie intraveineuse, imaginons qu'on donne à chaque groupe le traitement prévu plus un placebo de l'autre. Ce souci d'égalisation, souhaitable pour une attitude explicative, serait une erreur dans cet essai de type

pragmatique, qui vise à comparer des traitements dans leurs conditions réelles d'utilisation. Les conclusions de l'essai seraient sans intérêt.

Deuxième exemple : Dans la prévention des thromboses post-opératoires, les études, longtemps quiescentes, ont été rajeunies par la mise au point du test au fibrinogène marqué ; l'apparition d'une thrombose se traduisant par une augmentation de la radioactivité locale, on peut ainsi déceler de manière très sensible des thromboses infra-cliniques qui, dans la plupart des cas, n'aboutissent pas à une phlébite clinique : elles sont qualifiées de « phlébites isotopiques ». Si on veut répondre à la question : « un traitement est-il *efficace* ? », ce critère peut être tout à fait indiqué : sensible, demandant par conséquent peu de sujets, de nature biologique permettant d'« expliquer » des phlébites cliniques, il convient bien à une attitude explicative. Mais, s'il s'agit de comparer *l'intérêt* de deux traitements, il faudra choisir comme critère la phlébite clinique, sans compter le cortège des autres critères d'efficacité et surtout de coût (les hémorragies, par exemple). Le critère de phlébite isotopique serait alors tout à fait inapproprié.

Troisième exemple : Il arrive que des traitements destinés à des pays en voie de développement soient essayés dans des hôpitaux de pointe de pays industrialisés ; dans ceux-ci les patients seront hospitalisés, surveillés, et suivront parfaitement le traitement prescrit ; dans ceux-là ils seront seulement revus périodiquement, la compliance sera sans doute mauvaise ; en outre les populations sont foncièrement différentes à tous égards ; on voit mal comment les conclusions de l'essai pourraient être extrapolées. Cet exemple peut paraître évident, trivial, sans intérêt. Nous verrons plus loin qu'il n'en est rien.

Ces exemples auront suffisamment fait comprendre qu'une option pour un mot clef non conforme à l'attitude souhaitée risque de rendre l'essai difficilement, voire totalement ininterprétable. En fait, on peut dire que si des essais d'un coût – au sens large – élevé sont restés sans incidence sur le progrès thérapeutique, la cause en était bien souvent un mauvais choix au niveau d'un ou plusieurs des mots clefs.

Rôles complémentaires des deux attitudes

Les deux attitudes occupent une place logique dans le décours de l'évolution des thérapeutiques : il faut d'abord tester l'efficacité de nouvelles molécules par des essais explicatifs ; il faut ensuite, pour une molécule nouvelle efficace, juger de son intérêt par un essai pragmatique : de son intérêt par rapport à des molécules classiques, s'il en existe, et s'il n'en existe pas, de son intérêt propre. On peut en donner un exemple très démonstratif.

Une vaste étude a été menée sur l'effet de l'aspirine dans la prévention de l'infarctus du myocarde. Sur près de 260 000 sujets consultés, 22 000, soit 8 % des sujets, ont été retenus pour l'essai (tableau 5).

Cet échantillon est, par principe même, en raison des clauses d'exclusion, non représentatif de la population des patients. On ne peut aucunement en faire grief aux auteurs, car il s'agissait d'un essai visant à démontrer l'efficacité d'un produit, donc

purement explicatif. Mais ce qu'on peut dire, c'est qu'une fois prouvée l'efficacité dans ces conditions de laboratoire, l'extrapolation à l'ensemble des patients ressemble (toutes proportions gardées) à l'extrapolation à la brousse d'essais conduits dans des hôpitaux de pointe. En particulier, la prise d'aspirine au long cours présente aussi des inconvénients, notamment un risque d'hémorragie. Le bilan des avantages et des inconvénients dépend du groupe envisagé. C'est bien pourquoi la communauté médicale n'a pas emboîté le pas d'emblée. Des essais de type pragmatique devraient suivre. Au fond, la logique est : l'animal, puis l'homme en groupe sélectionné (essai explicatif) puis l'homme en échantillon représentatif (essai pragmatique).

Tableau 5
La « Physicians'health study »
Prévention de l'infarctus du myocarde par l'aspirine

Population source :

Nombre de sujets

261 248 : Les sujets qui ont été sollicités : médecins, de sexe masculin, 40 à 84 ans, résidant aux USA en 1982.

112 528 : (43 %) des sujets ont répondu.

59 285 : (23 %) des sujets étaient disposés à participer.

33 223 : (13 %) des sujets répondaient aux critères cliniques d'éligibilité : pas d'antécédent d'infarctus, d'accident vasculaire cérébral permanent ou transitoire, de cancer, d'affection hépatique ou rénale, d'ulcère peptique, de goutte, pas de contre-indications à l'usage de l'aspirine ou d'utilisation habituelle, pas de prise de médicament antiplaquettaire, d'anti-inflammatoire non stéroïdien, et pas de supplémentation du régime en vitamine A (*).

22 071 : 8 % des sujets sont restés après les dix-huit semaines de période de préinclusion (acceptabilité, observance, état de santé).

(*) Ceci du fait de l'existence concomitante d'un essai visant à apprécier l'efficacité du carotène contre le cancer.

Un démarrage difficile

En dépit de la logique précédente, c'est la démarche explicative (l'essai classique) qui a presque toujours été adoptée, bien souvent à tort. Le développement des essais pragmatiques a, au départ, été lent. Ceci pour plusieurs raisons.

D'abord l'absence de test. Après de nombreuses années d'effort, les médecins commençaient à admettre, à prôner la démarche statistique. Quand ils ont entendu dire « Les statisticiens proposent de conclure sans test », ils ont entrevu, avec un mélange de crainte et de joie, le retour à l'âge d'or. Aux yeux des statisticiens, la

démarche sentait le soufre. Quant aux revues, un travail ne se terminant pas par *p* était bon pour la poubelle. Mais surtout la démarche est d'une application malaisée. Une des difficultés, et non la moindre, est d'établir, à partir de nombreux critères d'efficacité et de coût, un indice – un score – unique.

Différentes méthodes ont été envisagées. L'une par exemple est l'établissement d'une combinaison pondérée des différents critères. Une autre, la méthode du handicap, considère un critère unique, qui est le principal, mais on ne conclut à la supériorité d'un des traitements que s'il dépasse l'autre, pour ce critère principal, d'une certaine quantité – le handicap – déterminée en fonction des autres critères, et notamment du coût. Ces quantifications sont certes difficiles, mais elles s'appuieront sur le consensus d'un groupe de médecins particulièrement compétents. Et elles sont incontournables, car quelle est l'alternative ? Comment peut-on apprécier l'intérêt d'un traitement, en propre ou comparé à d'autres, sans lui affecter d'une manière ou de l'autre, un score à finalité décisionnelle ? Si l'essai pragmatique n'est pas là pour le faire, c'est le médecin qui le fera, consciemment ou inconsciemment. Et ce sera moins bien.

En définitive, la démarche pragmatique, c'est l'évaluation de traitements dans les conditions habituelles d'administration, aux patients concernés, avec une méthode d'appréciation consensuelle, par une comparaison rigoureuse : c'est une rationalisation de la conduite médicale. N'est-ce pas le mieux qu'on puisse espérer ?

En dépit des difficultés, des essais ont été conduits par des épidémiologistes motivés, dans plusieurs domaines : en cancérologie, en diabétologie, et surtout en chirurgie. Mais c'est tout récemment qu'on a pu assister à une véritable renaissance, dont la cause a été « l'effet SIDA ». Les malades, les médecins, la communauté scientifique se sont insurgés contre la lenteur imposée à la consécration de nouvelles molécules. Ne faut-il pas, entre la découverte d'une molécule et son autorisation de mise sur le marché, quelque dix ans ? Comment, dans le cas du SIDA, accepter pareil délai ? Dans ce délai, la « machinerie » de l'essai thérapeutique occupe une place importante ; on a exigé son allégement. Diverses méthodes ont été proposées, les unes inacceptables, les autres raisonnables. C'est parmi celles-ci que se situe l'utilisation de l'essai pragmatique, découvert soudain comme une panacée, en raison de la position adoptée pour les quatre mots clefs.

Les *patients*, d'abord. L'exemple de l'essai sur l'aspirine montrant la nécessité, après des essais sur groupe choisi, de nouveaux essais pour l'extrapolation à la population générale, fait aisément comprendre qu'on a préféré passer tout de suite au second stade. Cette exigence a été formulée, mais elle a ses inconvénients, et n'a guère été suivie.

Les *critères*, ensuite. La plupart des molécules étant toxiques, on a demandé que leur évaluation tienne compte d'un bilan Coût-Avantage.

Le *mode de comparaison*. Il n'a pas paru tolérable que des essais éventuellement très longs, portant sur des milliers de malades, risquent d'aboutir à la non-conclusion à la suite d'un test.

Enfin et surtout, les *traitements* posent un réel problème. En raison de leur toxicité, un grand pourcentage de malades ne les supportent pas et doivent les abandonner plus ou moins vite, ce pourcentage pouvant aller jusqu'à 30 % et davantage. En général, dans la comparaison des deux traitements A et B, ceux qui ne supportent pas A l'abandonnent au profit de B, et réciproquement. Devant cette difficulté, on a proposé deux modes d'analyse :

- *L'analyse en intention de traiter*. Les malades initialement traités par A ou par B, sont considérés comme traités par A et par B respectivement, même s'ils ont été traités différemment par la suite. C'est certes un scrupule de rigueur, mais s'agissant de comparer les effets de A et de B, le test manque de puissance, et dans la majorité des cas on aboutit à une différence non significative.
- *L'analyse* dite *standard*. On compare les sujets qui auront reçu A ou B jusqu'à la fin de l'essai. Ces deux groupes sont évidemment non comparables, et les conclusions sont affectées d'un biais important.

C'est ici que l'attitude pragmatique apporte un avantage fondamental. Ce qui arrivera, en dehors de tout essai, dans les conditions de la pratique courante, c'est que les sujets intolérants à A recevront B, et réciproquement. On décide donc qu'on va comparer, non A et B, mais les stratégies « A, et si intolérance B » à « B, et si intolérance A ». Les changements de traitement ne sont pas des incidents gênants, ce sont des réalités, ils font partie des stratégies à comparer. Cette attitude peut paraître un tour de passe-passe ou un jeu de mots. Il n'en est rien pour plusieurs raisons.

D'abord, les changements de traitements étant considérés comme gênants dans l'attitude explicative, on essaie de les éviter au maximum, alors que dans l'attitude pragmatique, on les effectue comme on ferait dans la pratique, ce qui est bien préférable.

Ensuite, ces changements, dans l'attitude pragmatique, parce qu'ils font partie des stratégies à évaluer, sont clairement précisés dans le protocole, ce qui n'est pas toujours le cas dans la démarche explicative.

Enfin et surtout, la perte énorme de puissance qui existe dans l'analyse *en intention de traiter* dans l'attitude explicative, disparaît totalement en attitude pragmatique puisque, opérant sans test, on a une puissance de 100 %.

Pour toutes ces raisons, l'essai pragmatique est maintenant, dans le SIDA, prévu, discuté, et souvent adopté, sinon pour tous les mots clefs du protocole, au moins pour le remplacement des traitements par des stratégies thérapeutiques prévoyant clairement les passages d'un traitement à d'autres.

Remis en selle par le SIDA, l'essai pragmatique a vu le jour dans d'autres domaines : dans l'asthme, dans l'évaluation d'antibiotiques, dans des comparaisons médico-économiques de divers traitements.

Parallèlement, le principe même de l'essai pragmatique a été cité dans la littérature et discuté. En 1990, un auteur publiait dans le *Lancet*, un des journaux médicaux les plus illustres, une note où il se déclarait stupéfait : « Comment ! – écrivait-il – les statisticiens s'efforcent de minimiser le risque de première espèce ! Mais, si deux

traitements sont équivalents, choisir l'un d'eux plutôt que l'autre est sans importance ! » Il découvrait alors que le seul but était de minimiser l'erreur de signe, qu'il appelait risque de type 3, et ajoutait que l'essai demandait alors un effectif beaucoup plus faible. Ainsi découvrait-il, après plus de vingt ans, l'essai pragmatique ! Quant à la réduction du nombre de sujets nécessaires, il est certain qu'à vouloir minimiser un seul risque au lieu de deux, on aboutit, pour des mêmes valeurs (5 % par exemple), à une réduction massive. Le problème est peut-être qu'une erreur de signe exige un risque plus faible qu'on ne fait pour α et β. Mais il est clair que, même pour une valeur exigée faible pour γ, il y a un gain notable. Cet avantage n'avait pas échappé à ceux qui ont prôné l'essai pragmatique pour le SIDA.

D'autres articles ont discuté la méthodologie de l'essai pragmatique. Il est devenu peu à peu universellement connu (souvent de manière incorrecte). Des livres récents lui accordent une place privilégiée. Témoins les deux phrases sans équivoque citées dans l'encadré. Enfin récemment, il a été impliqué officiellement en France dans l'évaluation des médicaments. La commission qui consacre la naissance d'un médicament – la Commission d'Autorisation de Mise sur le Marché – s'intéresse en priorité à l'efficacité et par conséquent aux essais explicatifs. Mais le médicament doit ensuite être jugé, pour les problèmes de remboursement en particulier, par la Commission de Transparence. Celle-ci doit quantifier *« le progrès thérapeutique apporté par le nouveau médicament, dans un bilan Coût-Efficacité, en situation, c'est-à-dire sur les patients auxquels il est destiné et dans les conditions naturelles d'emploi et d'environnement. »* C'est exactement la définition de la démarche pragmatique – et la commission recommande la conduite d'essais pragmatiques.

G. D. MURRAY

Br. J. SURG. 1991 ; vol. 78 ; July, 777-81.

STATISTICAL ASPECTS
OF RESEARCH METHODOLOGY

The only crucial decision that needs to be made in the writing of a study protocol is the choice between an explanatory and a pragmatic approach.

BJORN ANDERSEN

Blackwell Scientific Publications. 1990 ; 70-71.

METHODOLOGICAL ERRORS
IN MEDICAL RESEARCH

The fundamental issue is a distinction between pragmatic (managerial) and causal (explanatory) trials.

2. L'épidémiologie

Un secteur essentiel de l'épidémiologie est l'enquête étiologique. Une enquête étiologique vise à *analyser* le rôle de *facteurs* d'une *maladie* chez des *patients*. Ces quatre mots clefs doivent être définis d'une manière tout à fait différente selon le but poursuivi.

Un premier type d'enquête est destiné à tester une hypothèse biologique, notamment le rôle causal de l'exposition à un facteur, disons à titre d'exemple le rôle du tabac dans le cancer bronchique. S'agissant, dans ce vieil et toujours illustre exemple, de mettre en évidence une relation causale, les quatre mots clefs seront logiquement définis comme suit :

a) *Maladie :* une entité simple, bien définie et identifiable (un diagnostic histologique est souhaitable, éventuellement une autopsie).

b) *Facteurs de risque :* celui incriminé (le tabac) et divers tiers facteurs envisagés uniquement pour leur contrôle dans l'analyse.

c) *Population :* un groupe choisi pour une raison d'opportunité (on a choisi par exemple le corps médical anglais), le souci de représentativité étant secondaire en raison de la stabilité des conclusions.

d) *Analyse des résultats :* test statistique comparant les exposés et non exposés pour le facteur incriminé à égalité de tiers facteurs.

Toute différente est la situation où on veut prendre une décision pour une pathologie donnée, sans nécessairement faire avancer la connaissance sur le rôle des facteurs de risque. Un bon exemple est, en périnatologie, celui des grossesses à haut risque.

De très nombreux facteurs permettent de prédire une mauvaise issue de la grossesse : on peut, à partir d'un petit nombre d'entre eux bien choisis, identifier des femmes à haut risque, pour lesquelles sera adoptée une mesure sans rapport avec les facteurs de risque (par exemple surveillance plus fréquente), ce qui évite toute recherche causale ; on peut proposer le protocole suivant :

a) *Maladie :* une entité éventuellement polymorphe (« mauvaise issue » de la grossesse comprenant mortalité périnatale, prématurité et hypotrophie).

b) *Facteurs de risque :* de nombreux facteurs auxquels on suppose un rôle prédictif.

c) *Population :* un groupe représentatif, l'extrapolabilité de la décision étant sujette à réserves.

d) *Analyse des résultats :* de type décisionnel, comportant par exemple le calcul du risque à partir des facteurs prédictifs, et la détermination du seuil entraînant la décision.

Bien que les deux types d'enquête décrits ci-dessus portent sur l'étude de facteurs de risque, on voit clairement qu'il s'agit de deux types de problèmes entièrement différents. Le premier exemple est une quête de connaissance et peut, comme dans

les essais thérapeutiques, être qualifié d'*attitude explicative ;* le second vise à une prise de décision, sans quête de connaissance, illustrant ce constat heureux et souvent dénié qu'on peut agir sans comprendre. Il est tout à fait l'analogue de l'attitude qualifiée, à propos des essais, de *pragmatique.*

Le choix optimum des quatre mots clefs du protocole dans les deux attitudes est récapitulé au tableau 6.

Tableau 6
Épidémiologie (enquête étiologique)

Maladie	Simple, bien définie	Éventuellement complexe, mal définie
Facteurs	Celui qui est incriminé et tiers facteurs	Nombreux facteurs prédictifs.
Échantillons	Groupe choisi	Groupe représentatif
Analyse	Test à égalité des tiers facteurs	Calcul du risque, seuil de décision
	ATTITUDE EXPLICATIVE	ATTITUDE PRAGMATIQUE

À l'inverse de ce qui se passe pour les essais thérapeutiques, l'épidémiologie pragmatique est beaucoup plus pratiquée que l'épidémiologie explicative, ce qui se comprend facilement : la recherche causale est peu « gratifiante », car la preuve de causalité étant de principe exclue dans une enquête d'observation, on ne peut espérer qu'une forte présomption au prix d'une accumulation d'études variées. L'épidémiologie pragmatique, en évitant cette recherche, est plus aisée – du moins en apparence. On peut s'étonner que les épidémiologistes, soucieux d'accumuler les définitions de leur discipline, aient laissé de côté cette distinction entre deux types si différents d'épidémiologie, distinction qu'ils opèrent, mais sans le savoir.

L'absence de cette explicitation est regrettable, car elle entraîne souvent des enquêtes mal formulées et, en outre, un manque de recherches appropriées sur le plan méthodologique.

Un autre secteur, voisin des enquêtes étiologiques, où apparaît nettement la distinction entre les deux attitudes, est la considération du « normal » dans le domaine médical. On connaît la distinction entre les deux définitions du normal : la définition « fréquentielle » où le normal est le voisinage de la valeur la plus fréquente, et la définition « opérationnelle », normative, où une valeur sort de la normale dès lors qu'elle justifie une intervention. Les facteurs du « normal » à prendre en considération dans ces deux situations sont, bien sûr, différents. Citons, à titre d'exemple, celui des

bébés dits *hypotrophiques*, c'est-à-dire de poids trop faible pour une durée de gestation donnée. Parmi les nombreux facteurs du poids du bébé figurent notamment le sexe, le tabagisme de la mère, la parité. Si l'on veut « expliquer » le poids de naissance, tous ces facteurs sont intéressants. Mais si, dans une attitude pragmatique, on veut établir un seuil entraînant une décision, le sexe de l'enfant sera sûrement à prendre en considération, le tabagisme de la mère ne le sera sûrement pas ; la parité le sera-t-elle ? La réponse n'est pas évidente, mais ce qui l'est, c'est que le problème ne pourra être résolu correctement que si on a clairement opté pour l'attitude explicative ou l'attitude pragmatique.

Ainsi, en épidémiologie, d'une manière générale, comme dans le domaine des essais, un exposé précis de la distinction entre les deux attitudes, en tant qu'outil de réflexion, devrait être un objectif pédagogique de premier ordre.

3. Généralisation

La distinction entre les deux attitudes est un outil de réflexion dépassant largement le cadre des essais thérapeutiques et des enquêtes épidémiologiques. Il s'étend d'abord à de très nombreux problèmes d'expérimentation ou d'observation en dehors de la médecine.

La plupart des problèmes *d'expérimentation* peuvent être envisagés, comme l'essai thérapeutique, sous l'angle explicatif ou pragmatique, qu'il s'agisse ou non de traitements. Soit par exemple, en agronomie, à comparer deux races A et B de haricots : on peut imaginer une comparaison à égalité de compacité, d'engrais, etc. Cette manière explicative d'aborder le problème serait une erreur, car il est essentiellement pragmatique : il ne s'agit pas de tester l'hypothèse, tout à fait invraisemblable, d'égalité des deux races, mais bien de choisir la meilleure. Il faut donc donner à chacune ses chances optimales de croissance et comparer les résultats dans un bilan Coût-Avantages.

De même, dans les problèmes *d'observation* visant à étudier l'apparition d'un phénomène, on peut envisager, soit de mettre en évidence des facteurs explicatifs, de type causal, soit de prédire, sans nécessairement comprendre, pour agir.

Quant à l'exemple décrit plus haut de la définition du normal, il dépasse bien entendu le cadre de la médecine, et s'étend à tous les domaines où il est question de normes ; ainsi, dans le contrôle de fabrication, il est intéressant de connaître les normes fréquentielles, mais s'il s'agit de déterminer les lots à rejeter, c'est l'attitude opérationnelle qui s'impose.

Quête de connaissance et prise de décision, ces opérations ne sont-elles par les deux pôles autour desquels s'oriente notre vie ? Et ceci aux échelons du conscient et de l'inconscient, du macroscopique comme du microscopique : quête d'information et action, les molécules du vivant font-elles autre chose ? Sans doute ce constat est-il un lieu commun, mais ce qui l'est moins, c'est que ces deux opérations, en revanche, nécessitent des protocoles radicalement différents, donc un choix clair de la formulation au départ.

BIBLIOGRAPHIE

SCHWARTZ D. *Connaissance et décision* ou *Attitude explicative et attitude pragmatique*, Encyclopédie médico-chirurgicale, 16900 A 10, 1-1986, 4 p.

SCHWARTZ D., LELLOUCH J. "Explanatory and Pragmatic Attitudes in Therapeutic Trials", *J. Chron Dis.* 1967, 20, 637-48.

SCHWARTZ D., FLAMANT R., LELLOUCH J. *L'essai thérapeutique chez l'homme*, Paris, Flammarion Médecine Sciences, 1970.