

J. P. BENZÉCRI

**Efficacité thérapeutique et efficacité  
méthodologique ou : la révolte d'Esculape. II :  
la doctrine des essais thérapeutiques**

*Les cahiers de l'analyse des données*, tome 11, n° 3 (1986),  
p. 259-280

[http://www.numdam.org/item?id=CAD\\_1986\\_\\_11\\_3\\_259\\_0](http://www.numdam.org/item?id=CAD_1986__11_3_259_0)

© Les cahiers de l'analyse des données, Dunod, 1986, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/>

## EFFICACITÉ THÉRAPEUTIQUE ET EFFICACITÉ MÉTHODOLOGIQUE OU : LA RÉVOLTE D'ESCULAPE. II : LA DOCTRINE DES ESSAIS THÉRAPEUTIQUES

### [ESCULAPE II]

par J.P. Benzécri

#### 1 L'expérimentation contrôlée

Tout savoir humain, pratique ou spéculatif s'acquiert par l'observation et le raisonnement. Or le raisonnement n'est certain que dans la mesure où il se fonde sur la nature des choses, qui nous est mal connue; et l'observation est ambiguë en ce que d'un fait réel infiniment complexe nous ne retenons que quelques éléments auxquels nous attribuons la dignité de causes ou d'effets. Il faut allier la patience dans l'expérimentation à l'imagination dans la critique, pour séparer le contingent de l'essentiel et découvrir des principes généraux au travers des faits particuliers.

Comment le médecin qui sait que tous les malades ne sont pas destinés à guérir, attribuera-t-il à un traitement qui ne peut être infaillible, une ou plusieurs guérisons, plutôt que les échecs qui s'y mêlent? Comment dans le traitement, fera-t-il la part des éléments utiles, et de ceux qu'il faudrait supprimer? La science expérimentale contemporaine s'en remet à la statistique du soin de tirer de cas assez nombreux des conclusions d'une validité mesurée.

Selon CL. Chastang & coll \*\*: (*in* Medical Computing ; Ed.M Laudet, J. Anderson, F. Bégon ; Taylor & Francis ; 1977 ; pp. 361-373) :

" La méthodologie de l'essai thérapeutique, développée depuis 1945, est bien codifiée : la Phase I vise à apprécier la toxicité du nouvel agent thérapeutique ; la Phase II permet d'évaluer l'efficacité et, en fonction de ces résultats on entreprend ou non un essai comparatif randomisé (Phase III).

(pour ne rien dire de la Phase IV de pharmacovigilance, ou contrôle des accidents survenant après la mise du médicament sur le marché).

Même si, comme le soulignent à juste titre les auteurs cités (cf. *infra* § 4.1) un essai non comparatif (de Phase II) suffit dans certains cas à établir la supériorité d'une thérapeutique nouvelle, c'est l'essai comparatif qui sera le thème central du présent exposé.

---

(\*) *Professeur de statistique. Université Pierre et Marie Curie.*

(\*\*) *De ce travail, souvent cité dans la suite, nous donnons ici la liste complète des auteurs dont plusieurs nous sont connus personnellement ; et certains ont une notoriété qui les signale à l'attention du lecteur. Ce sont Cl. Chastang, J.P. Nakache, J.P. Chantalou, F. Grémy, M. Weil, M.F. Gémon-Auclerc, G. Auclerc, et Cl. Jacquillat.*

"Il y a maintenant 50 ans [écrit T.J. Steiner : *in* JIM Vol 9 ; n° 52 ; Sept 1984 ; pp. 291 sqq], que Mapother \* a énoncé les conditions requises pour l'organisation d'un essai, en notant qu'il existe parmi ceux qui affirment avoir été guéris d'une maladie par un nouveau traitement :

- " ceux qui n'ont jamais eu la maladie ;
- ceux qui l'ont encore ;
- ceux qui n'ont jamais pris le traitement ;
- ceux qui auraient tout aussi bien guéri sans lui".

Les notions de sélection, d'évaluation, d'observance et de groupe témoin ont été ainsi définies d'une manière particulièrement concise".

Ayant dans un précédent article ([*ESCALAPE I*], cité ici I), présenté avec quelques détails le protocole et les résultats de plusieurs essais en prévention cardiovasculaire, nous voulons ici expliquer pour elles-mêmes ces notions de base, tout en frayant les voies d'une méthodologie nouvelle fondée sur l'analyse des données multidimensionnelles. Plutôt que de suivre l'essai dans ses étapes successives, notre exposé ira du simple au complexe mêlant des suggestions aux témoignages de nombreux auteurs (glanés dans les revues médicales déjà citées dans I); et répétant souvent les thèses essentielles.

Si la notion d'*observance* ne se rencontrera dans la suite qu'occasionnellement, *groupe témoin*, *sélection* et *évaluation* retiendront notre attention. L'effet placebo objet du § 2, conditionne l'administration du traitement aux sujets effectivement traités et aux témoins. La sélection des sujets et leur répartition en groupes fait l'objet du § 3. Il faudra nous arrêter au choix des traitements, encadré par les règles de la science et de la morale (§ 4). Enfin l'analyse multidimensionnelle que nous prônons, est au fond une évaluation conjointe de l'état basal des sujets et de leur réponse au traitement (§ 5).

## 2 Placebo et double insu :

2.1 Réalité de l'effet placebo : Le seul fait de suivre un traitement, d'être l'objet de soins attentifs peut, chez certains patients, orienter le cours de la maladie vers la guérison naturelle ou au moins hâter celle-ci ; voire, si la guérison de l'individu par ses seules ressources est impossible, procurer momentanément une amélioration, un soulagement. C'est ce qu'on appelle l'effet *placebo* ; usant d'un mot dont le sens est "je plairai", comme si l'agrément avait valeur de traitement.

Même si l'on n'a jamais observé de septicémie ou de cancer regressant sous l'effet placebo, on ne saurait minimiser l'amplitude de cet effet non seulement sur la douleur subjective mais aussi face à tout déséquilibre entretenu par des messages désordonnés, nerveux ou hormonaux.

Dans ces conditions pour apprécier à sa juste valeur l'effet des composantes actives d'un traitement (des médicaments surtout), il faut prendre pour terme de comparaison l'effet d'un traitement de même format dépourvu de composantes actives (dragées non médicamenteuses mais présentées comme un médicament...) ; sans oublier l'absence pure et simple de traitement.

Sans prétendre confirmer par l'autorité de références prestigieuses la réalité de l'effet placebo, nous en relevons deux exemples qui attestent l'attention que les médecins portent à cet effet.

(\*) T.J. Steiner cite de E. Mapother : *The Bradshaw Lecture ; 1936 ; non publiée.*

Comment traiter médicalement une lombalgie ? Sous ce titre, Ch. Brégeon publie dans l'édition française du JAMA (JAMA fr, Vol 8, n° 75, Nbre 1983 ; pp. 362 sqq) les pertinentes réflexions que lui inspirent des travaux américains. Nous citerons plusieurs fois Ch. Brégeon... Voici en quels termes celui-ci analyse, avec finesse, l'effet placebo.

"... s'il est aussi difficile d'opérer une discrimination objective entre les nombreuses approches thérapeutiques de la lombalgie, ne serait-ce pas en raison d'un facteur contaminant commun qui ne peut guère être évacué : l'effet placebo du médecin lui-même : R.A. Deyo souligne que le temps et l'attention accordés par le médecin et le personnel soignant ont à eux seuls des effets bénéfiques importants. Dans une affection dont l'évolution spontanée est favorable, la qualité de l'écoute, l'autorité dans la prescription, la confiance en soi dans l'exécution de certains traitements physiques, l'information donnée au malade sur la nature de ses troubles et les conseils pratiques d'hygiène de vie formulés et expliqués à partir de la physiologie rachidienne, l'art de rassurer ... n'auraient-ils pas souvent le même impact thérapeutique sur la perception de la douleur lombaire que les antalgiques ou les autres thérapeutiques médicales".

Que l'effet placebo ne concerne pas seulement la perception subjective d'une douleur, mais s'étende aux performances objectives, apparaît dans une deuxième note.

" Claudication intermittente : la contre révolution" est le titre d'une colonne de la Revue du Praticien, signée "L. C.", (RP ; T XXXVI ; n° 20 ; 9 avril 1986 ; p. 1128) et rendant compte en ces termes des déboirs d'un essai ...

"L'an passé, sous le titre "Claudication intermittente : une révolution en marche ?" je vous faisais part des résultats d'une étude sur la kétansérine ... Une équipe anversoise avait observé une amélioration spectaculaire de la distance de marche et des indices pléthysmographiques de la vascularisation des membres inférieurs chez les malades traités par ce médicament, comparativement à ceux qui avaient reçu un placebo ... Un groupe anglo-belge a refait un essai identique à celui des Anversois. Cette fois il n'est apparu aucun effet favorable de la kétansérine ... A côté de l'inaction de la kétansérine cette étude montre quelque chose de fascinant et qu'on ne se rappelle jamais assez : la puissante action possible d'un placebo. Sous placebo, la distance de marche mesurée sur un tapis roulant augmenta significativement d'une moyenne de 158 m à une moyenne de 205 m ...

2.2 Utilisation thérapeutique de l'effet placebo : Avec sa relative puissance et l'étendue de son domaine attestées par les courbes de résultats de nombreux essais comparatifs, le placebo peut apparaître comme une thérapeutique universelle, dépassée seulement par l'aspirine. Aussi a-t-on songé à mettre en vente des placebos sous emballage ; à quoi des persifleurs rétorquent que les laboratoires pharmaceutiques n'ont pas attendu pour cela d'autorisation de mise sur le marché ; le succès de ce que l'on appelle "médécines douces", confirmant la faveur du public pour de tels traitements.

S'interrogeant sur "l'influence de la personnalité dans la relation médecin-malade", O. Kerneis (RP ; T XXXV, n°s 59-60 ; 24 Décembre 1985, pp. 3429 sqq) propose cette statistique approximative.

Un courant de la psychologie médicale, constitué depuis une trentaine d'années, s'est organisé à partir d'une constatation paradoxale : malgré un perfectionnement dans le dépistage des lésions, 50 à 75% des consultants de médecine générale n'offrent aucune maladie lésionnelle objectivable.

Sans faire fi de la proposition d'O. Kerneis, de répondre à cette attente diffuse des patients par un approfondissement et une systématisation de la prise en compte de la personnalité du malade par son médecin, il faut dire que l'effet placebo, sinon le placebo au sens strict est la réponse ordinaire à la maladie non objectivable.

Aussi ne lisons-nous pas sans surprise cette réflexion à propos des essais thérapeutiques en médecine générale de ville (J. Ankri ; RP T XXXVI n° 7 ; 1-er Févr. 1986 ; pp. 378-379):

"L'utilisation du placebo dans certaines études comparatives pourrait être considérée comme peu possible en médecine générale ; l'expérience nous a prouvé le contraire, dans la mesure où cette utilisation a été reconnue par tous les participants comme éthique".

Quelque réserve que nous fassions devant la valeur morale que l'accord universel des participants est censé conférer à une pratique, il est clair qu'une possession invétérée a fait au placebo sa large place en médecine générale !

Nous ne songeons pas à l'en déloger par la violence ; suggérant au contraire d'introduire dans les essais thérapeutiques, une transition vers zéro qui s'accorde avec la pratique implicitement acceptée (§ 2.4).

2.3 Effet placebo et essai en double insu : L'effet placebo montre que dans l'action d'un traitement, l'idée que s'en fait le patient peut jouer un grand rôle. C'est pourquoi il est d'usage dans les essais comparatifs de deux traitements de ne pas faire connaître à un patient lequel de ceux-ci lui est attribué ; pratique particulièrement de règle si, comme il est fréquent, l'un des deux traitements est un pur placebo ; mais un traitement inefficace ou peu efficace est, plus généralement un placebo de fait ... Comme d'autre part, ainsi que l'a souligné Ch. Brégeon (cf. *supra* § 2.1), le médecin lui-même agit comme un placebo, il convient que celui-ci soit dans les mêmes dispositions, quel que soit celui des deux traitements qu'il administre ; ce qui ne se réalisera que dans la mesure où le médecin sera comme un patient tenu dans l'ignorance du contenu réel des drogues administrées ... C'est ce qu'on appelle "le double insu" ou parfois "le double aveugle".

Même, sans parler de la tâche délicate de distribution des drogues dévolue à l'organe central de l'essai, la pratique du double insu présente de multiples inconvénients.

D'abord l'ignorance du traitement a sur le malade et sur le médecin un effet placebo négatif : comment croire, en effet, à ce dont on ne connaît pas le nom ? Ainsi B.M. Rifkind, parlant au nom des organisateurs de l'essai LRC-CPPT (cf. I § 2.4), prend lui-même argument de "l'inaptitude dans les essais à longs termes randomisés et particulièrement en aveugle, à maintenir une réponse [physiologique à de hauts] niveaux". Et T.J. Steiner dans son bel article consacré aux essais cliniques pour les accidents vasculaires cérébraux (JIM Vol 7, n° 52, Sept 1984 ; pp. 291-296), après avoir noté que les malades, au sortir de l'hôpital, observent généralement bien le traitement auquel ils attribuent l'amélioration de leur état, constate que

"... si l'éventualité d'un traitement placebo a été clairement exposée au malade, ce bénéfice disparaît habituellement totalement et l'observance a long terme risque de s'en ressentir et de devenir un problème majeur".

Ensuite le médecin ne peut ignorer longtemps qu'il administre à son patient un traitement inactif, si l'alternative au placebo a un effet physiologique constatable (par exemple sur la pression artérielle),

"- ou on se tient à un double aveugle strict et la recherche de la posologie idéale et du médicament idéal nécessitent des acrobaties méthodologiques que nous ne savons pas faire ;

- ou on adopte un protocole à produit et posologie imposés et on s'écarte des conditions habituelles de la prescription anti-arythmique;

- ou encore on réalise des protocoles ouverts et personne ne peut en tirer des conclusions valides et sûres".

réflexions judicieuses, qui débordent le problème du double insu, pour aborder celui du choix des protocoles, et sur lesquelles nous reviendrons (§ 4.3).

T.J. Steiner (*op; laud*) considère qu'est parfois inévitable "le recours à un observateur séparé non aveugle ayant pour rôle de contrôler le protocole thérapeutique" ... de recenser les effets secondaires..

Enfin en dehors des thérapeutiques médicamenteuses, il n'est presque jamais possible de pratiquer le double insu ! (simuler une opération chirurgicale ...!).

4.4 Placebo et transition vers zéro : Supposons qu'un essai fournisse des résultats tels que :

Placebo : efficacité 40% ; Drogue : efficacité 45% ;

avec éventuellement : Placebo  $\geq$  Drogue, dans un large secteur d'indications que délimitera l'analyse multidimensionnelle (cf. § 5.3). La question est alors de savoir quelle conclusion tirer d'un tel résultat pour la pratique médicale.

Il importe d'abord de noter que le résultat lui-même n'a de sens que si l'on dispose d'un troisième chiffre relatif à l'absence pure et simple de traitement, sans effet placebo ; par exemple :

Absence : eff. 10% ; Placebo : eff. 40% ; Drogue : eff. 45%

Ceci étant, comme les conditions de prescription d'un placebo dans les essais, ne peuvent être étendues à la pratique générale (même hospitalière), le résultat, pour frappant qu'il soit n'a pas d'application thérapeutique.

Une autre voie s'offre, qui a de plus, à nos yeux, le mérite de faire rentrer l'étude de l'effet placebo dans le cadre des essais thérapeutiques en libre choix sans cécité simulée, tels que nous les proposons (cf. *infra* §§ 3.4 et 4.3). On proposerait comme alternative à la dose usuelle du produit actif la dose (1/2) ou (1/4). Le discours du médecin étant : "On s'interroge sur l'opportunité de conserver le dosage actuel, un dosage moindre pouvant être aussi efficace et plus léger quant aux effets latéraux éventuels : quant à vous, désirez vous prendre la dose 1 ou la dose (1/2) ?"

La conclusion d'un tel essai pourrait être que (1/2)  $\approx$  1 ; (voire seulement 1  $\approx$  (1/2) dans un secteur d'utilisation) ; résultat immédiatement applicable à la pratique médicale ; le produit étant désormais conditionné à la dose (1/2).. (Quitte à poursuivre les essais de réduction ..., jusqu'à la commercialisation de quasi placebo ... (cf *supra* § 2).

La recherche de la dose active minima est d'ailleurs pour la pharmacologie clinique un thème de recherche important. Dans un article consacré aux conséquences à long terme de la suppression de la sécrétion d'acide gastrique (PGM ; special report Cimetidine 85 ; Nbre 1985;

pp. 65-75), Tadataka Yamada écrit :

"Du fait des risques potentiels liés à une suppression chronique excessive de la suppression d'acide gastriques, la voie de la prudence pour les cliniciens semble être de choisir le régime thérapeutique le moins puissant qui soit encore efficace" [contre les ulcères].

### 3 Recrutement des patients et randomisation

3.1 Recrutement et diagnostic : Commentant un article de synthèse américain dont il se dégage l'impression que "les études thérapeutiques sur la lombalgie sont mal conduites et mal exploitées", Ch. Brégeon s'interroge (JAMA fr Vol 8, n° 75, Nbre 1973, pp. 362 sqq) :

"Est-il légitime, tant sur le plan nosologique que thérapeutique, de regrouper toutes les lombalgies sous une même rubrique ? Le symptôme "lombalgie" recouvre en fait un ensemble disparate de situations, et on peut ici faire une comparaison avec l'entité "anémie". Il y a longtemps que l'on ne considère plus de façon globale "les anémies", mais que l'on s'efforce de préciser leur étiologie et leur contexte avant de mettre en oeuvre une thérapeutique particulière à chaque type : anémie d'une hypothyroïdie, d'un saignement chronique, d'une inflammation sévère, anémie biernérienne etc. Il ne viendrait également à l'idée de personne, de nos jours, d'inclure systématiquement la prescription de fer ou de vitamines B12 dans un traitement en aveugle des diverses catégories d'anémies. De la même manière, il conviendrait de rechercher avec beaucoup plus d'attention les conditions étiologiques particulières à chaque lombalgie avant d'instituer un traitement, fût-il randomisé".

Et après avoir campé, non sans humour, quelques silhouettes de lombalgiques, Ch. Brégeon poursuit :

"Les progrès thérapeutiques sont certainement liés à un affinement du diagnostic étiologique, comme le souligne R.A. Deyo [l'auteur de l'article commenté] en insistant sur l'importance d'une bonne description de symptomatologie et de l'origine des groupes de patients, alors qu'une application standardisée des mêmes méthodes thérapeutiques à des lombalgies tout-venant, sans tri préalable, revient à introduire de nombreux biais méthodologiques, ou à multiplier les thérapeutiques placebo".

Voilà posé le maître problème des essais thérapeutiques : efficacité absolue d'un traitement, ou efficacité comparée de deux traitements sont des notions qui n'ont de sens que relativement à une pathologie déterminée ; et même relativement à un degré de gravité fixée. Affirmons d'emblée qu'aucun artifice de statistique (cf. § 3.3) ne permet d'éluider la question du diagnostic ; au contraire, selon nous, des essais ouverts, prenant en compte explicitement la diversité des individus (§ 3.4), sans borner les comparaisons à des moyennes de groupes généralement constitués *a priori* s'accrochent de critères d'entrée imparfaits ; et l'analyse multidimensionnelle permet au cours de l'essai, de progresser dans la compréhension de l'état basal même des sujets, et de calculer sur des groupes constitués *a posteriori* efficacités absolues ou relatives.

Même dans le domaine de la prévention primaire du risque cardiovasculaire, où les critères de sélection semblent faciles à fixer sans qu'il soit besoin d'aucun diagnostic, on a vu que des conclusions claires et utiles ne ressortent que pour autant qu'on a défini adéquatement des groupes de sujets courant des risques équivalents (cf. I §§ 3.1.2 & 3.2.3).

3.2 Recrutement rétrospectif de témoins : "Prospectif" est un des maîtres mots de la doctrine codifiée des essais. Cependant, il est naturel au médecin de juger des cas nouveaux d'après les cas anciens ; et on souhaiterait que de semblables comparaisons soient faites systématiquement sur une grande échelle avec le secours de l'analyse multidimensionnelle. Surout quand il apparaît immoral d'attribuer à des patients dont la vie est en danger, un placebo ou un traitement inefficace, à seule fin de confirmer la supériorité d'un traitement nouveau dont l'efficacité n'est guère douteuse (cf. *infra* § 4.1) ; Chastang et coll. (*in* Medical Computing ; éd Laudet et coll ; Taylor & Francis ; 1977 ; p. 364) osent même affirmer que dans certaines conditions "la formation d'un groupe témoin est non seulement éthiquement condamnable, mais encore retarde la recherche thérapeutique par une mauvaise utilisation (sic) des malades disponibles".

Il convient donc de considérer les objections faites aux études rétrospectives, afin d'assigner à celles-ci des limites raisonnables.

Un article de H.E. Lessner et coll consacré aux Thérapeutiques adjuvantes du cancer du colon (New Engl. J. of Med ; traduit dans JIM n° 70, suppl gastro. ; 30 Mars 1986 ; p. 27) offre un exemple qui mérite d'être cité dans ses détails

"L'inefficacité de la chimiothérapie et de l'immunothérapie utilisées dans notre étude a été démontrée parce qu'il existait un groupe témoin non traité, sélectionné de manière prospective. La survie après chirurgie, dans ce travail multicentrique a été nettement supérieure à celle que la littérature permettait de prévoir. Si une population historique avait été choisie pour servir de groupe "témoin" aux malades traités par chimiothérapie ou par immunothérapie, ou par l'association des deux, on en aurait tiré la conclusion erronée que l'association fluorouracile/rémustine et le MÉR [fraction du BCG extraite par le méthanol] ont bien une efficacité antitumorale adjuvante dans le cancer colique. Une telle conclusion, basée sur une expérience de 572 cas aurait très vraisemblablement modifié les habitudes thérapeutiques et encouragé l'utilisation sur une grande échelle d'un traitement inactif. Ceci aurait été particulièrement fâcheux au vu des 6 cas de leucémie survenus après chimiothérapie. L'inadéquation des groupes témoins historiques et la nécessité absolue d'une population prospective sans traitement dans un essai clinique évaluant les thérapeutiques adjuvantes du cancer colique semblent d'après cette étude évidentes".

Dans un panorama historique de la chimiothérapie du cancer (PGM, Vol 77 n° 6, May 1, 1985 ; pp. 165-174) Carl G. Cardinal se prononce dans le même sens que Lessner et coll.

"Au début des années 70 ... se répandit l'usage d'injection du BCG chez les patients atteints de mélanomes aux stades I et II. Ces essais contrôlés sur des témoins historiques semblèrent démontrer un bénéfice ... Mais quand l'immunothérapie non spécifique fut comparée au placebo en essai randomisé, l'immunothérapie fut démontrée être sans valeur, et les concepts d'immunothérapie non spécifique et de validité des contrôles rétrospectifs disparurent pour ne jamais revenir-espérons-le!"

A quoi attribuer ces valeurs discordantes de la probabilité de guérison ( d'amélioration, ou de survie à un terme donné ...) calculées pour des malades traités de la même façon mais à plusieurs années de distance ? On peut penser que même sans aucune modification spectaculaire de la thérapeutique, le temps apporte par degrés insensibles des améliorations importantes. Un autre explication, d'une portée plus générale, (car elle s'applique aux variations de tout sens en mieux ou en pire ; et vaut aussi contre les essais multicentriques ...) est donnée par D. Lane dans une lettre au JAMA (JAMA fr, Vol 10, n° 107 15 Sept 1985 ; p. 996).

"L'article de Oye et Shapiro sur l'évaluation des essais thérapeutiques de chimiothérapie des cancers ne fait que refléter les limites habituelles des analyses statistiques rétrospectives portant sur des études antérieures de réponse et de survie. Le principal défaut en est que les auteurs tentent d'évaluer les données publiées par d'autres, sans avoir fait eux-mêmes les observations ..."

Cependant Terra Ziporyn, rend. compte (JAMA 1983 ; 250 ; pp. 1376-7 ; et JAMA fr ; Vol 8 n) 76 ; Déc 1983 ; pp. 473-479) des efforts de K.B McCredie et coll (de Houston) pour pratiquer "Un traitement personnalisé de la leucémie aiguë. Partant du principe

" ... qu'il n'est pas question d'appliquer un traitement très agressif, pouvant être mortel, chez un patient devant être inclus dans le groupe de survie prolongée ...",

l'équipe de Houston a recherché avec succès par l'étude rétrospective de 325 dossiers "des facteurs prédictifs de la survie à long terme des patients en rémission".

Notre conclusion sera, qu'il convient de normaliser au maximum les observations, afin de pouvoir les exploiter statistiquement ; l'analyse factorielle pouvant donner une vue d'ensemble utile de la diversité des cas, même si sont illusoire les comparaisons entre fréquences calculées sur des groupes étrangers l'un à l'autre.

3.3 Groupes a priori et randomisation : La comparaison de deux ou plusieurs traitements  $T_r$  (dont l'un peut être un placebo) se fait en appliquant chacun de ceux-ci à un ensemble  $I_n$  de patients. Pour que la comparaison en termes de pourcentages de guérison ou d'amélioration (etc.) ait un sens, il faut que tous les patients relèvent d'un même diagnostic (cf. § 3.1) et que les niveaux de gravité de l'atteinte soient semblables pour tous les groupes : on dit alors que les ensembles  $I_n$  constituent des *groupes comparables*. On conçoit que la comparabilité soit rendue moins accessible par la diversité de la pathologie, la difficulté du diagnostic, la multiplicité des facteurs de gravité (ou de risque).

3.3.1 Les principes reçus : Selon la pratique actuelle, la comparabilité peut être assurée soit en appariant des patients, soit en conjuguant stratification et affectation aléatoire (randomisation). Supposons d'abord pour nous borner au plus simple qu'il n'y a que deux traitements  $T_1$  et  $T_2$ .

On dit que les patients sont appariés si à chaque individu  $i_1$  du groupe  $I_1$  recevant le traitement  $T_1$ , il correspond un individu  $i_2$  identique à  $i_1$  non seulement quant au diagnostic général mais quant à la variance étiologique et à toutes les variables de gravité ;  $i_2$  étant affecté au groupe  $I_2$  traité par  $T_2$ . Appliquer cette méthode est impossible à double titre : d'une part les caractères pertinents qui définissent un cas individuel sont incomplètement élucidés (ceci est particulièrement vrai des facteurs génétiques, que les études américaines prennent souvent en compte en distinguant blanc de noir) ; d'autre part les caractères connus sont en général assez nombreux pour qu'il soit impossible de trouver un  $i_2$  identique à  $i_1$  sous ce rapport.

Dans ces conditions, on se borne à utiliser un petit nombre de variables pour caractériser l'état basal des sujets. Chaque combinaison de modalités de ces variables (par exemple, un sexe ; une tranche d'âge ; et une localisation de l'atteinte déterminée, définit une *strate*. Au sein de chaque strate, la répartition égale des individus entre les traitements se fait par tirage au sort (*randomisation*) ; avec l'espoir que l'effet de toutes les variables non prises en comptes (connues

ou inconnues) est ainsi globalement neutralisé ; ce qu'on pourra vérifier *a posteriori* pour une variable particulière (en comparant, e.g. les poids des sujets de I1 et de I2 ...).

L'application conjuguée des règles de stratification et de randomisation, varie grandement selon le domaine de la pathologie et les vues des organisateurs de l'essai : c'est pourquoi nous considérons 3 exemples.

### 3.3.2 Efficacité sur l'ulcère duodénal de la Cimétidine, administrée

en une seule prise le soir : K.E Peace & coll ; PGM ; special report Nbre 1983 ; pp. 27-33. L'objet de l'essai était de comparer l'efficacité des 4 doses : 1600 mg, 800 mg, 400 mg, et 0 mg (placebo). En cette matière, le diagnostic est certain et sans variantes possibles, l'ulcère étant vu à l'endoscopie. Comme facteur de gravité, de risque, ou de diversification, les auteurs ont considéré, la taille de l'ulcère, le fait de fumer ou non, le centre où ont été effectués le traitement et l'observation. Il semble que dans l'esprit des auteurs la variable centre était destinée à prendre en compte globalement la diversité anthropométrique ou sociologique des sujets (âge, sexe, corpulence ...). La stratification a été faite d'après le tabagisme et le centre. *A posteriori* les auteurs ayant constaté peu de différences entre centres, ont regretté de n'avoir pas stratifié suivant la taille de l'ulcère et le tabagisme (e.g. les patients non fumeurs, dont l'ulcère a un diamètre compris entre 0,5 cm et 1 cm, seraient partagés aléatoirement et également entre les 4 traitements).

Quant aux effectifs, on avait décidé d'aller jusqu'à 164 sujets par traitement, en prenant pour référence le cas modèle de la comparaison d'un placebo, pour lequel le taux de guérison est de l'ordre de (1/4) et d'un traitement dont l'efficacité est celle du placebo + 20% (référence non dépourvue d'intérêt ; mais de laquelle il est illusoire de déduire autre chose qu'un ordre de grandeur ...).

### 3.3.3 Essais cliniques pour les accidents vasculaires cérébraux :

T.J. Steiner ; JIM Vol 9, n° 52, Sept 1984 ; pp. 291-296. Selon l'auteur le domaine est d'une grande complexité nosologique, et les facteurs de pronostic multiples et difficiles à pondérer.

"...Classiquement, nous pourrions relever un certain nombre de données dont la valeur prédictive vis-à-vis du pronostic total est connue... [ainsi qu'] un grand nombre d'éléments connus pour affecter les possibilités de récupération ... Pour chacun des critères ..., seul un jugement qualitatif et non quantitatif peut être porté : il est impossible de dissocier les valeurs péjoratives respectives de l'appartenance au sexe masculin ou de l'atteinte de l'hémiplégie mineure. Sont-elles équivalentes ou l'une est-elle plus importante que l'autre ? ... Même lorsque la certitude du diagnostic est acquise, les mécanismes physiopathologiques qui interviennent sont généralement inconnus (ainsi la nature embolique d'un infarctus [du cerveau] est difficile à affirmer tout autant que la nature des emboles. Deux patients, tout à fait comparables sur le plan clinique (et apparemment pronostique) au moment de leur inclusion dans l'essai, peuvent dans les quelques heures qui suivent connaître une évolution diamétralement opposée, l'un récupérant, l'autre s'aggravant rapidement.

Pour l'auteur, la randomisation peut apporter une solution statistique au problème de la constitution des groupes de traitement. Mais, à la vérité, même si les deux groupes sont issus d'une même population parente, celle-ci n'en est pas moins un mélange suivant des proportions *contingentes* de types cliniques très différents ; non reconnaissables (et même, dans une certaine mesure non inventoriés ; ce

qui laisse peu de signification *essentielle* aux taux moyens de succès calculés sur un tel mélange. L'auteur, au fond, le sent bien.

Envisageant des critères de sélection strictes d'après le territoire cérébral atteint, l'auteur constate qu'en assurant l'homogénéité, on réduit la portée des conclusions, car :

"est-il certain que les résultats obtenus au niveau du territoire de l'artère silvienne, permettent de déduire par extrapolation ce qu'on obtiendrait au niveau de celui de la cérébrale postérieure ?"

D'autre part "une sélection rigide a pour effet d'exclure 90% des malades susceptibles de participer à l'essai", d'où la nécessité de recourir à un recrutement multicentrique où Steiner voit l'occasion de difficultés supplémentaires.

3.3.4 Essai comparatif randomisé en carcinologie : Dans un travail déjà cité (*in* Medical Computing ; éd Laudet et coll ; T & Francis ; 1977, pp. 364 sqq) Cl. Chastang et coll, après avoir opportunément réduit le domaine des essais randomisés, en définissent ainsi la place tout en critiquant les modalités.

"La randomisation sera de règle si on veut tester une addition thérapeutique, ..., une thérapeutique adjuvante, ou comparer deux associations thérapeutiques ... alors que le traitement de référence possède une efficacité certaine ... [si] se pose ... un problème de choix entre deux thérapeutiques dont les résultats prévisibles sont estimés voisins ... La méthodologie des essais randomisés est bien codifiée ; cependant, cette méthodologie, très stricte, apparaît inadaptee dans certains domaines ..."

Dans la plupart des essais thérapeutiques, les critères d'inclusion conduisent à un échantillonnage très hétérogène, parfois inadaptee au but de l'essai ... Ainsi, certains protocoles de leucémie aiguë lymphoblastique prévoient la même thérapeutique pour les enfants et les adultes alors qu'on sait que l'âge est un paramètre pronostique important ... Il nous semble que les critères d'inclusion dans l'essai doivent définir un contexte pathologique homogène où se trouve un choix thérapeutique ... Ainsi, comme à d'autres étapes de la planification de l'essai thérapeutique, il apparaît essentiel de disposer d'une bonne connaissance des facteurs pronostiques d'une affection ..."

"Selon certains, la randomisation aboutit par fonction, à une équirépartition des paramètres au sein des groupes dès que le nombre de sujets est suffisant. La vérification systématique de la comparabilité des groupes résultant d'un tirage au sort montre que le hasard aboutit de façon non exceptionnelle à des répartitions significativement différentes, d'où des difficultés dans l'analyse si ces différences concernent des paramètres pronostiques ..."

[Notons au passage qu'un tel déséquilibre entre groupes est apparue dans une étude (de J. Valenzuela & coll) sur le traitement de l'ulcère duodénal parue dans le même rapport que celle considérée au § 3.3.2 : dans les groupes recevant la dose de 400 mg, la taille des ulcères était plus petite que dans le groupe recevant 800 mg. Les auteurs ne voulant pas conclure que l'efficacité de 400 mg était  $\geq$  à celle de 800 mg, ont corrigé le biais du déséquilibre des groupes, en appliquant aux taux de cicatrisation des ulcères une technique de correction par régression, dont la sûreté peut être mise en doute ...]

C'est pourquoi Cl. Chastang & coll, affirment l'intérêt de la stratification. Cependant dans celle-ci,

"La prise en compte d'un grand nombre de paramètres pronostiques est impossible car le nombre de strates qui en résulte augmente très rapidement. Or, les études pronostiques aboutissant à la mise en évidence de la valeur pronostique de très nombreux paramètres. Ainsi, convient-il de ne retenir que les variables pronostiques indépendantes et les paramètres qui stratifient les malades (e.g. thérapeutique antérieure).

"L'emploi des méthodes multidimensionnelles permet, le plus souvent, de résumer plusieurs dizaines de paramètres pronostiques en un nombre réduit de facteurs pronostiques indépendants. Donc, il serait possible d'utiliser dans la randomisation une stratification selon les facteurs pronostiques issus d'études multidimensionnelles antérieures ou selon la typologie pronostique : tout malade à randomiser est projeté en élément supplémentaire sur le sous-espace des facteurs pronostiques".

Selon nous, le recours aux méthodes multidimensionnelles peut aboutir à mieux qu'une simple amélioration des stratifications. Il permet d'éliminer à la fois la randomisation et le double insu. Telle est la thèse du § suivant.

3.4 Essais ouverts et choix individuels : Supposons que suivant des voies dont I § 4.3.2 offre un exemple, et qui seront précisées au § 5.3, on ait construit un espace typologique E dont les coordonnées (ou facteurs), offrent sous une forme condensée, l'information pertinente apportée par les variables décrivant l'état basal des sujets. Si l'on tient compte de la diversité étiologique, ou même de la subdivision du diagnostic (cf. § 3.3.3) l'espace E aura plusieurs dimensions (comme le suggèrent Cl. Chastang et coll). Si on se borne à considérer comme pertinente, l'information qui définit un niveau global de risque, E sera simplement une droite, qu'on peut appeler "l'axe de gravité", ou "axe de risque". L'espace E pourra avoir été construit d'après des informations rétrospectives (cf. § 3.2), l'essai en cours permettant de revoir et de perfectionner cette construction. Quoiqu'il en soit, à tout individu ancien ou nouveau, pour l'état basal duquel on connaît les variables retenues dans la construction de E, sera associé un point de E, (défini comme il est classique en a. des correspondances pour l'adjonction d'un élément supplémentaire). De même à tout sous-ensemble de patients défini par un caractère contingent (avoir été suivi par tel médecin ...) ou essentiel (avoir présenté tel accident dans telle période après son admission ...) il correspondra un ensemble de points de E, dont on pourra considérer la position moyenne (centre de gravité) et la dispersion ; etc. .

Dans la mesure où la construction de E a été réussie, (synthétise véritablement l'information pertinente), c'est dans chaque zone de E qu'il convient de répondre à la question de l'efficacité absolue et relative des traitements. Prétendre calculer une efficacité globale étendue à E tout entier, relève de cet esprit de confusion légitimement condamné par Ch. Brégeon (cf. § 3.1). Quant au partage de E en zones, il peut être fait de multiples façons, soit en recourant à la classification ascendante hiérarchique ..., soit (particulièrement si E est l'axe de gravité), en plaçant des bornes selon le jugement du clinicien. La définition des zones de E peut être modifiée autant de fois qu'on le voudra pour un même essai ; car à la différence de strates (cf. § 3.3.1) elles ne servent pas à constituer des groupes de traitement, mais seulement à dépouiller les résultats de l'essai.

En effet, dans la mesure où les cas individuels sont situés à leur place dans E, point n'est besoin de randomisation ni de double insu. On rentre dans la pratique classique du libre choix du traitement par le médecin en accord avec son malade. La seule contrainte de l'essai consiste en ce que les informations doivent être recueillies

sous un format fixé à l'avance ; et que l'on doit se borner à un petit nombre de traitements (cf. § 4) afin que chacun d'eux soit échantillonné sur un nombre suffisant de cas. L'essai devient libre et ouvert. Qu'il tombe dans une zone donnée de E des nombres différents de cas ayant reçu chacun des traitements, ne nuit pas à la validité d'une comparaison, pourvu que ces nombres soient assez élevés : exigence que le caractère ouvert de l'essai permet de satisfaire progressivement. Tout autre est la condition des essais randomisés ; lesquels pour définir des groupes de traitement qui sont au fond comme des mélanges en proportions égales de constituants inconnus (§ 3.3.3), suivent un protocole rigide.

Ainsi l'individu ne disparaît jamais dans le groupe ; c'est au contraire la constitution de groupes (définis par des zones etc.) qui est provisoire et réversible. On est aux antipodes de ce par quoi J. Ankri, s'adressant à des généralités (RP Vol 36 n° 7 1-er Févr 1986 ; p. 378) caractérise, à juste titre les essais randomisés, quand il écrit :

"Parler d'essais thérapeutiques en médecine générale, ... c'est surtout pour le médecin généraliste ne plus se comporter face à son patient dans un dialogue régulier, mais considérer son patient comme élément d'un groupe qui lui même permettra la comparaison de deux traitements ..."

On retrouve dans l'espace E, en une vision explicite, le traditionnel raisonnement par analogie sur des cas individuels.

Il importe de noter ici que des essais non randomisés sont parfois cités dans la littérature ; et que le libre choix est toujours cher au médecin. Voici deux références.

"Efficacité de l'amygdalectomie dans le traitement des infections à répétition grave de la sphère rhinopharyngée chez l'enfant.

Résultats d'essais cliniques randomisés et non randomisés menés en parallèles" (JIM Vol 10 n° 60 Mai 1985 ; pp. 367 sqq).

"91 enfants ont été répartis au hasard selon qu'ils subissaient un traitement chirurgical ou non ; les 96 autres ont reçu le traitement que désiraient leurs parents" [qui avaient refusé de donner leur consentement à une affectation aléatoire]. "Dans les essais randomisés et non randomisés les effets ... ont été semblables".

Quant au libre choix, dans un article où il s'efforce d'arbitrer le débat entre partisans et adversaires de l'usage de  $\beta$ -bloquants face aux hémorragies digestives (JIM n) 49 suppl. Gastro. Avril 1984, pp. 13-17), Telfer B. Reynolds rejette l'utilisation dans les essais de thérapeutiques exposant la vie du malade ; et il fait référence explicite à la volonté de celui-ci quand il écrit :

"De nombreux cliniciens prescrivent ... le propranolol dans l'hypertension portale, arguant de la ... tolérance de ce traitement et de la possibilité de recourir à un autre ... en cas de récurrences d'hémorragie. Cette attitude omet de prendre en compte le risque vital d'une récurrence ... Personnellement j'estime que ... le propranolol ne doit être utilisé dans les études contrôlées que comme une thérapeutique d'appoint, ou chez les malades qui refusent une autre forme de traitement..."

avec, en conclusion, cette référence personnelle au libre choix :

"Si j'étais moi-même atteint de cirrhose ; et avais survécu à un épisode d'hémorragie par rupture de varices .... je demanderais à subir une sclérothérapie par un endoscopiste expérimenté ... si les épisodes

hémorragiques récidivaient, ... je choiserais une anastomose chirurgicale ..."

Cette référence explicite au choix du traitement nous servira de transition vers la § 4.

#### 4 Choix des traitements et moralité des essais

4.0 Un témoin historique : Peut-être les essais thérapeutiques sont-ils aussi anciens que la médecine. L'illustre Tournefort, dont on se souvient aujourd'hui comme d'un grand botaniste, ne manquait pas de considérer les plantes en pharmacien et en médecin. Du récit d'un voyage qu'il fit en Orient à la fin du XVII-ème siècle, nous citerons un épisode qui concerne notre propos (d'après P. Guiral\* ; *in* Tournefort : édité par le Muséum National d'Hist. Nat. 1957).

"Nous fûmes sur le haut de l'Olympe autant que les neiges nous le permirent : ce haut est tout de sapins ; nous y trouvâmes d'assez belles plantes, parmi lesquelles j'estime beaucoup le véritable hëllébore noir des anciens, qui est bien différent de celui à qui on donne ce nom en Europe ; vous en recevrez de la graine ; nous en avons fait de l'extrait et même quelques expériences sur des Juifs dont les faubourgs de Pruse sont remplis. Il purge très bien à demi gros, mais il fatigue furieusement l'estomac et trouble le cerveau par des élancements considérables : ainsi je crois qu'il est mieux de n'en donner que 15 ou 20 graines et le mêler avec quelque autre purgatif. Les premiers Grecs ou Arméniens qui se présenteront m'en donneront des nouvelles".

Ne nous hâtons pas de qualifier Tournefort de raciste : il apparaîtrait seulement que voyageant dans l'Empire Ottoman notre botaniste se garde de procéder à des essais thérapeutiques sur les Turcs. Émerveillons-nous plutôt de la franchise de l'expérimentateur, et de la naïveté avec laquelle les habitants de la contrée qu'il visite confient à un Docteur étranger le soin de les médicamenter. Sommes-nous aujourd'hui plus justes ? appliqués que nous sommes à nous justifier ou à nous cacher ?

#### 4.1 Essais portant sur un seul traitement, sans comparaison à un placebo : Cl. Chastang et coll (*op. laud.*) affirment :

"En carcinologie, la décision thérapeutique ne requiert pas systématiquement la réalisation d'un essai randomisé [par quoi les auteurs entendent : comparatif]. En effet pour des affections dont le pronostic, en l'absence de thérapeutique efficace, est très péjoratif, la mise en évidence de l'intérêt d'une nouvelle thérapeutique sur un échantillon réduit de malades peut être suffisante pour généraliser cette thérapeutique. La chimiothérapie a fourni ces dernières années de nombreux exemples d'une telle recherche thérapeutique ....

Ces exemples soulignent l'importance décisionnelle de la phase II [évaluation d'efficacité non comparative] qui permet parfois de conclure d'emblée ... En outre, la valeur thérapeutique démontrée en phase II ne sera pas limitée à une population dont est issu l'échantillon de malades traités, mais sera, le plus souvent, étendue à une population bien plus large pour laquelle on dispose déjà de thérapeutiques efficaces (réinductions dans les leucémies aiguës).

La réticence des praticiens face à l'utilisation de placebos ou de traitements dépréciés est d'ailleurs bien connue. Un exemple nous en est fourni par le compte rendu du "Traitement de la colique néphrétique par un anti-inflammatoire non stéroïdien : le kétoprofène (par J.-P. Archimbaud et SOS Lyon Médecins ; *in* JAMA fr supp au n° 91 Vol 9 15 Déc 1984 ; pp. 12-15). Avec 90% de succès sur 189 observations

(\*) P. Guiral renvoie à la p. 508 du Manuscrit n° 395 de la Bibliothèque du Muséum ; manuscrit qui contient des lettres de Tournefort.

rapidement collationnées dans le cadre particulièrement favorable d'un organisme spécialisé dans les urgences, les auteurs croient pouvoir affirmer l'efficacité de la thérapeutique qu'ils ont adoptée. A un auditeur qui suggère la nécessité d'une expérimentation en double aveugle, J.-P. Archimbaud répond :

"... nous avons eu déjà beaucoup de peine à réaliser cette enquête et ... je n'ai pas cru devoir imposer aux médecins de SOS Lyon Médecins une expérimentation en double aveugle. Ils employaient ce produit depuis 1976 et *avaient constaté une très grande différence d'efficacité avec leur conduite à tenir précédente (analgésique et antispasmodique)* et n'auraient pas accepté de faire un traitement dont le résultat aurait été une fois sur deux aléatoire".

4.2 Comparaison interrompue : Eventuellement un essai comparatif peut être interrompu du fait de la supériorité manifeste d'un traitement. Tout en se faisant l'écho de la réserve qui est de mise face aux analyses en cours d'essai, Cl. Chastang et coll n'en marquent pas moins l'intérêt.

"La rigueur méthodologique s'oppose à la réalisation d'analyses alors que l'essai n'a pas atteint les conditions prévues pour le protocole (nombre de sujets nécessaires, reculs) ...

Pendant, la réalisation d'analyses intermédiaires apparaît une nécessité éthique : elles permettent de détecter une éventuelle toxicité de la nouvelle thérapeutique, voire son inefficacité. Et inversement, elles conditionnent une mise en évidence précoce de l'efficacité de la nouvelle thérapeutique et la généralisation du meilleur schéma thérapeutique à tous les malades qui en présentent l'indication".

Voilà un exemple d'essai interrompu (N. Borhani ; *in* JAMA fr Vol 10 n° 110 suppl. Cardio. 30 Oct 1985 ; p. 23) dans le domaine de la prévention secondaire de la maladie coronarienne (après infarctus) :

"Le BHAT a été l'un des plus importants essais cliniques à long terme sur les  $\beta$ -bloquants ; il a porté sur le chlorhydrate de propranolol (cf. :  $\beta$ -Blocker Heart Attack Trial Research Group ; JAMA 1982 ; 247 ; pp. 1707-1714). Cet essai a été interrompu de façon prématurée à cause d'une réduction significative de la mortalité (26%) dans le groupe propranolol comparé au groupe placebo. L'efficacité était maxima chez les malades de plus de 60 ans et ceux porteurs d'une arythmie ou d'autres complications.

On notera au passage que l'indication du propranolol est apparue dans une zone de gravité élevée.

4.3 Liberté thérapeutique et discipline des essais comparatifs : Selon l'usage présent, le protocole une fois adopté par les experts doit, (sauf exceptions : cf. § 4.2 et § 3.4 *in fine* amygdaléctomie) s'exécuter, sans que malades et médecins aient d'autre possibilité que de refuser d'y entrer ou de l'accepter en bloc ; le malade ayant donné, selon une formule sur le sens de laquelle on ironise souvent, son "consentement éclairé".

A l'attribution par tirage au sort (randomisation) de traitements administrés en double insu, nous proposons de substituer le libre choix (§ 3.4). Reste à préciser dans quel cadre s'exerce ce choix.

Il faut d'abord que l'état basal du patient et son évolution ultérieure soient décrits par un ensemble fixé de variables, sans lesquelles, la place du cas individuel dans l'espace typologique E, sera inconnue (éventuellement, on prévoira deux protocoles d'examen, l'un léger, l'autre lourd ...).

Ensuite les traitements à comparer doivent avoir été bien définis: exige sur laquelle nous reviendrons (§ 4.4), car dans certains domaines, le traitement apparaît comme un dialogue, où chaque réponse physiologique du malade à la thérapeutique reçue doit être suivie d'une réplique du médecin traitant. En sorte que, plutôt que des traitements rigides à dosage strictement préétablis, on doit proposer comme termes de comparaison des stratégies de traitement. Par exemple (cf. I § 3.1) certains essais d'antihypertenseurs, assignent au médecin un niveau de pression diastolique à atteindre en précisant les médicaments à utiliser en 1-ère, 2-ème, voire 3-ème intervention, mais sans prévoir (et pour cause) par quelles prescriptions on parviendra au terme fixé.

Supposons qu'aient été arrêtés par les experts, les traitements (ou stratégies de traitement) à comparer : disons T1 et T2. Il reste à attendre que les binômes malade-médecin choisissent de suivre T1 ou T2. Dans les essais randomisés le "Tu auras ce que le sort te donnera" est de règle ; le nombre de ceux qui vont à T1 et à T2 est préétabli (même si l'une des difficultés majeures des essais est de recruter des patients satisfaisant aux critères d'admission, et acceptant le protocole). Selon nous, il appartient aux experts de mettre en valeur l'efficacité et les risques inhérents à T1 et T2. Dans une profession où au "Oui" d'Hippocrate, n'a jamais manqué le "Non" de Gallien, thérapeutique nouvelle, (voire agressive), et thérapeutique éprouvée (voire abstention) trouveront l'une et l'autre des adeptes ... Reste que si T1 est beaucoup plus apprécié que T2, l'ensemble I2 des sujets qui reçoivent cette dernière thérapeutique sera de faible effectif ; et la supériorité éventuelle de T2, dans des zones de E qui lui sont propres sera lente à se manifester.

Il est vrai que réticences et engouements injustifiés ont de tout temps entravé le progrès de la médecine. Mais d'une part ces attitudes sont inévitables en une matière où l'homme met en jeu sa vie ; d'autre part, présentes au niveau des choix individuels, elles le sont aussi au niveau des experts qui décident de l'opportunité d'un essai randomisé. Et nous croyons qu'il n'y a finalement rien de plus juste ni de plus moral (cf. § 4.5) que d'accorder dans les essais à chaque thérapeutique un poids (c'est-à-dire un nombre de sujets) déterminé par la confiance que celle-ci aura pu inspirer. Le caractère ouvert des essais, assurant d'autre part qu'à la faveur des fluctuations de l'opinion médicale, toutes les zones de l'espace E, finiront par se peupler en points individuels relevant des deux traitements ; à moins qu'un de ceux-ci ne tombe dans un discrédit durable (dont on pourra toujours se demander s'il est justifié ...).

4.4 Traitements et stratégies de traitement : Dans l'article déjà cité consacré au traitement médical des lombalgies (JAMA Vol 8, n° 102, 30 Mars 1985 ; p. 363) Ch. Brégeon note que, dans les essais :

"la monothérapie est privilégiée et les associations sont considérées plus ou moins comme des contaminations compliquant l'appréciation de chacune des méthodes" ; et il se demande :

"Pourquoi faut-il *a priori* exclure l'emploi simultané ou successif de plusieurs techniques ?"

La réponse est évidemment méthodologique. Et T.J. Steiner (*op. laud.*) ne dissimule pas que :

"L'équilibre entre souplesse et rigidité dans l'élaboration du protocole thérapeutique résulte d'un dosage qui nécessite une connaissance très approfondie de la pathologie en question et de son traitement".

Pathologie et traitement sont d'une complexité très inégale selon les domaines. Sur "Le choix des médicaments anti-arythmiques" Ph. Coumel écrit (JIM ; suppl. Cardio. au n° 54 ; Nbre 1984 ; pp. 29-34) :

"L'important est de réaliser que nous sommes en permanence face à des problèmes multiparamétriques, que nous utilisons des médicaments dont les propriétés ou tout au moins les effets ne sont pas nécessairement spécifiques, et que les démarches thérapeutiques sont probablement beaucoup plus empiriques qu'elles ne le paraissent.

Les combinaisons thérapeutiques, sans doute parce qu'elles sont finalement mieux adaptées, ont le secret du contrôle des arythmies rebelles, mais par essence elles ne se prêtent guère à des protocoles stricts d'évaluation" ...

Et à propos de la même question de l'évaluation des anti-arythmiques A. Castaigne (*op. laud.* : RP ; TXXXV ; n° 52 ; 1985 ; p. 3121) ayant écarté les protocoles stricts contrairement aux conditions naturelles de prescription, constate avec désenchantement que des protocoles ouverts, personne ne peut tirer de conclusions valides et sûres.

Quelque méthode qu'on utilise, il faut reconnaître que la comparaison, en tant que telle se fait entre deux termes, ou un petit nombre de termes, non entre une diversité infinie de chemins. *L'équilibre entre souplesse et rigidité* prôné par Steiner sera toujours de mise. Mais face à des problèmes multiparamétriques, l'analyse des données peut aider le praticien à prendre explicitement conscience de l'espace E où s'inscrit la diversité des cas ; à assimiler un cas nouveau à des cas déjà traités qui en sont proches ; à reconnaître dans l'*empirisme* de ses propres cheminements thérapeutiques ou de ceux de ses confrères, des voies frayées, d'une sûreté confirmée.

4.5 Ethique des essais thérapeutiques : Sous ce titre, dans la Revue du Praticien (TXXXV, n° 16, 15 Mars 1985 ; pp. 919-920) Jean BERNARD, s'adresse en ces termes aux médecins.

"L'évaluation d'un traitement est un devoir ...

En l'état actuel des connaissances, cette évolution nécessite :

- 1) une expérimentation sur l'homme sain ...
- 2) une comparaison entre un groupe de malades recevant le nouveau traitement et un groupe témoin soumis à un traitement classique, le meilleur en principe, ou non traité s'il n'existe pas de traitement actif reconnu.

Ces nécessités sont rudes.

La première a suscité de graves abus ... recours à de faux volontaires ... organisation d'un service de volontaires salariés ...

La seconde conduit, dans le traitement d'un malade, à ne pas tenir compte seulement du malade mais à tenter aussi de *contribuer au progrès de la thérapeutique*. Elle est encore aussi rendue plus rude encore par les exigences des statisticiens : 1) traitement des malades fixé par le tirage au sort ; 2) double ignorance (double insu) des malades et des médecins ; 3) poursuite de la comparaison jusqu'au nombre initialement fixé, alors même qu'une différence éclatante apparaît très vite entre les deux groupes.

... On doit beaucoup souhaiter la mise au point par les statisticiens de méthodes d'appréciation des résultats aussi efficaces médicalement et plus acceptables éthiquement".

C'est la *nécessité* d'opérer sur des groupes et non sur des individus qui est cause des *exigences* des essais randomisés. Nous croyons que l'analyse des correspondances, en préservant l'individualité des cas dans le dépouillement des essais, offre une nouvelle *méthode d'appréciation*. Au schéma d'une telle méthode, qui intègre les suggestions que nous avons déjà faites au cours de l'exposé, est consacré le § 5.

5 Espace typologique et essais ouverts en libre choix avec transition vers zéro : Par ce titre pesant nous prétendons répondre à un autre si souvent répété, et que nous citons dans la langue originale des auteurs.

A large prospective, multicentric, double blind, stratified, randomized, placebo-controlled, clinical trial ...

Avant d'exposer schématiquement la construction de l'espace typologique, nous écarterons diverses objections.

5.1 Inadéquation des modèles analytiques : Interrogé sur les progrès récents effectués dans le traitement des cancers, le P<sup>r</sup> G. Mathé (AMIF, T32 n° 329 Oct 1984) conclut en ces termes, en précisant l'opposition manifestée par lui contre "certaines méthodes statistiques" :

"Je souhaite en effet qu'un nouvel outil mathématique nous épargne les essais "randomisés". Lorsqu'un nouveau traitement est testé, les statistiques doivent nous permettre de comparer son efficacité à celles des anciens traitements. Trois moyens.

Traiter tous les malades avec le nouveau traitement et compulser les vieux dossiers pour apprécier les résultats des anciens. Méthode peu sûre [cf. *supra* § 3.2].

Deuxièmement la randomisation. Deux populations sont réparties au hasard et soumises respectivement aux deux traitements à comparer. Et tant pis pour les malades qui ont tiré le mauvais numéro !

Enfin la solution du "statisticien à visage humain". Comparer les résultats du nouveau traitement à un modèle mathématique" ...

Et d'évoquer le succès des instituts de sondage travaillant sur des échantillons éventuellement "améliorés par divers artifices mathématiques" ....

Pleinement d'accord avec G. Mathé pour proscrire la randomisation "cette perle suprême de la pharmacologie" (Quotidien du Médecin ; 17 Avril 1986 ; p. 26) nous tenons à mettre en garde contre certains "modèles et artifices" mathématiques, dont les médecins ne sont d'ailleurs pas dupes ...

Sous la signature A.C, voici comment en quels termes la Revue du Praticien (TXXXVI n° 24 ; ....1986 ; p. 1371) rend compte d'un travail paru dans *The New England Journal of Medicine* (1986 ; 314 ; pp.267-71).

"On discute ferme autour d'un problème important et ... non résolu : les digitaliques prescrits à des sujets en insuffisance cardiaque après un infarctus du myocarde sont-ils nuisibles ou utiles ?

La méthodologie utilisée pour aborder cette question est curieuse, et certainement inadaptée. Les uns après les autres, les équipes qui ont étudié des grands nombres de coronariens ... vont à la pêche dans leurs données et présentent la survie de deux groupes de patients : ceux qui reçoivent des digitaliques et ceux qui n'en reçoivent pas. Dans toutes les études, le résultat est clair : la mortalité

des malades ayant eu un infarctus et recevant des digitaliques est très supérieure à celle des sujets qui n'en reçoivent pas.

Il n'est pas nécessaire d'être un statisticien chevronné pour dire que les deux groupes de patients n'ont rien en commun et ne peuvent donc pas être comparés. Les patients qui reçoivent des digitaliques ont une fraction d'éjection plus basse, des antécédents plus lourds que ceux qui n'en reçoivent pas. Mais les statisticiens savent, partant de deux groupes dissemblables, ajuster et rendre théoriquement comparable ce qui ne l'est pas. En se livrant à ces petits jeux mathématiques, on aboutit à la conclusion que la surmortalité n'est probablement pas due à la digitaline mais à l'état cardiaque sous-jacent. Cependant, même après cette procédure d'ajustement, on ne démontre pas d'utilité de la digitaline chez ces patients".

Les statisticiens savent, le savent-ils vraiment ? En mathématique et dans certains domaines de la physique existent des fonctions dites "analytiques", en ce sens que de la connaissance de la fonction au voisinage d'un point  $x_0$ , on peut déduire sa valeur en tout point de  $x$  si éloigné soit-il de  $x_0$ . Rien de tel n'existe en Médecine, ni dans les autres sciences du complexe (sociologie, économie), où l'on ne connaît de façon sûre que ce qu'on a effectivement englobé. Ainsi les modèles analytiques ne s'appliquent-ils pas en médecine.

Quant à la prescription des digitaliques après infarctus du myocarde, nous recommandons de définir un "axe de risque" (cf. §§ 3.4 & 5.3.2) et d'examiner sur chaque intervalle de cet axe, le devenir des patients, digitalisés ou non, qui s'y projettent ; calculant par segment des taux de survie, d'accidents etc.

5.2 Résolution d'un sophisme statistique : Afin de défendre contre toute argumentation spé cieuse, l'analyse que nous proposons de la correspondance entre état basal et réponse au traitement, il importe de réfuter ici une note de J.S. Gill et coll (The Lancet, 1985, 1 ; pp. 565-569) dont le JIM (n° 60, suppl. Cardio ; Mai 1985 ; p. 73) rend compte sous le titre :

Niveau initial de pression artérielle et efficacité du traitement : une chausse-trappe statistique.

Ayant relevé que pour certains antihypertenseurs, plusieurs publications font état d'une baisse ( $\Delta PA$ ) de la tension artérielle initiale d'autant plus importante que son niveau avant traitement était plus élevé, J.S.Gill et coll rétorquent que "le coefficient de corrélation que l'on pourrait attendre entre  $PA1$  et  $\Delta PA$ , en dehors de toute efficacité du traitement serait de 0,71". Et ils proposent, "pour éviter de donner une signification à un artefact mathématique" d'étudier la relation entre la moyenne des valeurs avant et après traitement ( $(PA1 + PA2)/2$ ) et  $\Delta PA$ .

Sous le nombre 0,71, on reconnaît  $2^{-1/2}$  ; et le raisonnement de J.S. Gill et coll est facile à reconstituer d'après le C.R. du JIM. Les auteurs postulent que  $PA1$  et  $PA2$  sont deux variables aléatoires indépendantes de même loi  $y_1$  et  $y_2$  ; ou, ce qui revient au même, deux réalisations d'une même variable  $y$ . Or il n'en est rien !  $PA1$  et  $PA2$  mesurées sur un même sujet sont (sous l'hypothèse de l'inefficacité du traitement) une seule et même réalisation de la variable  $y$ , affectée toutefois de deux erreurs indépendantes  $e_1$  et  $e_2$  ; d'où si on note respectivement  $\sigma$  et  $\epsilon$  les écarts-type de  $y$  et  $e$  :

$$\text{Corr}(\Delta PA, PA) = \text{Corr}(e_1 - e_2, y + e_1) = \epsilon / (2(\sigma^2 + \epsilon^2))^{1/2}.$$

Or, si la mesure de la pression artérielle a quelque valeur :  $\epsilon \ll \sigma$  et  $\text{Corr}(\Delta PA, PA)$  est négligeable.

La proposition des auteurs de substituer à la valeur basale, la moyenne entre valeurs basale et finale, a, sans qu'il soit besoin de chercher d'autre grief, l'inconvénient de ne pas concerner directement la relation entre *état basal* et *réponse au traitement*. Quant à l'efficacité même du traitement, il va sans dire qu'elle pourrait exister alors que  $\text{corr}(\Delta PA, PA1) = 0$  ; par exemple si  $\Delta PA$  était la même pour tous les sujets...

### 5.3 Construction de l'espace typologique

5.3.0 Desiderata des praticiens : Un exemple nous rappellera la difficulté qu'ont les praticiens à tirer une réponse utile d'essais portant sur une pathologie multiforme.

Rendant compte des essais prospectifs des médicaments antiagrégants plaquétaires dans le traitement des TIA, "atteintes ischémiques transitoires", ou "épisodes réversibles de dysfonctionnement neurologique focal, attribués pour la plupart à l'embolisation de caillots résultant de l'agrégation de plaquettes au niveau des plaques d'athéromes des artères carotides ou vertébro basilaires", M. Ramirez Lassepos (PGM Vol 75 n° 5, Avril 1984, pp. 52 sqq) constate que seule l'aspirine (prise éventuellement en association) pourrait réduire le risque d'ictus, (et ce chez des patients sélectionnés, notamment ceux ayant déjà eu un ictus, ou plusieurs TIA carotidiennes) ; et il conclut en caractérisant en ces termes ce qu'il attend d'une grande étude américano-canadienne en cours, étude multicentrique, randomisée en double aveugle :

"On en pourra tirer de précieuses informations. Mais vu le grand nombre des variables susceptibles d'intervenir chez les patients ayant des TIA carotidiennes, il est peu vraisemblable qu'on obtienne une réponse définitive".

5.3.1 Format général des données de la construction : Face à un patient, le praticien doit déterminer son traitement d'après l'état basal qu'il observe. C'est en fonction des variables de l'état basal, qu'un patient nouveau "in" pourra être placé dans l'espace typologique E.

Mais l'espace E lui-même (cf. I § 4.2.3) sera généralement construit d'après un ensemble I d'individus dont l'histoire clinique a été suivie dans son évolution à partir de l'état basal.

De façon précise, notons VB l'ensemble des variables caractérisant l'état basal du patient, c'est-à-dire des variables biologiques, des signes cliniques, des données d'examen radiologiques ... (pour ne rien dire du signalement du patient par âge, sexe, données anthropométriques ..) dont le médecin est censé disposer au début du traitement. De même notons VR les variables de *réponse* ou d'*évolution*, par référence auxquelles on décide de considérer la valeur diagnostique et pronostique des variables de l'état basal : ces variables pourront être de nature très diverses et plus ou moins nombreuses selon les cas. S'il s'agit de malades dont la vie est en danger immédiat, ou si l'on s'intéresse à un pronostic de survie à moyen terme (par exemple à cinq ans) on pourra ne considérer qu'une seule variable de *réponse* : la survie, avec deux modalités : survivant, ou décédé (au terme choisi) (cf. § 5.3.2). Mais pour étudier à proprement parler la réponse à un traitement, il convient de caractériser cette réponse par les variations de plusieurs variables (biologiques, hémodynamiques ...) parmi celles observées à l'état basal (cf. § 5.3.3).

Sans entrer dans le détail des ensembles de variables VB et VR, on peut affirmer que celles-ci sont dans la quasi-totalité des cas très hétérogènes : il suffit pour s'en convaincre de considérer un dossier médical. L'analyse se fera en réduisant toutes les variables

à un format unique, celui de la question fermée à un nombre fini de modalités. Certaines variables peuvent avoir d'emblée ce format : c'est le cas de la variable survie, à deux modalités ; ou encore d'une variable comme "présence ou absence d'antécédent d'Infarctus du myocarde". Mais toute variable continue (concentration, pression etc.) sera découpée en une suite de modalités (telles que faible, moyenne, forte : cf. I § 4.2.3). De cette opération de codage dépend le succès de l'analyse : elle est tout l'art du statisticien ; nous ne faisons ici que la décrire schématiquement.

Ayant attribué à chaque variable  $v$  un ensemble noté  $J_v$  de modalités, on a deux ensembles  $JB$  et  $JR$ , comprenant respectivement les modalités des variables basales et celles des variables de réponse ; soit en notation ensembliste

$$JB = \cup \{J_v | v \in VB\} \quad ; \quad JR = \cup \{J_v | v \in VR\}.$$

On construit alors d'après l'ensemble  $I$  des individus, le tableau de correspondance entre modalités de l'état basal et modalités de la réponse. Tableau qu'on pourra disposer avec en ligne les modalités de  $JR$ , en colonnes celles de  $JB$  ; et à l'intersection de la ligne  $j_r$  et de la colonne  $j_b$  le nombre des individus  $i$  de  $I$ , qui rentrant à l'état basal dans la modalité  $j_b$  (e.g. avoir un antécédent d'IdM) ont dans leur réponse la modalité  $j_r$  (e.g. avoir une augmentation de l'Indice Cardiaque).

L'analyse de ce tableau fournit dans un espace  $E$  une représentation simultanée des deux ensembles  $JB$  et  $JR$ , ainsi que des individus  $i$  de  $I$  ou de tout individu nouveau  $i_n$ , adjoint au tableau  $JR \times JB$  comme une ligne supplémentaire (formée de 0 et de 1 avec dans la colonne  $j_b$ , 1 si l'individu rentre à l'état basal dans la modalité  $j_b$  et 0 sinon). L'espace  $E$  est engendré par les premiers axes factoriels issus de l'analyse du tableau  $JR \times JB$  ; le choix du nombre d'axe (1 seul =  $E$  est une droite ; deux =  $E$  est un plan ; etc.) dépendant de l'interprétation.

5.3.2 Axe de survie et axe de risque : Si la réponse est décrite par la seule variable "survie" à deux modalités, le tableau  $JR \times JB$  n'a que deux lignes et on sait que l'espace  $E$  est alors réduit à une droite : c'est l'axe de survie. La valeur prédictive de l'analyse est d'autant plus grande que sur cet axe les points figuratifs des patients survivants et décédés constituent des groupes peu empiétants.

Si la réponse est décrite par plusieurs variables mais que celles-ci ne sont que des mesures globales de la qualité de l'évolution (avec ou sans accidents ; etc. ; cf. I §§ 3.2.3 & 4.2.3) ; les résultats de l'analyse se verront clairement sur le 1-er axe ; ou éventuellement dans le plan des deux premiers axes.

Un exemple de construction d'axe de survie se trouve dans : Méthodes de discrimination appliquées à des données médicales en vue d'une décision thérapeutique [AORTE] C.A.D. Vol II n° 4 (1977) pp. 415 sqq) et Pratique de l'Analyse des Données ENSI pp. 406-414.

5.3.3 Référence à un exemple : les effets de la trinitrine Intra Veineuse : Nous ne reproduirons pas cette étude déjà connue des médecins et des statisticiens (Nouvelle Presse Médicale ; pp. 283 sqq, Janvier 1979 ; et C.A.D. Vol V n° 3 1980 pp. 341 sqq), mais tenterons d'en esquisser le schéma.

S. Baligadoo a démontré que l'insuffisance cardiaque peut être réduite par l'effet vasodilatateur de la trinitrine IV. Toutefois la

mise au point de cette thérapeutique a été délicate, car à côté de résultats encourageants on observait parfois de graves chutes brutales de la Pression Artérielle dans de gros vaisseaux. Alors que d'autres chercheurs tentaient de chiffrer l'efficacité globale obtenue sur un ensemble de malades et n'obtenaient que des résultats discordants, S. Baligadoo a pu le premier grâce à l'a. des données distinguer selon l'état initial des malades trois types principaux de réponse hémodynamique et circonscrire ainsi l'ensemble des cas auxquels s'applique la thérapeutique nouvelle.

Dans cette étude, les variables de l'état basal étaient mesurées par cathétérisme cardiaque ; les variables de réponse n'étaient autres que les *variations* de ces mêmes variables sous traitement.

Il importe de signaler à l'attention du praticien, que le domaine d'indication de la trinitirne IV défini par l'analyse comme une zone du plan (typologique à 2 dimensions) E, peut assurément être définie, après l'étude statistique, dans le langage usuel du cardiologue ; soit :

pression capillaire pulmonaire > 22 mm Hg  
 indice d'éjection systolique < 19 ml m<sup>-2</sup>/syst.  
 résistance artérielle systémique > 2300 dy s cm<sup>-5</sup>.

Un autre exemple se trouve dans [LEUCEMIE LCC] CAD Vol 1 n° 4 (1976) pp. 419-440 ; c'est-à-dire (cf. *supra* § 5.3) de l'espace où tout patient nouveau peut être représenté par un point dont la place a valeur de pronostique.

5.4 Résumé des suggestions méthodologiques : Nous voici parvenu au terme d'un exposé, dont la thèse principale, objet du § 3.4 est de fonder sur une représentation spatiale la considération explicite des cas individuels dans les essais. En découle une méthodologie nouvelle dont nous réappellons les caractères.

5.4.1 Essais ouverts : N'ayant pas à partager aléatoirement des strates en deux moitiés égales affectées à deux traitements, on accepte les cas individuels en aussi grand nombre qu'ils se présentent, et d'où qu'ils viennent, la seule règle stricte étant celle de la collecte des variables basales et des variables de réponse. L'analyse reposant sur la représentation spatiale des cas individuels, les anomalies de certaines observations, les différences de méthodes entre centres etc. ne passent pas inaperçues.

5.4.2 Essais libres : Le choix des thérapeutiques est laissé au médecin en accord avec le patient ; les thérapeutiques-type étant définies par les organisateurs de l'essai en gardant "un juste équilibre entre souplesse et rigidité". Sachant quelles drogues il administre, le médecin en contrôle l'effet. Chaque thérapeutique reçoit pour poids celui de la confiance qu'on lui accorde.

5.4.3 Transition vers zéro : il n'y a ni double insu ni placebo, mais l'effet placebo n'est pas ignoré ! Sont expérimentées des doses faibles concurremment aux doses fortes ; avec transition vers zéro aussi longtemps que l'effet bénéfique sur les malades d'une zone donnée de E\* semble préservé par une réduction de la dose.

---

(\*) C'est-à-dire (cf. *infra* § 5.3) de l'espace où tout patient nouveau peut être représenté par un point dont la place a valeur de pronostique.

5.4.4 Essais dépouillés par zone de gravité (ou plus généralement par zones de variantes de l'état basal) : il n'y a pas d'efficacité globale ; de moyennes de cas hétérogènes ; mais des zones où le praticien retrouve la notion familière de raisonnement par analogie entre les cas.

Pour expérimenter dans la voie que nous suggérons, point n'est besoin d'entreprendre de nouveaux essais : Nous avons au cours de l'expérimentation rencontré maints exemples d'essais déjà réalisés, n'ayant pas abouti à des conclusions utiles. Reprendre en se fondant sur une représentation spatiale l'exploitation des résultats de ses essais, pourrait permettre de délimiter des zones où se dégagent des conclusions nettes.

P.S. Dans son n° du 9 Juin 1986, le Quotidien du Médecin, sous la signature du D<sup>r</sup> C. Faber, évoque en ces termes l'activité du "Comité national d'éthique".

"A l'Académie des sciences, le Pr Jean Dausset a présenté un rapport sur trois années d'exercice du Comité National consultatif d'éthique, au cours desquels six avis ont été rendus publics."

Il s'impose de donner ici la teneur du sixième avis, qui concerne directement le thème du présent article.

"Les essais des nouveaux médicaments chez l'homme : les essais en double aveugle ne sont acceptables que s'il existe une notion d'équivalence entre les médicaments testés. Dans le cas contraire, si le bénéfice attendu pour la collectivité est très important et les inconvénients pour le malade minimes, l'essai pourra éventuellement être envisagé."