

J. P. BENZÉCRI

**L'élaboration statistique des données médicales
: de l'expérimentation thérapeutique à
l'expérimentation clinique**

Les cahiers de l'analyse des données, tome 5, n° 3 (1980),
p. 265-282

http://www.numdam.org/item?id=CAD_1980__5_3_265_0

© Les cahiers de l'analyse des données, Dunod, 1980, 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/>

L'ÉLABORATION STATISTIQUE
DES DONNÉES MÉDICALES :
DE L'EXPÉRIMENTATION THÉRAPEUTIQUE
A L'EXPÉRIMENTATION CLINIQUE
[ANA. MED.]

par J. P. Benzécri (1)

J'exposerai d'abord le thème du présent séminaire, et plus particulièrement, le titre de ma propre communication ; puis je présenterai l'ensemble du programme qui vous est proposé.

0 Objet de l'analyse des données : A la base de la plupart des études inscrites à ce programme, il y a un ensemble de dossiers médicaux. Partons d'un exemple typique, sur lequel je reviendrai : l'expérimentation d'un antibiotique cf. §1.1 et (ESSAI ANTIB.): on a quelque 200 malades, traités par cet antibiotique ; donc 200 dossiers. Chacun de ces dossiers est une observation médicale qui a ses particularités ; mais tous ont en commun un certain nombre d'informations (résultats positifs ou négatifs d'examens ; valeurs numériques de paramètres biologiques), auxquelles on peut donner le format uniforme de réponses à un même questionnaire clos : ce qui veut dire qu'à chaque examen, à chaque variable etc., on associera une question comportant un nombre fini de modalités de réponse. Ainsi à la question "SEXE" du malade, il y a deux réponses possibles ; pour l' "AGE", on ne distinguera pas tous les nombres d'années, mais compte-tenu du fait qu'il s'agit de patients adultes on distinguera seulement trois modalités de réponse : âge inférieur à 43 ans ; âge compris entre 43 et 63 ans ; âge supérieur à 63 ans. On procédera de même au découpage en classes de toute variable continue, telle que : dose totale prescrite, ou créatininémie ; le choix du nombre des classes et de leurs bornes étant, cela s'entend, laissé au spécialiste. La question "SEPTICÉMIE" admet deux modalités de réponse : présence et absence ; et de même pour la recherche des germes pathogènes, tels que *Pseudomonas*, *Proteus*, *Klebsiella*, etc. .

Imaginez que placés devant cette pile de 200 dossiers, vous tentiez d'en acquérir une vue d'ensemble : vous avez une très grande table, et y disposez les dossiers de façon à ce que ceux d'entre eux qui se ressemblent soient proches les uns des autres ; que les dossiers qui ont des caractères extrêmes soient à la périphérie de ce grand étalement. On peut prévoir que vous rencontrerez bien des difficultés ! Vous vous appliquerez à placer entre le dossier A et le dossier B, les dossiers des cas intermédiaires entre A et B : mais ce faisant, vous serez amenés à juxtaposer des dossiers qui diffèrent quelque peu.

Admettons cependant que vous ayez réalisé l'étalement des dossiers ; avec peut-être des espaces vides entre de grands groupes qui correspondent à des domaines bien distincts de la pathologie : par exemple, groupe des septicémies, groupe des infections mineures des voies urinaires... . Je vous demande maintenant de faire un travail de plus !

(1) Professeur de statistique. Université Pierre et Marie Curie.

(Conférence d'ouverture du séminaire du 18 Juin 1980 : Hôpital Tenon)

Chacune des variables comporte, on l'a dit, un nombre fixé de modalités ; préparez des étiquettes portant chacune le nom d'une de ces modalités. Par exemple pour l'âge il y aura les trois étiquettes : "moins de 43 ans", "de 43 à 63 ans", "plus de 63 ans" ; pour le *Pseudomonas*, deux étiquettes : "présence de P." et "absence de P." ; etc. . Il faut maintenant placer toutes ces étiquettes sur la table où sont étalés les dossiers. Cela ne semble pas impossible, parce que, par exemple, comme je viens de le suggérer, tous les dossiers de septicémie sont dans une même zone : on y mettra l'étiquette "Présence de septicémie" ; et il faudra placer ailleurs, au centre de l'étalement des autres dossiers, l'étiquette "absence de septicémie". Et de même pour les classes d'âge : car on peut penser que certaines affections frappent plutôt des sujets âgés ; et d'autres des sujets jeunes.

Finalement, vous aurez réalisé l'étalement simultané et cohérent de deux ensembles : l'ensemble des dossiers, et l'ensemble des étiquettes, ou ensemble des modalités des variables. Ce travail est déjà extrêmement embarrassant à faire pour 200 dossiers et 100 modalités de variables ; même si l'on dispose d'une grande table et qu'on consacre tout son temps à étudier les dossiers ; et avec 2000 dossiers et 300 étiquettes à disposer, ce serait l'impossible ! D'ailleurs, on ne se satisfèra jamais du résultat : parce qu'on restera dans l'embarras, dans l'indécision sur la place d'un dossier ; et parce qu'au fond, par définition, opérer sur une table, cela veut dire, en termes mathématiques, utiliser deux dimensions, deux axes. Or il peut fort bien y avoir une troisième dimension, un troisième axe : ce qui vous amènera à suspendre les dossiers en hauteur ; et encore une quatrième dimension, que vous exprimerez en donnant aux dossiers des couvertures de couleurs différentes, etc.

Tout ce travail, bien qu'il évoque un amoncellement de papiers, qui fait sourire, a un intérêt fondamental : il correspond à ce qu'on appelle l'esprit de synthèse ; (même si la vision de ceux qui ont un esprit de synthèse puissant n'est pas toujours spatiale). Et c'est le travail que nous faisons par l'analyse des données : nous le faisons par un processus mathématique ; qui certes doit être surveillé ; mais qui n'en est pas moins automatique, confié à l'ordinateur, effectué très rapidement, repris sur de nouvelles bases si besoin est...

Voilà donc expliqué ce que nous entendons par analyse des données.

Venons-en maintenant au thème du séminaire : *Analyse des données en pharmacologie et physiologie cliniques* et plus particulièrement au titre de mon propre exposé : *De l'expérimentation thérapeutique à l'expérimentation clinique*. La thèse inscrite dans ce titre, est que l'expérimentation thérapeutique telle qu'on la conduit actuellement (et qui aboutit principalement à produire, du moins dans sa phase ultime, des chiffres d'efficacité comparative de drogues décidant de l'adoption d'un médicament nouveau), devrait pouvoir être exploitée de façon beaucoup plus scientifique et être considérée comme une véritable expérimentation clinique. Nous savons tous qu'en médecine, l'expérimentation est dans une large mesure interdite, par le respect dû aux sujets que l'on traite ; mais il faut se rendre compte qu'un ensemble de cas très grand, fournit sous une certaine forme, la réponse aux questions qu'on serait amené à se poser en faisant de l'expérimentation proprement dite. A condition de donner aux informations qu'on recueille en grand nombre sur les malades une clarté suffisante, une bonne gestion de ces informations aboutirait à un travail scientifique ; qui serait certes soumis à des contraintes bien plus grandes que celles que rencontre un biologiste opérant sur des animaux ; mais qui aurait en même temps pour lui outre l'intérêt exceptionnel de son objet (l'homme même), l'appui d'une masse immense de travaux qu'on effectue de toute façon même si l'élaboration est présentement rudimentaire : le labeur des dizaines de milliers de médecins français étant d'un ordre de grandeur bien supérieur à celui

des travaux de laboratoire ; pour nombreux que ceux-ci puissent être.

1 L'essai thérapeutique : En parcourant le programme de notre séminaire, je commencerai par 3 exemples d'essai thérapeutique. Il s'agit d'essais en phase III (évaluation de l'efficacité d'une drogue dont l'intérêt est déjà reconnu), ou IV (pharmacovigilance, après libre introduction de la drogue sur le marché). Il ne faut pas attendre de telles études de véritables découvertes scientifiques (nous tenterons d'en proposer ensuite). Mais ce sera pour nous l'occasion de rencontrer les méthodes statistiques multidimensionnelles.

1.1 Essai d'un antibiotique : [ESSAI ANTIB.] : Il s'agit, je l'ai dit, de 200 dossiers, (plus précisément 187) étiquetés par quelques dizaines de modalités de variable (exactement 69). Le graphique plan présenté * ici, est rapporté aux deux premiers axes construits par l'ordinateur ; c'est-à-dire qu'il représente en son sens mathématique précis qu'on vous expliquera (cf l'exposé [INT. CORR. MED.], de F. Benzécri), la meilleure représentation plane qu'on puisse donner des deux ensembles en correspondance : l'ensemble des 187 dossiers, et l'ensemble des 69 modalités descriptives. En fait le graphique ne comporte que les modalités (on ne les a même pas toutes indiquées, par souci de clarté). Les points figurant les dossiers (cas individuels) manquent ; nous en reparlerons.

Le quadrant supérieur droit offre un tableau clinique grave : le malade a été hospitalisé en état d'inconscience ; il s'agit d'une septicémie, traitée avec un autre antibiotique associé. L'étiquette Gram +, indique la classe des germes qui sont en cause dans la plupart des cas. Le traitement aboutit à un état clinique amélioré ; etc..

Dans le quadrant inférieur droit on trouve d'autres germes : *Pseudomonas*, *Klebsiella*, associés à un échec clinique.

Dans la moitié gauche du plan, (plus précisément à gauche de l'axe 1) prédominent les infections urinaires, évoluant vers une guérison totale.

J'appelle l'attention sur la dose totale prescrite, modérée ou forte à gauche (de 1,5g à 3,5g) ; très forte à droite sur l'axe (supérieure à 4,5g) ; très faible dans le quadrant inférieur droit (inférieure à 1,5g). Des associations intéressantes apparaissent entre la pathologie d'une part, l'âge et le sexe de l'autre. Sur le deuxième axe s'étagent de haut en bas dans leur ordre naturel les trois modalités : AIGU, RECIDIVANT, CHRONIQUE ; etc. Il appartient au Dr Assouly d'exposer le détail de cette étude ; mon but est seulement de vous soumettre un graphique, et de vous expliquer à quoi il peut servir.

Nous avons dit que l'ensemble des 200 malades n'est pas figuré ici. Qu'apprendrait-on à le regarder. D'abord la répartition des points, dense ici, clairsemée là, révélerait l'isolement relatif du bloc des septicémies ; montrerait que la plupart des cas sont à gauche, dans la zone des guérisons totales. Mais il y a plus : à chaque point, correspond dans le dossier une observation médicale dont la lecture peut, e.g. expliquer pourquoi le médecin a eu recours à une thérapeutique associée ; comment il a choisi la dose prescrite ; pourquoi il a renoncé à obtenir une guérison complète par la nêtilmicine ...

Cependant l'étude a un objectif pratique immédiat, qui n'est pas tant d'acquiescer une vue d'ensemble de la pathologie infectieuse, que

* Le lecteur trouvera ce graphique dans l'article de P. Assouly, [ESSAI ANTIB.], auquel nous renvoyons pour plus de détails.

de trouver au sein de cette pathologie un domaine propre à cet antibiotique. Cet antibiotique est un aminoside. Les drogues de cette famille (parmi lesquelles on peut citer la gentamicine) sont remarquables par leur forte efficacité ; elles présentent une toxicité rénale assez élevée, et peuvent provoquer des incidents cochléovestibulaires (atteintes de l'oreille interne). L'antibiotique ayant montré chez l'animal une néphrotoxicité quasi nulle, on l'a introduite en clinique humaine. Et là, on a eu à déplorer quelques accidents : vous voyez en effet sur le graphique le point atteinte rénale ! Globalement, i.e. sur l'ensemble des cas traités, cet antibiotique est d'une toxicité comparable à celle de la gentamicine (antibiotique analogue, on l'a dit) mais plutôt un peu supérieure : à se borner à un dépouillement classique des essais, on aurait dû conclure à l'abandon pur et simple de l'antibiotique. Mais le graphique suggère un dépouillement par strates : il y a d'une part le domaine des septicémies, où l'antibiotique est inférieur à d'autres drogues ; et d'autre part le reste de la pathologie : où le nouvel antibiotique se distingue par une bonne efficacité et une toxicité modérée (relativement à celle des autres aminosides). L'essai sur l'animal n'avait pas révélé de toxicité, parce que comme d'usage, on avait procédé sur des animaux sains (non atteints de septicémie) pour étudier la diffusion, l'élimination, la toxicité de la drogue (essai dit de phase I).

Je signale encore un détail, qui a rapport avec la toxicité : pour quoi dans certains cas a-t-on prescrit des doses relativement très faibles (total 1,5g) : la réponse n'apparaît pas sur le plan 1 x 2 ; on pourrait la chercher en lisant les observations individuelles ; mais il suffit de regarder le plan 2 x 3 : "dose totale inférieure à 1,5g" y apparaîtrait associé à "antécédent rénal" (i.e. atteinte rénale antérieure) : voilà ce qui avait incité le médecin à la prudence. Ce détail montre à l'utilisateur de l'a. des données, qu'il doit avoir la patience de regarder d'autres graphiques que le plan 1 x 2 présenté d'abord !

1.2 Un pansement de la muqueuse digestive : Ce produit (*) (une argile donnée au malade sous forme de suspension chimiquement inerte ; et destinée à isoler la muqueuse digestive) est déjà sur le marché, et a largement pénétré dans le domaine de la gastroentérologie. L'étude que présente J. Maccario est un exemple de pharmacovigilance, ou étude de la phase IV. Il s'agit de l'analyse d'un très grand nombre d'observations, recueillies auprès de plusieurs centaines de praticiens, travaillant dans toute la France en médecine de ville. Ici encore, on a un ensemble de dossiers qu'on peut décrire chacun, par l'ensemble des modalités qu'il possède d'un certain nombre de variables. Mais cette fois le nombre des dossiers ne permet pas du tout d'envisager qu'on les dispose manuellement sur une table !

De façon précise, l'étude statistique a été scindée en trois parties :

- I 1332 cas d'affections oeso-gastro-duodénales (≈ estomac)
- II 2265 cas d'affections entérocoliques (≈ intestin)
- III 3614 cas d'affections digestives en pédiatrie.

Dans chacune de ces parties, les variables descriptives qu'il convient de retenir sont en effet différentes.

Qu'apprend une telle analyse ? D'abord le domaine d'utilisation du produit ; et son champ d'efficacité propre. Ce pansement gastrique réussit particulièrement dans les diarrhées aiguës : il a peu d'effet contre la constipation (en fait le produit est susceptible de provoquer des constipations : et c'est là le seul accident enregistré dans son

(*) Le rapport de l'expertise, conduite par le Pr. H. Sarles, et publié par le laboratoire IPSEN s'intitule "7200 cas traités par le produit".

emploi). Il est plus original de noter que l'analyse du domaine entérologique isole dans un quadrant deux maladies qui se distinguent par le fait qu'elles sont associées à un traitement anti-inflammatoire par les corticoïdes : ce sont l'iléite de Crohn et la rectocolite hémorragique ; seule cette dernière maladie bénéficie de l'action du produit.

Ce sont là des conclusions thérapeutiques acceptables, et en partie prévisibles. Mais il y a plus : on acquiert une vue d'ensemble ordonnée de la pathologie digestive : et comme dans l'exemple de la pathologie infectieuse traitée par cet antibiotique, il y a une distinction très nette entre hommes et femmes. Autant qu'une étude clinique différenciée par le sexe, c'est une sorte de sociologie qui ressort de ces analyses ; (les conditions de vie, l'alimentation, l'alcool, le tabac, et les états psychosomatiques, conditionnent la pathologie digestive). De ce point de vue l'analyse de 2000 cas traités par la nétilmicine aux U.S.A., et confiés à P. Assouly par le laboratoire Shering, révélera sans doute non seulement des lois universelles de la pathologie humaine et de la sensibilité des germes (dont toutefois les souches ne sont pas également distribuées dans ce monde) ; mais aussi une distribution des maladies propre aux U.S.A., et surprenante pour le chercheur français.

1.3 La toxicité de la digitaline : Voici un médicament qui a une histoire de plus d'un siècle, mais à propos duquel les questions restent posées : quand l'imprégnation digitalique survient-elle ? Comment dépend-elle du terrain ? de la dose prescrite ? Dans un service hospitalier spécialisé, il est possible de collationner plusieurs milliers de fiches d'observations relatives à l'utilisation de la digitaline ; parmi lesquels, nécessairement on aura eu à déplorer des accidents plus ou moins graves. C'est à une telle étude que s'est livré S. Baligadoo, sous la direction du Pr. P. Chiche, en vue d'établir une typologie des effets secondaires de la digitaline, et reconnaître les facteurs prédisposant à la toxicité. L'imprégnation digitalique (avec des signes tels que "présence d'extrasystoles") apparaît associée aux plus fortes doses. Il ne faut toutefois pas conclure immédiatement que la forte dose est la cause même de l'accident ; car peut-être le terrain (la gravité du cas) a incité à prescrire une forte dose, et l'accident serait survenu sur un tel terrain, même avec une dose moindre. Il ne m'appartient pas de répondre à ce doute : mon propos est seulement de montrer sur un exemple que, si la statistique multidimensionnelle ne fournit pas toujours une réponse, elle donne du moins le moyen de poser plus rapidement les questions pertinentes.

Cette étude en recherchant les rapprochements entre les examens biologiques initiaux et la présence de toxicité permet également de déceler plusieurs facteurs, connus avant l'administration de la digitaline, qui seraient liés à une toxicité plus fréquente.

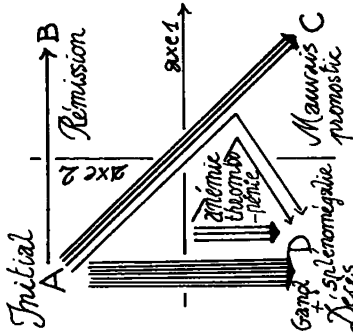
2 Nosologie et pronostic : Nous abordons maintenant un ensemble d'études qui, à la différence de celles présentées jusqu'ici, concernent exclusivement des cas très graves. Dans de tels domaines, faire une typologie précise des états des malades et de leur évolution peut non seulement fournir des éléments de pronostic et des indications thérapeutiques, mais aussi révéler des distinctions au sein d'entités nosologiques sous lesquelles on range, présentement des formes de maladies très diverses. De nos jours en effet, la multiplicité et la puissance des agents thérapeutiques est telle que la réponse du malade aux traitements peut affiner un diagnostic : S. Baligadoo nous a cité l'exemple de l'angine de poitrine, où l'on distingue deux formes : l'une où la douleur cède à la trinitrine, et l'autre où elle cède à un spasmolytique.

2.1 Les leucémies lymphoïdes chroniques * : Rappelons pour les statisticiens non initiés à la médecine, que cette maladie se caractérise par un accroissement du taux de cellules lymphocytaires matures dans le sang et la moelle osseuse. Les lymphocytes étant une variété de globules blancs dont la fonction reconnue est de produire des anticorps,

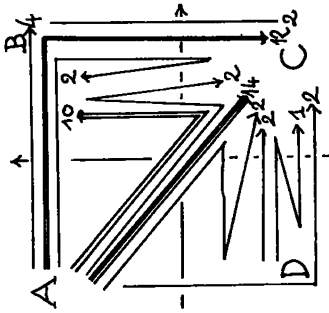
* cf Cahiers Vol I n° 4 pp 419-440 ; 1976 : nous citons largement cet article.

وكلَّ سَعْمٍ فَلَهُ أَوْقَاتٌ
يَكُونُ فِيهِ الْمَوْتُ وَالْحَيَاةُ
"أبو سينا - Avicenne"

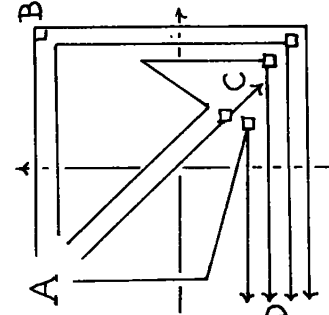
Décès avant 4 ans.



Survie à 4 ans.



Décès après 4 ans.



Analyse du tableau des évolutions de 90 malades atteints de L.L.C. : 392 états x 42 modalités (dont 4 en ét. supplémentaire). Schéma des trajectoires des malades.

à gauche : les malades morts moins de quatre ans après le diagnostic
au centre : les malades encore vivants 4 ans après le diagnostic ; une flèche manquée d'un nombre sténatise la trajectoire de plusieurs malades ayant connu des évolutions analogues.
à droite : des malades décédés après plus de 4 ans de maladie ; un carré □ représente leur position au bout de 4 ans
D'après Ch. Bastin, L. Depps, J.-D. Rain & coll.

leur prolifération peut apporter dans cette production des perturbations graves : absence d'anticorps, ou au contraire production d'anticorps contre des éléments vitaux de l'individu : globules rouges, plaquettes. De tels anticorps (dont pour les hématies la présence est reconnue par l'épreuve de Coombs) provoquent un déficit en globules rouges (anémie) ou en plaquettes (thrombopénie) dit cytopénie *périphérique* par opposition à la cytopénie *centrale* (dont la cause est au contraire la faible production d'hématies ou de plaquettes et non la destruction de ceux-ci par des anticorps). La maladie est très hétérogène, et les délais de survie sous les traitements actuels, très variables.

L'étude que présente L. Degos, concerne initialement 102 malades. Une première analyse porte sur les états initiaux de ces patients, décrits par un ensemble de 91 modalités de variables. On en déduit un ensemble de syndromes principaux qui sont : la prolifération ganglionnaire (g. lymphatiques) ; l'infiltration lymphocytaire (e.g. dans la moelle osseuse) ; et la cytopénie (soit centrale, soit périphérique ; distinction expliquée ci-dessus).

Une première conclusion très utile apparaît ici : les nombreux examens effectués, dont certains sont très douloureux voire dangereux, fournissent des résultats fortement corrélés entre eux. Ainsi on peut se dispenser de recourir à la lymphographie (examen radiologique des voies lymphatiques rendues opaques au rayon X par injection d'un produit) ou à la biopsie médullaire lorsqu'on a effectué d'autres examens simples dont les résultats leur sont très liés : hémogramme, ponction de moelle, examen clinique. Ces examens, avec l'épreuve de Coombs (cf *supra*) et le taux de γ -globuline (anticorps du sérum sanguin) suffisent à un diagnostic complet sur les trois syndromes reconnus à l'analyse multidimensionnelle. Par contre il n'apparaît aucune classification tranchée des malades, permettant de définir des subdivisions de la maladie.

La deuxième analyse, dont les graphiques présentés ici donnent les résultats schématiques, concerne l'évolution des malades. Pour 90 malades, on dispose de plusieurs fiches consignant les résultats d'examens successifs, moins complets en général que l'examen initial, mais réductibles à un format unique, comportant 42 modalités de variables. Au total on a 392 fiches-état : répétons que cette fois nous n'avons pas un dossier unique correspondant à chaque malade, mais plusieurs fiches. Et ce sont ces fiches que nous classons, que nous étalons sur un plan : cet étalement étant fait sans tenir compte de ce que deux fiches concernent ou non le même malade, mais seulement en se fondant sur les similitudes que les fiches ont entre elles ; avec comme dans les analyses précédentes pour étiqueter l'ensemble des fiches, les modalités des variables. Ainsi sur le graphique obtenu, chaque malade est représenté par autant de points différents qu'on a effectué d'examens sur lui ; et la succession de ces points jalonne sur le graphique la trajectoire suivie par le malade.

Pour schématiser ces trajectoires, on a marqué des lettres A, B, C, D les quatre quadrants du plan des axes 1 et 2. Nous considérerons trois graphiques similaires. Le premier concerne les malades décédés moins de quatre ans après le début du traitement. Les flèches se terminent au point représentant le dernier examen antérieur au décès ; l'épaisseur d'une flèche indique le nombre de cas analogues que cette flèche schématise. On voit que la plupart des décès se produisent dans la zone D. Il y en a quelques uns dans la zone C ; un seul dans B. Il n'y en a pas dans A, parce que cette zone correspond à un état initial où le mal n'est pas encore développé.

Le deuxième graphique concerne les malades qui sont survivants ans après le début du traitement : 16 d'entre eux sont dans la zone B,

qu'on peut appeler une zone de rémission, car on n'y a vu qu'un cas de décès. Beaucoup sont dans la zone C : ce qui n'est pas d'un bon pronostic, comme on le voit sur le troisième graphique, qui concerne les malades morts après plus de 4 ans de traitement, avec marqués d'un carré blanc, les points intermédiaires décrivant leur état après 4 ans de traitement.

Soulignons que la préparation des données en vue d'une telle analyse n'impose pas au clinicien de contrainte excessive : il suffit que toutes les fiches d'observations soient réductibles à un format unique; mais aucun rythme régulier d'espacement des examens n'est imposé.

Quant aux conclusions cliniques de cette analyse évolutive, les plus importants concernent la caractérisation de la zone D : on y lit anémie, thrombopénie, épreuve de Coombs positive, splénomégalie (grosse rate). La gravité de ces signes est ainsi affirmée : l'importance de la cytopénie était suggérée par certains médecins, mais n'avait pas été montrée aussi nettement. Sur la splénomégalie on peut tenter une thérapeutique symptomatique : e.g. irradiation de la rate.

2.2 Réponses hémodynamiques à la trinitrine intraveineuse * : La trinitrine (composé nitré de formule $C_3H_5(ONO_2)_3$, qui n'est autre que la nitroglycérine, explosif de base de la dynamite !) est un vasodilatateur puissant, d'action instantanée couramment prescrit *per os* dans le traitement de l'angine de poitrine. Depuis quelques années, on a également voulu administrer la trinitrine (ainsi que d'autres dérivés nitrés) par voie *intraveineuse* (I.V.) afin de réduire l'insuffisance cardiaque par vasodilatation et diminution des résistances des artérioles et des veinules. Toutefois la mise au point de cette thérapeutique a été très délicate, car à côté de résultats encourageants on observait des cas parfois graves de chute brutale de la pression artérielle dans de gros vaisseaux tels que l'artère humérale. Aussi trouvait-on sur la trinitrine I.V., des publications pleines d'espoir et d'autres décevantes quant à sa capacité d'augmenter le débit cardiaque; tandis que des chercheurs (notamment Gold et Williams aux E.U.), rapportaient des observations discordantes et tentaient de chiffrer par des épreuves statistiques l'efficacité globale obtenue sur un ensemble de malades.

On en était là quand S. Baligadoo (avec la collaboration statistique de M. Bureau) a pu le premier, grâce à l'Analyse des Données, distinguer selon l'état initial du malade, trois types principaux de réponse hémodynamique, et circonscrire ainsi l'ensemble des cas auxquels s'applique la thérapeutique nouvelle.

En bref, S.B. a démontré que le débit cardiaque n'est augmenté par la trinitrine que dans les cas où, à l'état basal, la pression capillaire pulmonaire CP est élevée, les résistances artériolaires RS sont élevées, l'indice cardiaque ** IC est bas.*** En bref quand les RS sont élevées, la trinitrine induit une diminution des résistances et favorise une augmentation du débit cardiaque : c'est le premier type de réponse, qui justifie la prescription de la T. Mais si à l'état de base les RS sont normales (2° type) ou basses (3° type) la trinitrine ne les abaisse pas davantage, le débit cardiaque n'est pas augmenté ; et dans le dernier cas (3° type) des accidents graves peuvent survenir.

* Les travaux de S. Baligadoo sur ce sujet sont publiés notamment dans la Nouvelle Presse Médicale n° 4 pp 283 sqq Janv. 1979 ; un exposé destiné aux statisticiens se trouve dans ENSI (Prat. de l'A. des D.) partie IV § 10.

** Cet indice, ou volume du sang chassé par le cœur en une minute, rapporté à la surface du corps, est considéré comme un bon indice global de l'oxygénation du malade.

*** L'influence de la valeur initiale de la pression capillaire pulmonaire sur le débit cardiaque était connue mais l'intérêt de la mesure de la valeur initiale des résistances n'était pas signalé.

Il est parfaitement logique, que dans le cas de résistances peu élevées, une vasodilatation (augmentation du diamètre des conduits) puisse aboutir à une baisse de pression telle que la circulation n'en est pas exaltée, mais au contraire ralentie.

Et ce fait maintenant confirmé, aurait pu être établi quatre ans plutôt si certains travaux d'auteurs étrangers avaient été analysés statistiquement suivant la voie nouvelle adoptée par S. Baligadoo.

Du point de vue méthodologique nous ferons deux remarques :

I On reproche parfois à l'A. des D. d'aboutir à des résultats qui pour le spécialiste étaient évidents *a priori*. Dans la présente étude, il s'agit de résultats évidents (même si la délimitation précise des types de réponse n'est aucunement évidente!) en ce sens que S.B. peut *a posteriori* les déduire des lois élémentaires de la physiologie circulatoire. Mais avant S.B. personne ne s'en était avisé! (à l'exception de l'étude de l'influence d'un paramètre sur la pression capillaire pulmonaire). Tout raisonnement est une suite d'évidences; la difficulté est de prendre ces évidences, de les choisir et de les ranger dans l'ordre logique jusqu'à la conclusion. Le statisticien ne peut faire cela par lui-même, mais dans la mesure où les données, les faits sont convenablement rangés, leur ordre suggère au médecin, l'ordre d'un raisonnement.

II Pour rapporter la réponse des malades à leur état de base, il faut construire et analyser des tableaux croisant les modalités des variables descriptives de l'état de base, avec les modalités des variables descriptives de la réponse; i.e. dénombrer combien de fois à telle modalité de base (e.g. pression capillaire pulmonaire supérieure à 22 mm Hg), correspond telle modalité de réponse (e.g. pression auriculaire droite diminuée de 1 à 2 mm Hg). Même si la correspondance entre les deux groupes de modalités n'est pas régie par une formule stricte, l'analyse révèle des associations prédominantes qui, comme des lois approchées, servent de maillons dans le raisonnement du médecin. C'est ce qu'a fait ici S. Baligadoo; et qu'il convient de faire à l'avenir dans toute analyse d'essai thérapeutique, telle que celle du pansement gastrique rappelée ci-dessus (§ 1.2). A la vérité la construction de tableaux à partir des données primaires est un art logique nouveau, qu'on ne peut enseigner en une conférence d'introduction; mais dont nous tenons à souligner l'importance.

2.3 Indications de la contreimpulsion par ballon intra-aortique (CPBIA)*:
Depuis quelques années, les malades atteints d'un infarctus myocardique (i.e. mauvaise irrigation sanguine du muscle cardiaque) aigu, compliqué d'une défaillance sévère de la pompe cardiaque peuvent bénéficier de certaines techniques d'assistance circulatoire mécanique, telles que la CPBIA. Cette technique consiste à introduire par l'artère fémorale jusque dans l'aorte thoracique un ballon de quelque 24 cm de long et d'un diamètre inférieur à celui de l'aorte. Le ballon est relié par un cathéter à une pompe d'animation externe qui en assure le gonflement et le dégonflement alternatif synchronisé sur l'électrocardiogramme du patient, de telle sorte qu'en résulte une meilleure oxygénation de tout le corps et du cœur lui-même, et une diminution du travail du ventricule gauche, donc des besoins en oxygène du muscle cardiaque.

Une telle thérapeutique nécessite l'intervention d'une équipe médicochirurgicale entraînée avec un matériel important. On ne recourt

* cf Cahiers Vol II n° 4 pp 415 sqq (1977); et ENS1, partie IV § 16. Le traitement statistique a été effectué au laboratoire du Pr. F. Grémy (groupe de recherche INSERM U 88) par J.P. Nakache et J. F. Chastang.

donc à la contrepulsion que si, sans elle, les chances de survie du patient paraissent réduites. D'où un problème de décision que P. Lorente a résolu par une règle de discrimination fondée sur l'a. factorielle.

Les données analysées sont issues des dossiers de 101 patients reçus à l'Unité de soins intensifs de l'hôpital Lariboisière, alors que la contrepulsion n'y était pas encore pratiquée. Le but de l'étude étant, si un nouveau patient se présente, de rapprocher immédiatement celui-ci de quelques uns des 101 cas de base, on choisit comme variables descriptives les résultats d'examens initiaux effectués à l'entrée des patients, soit 22 variables ou questions, totalisant 83 modalités (e.g. : présence ou absence d'infarctus dans les antécédents, 2 modalités ; pression aortique moyenne ; 3 modalités : inférieure à 70 mmHg ; de 70 à 81 ; supérieure à 81 ; etc.).

Dans le plan 1×2 on voit d'abord s'opposer suivant la direction de l'axe 1, les malades dont la fonction ventriculaire gauche est très atteinte, à ceux chez qui cette fonction se réalise à peu près. Une discrimination quasi parfaite (avec seulement quelque confusion dans la zone limite) apparaît ainsi entre les patients les plus atteints qui décèdent au cours du mois qui suit leur admission, et ceux qui survivent (du moins au-delà de ce terme).

D'autre part l'axe 2 s'interprète par l'état de la fonction ventriculaire droite. Cette fonction étant moins chargée que celle du ventricule gauche on conçoit qu'elle ne joue pas le premier rôle dans la discrimination des cas dont l'issue est fatale.

L'analyse dont nous venons de rendre compte présente un étalement suivant le pronostic pour 101 cas déjà traités. Or quand un cas nouveau se présente, on peut aussi l'insérer dans le plan 1×2 issu de l'analyse des 101 premiers cas (et cela, par un calcul très simple, qui peut se faire sans le secours de l'ordinateur : c'est ce qu'on appelle adjoindre un élément supplémentaire). Si le nouveau cas se trouve dans la zone où la survie est assurée sans contrepulsion, on n'applique pas cette technique. Si au contraire le cas tombe dans la zone de risque maximum (ou seulement dans la zone intermédiaire) on tente dans la mesure du possible de recourir à la CPBIA.

C'est ce qu'on a fait pour quelques dizaines de cas : et les résultats ont été conformes à ce que l'analyse des 101 premiers dossiers laissait prévoir. De façon précise voici les résultats d'une première statistique : 21 patients situés nettement dans la zone de survie ont reçu seulement le traitement médical classique et tous ont survécu (au-delà d'un mois). Pour 36 cas tombant dans la zone de décès, la CPBIA était indiquée ; 25 patients ont pu en bénéficier ; 9 d'entre eux ont passé le cap de la phase aiguë ; les 16 autres sont décédés avant un mois, ainsi que les 11 sujets qui n'ont pu bénéficier de l'assistance circulatoire indiquée dans leur cas.

Ainsi tout en acquérant par l'A. des D. une vision d'ensemble d'un domaine de la pathologie, on a résolu un problème de discrimination (entre deux classes de patients : d'une part ceux qui survivent grâce au traitement médical classique ; et d'autre part ceux qui décèdent) ; et donc un problème de décision (quant à l'indication d'une technique nouvelle d'application délicate : la contrepulsion).

3 Exploitation des données physiologiques et biologiques : Ici, le but poursuivi est double : d'une part en découvrant les liens qu'ont entre eux les résultats de l'ensemble des examens, (on dit encore : leur redondance) concevoir des protocoles simplifiés apportant au clinicien les mêmes informations que l'examen général, tout en dispensant le malade de certaines épreuves douloureuses ou dangereuses ! (ainsi dans la leucémie lymphoïde chronique (cf § 2.1), L. Degos a pu montrer

que ni la biopsie médullaire ni la lymphographie n'étaient indispensables). D'autre part, extraire des données de la clinique ou des expériences de laboratoire, des dimensions et des relations nouvelles indiscernables avant l'analyse multidimensionnelle.

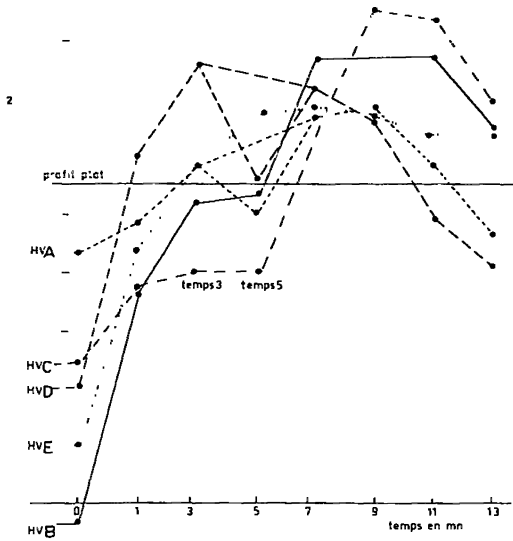
3.1 Electrocardiographie (ECG) : Ici, je me bornerai à des perspectives de recherches sans rendre compte de résultats précis sur l'ECG. On sait que cette technique consiste à enregistrer les variations temporelles de l'activité électrique du coeur. Dès 1903, Einthoven interprète les différences de potentiel mesurées entre des électrodes placées en divers points de la surface du corps, comme si ces D, de P, étaient créées par le champ d'un dipole, ou système de deux charges électriques opposées situées sur un axe de part et d'autre du centre du coeur. De ce point de vue, les courbes de l'électrocardiogramme ne sont que les projections d'un dipole (variable au cours du temps) vu sous des angles différents : on enseigne classiquement à construire le vecteur représentant ce dipole, au moins en projection sur le plan frontal. Mais pratiquement, le cardiologue ne recourt pas à ces constructions, qu'il juge longues et délicates : il interprète l'ECG par inspection simultanée des courbes tracées. Cependant, depuis une quinzaine d'années, les ressources nouvelles du traitement électronique des informations, permettent de présenter directement un vecteur (dont les trois composantes spatiales sont calculées avec précision) : c'est ce qu'on appelle l'électrovectorcardiographie, ou en bref la vectographie (ECVG).

Cette nouveauté a été diversement accueillie : l'ECVG a ses enthousiastes ; tandis que beaucoup de cardiologues compétents, formés avant l'ECVG, affirment que toute l'information contenue dans l'ECVG leur est déjà accessible sans qu'il soit besoin d'une présentation vectorielle. Du point de vue de la physique, on remarquera que l'assimilation du coeur à un dipole ne peut être qu'une première approximation : au-delà du moment dipolaire (vecteur), on doit considérer pour tout système de charges électriques son moment quadrupolaire (tenseur), etc. Il est possible qu'un cardiologue acquière, notamment d'après les dérivations précordiales, des informations qui, qualitativement au moins, vont au-delà de la représentation dipolaire (vectorielle).

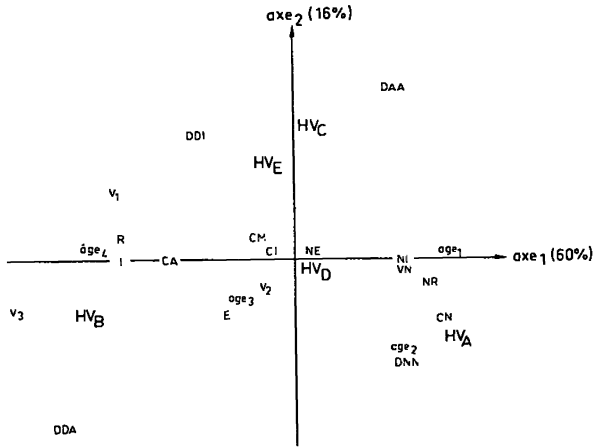
Soumettre directement au médecin les composantes d'un tenseur n'est pas facile. Mais... ici peut intervenir l'A. des D. : certains affirment qu'en analysant des électrocardiogrammes adéquatement décrits par des cardiologues, ils ont obtenu des facteurs équivalents aux résultats qu'on acquiert présentement à grand prix par une angiographie coronarienne. Je puis seulement annoncer ces recherches qui sont en cours.

3.2 Modification de la conduction auriculo-ventriculaire après injection d'Ajmaline * : On sait que le muscle cardiaque, plus particulièrement le système cardionecteur (ou partie du myocarde, histologiquement différenciée et chargée normalement des fonctions d'excitation et de conduction) est le siège de phénomène de conduction (dans le sens oreillette-ventricule) se manifestant par des potentiels électriques enregistrables. On désigne sous le nom général de bloc, les troubles de cette conduction, (ralentissements, interruptions...) dont certains, extrêmement graves, exposent le sujet à une mort subite. On considère ici l'influence d'une drogue, l'Ajmaline, provoquant normalement l'allongement des différents espaces (i.e. délais temporels) de la conduction auriculo-ventriculaire, mais avec, suivant les individus, une diversité de réponses, permettant dans certains cas d'affirmer qu'un sujet est susceptible de présenter ultérieurement des phénomènes de blocs les plus graves (ce qui justifierait chez ce sujet l'utilisation d'un stimulateur électrique du coeur, ou "pace maker").

* Cf Dentzeau, Lellouch, Baligadoo, Haiat et Chiche ; in Ann. Cardiol. Angéiol., T. 26 (n°2), pp 109-115 (1977). Les graphiques illustrant le § 3.2 proviennent de cet article.



— Etude statistique : les différents profils évolutifs de l'espace HV sous a)maline.



— Etude statistique répartition des profils de HV sous a)maline avec les différents paramètres cliniques.

- 1) Cardiopathie : cœur normal (CN), ischémique (CI); Infarctus du myocarde (CM); autres (CA)
- 2) insuffisance cardiaque (I), sans insuffisance cardiaque (NI)
- 3) Volume cardiaque: normal (VN), V1, V2, V3
- 4) Traitement : rien (DNN); digitaline(DDI); digitaline + anti arythmique (DDA); autres (DAA); diurétiques (R); sans diurétiques (NR)
- 5) Sans bloc ou hémibloc de branche (NE); avec bloc de branche (E).
- 6) Age : entre 20 et 55 ans (Age 1, 2); entre 55 et 80 ans (age 3, 4).

S. Baligadoo et coll. ont pu étudier en détail la réponse de 45 sujets à une injection intraveineuse de 100 mg d'Ajmaline. Quant à l'analyse des données, nous concentrerons notre attention sur une seule des variables de conduction étudiées : l'espace HV (ou temps séparant le début de la déflexion Hissienne, du début de l'onde ventriculaire), mesuré chez chaque sujet (en enregistrant les potentiels gauche à une sonde) d'abord à l'état basal, puis après injection d'A, à des temps successifs séparés de celle-ci par 1 minute, 3', 5', 7', 9', 11' et 13'.

Ainsi l'information recueillie prend la forme d'une courbe de variation de l'espace HV en fonction du temps, courbe définie par 8 points. Ces courbes (cf fig.) sont très diverses : et on conçoit *a priori* qu'elles doivent l'être. Car dans une courbe de réponse, il y a d'abord un temps de latence plus ou moins long précédant la réponse proprement dite ; puis une phase de croissance plus ou moins prononcée ; l'arrivée au maximum, un palier éventuel avant la décroissance et le retour à la normale ; à moins qu'il n'y ait une tendance asymptotique vers un niveau supérieur au niveau initial.

On voit combien une modélisation mathématique, forcément primitive à moins qu'elle ne soit trop complexe, est ici hors de mise. Ce qu'il faut faire, c'est accepter les courbes telles qu'elles sont et les ranger d'après les similitudes qu'elles ont entre elles : on reconnaît ici le point de vue général de l'analyse des données. C'est ainsi qu'a procédé S. Baligadoo ; et il a reconnu 5 types de courbes qu'il a appelées A, B, C, D, E. Il ne faudrait pas voir dans ces 5 types des entités définitives, nettement délimitées : ce sont seulement des représentants adéquats auxquels on peut rattacher chacune des 45 enregistrées dans l'étude.

A partir de cette typologie des courbes, S. Baligadoo a créé un nouveau tableau de données, mettant en correspondance les cinq types A, B, C, D, E avec un certain nombre d'informations cliniques caractérisant l'état du cœur, le traitement éventuel (e.g. par la digitaline), sans oublier l'âge du patient.

D'après le plan 1 x 2 issu de l'analyse de ce tableau, on voit notamment que le type A qui correspond à une réponse modérée à l'Ajmaline est associé à un cœur normal, sous traitement, avec âge moyen (classe 2) etc. . Au contraire les courbes B, (pour lesquelles l'augmentation du temps de conduction, peut à son maximum aller jusqu'au doublement du temps de base), correspondent aux anomalies les plus graves. Toutefois, dans certains cas, le doublement de l'espace HV, ne semble explicable par aucun facteur pathologique ; en sorte que S. Baligadoo et coll. concluent que (contrairement à ce qu'on avait supposé) ce signe n'est pas d'une fiabilité suffisante pour permettre à lui seul de poser un diagnostic.

3.3 La morphologie des leucocytes : Le dénombrement des éléments figurés du sang (globules rouges, ou hématies ; et globules blancs, ou leucocytes parmi lesquels il importe de distinguer divers types) aboutit à la formule sanguine ou hémogramme : c'est un examen d'usage courant, pratiqué en laboratoire en observant au microscope le sang convenablement préparé.

Certaines modifications de la formule sanguine résultent de maladies bénignes ; mais dans d'autres cas sont en cause diverses formes de leucémies. Parmi celles-ci on distingue notamment des leucémies lymphoïdes chroniques (LLC : cf *supra* § 2.1) des leucémies aiguës lymphoblastiques (LAL) ; etc. . Mais ces grands types LLC, LAL, ... entre lesquels la discrimination n'est pas toujours facile, sont loin d'être homogènes et les hématologues cherchent à y reconnaître des subdivisions.

Dans ce but on s'attache notamment à automatiser et à perfectionner l'observation des leucocytes. C'est ainsi que l'hôpital Saint-Louis, dispose d'une chaîne de traitement comprenant l'étalement du sang par centrifugation, la coloration automatique, l'observation optique par un microscope asservi, avec analyse sur ordinateur de l'image digitalisée; ce qui d'une part fournit automatiquement un hémogramme classique, et d'autre part donne sur chaque leucocyte un ensemble de mesures telles que surface du noyau, surface du cytoplasme, rapport du carré du périmètre à la surface, indice de proximité du noyau au bord de la cellule, et absorption du noyau et du cytoplasme pour les radiations bleues et les radiations jaunes.

Ces paramètres morphologiques, ne sont pas en eux-mêmes inaccessibles à l'observateur humain (non armé d'un ordinateur) mais il lui est à peu près impossible d'en faire la synthèse. L'oeil entraîné reconnaît bien le noyau polylobé d'un polynucléaire, ou les granulations du cytoplasme d'un granulocyte convenablement colorés. Mais l'observateur ne remarquera pas sur une plaque un pourcentage anormal de cellules de forme oblongue ou dont le noyau occupe une surface relative supérieure à la moyenne. L'importance d'une gradation continue de mesure est pourtant déjà reconnue en hématologie : par exemple on construit pour les hématies une courbe dite de Price-Jones qui sert à caractériser l'anisocytose, i.e. la dispersion anormale du diamètre de ces cellules (e.g. 4μ à 10μ ; alors que la normale est de $6,2\mu$ à $8,2\mu$). En terme statistique cette courbe de P.-J. n'est autre que l'histogramme des tailles des hématies. Mais pour les leucocytes, avec e.g. 8 mesures par leucocyte la notion usuelle de formule sanguine, et celle d'histogramme ne suffisent pas : il faut passer par une analyse multidimensionnelle pour restituer au cytologiste une synthèse des informations recueillies d'abord automatiquement.

Ce sont de telles analyses que Ch. Bastin a effectuées pour G. Flandrin. Et voici un aperçu des premiers résultats obtenus * tels que les a expliqués G. Flandrin.

1°) Quant à la corrélation des paramètres de cellules individuelles on retrouve des données classiques de cytologie. Les cellules à cytoplasme peu étendu sont celles dont le noyau est le plus volumineux, et réciproquement. Un noyau volumineux est relativement peu absorbant (la chromatine y est plus claire); ...

2°) Quant à ce qu'on peut appeler la *formule leucocytaire morphologique* (réalisée d'après les mesures prises sur 100 cellules par sujet) on retrouve d'abord une bonne discrimination entre LLC et LAL ; puis viennent des résultats originaux. D'une part, les formules des sujets sains (étroitement groupées sur le graphique) apparaissent beaucoup moins dispersées que celles des sujets atteints de leucémie (qui s'étalent sur tout le plan). D'autre part s'isolent nettement les formules des sujets atteints d'une variété de LAL, singulière du point de vue immunologique et ne bénéficiant présentement d'aucune rémission par la chimiothérapie : les leucémies lymphoblastiques à cellules de Burkitt.

Ainsi l'on peut attendre que la *formule leucocytaire morphologique* contribue à un progrès nosologique.

3.4 Facteurs chimiques du pouvoir mutagène révélé par l'épreuve d'Ames:

Cette épreuve caractérise les substances chimiques par leur action sur des souches bactériennes appartenant au genre de Salmonelles. On sait en effet qu'en France et dans le monde une législation s'instaure pour

* Dans cette rédaction, nous précisons notre exposé oral d'après l'enregistrement de la communication de G.F. . Pour un exposé des travaux ultérieurs, cf [TYP. LYMPH.].

surveiller l'introduction sur le marché de substances à risques cancérigènes. L'analyse dont je vais parler rentre dans le cadre de cette surveillance : mais on conçoit que ceux qui y procèdent ne désirent pas la divulguer par crainte d'alarmer le public. Je puis vous dire toutefois qu'on a étudié sur plusieurs souches de Salmonella, le pouvoir mutagène de quelque 20 substances de structure chimique analogue.

Les résultats de ces essais peuvent être rangés en un tableau rectangulaire dont les lignes sont les produits chimiques, et les colonnes correspondent aux modalités d'effet ou d'absence d'effet sur chacune des souches considérées (donc $2n$ colonnes, s'il y a n souches). L'analyse de ce tableau montre en bref qu'il y a 2 types de souches : nous dirons A et B ; et quatre types de produits distingués par la présence ou l'absence d'effets sur les souches des types A et B : produits AB agissant sur à peu près toutes les souches ; produits Ab agissant sur A non sur B ; produits aB, actifs sur B seul ; et produits ab non mutagènes, au moins pour les souches étudiées, car il est à craindre que des potentialités mutagènes restent à révéler par d'autres épreuves !

Ceci posé on a cherché les déterminants chimiques qui font qu'une substance est de type AB, Ab, aB ou ab. Pour cela les chimistes se sont livrés à un travail de mécanique quantique considérable : ils ont caractérisé la distribution des charges du nuage électronique sur les atomes qu'on suppose constituer le site mutagène. Ces données physiques ont été croisées avec les propriétés de mutagénèse, les réponses des souches. L'analyse a montré une bonne discrimination des produits des quatre types, fondée uniquement, répétons-le sur des propriétés physiques.

Encouragés par ce succès, les chimistes ont calculé les configurations électroniques d'une dizaine d'autres molécules du même type, et on a pu vérifier sur ces cas nouveaux que le type mutagène était également bien prédit d'après la structure. Remarquons que formellement la chaîne des opérations statistiques (analyse, discrimination, prédiction) est la même que dans l'étude sur la contreimpulsion (§ 2.3) présentée par P. Lorente.

Prédire les propriétés biologiques à partir des propriétés chimiques (plus exactement de la structure électronique) nous paraît offrir de grandes perspectives dans la chimiothérapie du cancer. Présentement en effet, la chimiothérapie procède par tâtonnements, sans être guidée par des principes (les tâtonnements étant d'autant plus lents que le seul critère d'efficacité est la survie des malades qui - fort heureusement - s'accroît et ne peut donc être évaluée rapidement). Les principes apparaîtraient sans doute si, d'une part, on faisait une typologie des effets biologiques des substances utilisées (d'après une suite d'épreuves sur bactéries, sur cultures de tissus, sur animaux vivants...); et que d'autre part on mettait ces effets biologiques en rapport avec la structure.

4 Le sujet sain : Faire la typologie des sujets sains est un problème posé depuis Hippocrate. Or bien que des différences entre sujets soient admises, la thérapeutique en tient assez peu compte. On désirerait avoir une batterie d'épreuves nous apprenant comment les sujets répondent aux maladies et aux drogues et aussi sur quelle pente dangereuse ils sont engagés, par exemple: quel est pour chacun le risque d'accident cardiovasculaire.

4.1 Analyses de bilan de santé : Dans cette voie je rendrai compte d'une analyse très suggestive effectuée sur les bilans de quelque

20.000 sujets sains * . Les données prises en compte sont classiques (cholestérolémie, glycémie, uricémie ; hémoglobine ; phosphatases alcalines, spirométrie ; pression artérielle systolique ; et aussi surcharge pondérale et tabagisme). Sur le 1° axe est apparu un échelonnement des sujets fortement corrélés à leur âge. Mais en réalité ce 1° facteur, peut être considéré lui-même comme une définition biologique de l'âge** (ou du moins il suggère qu'une telle définition est accessible... il en est une préfiguration...). Il s'impose donc de comparer cet âge fictif calculé d'après les données du bilan de santé, à l'âge réel du sujet. Si par exemple un sujet de 40 ans se trouve sur le graphique entouré de sujets dont l'âge moyen est 80 ans, on en pourra conclure que ce sujet de 40 ans est précocement usé, ce qui justifie une surveillance particulière. Il semble donc utile de perfectionner la batterie d'examen servant à définir cet âge biologique. A ce propos le Dr. J.L. Parier (directeur scientifique d'IPSEN) nous a signalés les travaux du Pr. Bouissou et coll. (la *Nouvelle Presse Méd.* 1975, 4, n° 44 ; p. 3144) d'où il résulte que le taux des stérols cutanés est un bon indicateur, relativement accessible, de l'état des artères, du stade athéromateux.

4.2 Les variables lentes : Ces variables sont intermédiaires entre les caractères typologiques qui sont des constantes du sujet, et les variables biologiques décrivant son état apparent momentané. Voici comment nous avons rencontré ce problème dans de nombreuses analyses de données. Supposons e.g., qu'il s'agisse de malades présentant des syndromes inflammatoires d'origines diverses. Le médecin applique immédiatement une thérapeutique anti-inflammatoire qui dans une large mesure peut faire disparaître le syndrome. Cela ne signifie pourtant pas toujours que la guérison soit obtenue. On désirerait donc connaître l'état profond du malade, mesurer des variables lentes, c'est-à-dire des paramètres biologiques sur lesquels la thérapeutique n'a pas immédiatement prise mais qui sont de bons indicateurs de l'état réel du malade. C'est là un problème inspiré des recherches de R. Thom, sur lequel a attiré notre attention***, un éminent statisticien, un grand ingénieur, décédé le 13 Mai 1980 : Robert Gibrat.

4.3 Thèmes divers : Nous signalerons brièvement quelques thèmes de recherche se rapportant à divers titres au sujet sain.

4.3.1 On découvre aujourd'hui que le type H.L.A. (en bref caractères génétiques, observables pour les antigènes propres à la paroi des leucocytes) est en rapport avec le risque de certaines maladies : il y aurait d'amples données à recueillir et à analyser.

4.3.2 Dans l'entraînement des sportifs, se conjuguent choix des exercices, caractéristiques biologiques et performances : des données s'y rapportant ont été analysées par J.P. Fouillot et F. Tékaïa (cf *Cahiers Vol IV n° 2 pp 231 sqq ; 1979*).

4.3.3 Est à notre programme une intéressante étude de F. Picard qui concerne la réponse physiologique au régime alimentaire, non de sujets humains, mais de poules pondeuses.

* Ces bilans sont ceux des Investigations Pré-Cliniques (I.P.C.) organisée que dirige le Dr. J-R Debray de l'Institut. Ces analyses ont été effectuées par F. Benjelloun, sous la direction de L. Guize.

** Dans d'autres analyses on a, en bref, analysé l'écart des données relatives à chaque sujet par rapport à la moyenne de sa classe d'âge, ce qui donne directement une fonction s'interprétant avec l'excès de l'âge physiologique sur l'âge réel.

*** Cf Causalité et Analyse des Données en Médecine : Conférence aux secondes journées internationales A. des D. et Informatique, IRIA, Oct. 1979.

5 Médecine et statistique

5.1 Introduction à la représentation synthétique des données multidimensionnelles

: Tel est le titre d'un exposé * figurant au programme du présent séminaire. On ne tentera pas d'exposer aux médecins la description d'un programme d'analyse des données, avec les méthodes mathématiques qu'on y applique. Mais prenons une analogie. Vous savez tous que $2 \times 2 = 4$, donc $2 = \sqrt{4}$; et $3 \times 3 = 9$, donc $3 = \sqrt{9}$; quant à calculer $\sqrt{15}$, vous prenez dans votre poche l'instrument approprié et appuyez sur quelques touches... C'est un peu comme cela que vous procéderez en A. des données. Le médecin doit savoir ce que signifie la proximité de deux points sur un graphique; il devra apprécier la qualité de la représentation fournie sur un plan; prendre garde, e. g. qu'il y a une troisième dimension de profondeur (et d'autres au-delà) non figurée sur ce plan. Qui sait cela, est capable de bien utiliser l'analyse des données, même si quant aux opérations statistiques proprement dites, il ne fait qu'appuyer sur une touche. Il reste, et c'est le véritable intérêt de l'analyse des données, qu'il est plus difficile de choisir les tableaux à analyser (les individus à écarter; ceux à adjoindre en éléments supplémentaires; etc.), que de décider d'extraire la racine d'un nombre. Comme je vous l'ai dit (§ 2.2 *in fine*) la construction des tableaux est un art logique nouveau, auquel ce séminaire devrait être pour vous une introduction.

5.2 L'exemple choisi : Dans cette introduction, on partira d'un exemple de données très schématique, mais qui, je le signale en passant, a pour origine une étude réelle importante. Il s'agit de la pénétration des médicaments. L'étude réelle considère plusieurs milliers de médicaments et plusieurs centaines de maladies délimitées suivant une nomenclature de l'O.M.S. : est recensé le nombre de fois que chaque médicament a été prescrit pour chaque maladie. On aboutit à réduire l'ensemble des médicaments à quelque 200 types. Pour le laboratoire pharmaceutique qui a demandé l'analyse (et n'entend pas la publier) il s'agit d'une étude de marché susceptible de révéler des zones de la pathologie où un produit nouveau bien conçu pourra pénétrer, sans rencontrer de concurrents trop nombreux... C'est aussi un moyen de confronter les indications des médicaments prônées par les laboratoires et publiées dans le fameux dictionnaire de Vidal, avec l'usage effectif des médecins (fruit de leur expérience? ou de leurs préjugés?). Le tableau qui vous est présenté, ne concerne quant à lui que 7 maladies infectieuses et 6 antibiotiques; et il a été construit sans dépouillement statistique, en se fiant seulement à la vraisemblance.

5.3 Le verdict de Claude Bernard : Je terminerai sur une question en forme de slogan : Pour ou Contre la statistique! Et citerai un grand Maître, sous l'ombre duquel ont été élevés nos pères.

Il y a dans l'*Introduction à la Médecine Expérimentale*, quelques phrases cruelles pour la statistique... "*En un mot, en se fondant sur la statistique, la médecine ne pourra être jamais qu'une science conjecturale; et c'est seulement en se fondant sur le déterminisme expérimental qu'elle deviendra une science vraie, c'est-à-dire une science certaine...* En effet, il suffit qu'un phénomène se soit montré une seule fois avec une certaine apparence, pour admettre que dans les mêmes conditions il doive se montrer toujours de la même manière. Si donc il diffère dans ses manifestations, c'est que les conditions diffèrent".

On reconnaît ici l'extraordinaire sagacité et aussi la persévérance de Claude Bernard; qui, e.g., ayant réussi du premier coup à rendre un lapin diabétique au moyen de la piqûre du plancher du quatrième ventricule, se heurte

* cf F. Benzécri : [INT. CORR. MED.]

ensuite à huit ou dix échecs ; mais sans songer à nier sa première expérience positive, poursuit jusqu'à parvenir à fixer le lieu précis qui convient pour que la piqûre entraîne inmanquablement le diabète. Et il affirme ailleurs sans complaisance :

"Ainsi dès que les circonstances de l'expérience sont bien établies, on ne peut plus faire de statistique : on n'ira pas par exemple rassembler les cas... ; pour savoir combien de fois il arrivera qu'en coupant le nerf sciatique on ait la paralysie des muscles auxquels il se rend. Les effets arriveront toujours et nécessairement, parce que la cause du phénomène est exactement déterminée".

Et voici un trait d'ironie, mais sur le ton du XIX-ème siècle, un ton qui n'est plus le nôtre !

"Une forme d'application très fréquente des mathématiques à la biologie se trouve dans l'usage des moyennes, ou dans l'emploi de la statistique, qui, en médecine et en physiologie, conduisent pour ainsi dire nécessairement à l'erreur... . Le sublime du genre a été imaginé par un physiologiste qui, ayant pris de l'urine dans un urinoir de la gare d'un chemin de fer où passaient des gens de toutes les nations, crut pouvoir donner ainsi l'analyse de l'urine moyenne européenne".

Mais ce que nous poursuivons en A. des D. ce n'est pas seulement de définir un point moyen (où peut-être sur le graphique aucun sujet, aucune modalité ne se trouve !) mais d'interpréter la dispersion, l'écart autour de ce point. Ce à quoi nous vous invitons, ce sont d'abord des rapprochements, des comparaisons ; pour ensuite raisonner par vous-mêmes. Varier, en multipliant les analyses, la présentation des données déjà acquises, c'est se livrer par la pensée à une expérimentation clinique prolongée. Répétons-le : si l'on a imposé aux observations un format assez précis, les cas anciens, provenant d'études rétrospectives peuvent se mêler aux cas nouveaux, leur servir de témoin ; l'expérimentation étant ainsi allégée et accélérée pour le bien des malades...

Certainement Claude Bernard n'épargnerait pas ses sarcasmes à cet abus systématique des moyennes indistinctes sur les essais thérapeutiques en vogue présentement. Mais peut-être, malgré son aversion pour la statistique, vous ferait-il crédit d'avoir recours à l'analyse des données, pour cerner les diverses apparences des phénomènes, et en définir les conditions.