

J.-P BENZÉCRI

F. BENZÉCRI

**Compte rendu d'une discussion entre
informaticiens et spécialistes du sida, après
un exposé d'analyse des données**

Les cahiers de l'analyse des données, tome 15, n° 3 (1990),
p. 305-322

http://www.numdam.org/item?id=CAD_1990__15_3_305_0

© Les cahiers de l'analyse des données, Dunod, 1990, 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/>

COMPTE RENDU D'UNE DISCUSSION ENTRE INFORMATIENS ET SPÉCIALISTES DU SIDA, APRÈS UN EXPOSÉ D'ANALYSE DES DONNÉES

[DISCUSSION PASTEUR]

rédigé par J.-P et F. BENZÉCRI.

Le mercredi 20 Juin 1990 se tint, dans la Salle Chamberland de l'Institut PASTEUR, une réunion de réflexion consacrée à la Recherche de facteurs pronostiques dans l'évolution du SIDA: après un exposé d'analyse de données vint la discussion dont on rend compte ici. Dans la note de F. Tekaïa, "Analyse statistique et recherche des facteurs pronostiques dans la pathogénèse du SIDA", le lecteur trouvera successivement: deux pages d'histoire bibliographique de la collaboration entre statisticiens et pastoriens; une description de la base de données 'CONSULT'; et une brève présentation des protagonistes de la discussion.

Dans la présente rédaction, les propos attribués aux médecins ont été soumis à la vérification des participants, mais rédigés par des statisticiens, à l'usage de statisticiens et au niveau de compétence de ceux-ci. On notera que les protagonistes sont désignés par leur profession suivie des initiales de leur nom. Pour la commodité du lecteur, on a introduit quelques titres de §; ces titres ne correspondent qu'approximativement au contenu des phases successives de la discussion. Sans s'interdire de déplacer ou résumer quelques propos, les rédacteurs n'ont pas voulu composer a posteriori un exposé strictement ordonné; car la ligne réfractée de la pensée est pour beaucoup dans l'intérêt d'un dialogue. Mais l'on a introduit quelques compléments d'après des travaux effectués après la discussion et, en partie suggérés par celle-ci.

1 Les stades et leur enchaînement

Médecin (Jean-Louis Virelizier): La cohérence du tableau que nous présentent les graphiques ne résulte-t-elle pas d'une sorte de pétition de principe dans la mesure où l'on retrouve la répartition des stades après avoir introduit l'information relative à ceux-ci?

Statisticien (Jean-Paul Benzécri): Il n'y a certainement pas de pétition de principe; car dans la première analyse, fondée sur les individus (plus précisément les observations), l'information relative aux stades n'est prise en compte ni directement ni même indirectement (la variable *virémie* étant absente, ainsi qu'*antigénémie*); cette information est seulement introduite après l'analyse, pour placer les centres de gravité des classes d'individus rentrant dans chacun des stades. Je répète que les axes sont construits et que l'on y place les modalités des variables aussi bien que les points figurant les consultations individuelles indépendamment de toute référence aux stades.

Au demeurant, je vous rappelle que la deuxième analyse, fondée sur le tableau des cumuls par stades pris comme tableau principal, fournit des facteurs 1 et 2 qui coïncident presque avec les facteurs 1 et 3 issus de la première analyse (le facteur 2, dont on a discuté l'interprétation, disparaissant seul). Cette 2-ème analyse, même si elle est moins probante que la première, ne peut elle-même être taxée de pétition de principe, dans la mesure où la notion de stade est introduite, mais non l'ordre des stades que propose l'analyse.

Immunologiste (Jean Michel Claverie): Or on constate d'abord que l'enchaînement des stades de 2 à 14 apparaît d'autant moins nettement que les sous-nuages de cas individuels afférents aux stades sont étalés et se recouvrent mutuellement. Cependant, les centres de ces sous-nuages ne coïncident pas avec le centre de gravité du nuage global: ils sont dispersés. Ce qui pose une première question:

Si dans la maladie, constamment, les mêmes pressions, produisant les mêmes effets, s'exerçaient pour détériorer l'état du patient, n'y aurait-il pas identité de la réponse physiologique, au lieu de l'étalement constaté?

Stat (JPB): Qu'on me permette de réfléchir ici sur l'origine et la nature de la notion de stade dans la classification **HIP**.

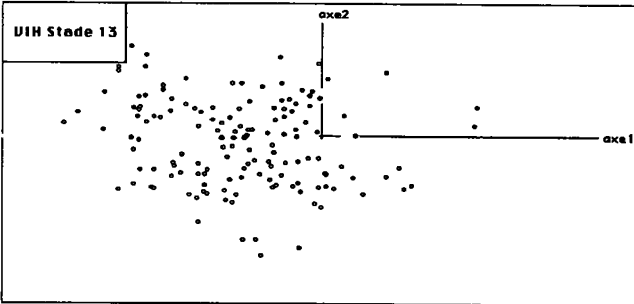
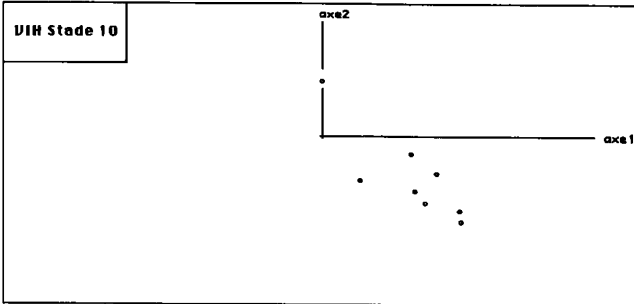
La classification en 14 stades a été créée, à l'Institut Pasteur, dans le but de rendre plus précise la nomenclature en 4 niveaux du **CDC** d'Atlanta. Cette nomenclature a le mérite de permettre aux médecins de se communiquer des statistiques sur la gravité de l'état des patients recensés et leur évolution. Elle n'est certainement pas assez fine pour suivre le progrès du mal dans un cas individuel. En particulier, le **CDC** se fonde sur des notions cliniques, à l'exclusion des variables biologiques dont on peut attendre beaucoup pour mesurer la réponse des patients aux traitements essayés.

Mais qui entreprend de créer, comme on l'a fait à Pasteur, une classification plus fine que celle du **CDC**, est soumis à de très fortes contraintes. Il est exclu, pour l'instant de faire autre chose que de subdiviser les classes du **CDC**. Car créer des classes enjambant sur celles du **CDC** rendrait impossible la communication.

Il y a plus: la notion même postulée d'enchaînement des stades doit être mise en rapport avec la réalité du développement clinique. La clinique ne montre pas une succession de stades par lesquels passeraient tous les malades, voire en parcourant à une vitesse inégale les intervalles séparant les stades successifs. Si l'on peut faire image, on dira que tous les malades partent d'un état initial asymptomatique, pour aboutir à la maladie sous sa forme déclarée et à la mort, mais que le chemin, le parcours entre les deux points extrêmes est différent pour chacun.

Une discussion s'engage alors entre clinicien (PhS) et statisticien (JPB) sur ce qu'on doit entendre par *enchaînement des stades* et *parcours clinique des malades*. Le médecin affirme d'abord la validité d'un schéma général:

Méd (PhS): Il y a un parcours qui est celui que tous les malades effectuent.



Le statisticien proteste en prenant pour argument l'inégalité des effectifs afférents aux divers stades dans l'échantillon analysé. Certains stades:

{8, 10, 11, 12},

fournissent au plus chacun deux dizaines d'observations; alors que d'autres, comme le 4, en fournissent des centaines. À l'appui de cette assertion, sont présentés quelques sous-nuages de stades; graphiques dont la très inégale densité est frappante.

(Le lecteur trouvera de tels graphiques dans [ANA. CLIN]).

Le médecin précise en proposant un parcours simplifié qui lui paraît convenir pour tous les malades.

Méd (PhS): Le patient est d'abord asymptomatique; puis il est en butte à un certain nombre d'affections; jusqu'à parvenir au stade terminal où les affections ne lui laissent plus de répit. La seule véritable exception à ce schéma simplifié est que certaines affections gravissimes (sarcome de Kaposi ou complications neurologiques) peuvent survenir d'emblée chez un patient asymptomatique et le conduire aux formes les plus graves de la maladie.

Le statisticien prend acte de ce schéma simplifié avec exceptions, mais rétorque:

Stat (JPB): Au niveau, du moins, des variables soumises à l'analyse, le tableau n'est pas celui d'une succession de stades, dans la mesure où, par exemple, aux stades 4 et 7 correspondent des sous-nuages qui se recouvrent presque exactement; ce qui implique que, quant aux variables considérées

(numérations de la formule sanguine et dosages de fractions des globulines), les deux stades désignent des réalités identiques...

Méd (PhS): Mais le 4 et le 7 sont une même chose!

Stat (JPB): Et le 5 et le 6, quant à eux, au contraire, vont avec le 2 et le 3.

Le médecin (PhS) acquiesce et précise:

Méd (PhS): La seule caractéristique propre à {5, 6} relativement à {2, 3}, ou à 7 relativement à 4 est l'adénopathie définie comme la présence dans deux territoires de ganglions dont le diamètre dépasse 1cm. Il s'agit là d'un seuil arbitraire. Pour une très faible différence de diamètre, deux patients se trouveront placés l'un dans 4 l'autre dans 7 (ou l'un dans 3, l'autre dans 5), sans qu'il y ait entre eux de différence clinique perceptible.

Médecins et statisticiens s'accordent sur la classification réduite suivante:

- {2, 3, 5, 6} = porteurs asymptomatiques;
- {4, 7} = détérioration du système immunitaire;
- {9, 13, 14} = états cliniques les plus graves;

les stades {8, 10, 11, 12} étant peu représentés.

Et le statisticien conclut:

Stat (JPB): la distinction entre 4 et 7, ou entre {2, 3} et {5, 6}, est héritée de la classification du CDC, qui fonde sur le critère d'adénopathie sa distinction entre classes II et III.

Le statisticien propose alors une transition entre le point de vue de l'estimation de l'état du système immunitaire et celui de l'histoire clinique.

Stat (JPB): Quant au système immunitaire, l'accord s'est fait sur un schéma triangulaire. Mais pour la clinique, au contraire, la survenue d'un sarcome de Kaposi ou de complications neurologiques gravissimes, ou simplement la survenue d'une infection opportuniste qui cède plus ou moins à un traitement classique, est un événement essentiel que tout malade redoute. Et de l'enchaînement de ces événements, du rythme auquel ils se présentent, résulte ce qu'on peut appeler une histoire clinique. Or on possède déjà, dans les dossiers, un nombre suffisant de telles histoires pour en tenter une typologie.

De même que l'état instantané du système immunitaire est, en quelque sorte représenté (après analyse multidimensionnelle) comme un spot sur un écran, l'histoire clinique est la trajectoire de ce spot; trajectoire ponctuée d'incidents cliniques. L'analyse factorielle et la classification automatique permettraient, en bref, de déduire de l'étude d'une centaine de parcours réels une dizaine de parcours-type, à l'un desquels tout cas concret se conformerait à peu près.

Le statisticien a souligné ici l'objectivité de sa méthode, objectivité garantie par le fait qu'étant ignorant en clinique, il ne peut ordonner les données suivant des vues personnelles: or, quant aux stades, il a pourtant abouti à des conclusions claires qui sont aussi celles du médecin.

Mais à la perspective d'une typologie des trajectoires de malades, un clinicien proteste:

Ce n'est pas si simple!

Et voici le tableau qu'il propose.

Méd (JLV): Un patient peut rester asymptomatique pendant deux, cinq, voire huit ans. Cependant, le patient est soumis à une évolution latente; son état potentiel se détériore. Cette détérioration se manifesterait tôt ou tard par des infections opportunistes; et il sera alors assez rapidement entraîné, après une sorte de rupture, vers le stade ultime de la maladie.

Le statisticien ne se déclare aucunement dérouté par ce schéma clinique. À l'objection du médecin il répond par une proposition faite *more geometrico*, en termes mathématiques.

Stat (JPB): Puisque, cliniquement, les patients des stades 4 et 7 sont asymptomatiques, et que l'on présume que leur état potentiel n'en est pas moins en lente détérioration, l'on peut, d'une part, suivre la variation de cet état comme la variation continue au cours du temps d'un point figuratif sur notre diagramme (ou un autre analogue, fondé sur des données complémentaires); et, d'autre part, constater si les incidents graves, ou plutôt les catastrophes, qui frappent les malades, surviennent, ou non, plus fréquemment lorsque ceux-ci ont un point figuratif correspondant à un état déjà fortement détérioré du système immunitaire. Autrement dit, en plaçant, sur la zone que couvrent les sous-nuages des stades 4 et 7, les points où était parvenu chaque malade lorsque la catastrophe s'est déclanchée pour lui; on verrait si ces points sont distribués uniformément sur cette zone très étendue, ou si, au contraire, les catastrophes frappent avec une densité beaucoup plus grande des patients que la détérioration de leur système immunitaire place du côté négatif de l'axe 1.

Le médecin insiste sur la complexité des phénomènes qui marquent l'histoire clinique d'un patient. Le statisticien répond en tentant de préciser le terme de *rupture* qui a été utilisé.

Stat (JPB): Pour représenter le système immunitaire, prenons l'exemple d'une corde à laquelle est suspendu un poids, et qui se détériore. À supposer que cette détérioration soit régulière, elle peut être décrite (pour poursuivre notre analogie) comme le mouvement d'un point dans un plan. Mais cette corde, de plus en plus fragile, se rompra de façon aléatoire. On sait seulement que, plus la corde est détériorée, plus elle est rongée, plus grande est la probabilité qu'un

accident quelconque en provoque la rupture. Il faut se souvenir qu'en mécanique, la rupture est un phénomène aléatoire, irrégulier dans sa ligne même, qui survient avec une probabilité déterminée dans un contexte de déformation régulière suffisamment avancée.

Je répète donc que, sur le graphique plan, on peut, d'une part, suivre le mouvement du point figuratif du système immunitaire, mouvement qui correspond à la détérioration de la corde; et, d'autre part, noter la rupture, c'est-à-dire la position où était parvenu le système quand s'est déclanchée la catastrophe.

Un immunologiste (JMC) remarque que s'il apparaît que ces points de rupture sont uniformément distribués sur tout le domaine du stade considéré, la coordonnée sur l'axe 1 étant aussi bien positive que négative, ce sera l'échec de l'étude suggérée par le statisticien. Celui-ci hésite un instant avant de répondre:

Stat (JPB): On ne peut accepter l'hypothèse que la position du point figuratif du système immunitaire soit sans lien avec la survenue d'une catastrophe; car, quelque réserve qu'on ait pu formuler sur l'enchaînement des stades, il est clair que la distribution régulière de ceux-ci le long de l'axe 1 démontre la pertinence quant à la clinique des informations retenues pour décrire l'état du système immunitaire; informations dont la coordonnée d'un examen sur l'axe 1 offre un résumé.

2 Variables spécifiques et variables génériques

Un immunologiste (JMC) oriente alors la discussion dans une autre direction: la recherche d'informations biologiques nouvelles beaucoup plus fines et spécifiques que celles traitées jusqu'à présent.

Imm (JMC): Il faut distinguer entre connaissance épidémiologique et connaissance scientifique. De données générales, qu'on les traduise en terme de stade ou de point figuratif de l'état du système immunitaire, on ne peut tirer plus que des statistiques chiffrant l'extension de la pandémie, ou servant aux assureurs à fixer les primes en fonction de l'état du patient, et aux services d'immigration à réglementer l'entrée des sujets séropositifs... Que la détérioration de l'état du patient soit manifeste, même si on ne l'évalue que d'après des variables grossières, n'est pour le clinicien qu'une évidence dont il n'y a rien à conclure de profond ni d'utile sur le déterminisme de la maladie. C'est sur des informations fines, éventuellement coûteuses à acquérir, que peut seul se fonder le progrès scientifique et thérapeutique.

Le statisticien rétorque que l'*homéostasie du système immunitaire* est un véritable thème de recherche scientifique.

Stat (JPB): Même si l'on se borne à suivre des variables non spécifiques du SIDA, il faut comparer la réponse des sujets séropositifs et non séropositifs

aux agressions pathologiques ordinaires: du mouvement de ce qu'on a, pour faire image, appelé "le spot" il me semble que tout est à apprendre.

Imm (JMC): Sans prendre à parti qui que ce soit, je dois formuler les plus expresses réserves vis-à-vis des variables qui sont à la base de l'analyse présentée ici. Par exemple, je défie quiconque de donner une signification précise de l'élévation de la concentration en $\beta 2$ -microglobulines...

Sans pouvoir aucunement suggérer un processus biologique, le statisticien rappelle la pertinence de la variable incriminée; laquelle monte avec tous les taux d'immunoglobulines quand se détériore l'état du système immunitaire. L'immunologiste, toujours réservé, prononce le terme de *facteur pronostique*.

Imm (JMC): Que l'état du patient se détériore et qu'alors certaines concentrations augmentent, ne nous apprend rien d'utile. Nous avons besoin de faits précis et fins qui signent la détérioration de l'état et permettent d'en prédire l'évolution.

Imm (Yves Rivière): Notre équipe a choisi de rechercher dans les lymphocytes périphériques de sujets séropositifs une activité cellulaire cytotoxique dirigée contre des cellules présentant des antigènes du VIH, en utilisant des cellules B autologues (du même donneur) transformées par le virus d'Epstein-Barr (B-EBV), infectées ou non par des virus vaccine recombinants; ce qui permet d'exprimer à volonté telle ou telle protéine virale du VIH, autorisant ainsi l'étude de la spécificité des cellules tueuses vis-à-vis des différents antigènes viraux... En collaboration, notamment, avec F. Tekaiä et Ph. Sansonetti, ici présents, nous tentons d'évaluer la réponse immunitaire de type cellulaire contre différentes protéines virales et régulatrices. Ces recherches n'ont pas seulement un intérêt descriptif; car, comme le dit le pr. Luc Montagnier, un grand nombre de types cellulaires répondent... reste à trouver ceux qui sont importants pour un vaccin...

Le statisticien veut préciser ce qui constitue, pour lui, le thème même de la rencontre et l'objectif par excellence de la recherche.

Stat (JPB): Mettons à part la recherche fondamentale qui doit aboutir à la découverte de drogues nouvelles ou à l'observation plus fine du processus pathologique. Le problème que la statistique peut contribuer à résoudre est celui de l'évaluation de l'état du malade. Il ne s'agit pas ici de prédire une évolution à venir mais de constater, aussi tôt que possible, en termes chiffrables, les effets des traitements essayés. Si l'on pouvait se fier à une coordonnée pour mesurer l'état du malade, on pourrait conclure à l'effet favorable d'un traitement qui ferait déplacer le "spot" dans la direction souhaitée. Vous savez qu'on évalue classiquement l'efficacité d'un traitement en fonction de la fréquence d'événements bien définis: décès du malade ou guérison. Avec le SIDA, maladie d'évolution lente (et pour laquelle on n'a pas d'exemple de guérison) il est exclu

d'attendre l'accumulation de telles statistiques pour autoriser l'usage d'une drogue utile.

Reprenons l'exemple de la rhinite allergique (présenté dans notre exposé). Les critères classiques fondés sur le dénombrement des guérisons sont inopérants, car il s'agit d'une affection chronique dont on peut seulement observer les accès ou la rémission. Au niveau de variables prises une à une, les trois drogues considérées ne différaient pas significativement. Mais en faisant la synthèse d'un ensemble de mesures, l'analyse des correspondances a permis de décrire l'état du patient par une coordonnée unique; dont les variations comparées, sous traitement et après arrêt de celui-ci, ont attesté la supériorité de l'un des produits quant à la permanence de ses effets.

Au statisticien qui souligne que l'étude de la rhinite prend en compte avec des notes subjectives grossières (éternuements, rhinorrhée...) des épreuves allergologiques fines, l'immunologiste rétorque que ce sont là des variables d'une médiocre spécificité et trouve dans le thème de l'allergie un exemple à l'appui de ses réserves face aux variables grossières.

Imm (JMC): On a d'abord constaté que la fraction IgE était liée à l'allergie; et pourtant, on trouvait, chez des sujets fortement perturbés par l'allergie, des niveaux d'IgE qui ne dépassaient pas notablement ceux de sujets indemnes. Le tableau s'est éclairci quand on a su doser les IgE spécifiques d'un allergène déterminé. On a alors trouvé, chez le sujet atteint, un pic d'IgE mille ou cent mille fois plus haut que la normale au niveau de l'IgE spécifique de l'allergène auquel il était sensible; et cela, même si le niveau global des IgE avoisinait la normale.

Le statisticien, soutenu par un médecin, répond en proposant un exemple complémentaire.

Stat (JPB): Le taux de cholestérol est depuis 30 ans incriminé comme un facteur de la pathologie cardiovasculaire. On sait aujourd'hui que ce qu'on a d'abord dosé sous le nom de cholestérolémie concerne des fractions diverses dont la signification clinique est non seulement inégale mais éventuellement opposée; l'attention portée au cholestérol n'en a pas moins été utile et le dosage global en est encore pratiqué. Ceci dit, on doit affirmer que le progrès de la Science est solidaire du progrès de l'observation; de ce qu'en termes d'ingénieur, on appellera l'acquisition de capteurs nouveaux. Le statisticien sait qu'il n'y a pas d'analyse sans données!

Méd (JLV): Il est commun de suivre une maladie sur des paramètres qui n'en sont pas spécifiques: on suit fort bien un lupus érythémateux disséminé sur la vitesse de sédimentation; alors que cette variable, élevée dans les processus inflammatoires, est la résultante de phénomènes complexes dont le détail n'est connu de personne.

Les médecins s'arrêtent quelque temps à une question qui reviendra souvent dans le débat: comment, sans en bouleverser la structure, compléter une base de données, fruit de cinq ans d'observation clinique? quels paramètres nouveaux y introduire? par exemple, les taux d'anticorps contre la protéine de régulation du virus... Bien qu'incompétent quant au choix des variables, le statisticien rappelle son point de vue, puis reprend le thème de l'homéostasie.

Stat (JPB): Le calculateur accueille volontiers toutes les variables et les traite de la même manière, que la mesure en soit facile ou coûteuse; seule importe la pertinence, dont médecins et biologistes sont *a priori* seuls juges, l'analyse n'apportant qu'une confirmation *a posteriori*. Le graphique présenté est une toile de fond sur laquelle on peut inscrire toute information nouvelle, non seulement relative à l'ensemble des sujets de la base, mais à des cas nouveaux, plus ou moins nombreux, dont on calculera les coordonnées sur les axes (selon la méthode usuelle d'insertion des *éléments supplémentaires*); ce qui permet aussi de placer toute modalité d'une variable nouvelle (en bref, comme centre de gravité du sous-groupe des cas rentrant dans cette modalité).

Mais, quel que soit l'intérêt des variables spécifiques du SIDA, le rôle des variables générales ne peut être minimisé. Le SIDA apparaît, fondamentalement, comme un dérèglement d'ensemble du système immunitaire: donc toute information relative à l'homéostasie naturelle (si peu étudiée) de ce système est précieuse à quelque niveau qu'on l'acquière. De ce point de vue, il semble même que les variables générales sont indispensables, comme offrant seules matière à des comparaisons. Prenons l'exemple d'une maladie infectieuse banale: j'imagine qu'il y a montée des immunoglobulines, baisse de l'albumine mais sans la pancytopenie observée dans le SIDA...

Ici un médecin introduit le thème du rapport des infections avec le SIDA.

Méd (JLV): On se demande si une infection opportuniste, qui n'est certes rendue possible que par la déficience immunitaire du patient, n'est pas en même temps cause d'aggravation dans l'état de celui-ci; si elle n'en est pas un cofacteur. Par exemple, on sait que L. Montagnier a noté l'amélioration de certains sidéens sous traitement anti-mycoplasme...

Imm (JMC): Des observations suggèrent que, même chez un patient asymptomatique, la charge de certains germes ou parasites soit nettement plus élevée que chez le sujet sain non infecté. Et l'on peut conjecturer que protéger les patients de tout contact avec des agents infectieux retarderait le progrès du SIDA.

Méd (PhS): On n'a pas de technique simple pour évaluer la charge en agents potentiels d'infections opportunistes que peut porter un patient asymptomatique déterminé.

Méd (JLV): Lorsque nous avons vu naître des enfants porteurs de VIH, donc fragiles quant à leur système immunitaire, nous avons tenté de les mettre sous bulle; ce qui, sans éliminer tous les contacts bactériens (puisque, notamment subsiste la flore intestinale), en réduisait grandement la fréquence: à notre grande déception, nous avons vu les enfants placés sous bulle évoluer aussi vite que les autres vers les états les plus avancés du SIDA.

Stat (JPB): Cette question de la sensibilité aux chocs infectieux du sujet séropositif nous ramène au thème de l'homéostasie, de l'équilibre; et suggère de considérer, du point de vue de la cybernétique, l'analogie avec un objet vertical. Si cet objet dispose d'un système de rétablissement, il répond à un choc par quelques oscillations, puis s'arrête à sa position initiale; sinon, il tombe pour ne pas se relever.

Méd (PhS): Il faut noter que, parmi les milliers d'examen effectués, très rares sont ceux portant sur des patients qui se trouvaient précisément soumis à une affection de quelque nature que soit... car il s'agit, ordinairement, de patients asymptomatiques. Prenons toutefois l'exemple d'un patient, dont l'état immunitaire est déjà délabré, qui entre dans une pneumocystose avec 150 T4/ μ l; il en ressortira peut-être avec 90 et ne remontera éventuellement que très lentement à partir de ce chiffre. Il s'agit, assurément d'un coup de boutoir terrible; et, quand ces chocs se multiplient, on entre dans la phase terminale; où, hélas, le clinicien doute qu'on puisse trouver aucun enseignement utile.

C'est de ce même point de vue de l'utilité pour la clinique que les médecins envisagent non seulement d'introduire des variables nouvelles, mais aussi d'éliminer des examens celles des variables présentement recueillies qui ne sont pas spécifiques.

Imm (JMC): La structure d'ensemble mise en évidence par les analyses statistiques ne subsisterait-elle pas si on ne gardait que les trois variables: taux de T4, de T8 et antigénémie? Si cela se faisait sans perdre d'information pertinente, ne serait-il pas commode de supprimer tous les autres dosages, notamment celui des β 2-microglobulines, dont j'ai déjà noté l'absence de spécificité?

Stat-Inf (Fredj Tekaïa): Ce paramètre est néanmoins donné (avec le nombre des T4) pour être des plus pertinents par toutes les analyses discriminantes qu'on a pu tenter pour définir la notion de stade ou retrouver celle-ci.

Méd (PhS): Si la mesure d'un paramètre hautement spécifique (tel que le taux d'anticorps contre tel composant protéique du virus...) est d'un intérêt supérieur pour suivre le cours de la maladie, je me rallierai à ce nouvel examen. Mais s'il est prouvé que toute l'information accessible est donnée par quelques dosages peu coûteux, je serai encore plus heureux de me borner à ces dosages!

3 Cinétique et variabilité

Inquiet de restrictions en perspectives dans la collecte des données, le statisticien reprend la parole, pour suggérer de suivre, à de brefs intervalles de temps, la variabilité de l'état des patients.

Stat (JPB): À ceux qu'anime un légitime souci d'économie, je suis prêt à concéder que l'analyse statistique n'est vraisemblablement pas enrichie par la considération de milliers de cas plutôt que de centaines: en effet, les multiples analyses que nous avons considérées avec Fr. Tekai'a fournissent des résultats stables, indépendants de l'effectif de l'échantillon. Mais je verrais avec un extrême intérêt quelques dizaines de patients suivis avec une fréquence temporelle élevée, par exemple mensuellement...

Méd (PhS): Je ne puis accepter une telle fréquence: le SIDA est une maladie lente; il n'y a aucune utilité clinique à suivre les patients mensuellement!

Stat (JPB): Je laisse au médecin d'affirmer qu'il s'agit d'une maladie lente: son expérience lui en donne le droit. Mais, même si pour le clinicien un patient est dans un état stationnaire, je ne sais pas, au niveau du graphique plan, comment, de semaine en semaine, se déplace le point figuratif de l'état du système immunitaire de ce patient; je ne sais pas où il se trouve...

Méd (PhS): Mais si, vous le savez! il est là où vous croyez qu'il est, puisque l'évolution en est lente.

Stat (JPB): Pour que la variation de l'état du système immunitaire, mesuré (par exemple) par la coordonnée sur l'axe 1 calculée en fonction de toutes les variables, puisse servir d'indice de l'efficacité d'un traitement mis à l'essai, il faut connaître l'ordre de grandeur des fluctuations communes de cette coordonnée à l'échelle du mois: car seules sont susceptibles d'être interprétées (comme résultant d'un traitement) des variations supérieures à cet ordre de grandeur.

On s'interroge sur les ressources qu'offre la base de données pour une telle étude de variabilité. Pour plusieurs patients, on dispose d'une suite d'examens; toutefois, ces examens sont séparés par des intervalles de plusieurs mois. Le statisticien répète:

Stat (JPB): Je concède que le SIDA est une maladie lente. Mais chacun d'entre nous, aujourd'hui bien portant, n'est-il pas porteur de cette maladie lente qu'on appelle la vie, et qui conduit à la mort en moins de cent ans! Or la lenteur du vieillissement naturel n'empêche pas que d'un mois à l'autre, ou même d'un jour à l'autre on change d'état, qu'on soit tantôt fatigué tantôt dispos. Tout état stable est affecté de fluctuations.

Méd (PhS): Mais l'ensemble des mesures analysées (concentrations et numérations) est, à l'échelle d'un ou deux mois, tout à fait stable; aucun

médecin n'accepterait de répéter les examens à un rythme mensuel, même seulement chez quelques malades...

Stat (JPB): Je ne sais si la fixité absolue que vous percevez résisterait à l'application de nos méthodes. Vous considérez les variables une par une, et aucune des variations observées ne vous apparaît significative. Nous prenons les mesures toutes ensemble; nous en formons (pour calculer les coordonnées ou facteurs), des combinaisons déterminées; et, si les variations dont sont affectées toutes ces variables (ou du moins plusieurs d'entre elles) sont concordantes, se font dans le même sens (par exemple d'une amélioration), alors, apparaît sur le graphique un mouvement résultant dont l'amplitude sera perçue de chacun (statisticien ou médecin) comme significative; mais qui ne le sera véritablement que si elle dépasse l'ordre de grandeur des fluctuations communes.

Il convient de considérer ce que peuvent être les *variations concordantes* auxquelles je fais allusion. (Le tableau proposé ici a été construit après l'exposé pour donner toute précision à l'exemple esquissé oralement). Voyez, sur la ligne 'vrai', les mesures relevées sur un sujet au stade 14, et dont le point figuratif se place dans la marge gauche du nuage, avec ceux des états les plus graves. La ligne 'moy' donne les valeurs-pivot '≈' qui rentrent dans la moyenne des examens analysés. Enfin, la ligne 'Norm' pourrait, croyons-nous, donner les résultats d'examens effectués sur un sujet normal, en bonne santé.

Prenons pour exemple la variable IgG...

	LYMP	T4	T8	β	IgG	IgM	IgA	PLQ	HIP	->	F1	F2
vrai	950	20	700	400	2300	240	380	175	14		-788	-47
f-	1580	340	760	290	1820	150	270	195	?		-122	146
moy	1740	400	790	260	1700	130	240	200	?		52	233
f+	1900	500	820	230	1600	120	220	208	?		247	320
Norm	1600	800	600	180	1200	50	80	320	?		676	707

Méd (JLV): Effectivement, pour nous, le passage de 17 à 23 correspond au passage de ce qui peut être accepté comme limite supérieure du normal à une valeur nettement excessive qui est comme un signal rouge qui s'allume. Ai-je répondu à votre question?

Stat (JPB): Non, pas exactement; je n'ai pas encore posé la question! Considérez encore les lignes 'f+' et 'f-': ce sont des cas fictifs qu'on a construits en faisant à peu près le cinquième du chemin séparant 'moy' de 'vrai', respectivement dans le sens de l'amélioration (f+) et de la détérioration ('f-'). Pour les 4 lignes, on a calculé le facteur F1: et, entre 'moy' et 'vrai', la différence est d'environ 800; tandis qu'entre 'moy' et 'f+' ou 'f-' cette différence est de quelque 150, ce qui est tout à fait notable. Or, et c'est ma question, n'est-il pas vrai qu'entre 'moy' et 'f+' chacune des variations prise isolément, par exemple de 16 à 17 pour IgG, n'a pour le clinicien aucune

signification nette? Tandis que de la variation conjointe des paramètres résulte un déplacement net, observable, mesurable; et le problème est maintenant posé d'utiliser le nouveau moyen d'observation qu'est l'analyse multidimensionnelle pour aboutir à des conclusions...

Méd (JLV): Et la cinétique?

Stat (JPB): C'est justement là qu'est la cinétique; visible maintenant sur une échelle, sur un graphique; même si ce mouvement, je le répète, résulte de variations qui, considérées séparément pour chaque variable, ne sont pas interprétables.

Méd (JLV): on pourra, par exemple, suivre la variation sur 3 ans?

Stat (JPB): Sur 3 ans ou sur moins de 3 ans, sur un mois... Mais la question subsiste que les variations normales, même chez l'animal, ne sont pas véritablement connues; puisque, jusqu'à présent, les observations n'ont pas été considérées au niveau multidimensionnel.

Méd (JLV): Les seuils dont nous parlons (par exemple: 16 pour les IgG) n'ont pas été établis sur une seule mesure! mais en considérant une distribution statistique sur des personnes différant quant à l'âge et au sexe, observées à des heures diverses... C'est ainsi qu'on détermine un seuil et la signification qu'il convient d'accorder à une variation au delà du seuil.

Stat (JPB): Je ne mets pas présentement en doute l'usage que font des seuils les cliniciens. Mais la question est de savoir si des variations graduelles, toutes d'un ordre de grandeur non significatif, mais concordantes quant à leur sens (par exemple l'aggravation), peuvent apparaître sur nos graphiques comme des variations de l'état global qui méritent d'être interprétées et révèlent avec finesse les détails du processus que vous avez appelé justement *la cinétique*.

[N.B.: Nous appelons l'attention du lecteur sur le point 'Norm': ce point, introduit *a posteriori*, illustre l'interprétation de l'axe 2 suggérée au §3.3 de [ANA. CLIN.]: il se place nettement dans le quadrant supérieur droit du plan (1,2); associé aux modalités inférieures des concentrations, et aux valeurs supérieures des numérations cellulaires, à l'exception de celle des T8, 'T8+'; modalité dont on a dit qu'elle dépassait considérablement la normale. En descendant vers $F2 < 0$, à partir de 'Norm', on se dirige vers la modalité 'T8+'; ce qui pourrait correspondre à la phase initiale de l'infection par le VIH; phase dite asymptomatique, mais dont l'importance tend à être affirmée...]

4 Variables sociologiques et variables biologiques

Les médecins ne réservent pas tout leur intérêt à l'exploration de données biologiques nouvelles, spécifiques du SIDA; ils se penchent sur le contexte sociologique (auquel se rattache d'ailleurs la notion bien connue de *groupe à risque*).

Imm (JMC): Que sait-on du genre de vie des malades?

Méd (PhS): Sous sa forme actuelle, le dossier écrit comporte certaines informations; mais on doit s'interroger sur la qualité de données recueillies quasi exclusivement de la bouche des patients...

Stat (JPB): Même exactement relevée, une variable du genre de vie ne se prête que rarement à une interprétation univoque: car chacune est suivie d'un cortège d'autres, où se trouve peut-être l'explication réelle de telle corrélation observée. Par exemple habiter en pavillon de banlieue plutôt qu'à Paris, ne signifie pas seulement avoir un jardin personnel, mais aussi passer plus de temps en trajets; et implique des ressources différentes...

Imm (JMC): Je ne vois pas en quoi ce que vous dites de la variable sociologique "habiter en banlieue" ne s'applique pas, *mutatis mutandis*, à une variable biologique comme la β 2-microglobuline.

Stat (JPB): Certes, la β 2 n'est pas une espèce chimique unique mais une fraction complexe dont des causes différentes peuvent augmenter les diverses composantes. Mais il est aisé de se convaincre sur un exemple que l'ambiguïté des variables du genre de vie est d'un autre ordre de grandeur. Ainsi, plusieurs études statistiques ont suggéré (sans remettre en question les méfaits de l'alcoolisme), que l'absorption modérée de boissons alcooliques était un facteur favorable à l'état du système cardiovasculaire. Mais entre un sujet qui boit modérément et celui qui ne goûte pas au vin, dix grammes d'alcool par jour ne font pas toute la différence. Boire modérément, c'est savourer la vie sans excès ni tension; ne pas boire, sauf tradition familiale, c'est s'imposer une épreuve qui n'est sans doute pas la seule! Et si survient un accident cardiovasculaire, il ne faudra pas incriminer l'absence d'alcool...

Imm (JMC): ...mais le "stress"! Une étude récente sur le taux de malformations chez les enfants de travailleurs d'une centrale nucléaire anglaise suggère de semblables réflexions. Il est d'abord apparu que ce taux dépassait significativement la moyenne nationale; et l'on a incriminé les effets mutagènes des rayonnements. Mais une étude complémentaire a considéré les enfants vivant avec leurs parents dans des agglomérations semblables à celles où l'on place d'ordinaire les centrales nucléaires: petites villes, environnées de campagnes, au voisinage d'un cours d'eau... Et il est apparu que, pour ces enfants, le taux trouvé ne différait pas de celui observé auprès de la centrale nucléaire. Celle-ci est donc mise hors de cause, et l'on doit trouver d'autres hypothèses: par exemple, effets de l'endogamie liée à un tel milieu.

Stat (JPB): Il est déjà difficile de comprendre les relations qu'il y a au sein d'un groupe homogène de variables, en rapport direct avec le phénomène étudié; quand on invoque comme facteur explicatif des variables d'un tout autre ordre, les difficultés deviennent extrêmes. Certes, l'analyse des correspondances est

apte à traiter, après codage, des données très hétérogènes: par exemple, l'analyse d'un tableau de variables sociologiques de toute nature compilé par l'hebdomadaire "Le Point" pour déterminer dans quels départements *On vit le mieux en France*, (in *CAD*, vol XIII, n°3, 1988), a fourni de l'espace français un tableau cohérent. Mais il est grandement souhaitable d'avoir au moins des groupes de variables homogènes. La présente analyse en comporte deux: les dosages de protéines sériques et les dénombrements cellulaires.

Les médecins reviennent donc au projet d'introduire des variables spécifiques: doser les antigènes reconnaissant diverses composantes du virus VIH suscite un intérêt général. Mais seuls certains laboratoires, hautement spécialisés, sont à même de faire de tels dosages, d'ailleurs très coûteux: ce serait donc un projet de coopération requerrant un puissant patronage tel que celui de l'ANRS (Agence Nationale de Recherche sur le SIDA).

Imm (YR): De plus, il importe de noter que l'évaluation de variables particulières requiert une standardisation des techniques employées; afin que les résultats soient comparables d'un laboratoire à un autre; comme cela est le cas pour les dosages classiques (tel que le nombre absolu de T4).

Le statisticien souligne l'extrême importance de cette remarque! La compatibilité entre mesures analogues (mais non identiques) effectuées par des laboratoires distincts d'un même centre étant, elle-même, difficilement assurée.

On se demande si les échantillons de sérums conservés dans la sérothèque suffisent pour déterminer des variables nouvelles à adjoindre aux dossiers d'exams archivés dans la base de données. Quant au typage HLA, les spécialistes constatent que rien de ce qui a été publié jusqu'ici ne prouve une corrélation entre les groupes d'antigènes HLA et la susceptibilité au VIH.

Stat (JPB): Je rappelle que le graphique issu de l'analyse des variables déjà recueillies est comme une toile de fond sur laquelle on peut inscrire toute information nouvelle, non seulement relative à l'ensemble des sujets de la base, mais à des cas nouveaux, plus ou moins nombreux.

5 Devenir des lymphocytes et réponse du système immunitaire

Revient en discussion le thème de la représentation diachronique des états des patients.

Stat (JPB): Ma prédilection va à des mesures qu'on puisse répéter fréquemment, au moins sur quelques dizaines de sujets; afin de substituer une description mathématique précise à l'image du système immunitaire, objet suspendu qui oscille sous le choc...

Imm (JMC): ...objet suspendu à une corde qui se ronge!

Stat (JPB): ...objet rappelé à sa position d'équilibre par un ressort qui, progressivement se rouille. Toute ces images sont pour rappeler que la détérioration progressive d'un système se révèle par les réponses de ce système à des impulsions. Nous avons dit qu'il y a, d'une part, des phénomènes qui évoluent de façon régulière; et, d'autre part des phénomènes aléatoires qui se produisent brutalement: la rupture d'une corde...

Méd (JLV): c'est la maladie de Damoclès!

Stat (JPB): cas encore plus angoissant que celui du SIDA; dans la mesure où, pour Damoclès, il n'y a qu'un seul stade, et la menace est permanente, toujours au même niveau!

Afin de caractériser la réponse anormale d'un système immunitaire atteint, il faut avoir pour référence l'homéostasie d'un système normal soumis à une agression pathologique banale. Les médecins soulignent que cette réponse homéostatique est mal connue; les questions les plus simples, comme celle de la cinétique du nombre de cellules d'un type donné, sont présentement sans réponse. On prend pour exemple la grippe, maladie que le statisticien imagine, à tort, être la plus banale.

Méd (JLV): Il faut d'abord savoir que ce qu'on appelle vulgairement grippe n'est pas ce que scientifiquement on doit appeler de ce nom. La véritable grippe, qui a pour agent *Myxovirus influenzae A*, ne se rencontre en France que par épidémies espacées de quelque 4 ans. Or la grippe entraîne un effondrement du taux de lymphocytes circulants encore plus spectaculaire que celui observé dans le SIDA. Cet effondrement est suivi d'une prompt remontée; mais on ne sait si, dans la grippe, les lymphocytes sont proprement détruits ou seulement séquestrés; les lymphocytes pouvant quitter le système vasculaire par diapédèse et gagner d'autres compartiments.

Stat (JPB): Quelle que soit la cause profonde des variations observées, on peut comparer la manifestation de ces variations; d'une part, chez le sujet normal et, d'autre part, chez le sidéen. On peut conjecturer que, chez celui-ci, le retour à la normale se fait plus lentement.

Méd (JLV): Plus simple encore que l'étude de maladies, même bénignes, serait la comparaison entre les réponses immunitaires du sidéen et du sujet normal à ce stimulus que constitue un rappel de vaccination DTCoq. Y a-t-il, chez le sidéen, production normale d'immunoglobulines? la réaction fébrile a-t-elle la même amplitude et la même durée?

Stat (JPB): Il faut souligner que, même si les observations portent sur des sujets qui n'appartiennent pas à la cohorte suivie dans la base de données, les réponses immunitaires pourront être mises en rapport avec les résultats de l'analyse factorielle et interprétées dans ce cadre. Et, traduite en termes

mathématiques, la cinétique de la réponse immunitaire fournira l'analogie de la mesure de l'inertie d'un système oscillatoire ou de la dureté d'un ressort.

Un médecin porte la discussion sur la fonction normale des lymphocytes T cytotoxiques et sur leur rôle pathologique éventuel dans la diffusion du VIH.

Méd (JLV): Un lymphocyte T cytotoxique (CTL) déterminé est quiescent jusqu'à ce qu'il rencontre une cellule lui présentant l'épitope (site d'antigène) spécifique qu'il est apte à reconnaître. Se déclenche alors un processus complexe aboutissant, en bref, à la destruction, à la digestion en quelque sorte, du complexe auquel est lié cet épitope. Ce processus, normal chez le sujet sain, peut avoir, chez le sidéen, des conséquences catastrophiques. Car, si le CTL était précisément porteur d'un virion (unité virale) VIH, se trouve déclenchée la multiplication de celui-ci; et sont mis en circulation peut-être mille exemplaires du virus VIH dont un grand nombre se logeront dans des lymphocytes CTL de spécificités diverses. Ces CTL infectés seront certes quiescents; mais, chaque fois que l'un ou l'autre d'entre eux rencontrera l'antigène qu'il reconnaît, la catastrophe que nous venons de décrire se reproduira.

Les médecins s'interrogent ici sur la durée de vie des lymphocytes T. Certainement de l'ordre de plusieurs mois, cette durée (jusqu'ici mal connue) est peut-être de quelques années. Plusieurs hypothèses ont cours sur la transmission de la compétence (ou spécificité pour un épitope) de cellule à cellule...

Imm (JMC): Quoi qu'il en soit de cette permanence de la spécificité, le processus de diffusion que l'on vient d'évoquer conduirait à une croissance exponentielle de l'infection; ce qui ne semble pas en accord avec l'observation clinique.

Ici, médecins et statisticiens s'interrogent sur le tempo de l'infection par le VIH.

Stat (JPB): Quelle que soit la croissance du nombre des cellules infectées, ma préférence va à l'hypothèse que la maladie est due à une répression de la production de certaines classes de lymphocytes plus encore qu'à la destruction de ceux-ci; car on a observé que, dans le SIDA, les T4 sont en majorité indemnes de l'infection. C'est pourquoi je préconise avec tant d'insistance d'étudier les variations diachroniques du profil immunitaire chez le sujet sain et chez le malade; pensant que seule l'observation globale, au niveau du système, peut être révélatrice, ou, au moins, suggestive, pour la compréhension d'un désordre qui est, précisément, systémique. Et ce caractère systémique limite grandement la portée des études *in vitro*...

Méd (JLV): Quoique spécialiste des études *in vitro*, je souscris essentiellement à cette remarque...

Stat (JPB): Quant à moi, ne travaillant qu'*in ordinatio*, je suis encore plus loin que vous de ce que je crois être le vrai champ de bataille!

Méd (JLV): Une grande difficulté vient de ce qu'on ne connaît pas l'homéostasie des T4 chez le vivant, même chez un animal de laboratoire tel que la souris, pour ne rien dire de l'homme!

Stat (JPB): Même si les médiateurs biochimiques de cette homéostasie sont inconnus, on peut observer le système; noter ses caractéristiques de réponse; comparer sujet sain et sujet malade...

Méd (JLV): Il est vrai que les représentations géométriques, fournies par la statistique multidimensionnelle, permettent d'objectiver cet aspect purement quantitatif de l'homéostasie, en dehors de la connaissance des agents causaux. Je m'intéresse volontiers à l'hypothèse que la maladie soit particulièrement liée à une répression de la production des T4 plus encore qu'à la destruction pure et simple de ceux-ci; et demanderai au statisticien comment l'observation globale du profil pourrait étayer cette hypothèse.

Stat (JPB): Si, dans le cours normal d'une certaine maladie, à une certaine phase, le taux des T4 s'élève rapidement, cette élévation pourra faire défaut chez les sidéens...

Les médecins donnent pour peu vraisemblable une élévation naturelle rapide des T4; mais admettent que des différences apparues, dans l'homéostasie (la régulation du profil immunitaire), entre sujet normal et sidéen puisse suggérer que, chez celui-ci, la production des T4 est restreinte.

Stat (JPB): Bien que l'examen de la réponse du système, en termes de variations de profil, ne puisse suffire à en décrire, de façon univoque, la structure homéostatique, il est assuré que l'analyse multidimensionnelle sera un filtre très efficace pour éliminer les hypothèses incompatibles avec ce qu'on observe et pourra orienter la recherche biochimique.

L'outil statistique donne à ce qui n'était que vague une expression suffisamment quantitative et précise pour donner prise au raisonnement.

La discussion se termine: les médecins s'entretiennent du projet de recueillir de nouvelles données; et statisticien et informaticien se préparent à tirer de la base 'CONSULT' tout ce qu'elle recèle quant à l'histoire de cas individuels. [Alors que nous mettons au point le texte du présent compte rendu, certaines recherches ont déjà produit quelques fruits: cf., notamment [ANA. CLIN.], §3.3].

Note bibliographique

La note [RECH. PRON. SIDA], de Fr; Tekai'a, publiée dans ce cahier (CAD, Vol XV, n°3), offre au lecteur un panorama des travaux déjà rédigés ou en cours de rédaction.