

J.-P. BENZÉCRI

G. D. MAÏTI

E. BIARD

Effet hématoprotecteur de l'ADN-HP chez des patients traités par chimiothérapie anticancéreuse

Les cahiers de l'analyse des données, tome 16, n° 2 (1991), p. 179-200

http://www.numdam.org/item?id=CAD_1991__16_2_179_0

© Les cahiers de l'analyse des données, Dunod, 1991, 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/>

EFFET HÉMATOPROTECTEUR DE L'ADN-HP CHEZ DES PATIENTS TRAITÉS PAR CHIMIOTHÉRAPIE ANTICANCÉREUSE

[HÉMATOPROTECT.]

J.-P. BENZÉCRI

G. D. MAÏTI*, E. BIARD**

1 Introduction: origine et structure de l'étude

On sait que la chimiothérapie anticancéreuse, outre son action toxique souhaitée sur les cellules cancéreuses, exerce des ravages sur des lignées cellulaires saines, notamment sur les cellules sanguines; d'où une cytopénie (diminution du nombre des cellules), cause d'un déficit immunitaire qui peut être grave. Ainsi, la toxicité hématologique de certains antimétabolites expose les malades à un risque infectieux majeur lorsque la leucopénie va jusqu'à des taux de leucocytes inférieurs à 1000/ μ l (microlitre= millimètre cube). C'est pourquoi une leucopénie marquée impose de réduire la posologie des traitements antimétabolites et d'en espacer les cycles; ce qui limite l'efficacité de la chimiothérapie. On cherche donc à accompagner la chimiothérapie de traitements dits *hématoprotecteurs*, combattant la toxicité des antimétabolites.

La présente étude porte sur l'effet hématoprotecteur d'un produit, l'ADN-HP (ADN Hautement Polymérisé, du Laboratoire Biostablix), que nous désignerons par 'A'. Les données proviennent d'un essai contrôlé ouvert, effectué sur deux groupes de 30 malades; dont l'un, *groupe A*, reçoit le produit; tandis que l'autre, *groupe t*, comprend des témoins ne recevant aucun traitement antileucopéniant. Les sujets ont été recrutés en consultation de cancérologie, en respectant des critères que nous résumerons comme suit:

critères d'inclusion: sujets adultes des deux sexes, âgés de 18 à 65 ans; traités par chimiothérapie séquentielle pour un épithélioma ou un sarcome; présentant une cytopénie se manifestant par l'une des conditions sur les taux: {<3000 leucocytes/ μ l, <1500 granulocytes/ μ l, <100000 thrombocytes/ μ l};

(*) Expert Clinicien, Société Statmatic.

(**) C.I.D., Euroconsultant.

critères d'exclusion: soit à l'entrée: association, à la chimiothérapie, de radiothérapie ou de diverses drogues spécifiées (lithium,...); soit en cours d'essai: épisodes infectieux intercurrents (l'infection perturbant la formule sanguine par la réaction immunitaire qu'elle provoque).

L'essai étant ouvert, les sujets reçoivent, lors de leur admission, un numéro d'ordre (allant de 1 à 60); et, d'après une liste d'affectation aléatoire préparée avant le début de l'essai, les sujets admis sont incorporés à l'un des deux groupes: 'A' (administration du produit); ou 't' (témoin, non traité par un hématoprotecteur). Il importe de signaler, dès maintenant, que l'analyse des données ne nous permet pas de croire qu'une véritable affectation aléatoire ait été suivie: car les deux groupes 'A' et 't' diffèrent significativement quant à l'état des sujets à l'admission dans l'essai.

L'observation de chaque patient se poursuit pendant 9 semaines: de 'J-7' à 'J56'. Nous désignerons par le caractère '/' l'examen de l'état basal (J-7); et par les chiffres {1, 2, ..., 8} les 8 examens hebdomadaires échelonnés de J7 à J56. Il y a, pour tout sujet, deux cures de chimiothérapie anticancéreuse, débutant l'une lors de l'examen 1, l'autre lors de l'examen 4. Sans entrer dans les détails, nous dirons que, pour les sujets du groupe 'A', le traitement hématoprotecteur débute avant la première cure de chimiothérapie et se poursuit ensuite avec une posologie variable.

Dans la suite, les sujets seront désignés sur nos graphiques par des sigles de deux lettres: deux capitales pour le groupe 'A' (e.g. ZC, pour le sujet numéro 3); deux minuscules pour les témoins (e.g. fz, pour le sujet numéro 60). À un examen, nous affecterons un sigle de trois caractères, dont le premier désigne le rang dans le traitement et les deux suivants le sujet concerné: ainsi /ZC est l'examen du sujet ZC à l'admission (état basal); 6fz est l'examen du sujet fz à J42; etc...

On utilisera des sigles de 2 caractères pour désigner des profils moyens afférents à des groupes d'examens: {/, /A, /t}, respectivement pour les examens à l'admission (J-7) des 60 sujets, du groupe A et des témoins; {55, 5A, 5t}, pour les examens à J35; etc.

Des diverses informations consignées dans le dossier médical de chaque patient, nous ne retiendrons que l'hémogramme, ou *formule* sanguine, déterminé avec chaque examen. Désormais, un sigle tel que /ZC ou 6fz, désignera simplement la formule sanguine attachée à un examen. De façon précise, on a, pour une formule, les 11 variables suivantes, concernant respectivement les globules rouges, les globules blancs et les plaquettes.

{vHEM, vHMC, vHMG, vRET}:

vHEM: "hématies", numération des hématies (encore appelées érythrocytes, ou globules rouges);

vHMC: "hématocrite", rapport du volume des érythrocytes du sang au volume total;

vHMG: "hémoglobine", taux sanguin de cette protéine, contenue dans les hématies et véhicule de l'oxygène;

vRET: "réticulocytes", numération de ces cellules, ou érythrocytes très jeunes, conservant des ribosomes qui sont précipités en réseau (d'où le nom) par les colorations vitales;

{vLEU, vNE%, vÉO%, vBA%, vLY%, vMO%};

vLEU: "leucocytes", numération d'ensemble de tous les globules blancs; dont on distingue ici trois classes:

granulocytes (dits encore: polynucléaires, du fait de l'aspect de leur noyau), et subdivisés ici, selon les colorations que prennent leurs granulations, en *neutrophiles*, *éosinophiles* et *basophiles*;

lymphocytes, non subdivisés ici (mais qui diffèrent grandement entre eux quant à leur rôle dans la réponse immunitaire; et leur pathologie: cf., e.g., [DISCUSSION PASTEUR], in *CAD*, Vol XV, n°3, 1990);

monocytes;

vNE%: "neutrophiles", pourcentage de ces cellules sur l'ensemble des leucocytes;

vÉO%: "éosinophiles", pourcentage, calculé comme vNE%;

vBA%: "basophiles", pourcentage, idem;

vLY%: "lymphocytes", pourcentage, idem;

vMO%: "monocytes", pourcentage, idem;

il importe de noter ici que seuls, parmi les pourcentages, neu et lym sont donnés avec une précision relative qui permette d'en suivre les variations en cours de traitement; les autres pourcentages, étant arrondis à une valeur entière, sont souvent notés zéro; ce qui est d'autant plus regrettable que selon une communication personnelle du Dr Questel, le traitement hématoprotecteur A est particulièrement favorable aux éosinophiles;

{vTHR}:

vTHR: "thrombocytes", numération de ces éléments anucléés du sang circulant, encore appelés plaquettes, et contribuant à l'hémostase primaire et à la coagulation.

Dans la suite, les seules variables retenues seront au nombre de 10:

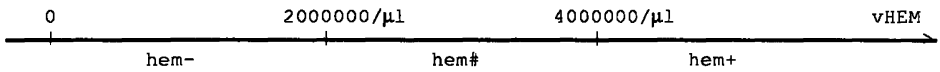
{vHEM, vHMC, vHMG, vRET, vLEU, vNE%, vNEU, vLY%, vLYM, vTHR};

on notera qu'à partir des pourcentages, {vNE%, vLY%}, on a calculé des numérations (vNEU, vLYM: rapportées au µl de sang), en les multipliant par vLEU. À partir de ces variables, on calculera, d'autre part, des quotients, afin de comparer à l'état basal les formules sanguines successives.

2 Codage des données et enchaînement des analyses

2.1 Tableau des variables et codage barycentrique

Avec 9 hémogrammes, ou *formules*, pour chacun des 60 sujets et 10 variables, on a un tableau (540×10); dont chaque ligne reçoit un sigles de trois caractères selon les principes posés ci-dessus. Tel quel, ce tableau hétérogène ne peut être soumis à l'analyse des correspondances: on code donc chaque variable suivant 3 modalités.

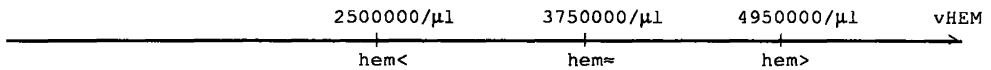


On sait qu'il est classique d'utiliser un codage dit *logique*, ou en (0,1), en découpant l'intervalle de variation de chaque variable en intervalles consécutifs. Par exemple, pour vHEM, on pourrait choisir les trois intervalles:

$$\text{hem-} = \{0, 2000000/\mu\text{l}\} ; \text{hem\#} = \{2000000, 4000000\} ; \text{hem+} = \{4000000, \infty\} ;$$

correspondant respectivement à une forte anémie, à une anémie modérée et à l'absence d'anémie. Dans le tableau ainsi codé, la variable vHEM est éclatée sur 3 colonnes; chaque individu (ici: chaque formule) ayant 1 dans la colonne de la modalité dans laquelle il rentre, et zéro dans les deux autres.

Cependant, avec un tel codage, les valeurs 2250000/μl et 3750000/μl, comprises dans hem#, sont toutes deux codées {0;1;0} alors que la signification clinique en est bien différente. Une telle imprécision serait, dans la présente étude, d'autant plus regrettable qu'il s'agit de suivre de semaine en semaine la variation continue d'une éventuelle cytopénie. En fait, le codage en (0,1) des variables continues n'est satisfaisant que si un très grand nombre d'individus permet de créer un nombre élevé de modalités (tout en ayant, dans chacune de celles-ci, un nombre d'individus qui suffise à en révéler le profil); ou si la multiplicité des variables rend tolérable la perte d'information au niveau de chaque variable.



C'est pourquoi (cf., e.g., [CODAGE LIN.], in *CAD*, Vol. XIV, n°2, 1989), on a adopté ici un codage linéaire par morceaux (dit encore *flou*, ou *barycentrique*), suivant 3 valeurs pivot, {hem<, hem≈, hem>} : ainsi, une valeur inférieure ou égale à hem<, premier pivot, est codée {1;0;0}; une valeur \geq 4950000/μl, deuxième pivot, est codée {0;0;1}; mais une valeur comprise entre deux pivots successifs est codée avec, sur chacun de ceux-ci, des composantes non nulles dont la somme est 1; une composante étant d'autant plus forte que la valeur à coder est plus proche du pivot correspondant. Ainsi, les valeurs 2250000/μl, 3250000/μl et 3750000/μl sont codées respectivement {1; 0; 0}, {0,4; 0,6; 0} et {0; 0; 1}; ce qui semble satisfaisant.

```

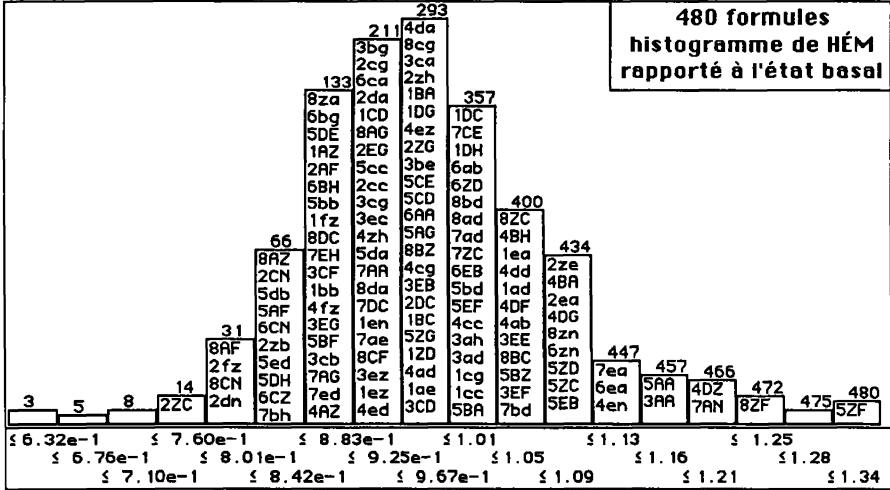
codage des variables (et non de quotients) sur 540 formules
A:DW:BIOWDcodx: pivots pour le codage barycentrique des variables
le nombre des variables est 10
vHEM (hématies/ $\mu$ l)      a 3 mod dont les sigles et pivots sont
hem< hem $\approx$  hem>      2.500e+6 3.750e+6 4.950e+6
vHMC (hématocrite %)      a 3 mod dont les sigles et pivots sont
hmc< hmc $\approx$  hmc>      2.580e+1 3.450e+1 4.410e+1
vHMG (hémoglobine g/l)    a 3 mod dont les sigles et pivots sont
hmg< hmg $\approx$  hmg>      8.200e-1 1.150e+0 1.490e+0
vRET (réticulocytes/ $\mu$ l)  a 3 mod dont les sigles et pivots sont
ret< ret $\approx$  ret>      2.616e+3 5.412e+4 1.343e+5
vLEU (leucocytes/ $\mu$ l)     a 3 mod dont les sigles et pivots sont
leu< leu $\approx$  leu>      1.000e+3 3.300e+3 7.100e+3
vNE% (neutrophiles/leuc)  a 3 mod dont les sigles et pivots sont
ne%< ne% $\approx$  ne%>      2.200e+1 5.400e+1 8.100e+1
vNEU (neutrophiles/ $\mu$ l)  a 3 mod dont les sigles et pivots sont
neu< neu $\approx$  neu>      3.400e+2 1.785e+3 5.092e+3
vLY% (lymphocytes/leuc)  a 3 mod dont les sigles et pivots sont
ly%< ly% $\approx$  ly%>      8.000e+0 3.900e+1 7.300e+1
vLYM (lymphocytes/ $\mu$ l)  a 3 mod dont les sigles et pivots sont
lym< lym $\approx$  lym>      2.945e+2 1.320e+3 2.750e+3
vTHR (thrombocytes/ $\mu$ l)  a 3 mod dont les sigles et pivots sont
thr< thr $\approx$  thr>      3.700e+4 2.120e+5 4.080e+5

```

Même si la continuité du codage assure que les nuances de l'information ne sont pas irrémédiablement perdues, il reste que le choix des pivots n'est pas indifférent; car les pivots extrêmes, hem< et hem>, sont des seuils en deçà ou au delà duquel, on ne fait plus de distinction; et qui fixent aussi l'échelle avec laquelle sont exprimées les variations de part et d'autre du pivot médian, hem \approx . [En toute rigueur, pour les sujets de sexe féminin, il conviendrait ici de coder la variable vHEM, avec des bornes de 10% inférieures à celles des hommes...]

Dans la présente étude, les pivots (dont le tableau est donné ci-dessus) ont été choisis au vu d'histogrammes présentés par le programme "zrang" (cf. [CODAGE LIN.], §2), en tenant compte des principes ci-dessus; dont l'application est assez sûre pour que deux analyses, fondées sur des découpages effectués indépendamment, aboutissent à des résultats très semblables. Le biologiste pourra les comparer aux seuils généralement reçus dans l'interprétation clinique des hémogrammes; le pivot supérieur des numérations correspondant généralement à une valeur normale. Mais il importe de signaler qu'ici les pivots doivent être adaptés à la distribution de la population des patients chez lesquels on veut suivre des variations; le recours à des variables quotient (cf. §2.2) réduisant toutefois la portée de l'analyse fondée sur un codage direct des valeurs mesurées.

Finalement, on a un tableau (540 \times 30), croisant les 540 formules (ou hémogrammes) avec les 30 modalités des 10 variables retenues. À ce tableau, sont ajoutées 27 lignes {/, /A, /t}, {11, 1A, 1t}, ..., {88, 8A, 8t}, construites en cumulant les lignes afférentes à l'examen d'admission, puis aux 8 examens suivants, soit pour l'ensemble des 60 sujets, soit pour les 30 constituant le groupe A (recevant un traitement protecteur) ou le groupe t (des témoins).

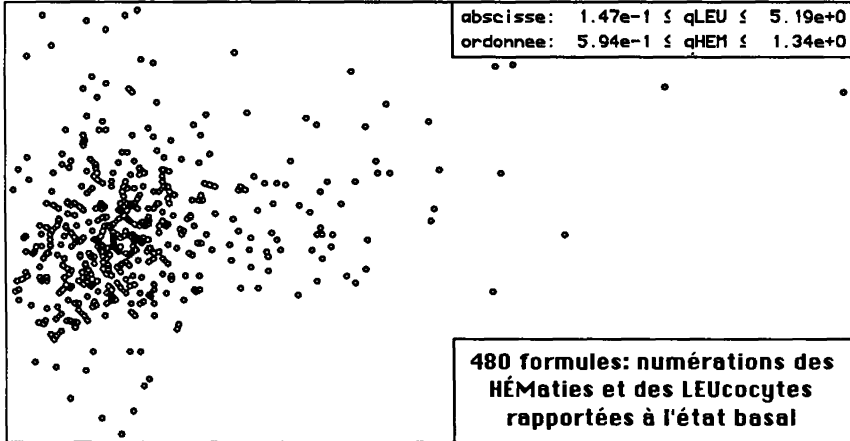


2.2 Variables rapportées à l'état basal

Pour chaque sujet, l'effet cumulé de l'antimitotique et d'un éventuel traitement cytoprotecteur se manifeste sur les rapports aux valeurs basales des valeurs successives prises, au cours du traitement, par les variables de l'hémogramme. On peut, pour chacune des 480 formules des examens de rang 1 à 8 (de J7 à J56), calculer une ligne de dix variables rapport (affectées de l'initiale 'q'). L'histogramme ci-dessus concerne le rapport qHEM. On voit que l'histogramme s'étale largement à droite et à gauche: afin de bien rendre compte des valeurs centrales, les pivots {HEM<, HEM=, HEM>} ont été choisis (à l'aide du programme "zrang") aux rangs {15, 240, 445}; ce qui correspond aux valeurs {0,7634; 0,9390; 1,129} (cf. tableau).

En codant de même les 10 quotients, on a un tableau (480 × 30); auquel on a adjoint 24 lignes supplémentaires: {11, 1A, 1t}, ..., {88, 8A, 8t}, construites en cumulant les lignes afférentes aux 8 examens suivants, soit pour l'ensemble des 60 sujets, soit pour les 30 constituant les groupes A (traitement) ou t (témoins); une dernière ligne, //, est créée en codant une suite de 10 quotients tous égaux à 1, afin de placer approximativement l'état basal.

Il vaut la peine de comparer deux variables rapport: prenons pour exemple qHEM et qLEU (globules rouges et globules blancs), bien connues de tout lecteur. Le croisement affiché par le programme "zrang", montre un nuage plutôt allongé dans la direction de la 1-ère diagonale: il y a, en effet, une corrélation positive (mais faible: .26) entre qHEM et qLEU. On notera d'autre part que l'amplitude de variation est nettement plus grande pour qLEU que pour qHEM; ce qui ne surprendra pas les cliniciens.



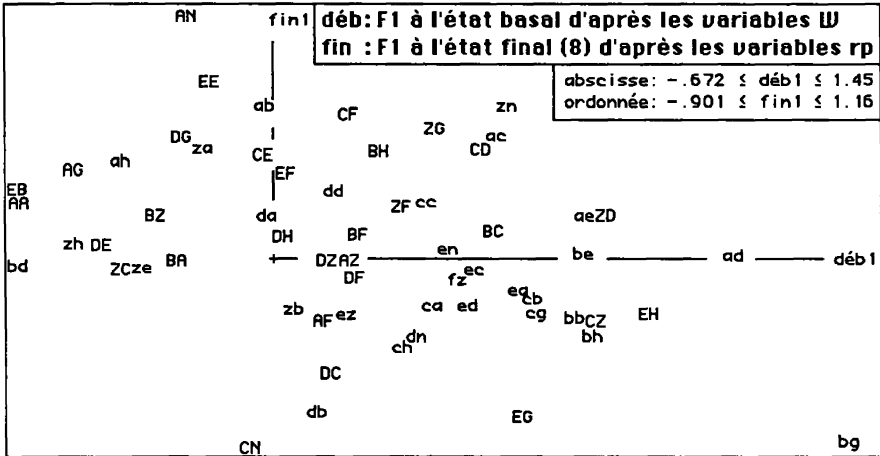
Cependant, seule l'analyse du tableau de correspondance (480 × 30) permettra de préciser l'existence d'un facteur général de cytopénie concernant toutes les classes de cellules: plus précisément, l'analyse des rapports montrera que toutes les variables de numération tendent à varier parallèlement (i.e. dans le même sens) d'un examen à l'autre.

```

A/TEMOIN chimiothérapie à J1 et J28 : rapports à l'état basal
A:D3:BIOrpDcodx:pivots pour le codage barycentrique des quotients
le nombre des variables est 10
qHEM (hématies)          a 3 mod dont les sigles et pivots sont
HEM< HEM= HEM>          7.634e-1 9.390e-1 1.129e+0
qHMC (hématocrite)      a 3 mod dont les sigles et pivots sont
HMC< HMC= HMC>          7.631e-1 9.463e-1 1.125e+0
qHMG (hémoglobine)      a 3 mod dont les sigles et pivots sont
HMG< HMG= HMG>          7.768e-1 9.499e-1 1.149e+0
qRET (réticulocytes)    a 3 mod dont les sigles et pivots sont
RET< RET= RET>          2.653e-1 8.456e-1 1.966e+0
qLEU (leucocytes)       a 3 mod dont les sigles et pivots sont
LEU< LEU= LEU>          2.888e-1 8.000e-1 1.517e+0
qNE% (neutrophiles/leuc) a 3 mod dont les sigles et pivots sont
NE%< NE%= NE%>         4.303e-1 9.830e-1 1.761e+0
qNEU (neutrophiles)     a 3 mod dont les sigles et pivots sont
NEU< NEU= NEU>          4.525e-2 7.844e-1 1.897e+0
qLY% (lymphocytes/leuc) a 3 mod dont les sigles et pivots sont
LY%< LY%= LY%>          3.137e-1 1.002e+0 1.882e+0
qLYM (lymphocytes)      a 3 mod dont les sigles et pivots sont
LYM< LYM= LYM>          4.584e-2 8.441e-1 1.680e+0
qTHR (thrombocytes)     a 3 mod dont les sigles et pivots sont
THR< THR= THR>          3.287e-1 8.726e-1 1.581e+0

```

Quant aux valeurs des pivots, on peut reprendre la remarque déjà faite, qu'elles sont le plus resserrées pour HEM, ainsi que HMC et HMG qui concernent directement l'ensemble des hématies; tandis que pour les réticulocytes (qui signent le renouvellement des hématies) elles sont très étalées.



Revenons à l'éventuelle liaison entre la variation relative de la cytopénie et la gravité de l'état basal. Anticipant sur la suite de l'exposé, nous disons que 'déb1' est d'autant plus élevé que l'état basal est meilleur (quant à la cytopénie); et que 'fin1' est d'autant plus élevé que l'évolution est plus favorable (ou, moins défavorable: une certaine cytopénie caractérisant, en moyenne, J56 relativement à J-7).

$$\begin{aligned} \text{corr}(\text{déb1}, \text{fin1}) &= -3.3824478e-1 \\ \text{fin1} - 6.7118536e-2 &= -3.0655433e-1 * (\text{déb1} - 2.1073571e-1) \\ \text{déb1} - 2.1073571e-1 &= -3.7321128e-1 * (\text{fin1} - 6.7118536e-2) \end{aligned}$$

Or le graphique, avec un nuage plutôt incliné suivant la diagonale descendante, suggère une corrélation négative entre déb1 et fin1: et le calcul (programme 'corel', cf. [CORREL. JUXT.], in *CAD*, Vol XIV, n°3, 1989), fixe cette corrélation à $-.338$; avec, entre déb1 et fin1, des formules de régression linéaire dont le coefficient est de l'ordre de $-1/3$. L'état final serait donc d'autant meilleur, relativement à l'état basal, que celui-ci était (dans l'absolu, variables W) plus mauvais: *a posteriori*, cette relation peut sembler naturelle parce que, à partir d'un état initial sans cytopénie, il n'y a pas lieu d'attendre une augmentation des numérations; laquelle signerait l'instauration d'une polycytémie...

Pour notre étude de l'effet hématoprotecteur, cette corrélation négative incite à chercher confirmation des résultats obtenus par l'analyse des rapports: ne doit-on pas craindre que l'efficacité apparue résulte simplement de ce que les sujets traités étaient, généralement, dans un état basal plutôt mauvais?

On peut répondre à cette question en concentrant notre attention sur l'analyse des valeurs (variables W) plutôt que sur celle des rapports (variables rp): d'après la fréquence des nombres de paires (A,t) pour lesquelles

$F1(A) < F1(t)$, il apparaît, en effet, que la différence entre traités (A) et témoins (t), très nette à l'état basal (J-7), n'est pas significative aux examens suivants: ce qui suggère que l'instauration du traitement hématoprotecteur, d'emblée efficace, compense l'infériorité initiale des sujets du groupe A.

Une autre voie a également été tentée: par un jeu de pondération des individus, guidé par l'histogramme de déb1, placé en tête du présent §, rétablir l'équilibre de la distribution de l'état entre traités et témoins. Les individus étant ainsi pondérés, les profils moyens /A et /t sont très proches l'un de l'autre sur l'axe 1 issu de l'analyse des valeurs (variables W); et l'on peut suivre avec confiance, dans l'une et l'autre analyse (W ou rp), la position relative des points moyens {1A, 1t}, {2A, 2t}, ..., {8A, 8t} afférents aux examens successifs des deux groupes A et t.

Il faut encore, s'interroger sur le choix des variables: d'une part, il y a redondance entre {HEM, HMC, HMG} qui, toutes trois, concernent les hématies; d'autre part, les pourcentages {LY%, NE%} (introduits d'abord fortuitement dans la présente étude multidimensionnelle) ne concernent pas directement la cytopénie, mais l'équilibre entre catégories de leucocytes.

En fait, l'analyse des variables W montre que, tandis que NE% est plus faible dans les formules marquées par la cytopénie, c'est le contraire pour LY% (qui est approximativement complémentaire de NE%; les catégories de globules blancs autres que LYM et NEU étant moins nombreuses que celles-ci). Et l'analyse des rapports (rp) montre que l'instauration d'une cytopénie s'accompagne communément d'une augmentation de LY% et d'une diminution de NE%.

Mais, quel que soit l'intérêt de ce résultat (sur lequel on reviendra), il semble que l'étude de la cytopénie, *stricto sensu*, demande qu'on fasse des analyses où figurent seulement en principales les modalités des 6 variables {HEM, RET, LEU, NEU, LYM, THR} et de leurs rapports.

Nous ne tenterons pas ici d'énumérer les nombreuses analyses effectuées pour la présente étude; et encore moins d'en rendre compte. Nous nous bornerons à quatre analyses: au §3, analyses des 30 modalités des 10 variables, pour les valeurs (W) et les rapports (rp), les individus recevant tous le même poids; au §4, analyses semblables mais avec 6 variables principales seulement, et des individus pondérés, afin d'équilibrer (comme on l'a dit) la distribution de l'état basal entre les deux groupes A et t.

Dans toutes ces analyses, on utilisera seulement les valeurs pivot publiées dans les tableaux des §§2.1 et 2.2 (pour les variables W et rp); valeurs choisies, ainsi qu'on l'a dit d'après les histogrammes construits pour l'ensemble non pondéré des examens. Nous avons parfois utilisé d'autres pivots, mais, les résultats étant concordants, nous ne retenons qu'un système de pivots.

3 Analyses avec 10 variables principales et l'ensemble des sujets non pondérés

Nous considérerons successivement l'analyse du tableau (540 × 30) décrivant les 540 hémogrammes (ou formules) par les modalités des variables mesurées (numérations et pourcentages) codées barycentriquement (variables W); et l'analyse du tableau (480 × 30) décrivant 480 formules par les modalités des variables quotient décrivant chaque formule par rapport à l'examen de base du même sujet (variables pr).

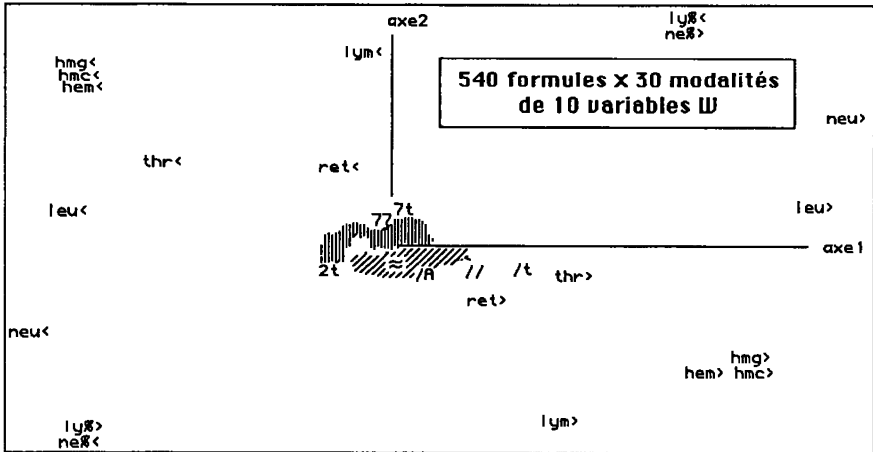
Dans les deux analyses, l'ensemble des 30 modalités sera projeté sur les graphiques. L'ensemble des individus (formules) ne peut être présenté en détail (sinon en se restreignant à l'état basal, comme on l'a fait pour l'histogramme du §2.3; ou, dans ce même §, pour le croisement de l'état basal avec l'état final). Mais on considérera, d'une part, des profils moyens représentant l'état d'un groupe (A ou t) à un examen de rang déterminé (e.g. rang 6, à J+42); et, d'autre part, afin d'apprécier le recouvrement de deux groupes, on dénumbrera la fréquence des nombres de paires (A,t) pour lesquelles $F1(A) < F1(t)$.

3.1 Tableau croisant 540 formules avec les 30 modalités des 10 variables mesurées (W)

A/TEMOIN chimiothérapie à J1 et J28:valeurs vraies codées barycentriquement
 trace : 9.054e-1 (540 examens)
 rang : 1 2 3 4 5 6 7 8 9 10
 lambda : 2355 1322 1067 895 728 582 450 439 324 283 e-4
 taux : 2601 1460 1179 989 805 643 497 485 358 313 e-4
 cumul : 2601 4061 5240 6229 7033 7676 8174 8658 9017 9329 e-4

SIGI	QLT	PDS	INR	F 1	CO2	CTR	F 2	CO2	CTR	F 3	CO2	CTR	F 4	CO2	CTR
ci	dessous	éléments	supplémentaires												
1A	190	56	2	-61	141	1	33	42	0	13	7	0	-6	1	0
1t	583	56	3	48	50	1	86	164	3	106	248	6	-75	122	3
2A	768	56	3	-170	604	7	-33	23	0	-3	0	0	-82	141	4
2t	909	56	4	-186	579	8	-100	166	4	53	48	1	-84	117	4
3A	467	56	2	-55	87	1	-64	120	2	86	216	4	39	44	1
3t	436	56	1	35	92	0	-60	275	2	3	1	0	30	67	1
4A	644	56	3	142	483	5	-39	36	1	30	21	0	66	104	3
4t	841	56	3	179	700	8	-30	19	0	-7	1	0	75	121	3
5A	652	56	2	-48	88	1	18	12	0	106	436	6	55	117	2
5t	299	56	2	17	11	0	86	263	3	8	2	0	-25	23	0
6A	718	56	2	-117	363	3	61	100	2	-78	164	3	-59	92	2
6t	794	56	3	-177	757	7	-11	3	0	-21	11	0	-31	24	1
7A	645	56	2	-102	322	2	55	92	1	-50	79	1	-70	151	3
7t	724	56	2	19	10	0	148	597	9	-54	79	2	37	37	1
8A	313	56	1	-18	25	0	17	24	0	-57	256	2	10	8	0
8t	415	56	1	72	300	1	36	74	1	7	3	0	25	37	0
/A	675	56	2	78	171	1	-112	352	5	-53	80	1	50	72	2
/t	955	56	9	344	824	28	-93	61	4	-89	55	4	45	14	1

D'après la suite des valeurs propres, nous bornerons au facteurs 1 et 2 l'interprétation de l'analyse.

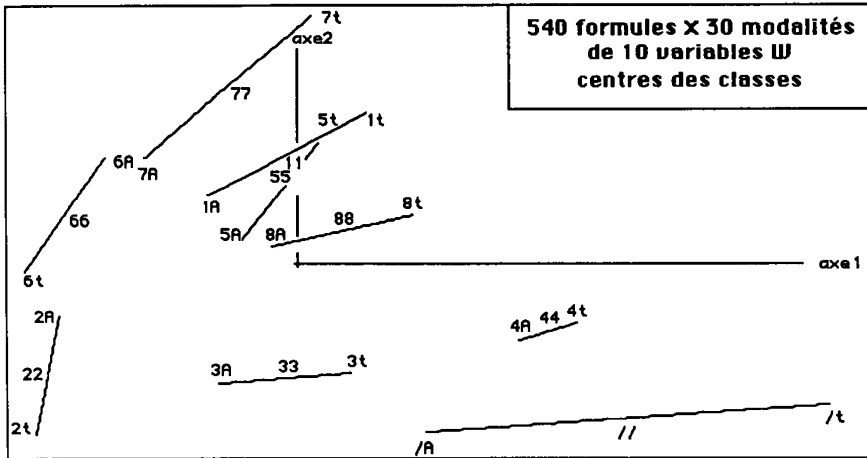


Sur l'axe 1, les modalités supérieures, situées du côté ($F1 > 0$), s'opposent aux modalités inférieures ($F1 < 0$): il n'y a, à cet énoncé général qu'une exception, créée par la variable LY%. Nous avons déjà dit que les faibles valeurs des variables de numération (ou de celles qui leur sont directement liées: HMC et HMG) caractérisent la cytopénie, d'autant plus marquée que $F1$ est plus faible (négatif); tandis que les pourcentages décrivent l'équilibre de la formule sanguine. Il apparaît ici que la cytopénie est généralement associée aux faibles taux de neutrophiles (rapportés à l'ensemble des leucocytes) et aux taux élevés de lymphocytes. Nous ne savons si cette relation est générale dans les cytopénies ou si elle est propre à celles induites par la chimiothérapie antimétabolique; nous ignorons aussi dans quelle mesure y participent les diverses classes de lymphocytes (T4, T8, ...).

Au centre du plan (1,2) sont confinées toutes les modalités moyennes (\approx); tandis que la qualité de la représentation est bonne pour toutes les modalités {<, >}, excepté pour les réticulocytes ($COR1 + COR2 < 8\%$). Les modalités de même nom des variables {HEM, HMC, HMG} sont quasi superposées; il en est ainsi pour les modalités opposées de NE% et LY%, dont on a noté le caractère complémentaire.

Dans le demi-plan ($F1 > 0$) s'opposent, suivant la direction de l'axe 2, {ly%<, ne%>} et {hem>, hmc>, hmg>}; de même, s'opposent dans le demi-plan ($F1 < 0$) {ly%>, ne%<} et {hem<, hmc<, hmg<}. Bien que l'ordre des modalités de LYM, {lym<, lym \approx , lym>}, soit le même, sur l'axe 1, que pour les autres numérations, c'est principalement suivant la direction de l'axe 2 que lym< s'oppose à lym>.

Au centre du plan (1,2) des modalités, on a marqué ceux des centres de classes de formules qui sont le plus écartés de l'origine {/t, 7t, 2t}: on



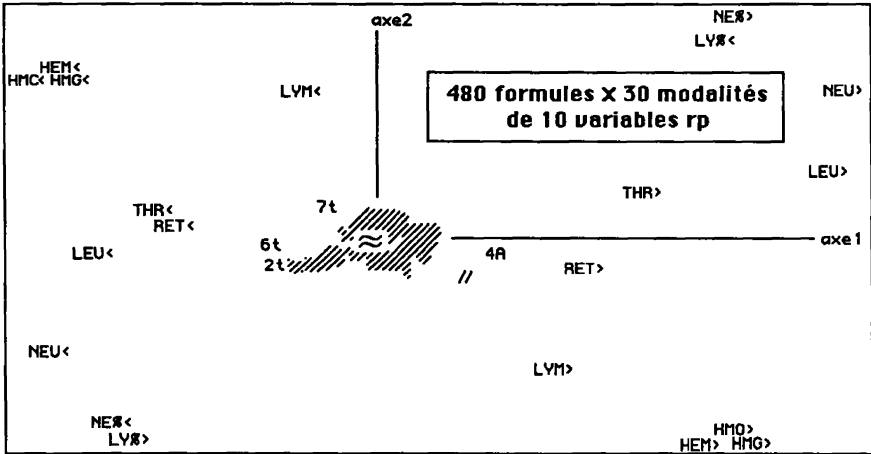
appréciera ainsi la faible dispersion du nuage des centres, relativement à celle du nuage des formules, qui est la même que pour le nuage des modalités.

Représenté seul sur toute l'étendue du plan (1,2), le nuage des centres montre que les deux groupes A et t sont proches (particulièrement en projection sur l'axe 1) dans toute suite des examens, ne se séparant nettement qu'à l'état basal. Le grand allongement du segment {/A, /t} suivant la direction de l'axe 1, atteste que, comme on l'a annoncé au §2.3, les sujets traités (A) se distinguent, en moyenne des témoins (t) par une cytopénie à l'état basal.

On notera qu'au cours du traitement, la cytopénie est la plus nette environ deux semaines après le début d'une cure antimétabolique, i.e. à J14 et J42 {22, 66}; au début de la deuxième cure, J28={44}, l'on est à peu près revenu au niveau basal {/}; mais à J56={88}, un résidu de cytopénie subsiste relativement à l'état basal moyen.

date de l'examen	J-7	J07	J14	J21	J28	J35	J42	J49	J56
paires F1(A)<F1(t)	593	503	428	498	482	459	416	515	487
≈fréq d'un écart >	≈5%	>33%	>50%	>33%	>50%	>80%	>50%	>25%	>50%

Afin d'apprécier dans quelle mesure sont significatifs les décalages entre centres de classes d'examen afférents aux deux groupes A et t, on a considéré la répartition sur l'axe 1 de ces classes elles-mêmes. Prenons pour exemple les examens à J35: on a un sous-nuage i5 de 60 formules, 30 pour A et 30 pour t: sur les 900 paires {t,A}, on en dénombre 459 pour lesquelles F1(A)<F1(t): ce nombre ne s'écarte certes pas significativement de 450; on peut donc poser F1(5A)≈F1(5t). Le tableau, construit d'après une simulation, indique que seul est significatif l'écart à l'état basal (J-7: /A ≠/t).



3.2 Tableau croisant 480 formules avec les 30 modalités des 10 variables rapportées à l'état basal (rp)

A/TEMOIN chimiothérapie à J1 et J28 : rapports à l'état basal

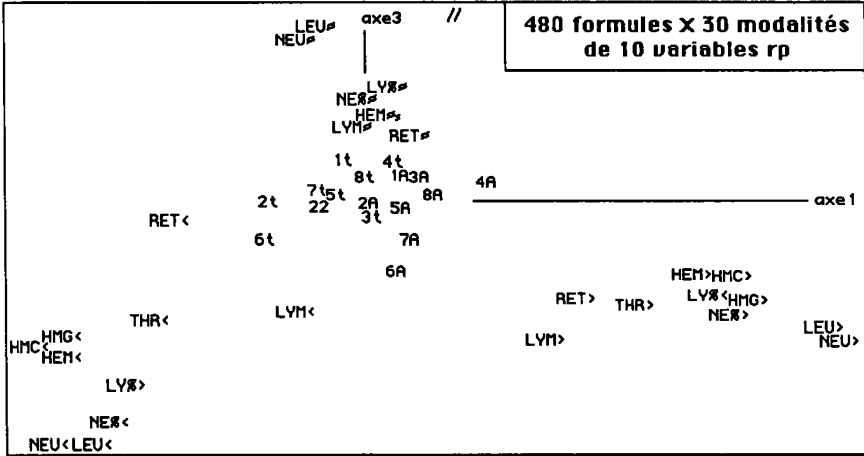
trace :	1.055e+0 (480 formules)									
rang :	1	2	3	4	5	6	7	8	9	10
lambda :	2669	1581	1132	1000	956	647	532	443	389	320 e-4
taux :	2530	1499	1073	948	907	613	504	420	369	303 e-4
cumul :	2530	4029	5102	6050	6957	7570	8074	8495	8863	9166 e-4

Dans le plan (1,2) la disposition des modalités diffère peu de celle trouvée au §3.1. Mais il faut souligner que la signification n'en est pas exactement la même; car il s'agit maintenant non des modalités des formules mais des modalités de leurs variations.

L'axe 1 exprime donc le fait qu'en moyenne, relativement à l'état basal, les numérations varient dans le même sens: soit vers une instauration ou une aggravation de la cytopénie, si la formule se projette du côté ($F1 < 0$); soit vers une réduction de la cytopénie ($F1 > 0$); l'amplitude de ces variations relatives étant donnée par les valeurs pivot qui définissent les modalités (ou, avec plus de précision, par les histogramme de rapports, tel que celui donné pour HEM au §2.2).

La qualité de représentation des modalités {<, >} est meilleure qu'au §3.1 pour les réticulocytes: ce qu'on peut expliquer parce que le taux de ces cellules jeunes caractérise directement le dynamisme de la formule sanguine, objet de la présente analyse.

Quant aux pourcentages, l'aggravation de la cytopénie s'accompagne d'une diminution de la part des neutrophiles dans les leucocytes du sang circulant; tandis que, par complémentarité, augmente la part des lymphocytes.



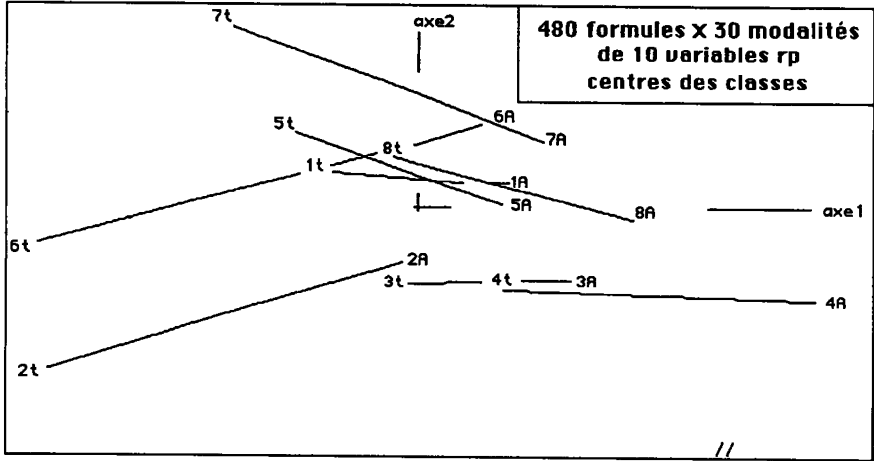
En analogie avec le §3.1, on voit que, du côté ($F1 > 0$) s'opposent, suivant la direction de l'axe 2, $\{LY\%<, NE\%>\}$ et $\{HEM>, HMC>, HMG>\}$; de même, s'opposent, du côté ($F1 < 0$), $\{LY\%>, NE\%<\}$ et $\{HEM<, HMC<, HMG<\}$; et, bien que l'ordre des modalités $\{LYM<, LYM\approx, LYM>\}$, soit le même, sur l'axe 1, que pour les variations des autres numérations, c'est principalement suivant la direction de l'axe 2 que $LYM<$ s'oppose à $LYM>$.

Bien que la considération de l'axe 3 ne nous ait rien apporté pour l'interprétation biologique, il vaut la peine de remarquer que les modalités moyennes $\{=\}$, situées du côté ($F3 > 0$), s'opposent à toutes les autres modalités $\{<, >\}$, situées du côté ($F3 < 0$). Dans le plan (1,3), on a ainsi une sorte d'effet Guttman, les modalités étant groupées en trois amas:

$$\{=\}: (F3 > 0, F1 \approx 0) \quad ; \quad \{<\}: (F3 > 0, F1 < 0) \quad ; \quad \{>\}: (F3 < 0, F1 > 0) \quad ;$$

font seules exception à ce schéma $LY\%<$ qui va avec $\{>\}$; et $LY\%>$ qui va avec $\{<\}$; parce que, comme on l'a dit, l'importance numérique relative des lymphocytes varie dans le même sens que la cytopénie.

Le nuage des centres de classes de formules est, comme au §3.1, beaucoup moins dispersé que celui des 480 individus: sur l'axe 1, l'inertie des $\{xA, xt\}$ ne représente que 8% de l'inertie totale (comme on peut le calculer en faisant la somme des CTR1, donnés par le listage des éléments supplémentaires); mais, comme on peut le voir sur le plan (1,3) et, plus clairement, dans le graphique du plan (1,2) où les centres de classes sont seuls figurés, l'on a pour tous les examens, de $x=1$ à $x=8$: $F1(xA) > F1(xt)$. Une telle concordance est très significative car, parmi les 256 combinaisons possibles des positions relatives de 8 paires, elle ne peut se réaliser que de deux manières (toujours $A < t$, ou toujours $A > t$).



On comparera aux centres de classes le point //, qui représente l'état basal, c'est-à-dire une suite de rapports tous égaux à 1.

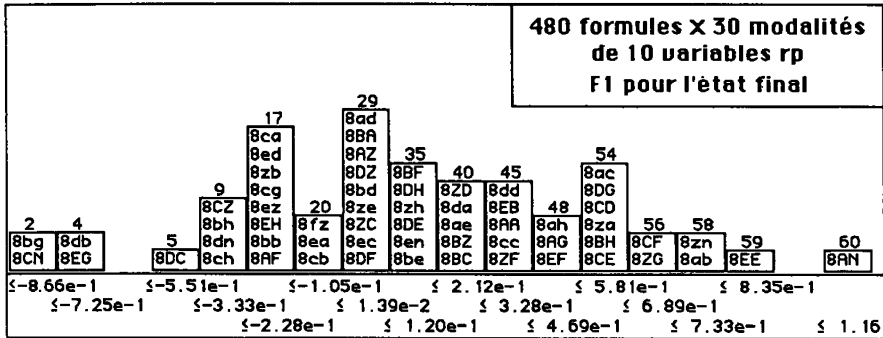
En suivant les projections sur l'axe 1, on voit que seul F1(4A) dépasse F1(//); ce qui correspond à une amélioration (diminution de la cytopénie) relativement à l'état basal: cette amélioration se manifeste également au §3.1, avec un autre codage des données, par le fait que $F1(4A) > F1(/)$.

À la fin de l'essai, à J56, le point 8A est un peu en retrait relativement à //: ce qui indique qu'en moyenne, malgré l'effet hématoprotecteur du traitement, les patients se ressentent encore de l'effet hématotoxique de deux cures antimétaboliques successives (débutées à J1 et à J28). Ici encore, les résultats du §3.2 concordent avec ceux du §3.1, où $F1(8A)$ est quelque peu inférieur à $F1(/)$.

date de l'examen	J07	J14	J21	J28	J35	J42	J49	J56
paires $F1(A) < F1(t)$	407	339	389	325	385	305	353	331

Comme au §3.1, on doit interpréter les décalages entre centres de classes en fonction de la dispersion des classes et du recouvrement, à chaque examen, des distributions des traités et des témoins. En dénombrant, à chaque date d'examen, les paires {traité, témoin} pour lesquelles $F1(A) < F1(t)$, on trouve des nombres toujours inférieurs à 450, nombre moyen correspondant à une distribution aléatoire, ce qui confirme le décalage systématique des centres de classes observé sur les graphiques.

De ces nombres, toutefois, seul celui à J42 a, avec 450, une différence significative au seuil de 5%: mais, d'une part, ainsi qu'on l'a dit plus haut, la concordance des signes des 8 différences est certes significative (même si il n'y



a pas indépendance entre ces différences, parce qu'il s'agit de l'évolution continue des deux mêmes groupes de 30 sujets); et, d'autre part, il est particulièrement satisfaisant qu'un net effet protecteur se manifeste précisément à J42, date à laquelle est le plus ressentie la cytotoxicité de la deuxième cure antimétabolique.

Un histogramme permet d'observer, dans ses détails, le recouvrement des distributions des deux groupes, considérés à J56 (examen n°8): nous rappellerons que les sigles des traités sont en capitales, et ceux des témoins en bas de casse.

On se souvient, cependant, (cf. §§2.3 et 3.1), que l'affectation des patients aux groupes (A) et (t) semble avoir été tout autre qu'aléatoire, les sujets traités ayant un état basal significativement moins bon que celui des témoins; et que l'état final est, en moyenne, d'autant meilleur, relativement à l'état basal, que celui-ci était (dans l'absolu, variables W) plus mauvais. C'est pourquoi, au §4, nous tenterons d'équilibrer les deux groupes en pondérant les sujets.

SIGJ	QLT	PDS	INR	F 1	CO2	CTR	F 2	CO2	CTR	F 3	CO2	CTR	F 4	CO2	CTR
ci	dessous	éléments		supplémentaires											
1A	555	62	2	66	144	1	17	9	0	78	205	3	-77	197	4
1t	327	63	4	-75	92	1	27	12	0	108	189	6	-45	33	1
2A	586	62	2	-3	0	0	-33	36	0	-34	40	1	-123	509	9
2t	906	62	6	-270	764	17	-106	119	4	-5	0	0	-47	23	1
3A	588	63	2	115	405	3	-47	67	1	61	117	2	1	0	0
3t	369	63	1	-20	26	0	-48	151	1	-35	83	1	-40	109	1
4A	906	63	6	289	816	20	-60	35	1	50	25	1	55	30	2
4t	481	63	2	56	111	1	-46	76	1	90	291	5	10	4	0
5A	524	62	1	69	191	1	2	0	0	-25	24	0	88	309	5
5t	619	62	2	-95	276	2	54	87	1	14	6	0	91	250	5
6A	783	63	3	56	55	1	60	61	1	-195	659	21	21	8	0
6t	919	62	6	-276	792	18	-26	7	0	-106	117	6	-18	3	0
7A	573	63	2	93	222	2	45	52	1	-108	298	6	-6	1	0
7t	691	63	3	-139	328	5	124	261	6	28	13	0	72	89	3
8A	664	62	2	155	595	6	-2	0	0	15	5	0	51	64	2
8t	291	62	1	-21	18	0	39	61	1	63	164	2	-34	48	1
//	829	125	46	213	116	21	-154	61	19	505	652	282	18	1	0

4 Analyses avec 6 variables principales et l'ensemble des sujets pondérés

Ainsi qu'on l'a dit au §2.3, on met en supplémentaire les deux variables de pourcentage, {NU%, LY%}, qui ne concernent pas directement la cytopénie; ainsi que {HMG, HMC}, qui font redondance avec la numération des hématies, HEM.

Quant aux poids des sujets, ils sont choisis d'après l'histogramme du §2.3, qui montre la distribution de l'état basal pour les deux groupes. Les poids varient de (1/4) à 2. À l'extrémité droite de l'histogramme, pour les états de base les meilleurs, on attribue pour poids (1/4) aux témoins et (1/2) aux traités; ainsi, d'une part, on rend égales les densités des deux groupes; et, d'autre part, on restreint l'importance des sujets dont l'état basal est bon et pour lesquels le traitement hématoprotecteur semble moins utile. À l'extrémité gauche, au contraire, pour les états de base les plus compromis, on attribue pour poids 2 aux témoins et 1 aux traités. Dans les zones centrales de l'histogramme, les pondérations nous servent aussi pour rendre égales et régulières les densités des deux groupes. Il va sans dire que les poids choisis pour les sujets sont attribués à chacun des examens de ceux-ci.

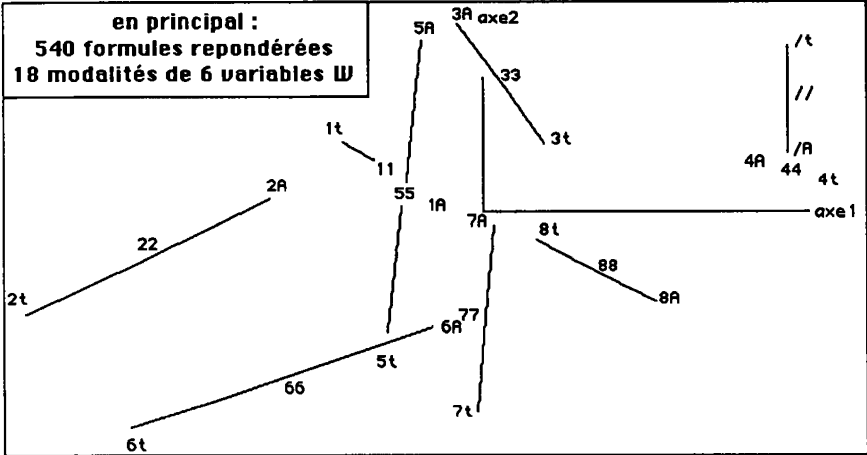
Il est clair qu'un tel choix des pondérations est dépourvu de rigueur: mais il nous offre, des données, une image complémentaire, prise sous un nouvel angle, et qui suggère des hypothèses sur ce qu'aurait pu apporter l'étude si les groupes avaient été bien équilibrés. Une autre voie aurait été d'entreprendre des analyses partielles, restreintes aux individus compris dans une bande assez étroite de l'histogramme de l'état basal; mais le faible nombre des sujets nous a dissuadé d'aller dans cette voie.

4.1 Tableau principal croisant 540 formules pondérées avec les 18 modalités des 6 variables mesurées (W)

A/TEMOIN chimiothérapie à J1 et J28:valeurs codées barycentriquement
 trace : 9.126e-1 individus pondérés Π
 rang : 1 2 3 4 5 6 7 8 12
 lambda : 2523 1268 1012 908 722 709 568 513 21 e-4
 taux : 2765 1389 1109 995 792 777 622 562 23 e-4
 cumul : 2765 4154 5264 6259 7050 7827 8449 9011 10000 e-4

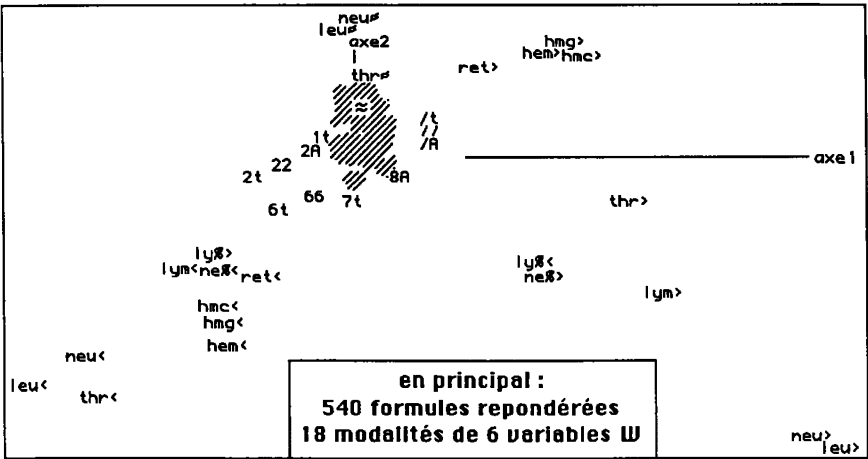
On notera qu'avec 6 variables codée chacune sur 3 modalités, la dimension du nuage est exactement 12.

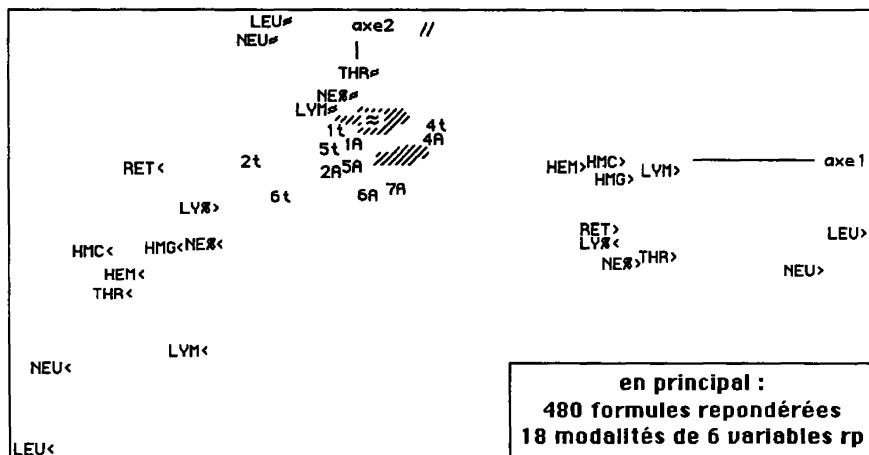
Quant à la répartition des modalités principales et supplémentaires, il apparaît que l'interprétation de l'axe 1 reste celle du §3.1; et que les mêmes agrégats de modalités se retrouvent dans le plan (1,2). Même quand les pourcentages sont en supplémentaires, on voit la complémentarité entre NE% et LY% ainsi que l'association de la cytopénie avec de faibles taux de neutrophiles (rapportés à l'ensemble des leucocytes). Et la redondance de {HEM, MMC, HMG} est confirmée.



Pour la distribution des formules sanguines, on observe d'abord que la pondération a produit l'effet attendu: les centres des examens initiaux, {/A, /t}, ne diffèrent pas quant au facteur F1, et se projettent même en des points voisins du plan (1,2).

Aux examens ultérieurs, le décalage entre témoins et traités ne se fait pas toujours dans le même sens; mais chaque fois que le décalage est net, il est en faveur des sujets traités: c'est, en particulier le cas à J14 et J42, alors que l'hématotoxicité de la cure antimétabolique est le plus marquée chez les témoins ($F1(2t) \ll 0$; $F1(6t) \ll 0$); et également à J56, alors que l'essai analysé s'achève avec une moindre cytopénie pour les traités que pour les témoins.



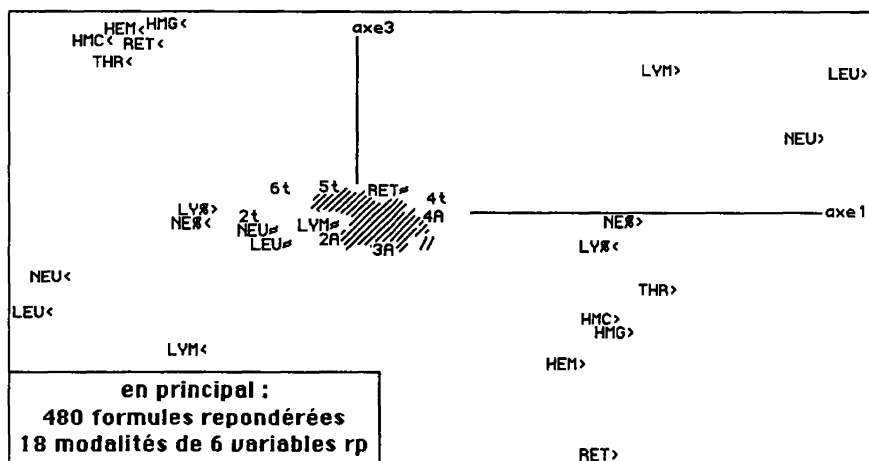


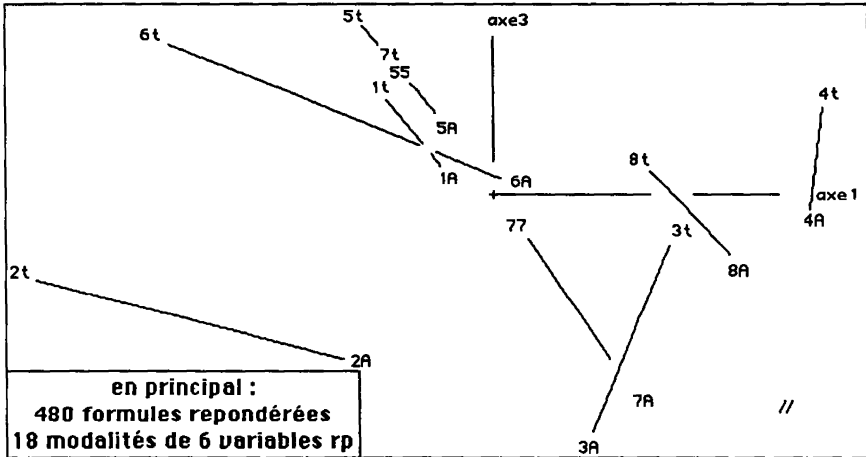
4.2 Tableau principal croisant 480 formules pondérées avec les 18 modalités des 6 variables rapportées à l'état basal (rp)

A/TEMOIN chimiothérapie à J1 et J28 : rapports à l'état basal

trace :	1.089e+0 formules pondérées : Π											
rang :	1	2	3	4	5	6	7	8	12		
lambda :	2987	1419	1242	984	940	834	746	662	69	e-4	
taux :	2742	1303	1140	903	863	765	685	608	63	e-4	
cumul :	2742	4045	5186	6089	6952	7717	8402	9010	10000	e-4	

L'interprétation de l'axe 1 est celle donnée au §3.1; et, avec trois amas de modalités, le plan (1,2) issu de la présente analyse reproduit le plan (1,3) sur lequel, au §3.2, on a reconnu un effet Guttman: on doit donc se placer dans le





plan (1,3) pour étudier les agrégats de modalités.

Quant aux centres de classes, on retrouve les conclusions du §4.1; d'autant plus clairement que, l'état initial des groupes étant maintenant équilibré, la position relative pour les profils rapportés à l'état basal (variables rp) est la même que pour les profils pris tels quels (variables W).

Nous répétons qu'aux examens successifs, de J7 à J56 (1 à 8), le décalage entre témoins et traités ne se fait pas toujours dans le même sens; mais que, chaque fois que le décalage est net, il est en faveur des sujets traités. C'est, en particulier le cas à J14 et J42, alors que l'hématotoxicité de la cure antimétabolique est la plus marquée chez les témoins ($F1(2t) < 0$; $F1(6t) < 0$). De façon précise, il n'y a que deux exemples d'inversion, $F1(3A) < F1(3t)$ et $F1(4A) < F1(4t)$; et tous deux concernent des centres situés du côté positif de l'axe 1, i.e. là où la cytopénie est absente ou modérée.

5 Conclusions et perspectives

Malgré le nombre restreint des sujets et leur répartition déséquilibrée entre les deux groupes de traités et de témoins, nous croyons pouvoir tirer de l'essai une conclusion très favorable au produit hématoprotecteur A. En effet, de multiples analyses, fondées sur des codages différents, et étayées, dans la mesure du possible, par des épreuves de validité robustes, ont toutes apporté des résultats concordants.

Quant à la méthodologie, l'analyse des correspondances, appliquée à des tableaux construits par codage barycentrique, a nettement mis en évidence le déséquilibre des groupes et aidé à y remédier.

Pour la biologie, le présent essai a le grand mérite d'offrir à l'analyse statistique, des suites d'examen hématologiques effectués sur le même patient à un rythme régulier, de semaine en semaine. Ce sont précisément de telles données que nous recherchions pour caractériser la cinétique du système immunitaire (cf. [DISCUSSION PASTEUR], §3, in *CAD*, Vol XV, n°3, 1990). L'analyse a fourni un premier facteur associant les numérations de toutes les classes de cellules; la cytopénie est, d'autre part, généralement associée aux faibles taux de neutrophiles (rapportés à l'ensemble des leucocytes) et aux taux élevés de lymphocytes

Les données analysées ici ne sont toutefois pas assez fines pour nous satisfaire, car elles ne distinguent pas entre les diverses classes de lymphocytes; et n'étudient la réponse du système qu'à une agression particulière, la cytotoxicité des antimitotiques, à l'exclusion de toute réponse immunitaire proprement dite (car les sujets ayant présenté une infection intercurrente sont exclus).

Pour la thérapeutique, on retiendra d'abord l'intérêt du produit A dans l'indication qui fait l'objet de l'essai. Mais, au-delà de l'essai, il nous paraît légitime de proposer le produit aux sujets atteints d'un *Syndrome ImmunoDéficient Acquis*, c'est-à-dire, en quatre lettres, du SIDA.

Références bibliographiques

[DISCUSSION PASTEUR]: "Compte rendu d'une discussion entre informaticiens et spécialistes du SIDA, après un exposé d'analyse des données"; in *CAD*, Vol XV, n°3, pp. 305-322; (1990).

[CODAGE LIN.]: "Le codage linéaire par morceaux: réalisation et applications", in *CAD*, Vol XIV, n°2, pp. 203-210; (1989).

[CORREL. JUXT]: "Calculs de corrélation entre variables et juxtaposition de tableaux", in *CAD*, Vol XIV, n°3, pp.347-354; (1989).

Remerciements

Les auteurs expriment leur reconnaissance à Monsieur Pierre FLECCHIA, Président Directeur Général et Pharmacien Responsable des Laboratoires BIOSTABLEX-URAP (89 Rue LAURISTON -75116, Paris), qui leur a confié les données de la présente étude.