PERGAMON

Journal of Steroid Biochemistry & Molecular Biology 74 (2000) 195–202

The Journal of Steroid Biochemistry & Molecular Biology

www.elsevier.com/locate/jsbmb

Deaths from breast cancer: tackling multidimensionality and non-linearity by correspondence analysis

Daniel Hans [a], Tiiu Ojasoo [b,*], Jean-Christophe Doré [c]

[a] *Académie Méditerranéenne d'Oncologie Clinique, Marseille, France*
[b] *4bis rue René Barthélémy, 92120 Montrouge, France*
[c] *ESA 8041 CNRS, Muséum National d'Histoire Naturelle, Paris, France*

Received 7 December 1999; accepted 4 July 2000

Abstract

Aim: We investigated the use of non-linear, multidimensional factor analysis for the study of observational data on death from breast cancer. These data were obtained in the context of a clinical practice and not in a clinical trial. We looked into the correlations between patient characteristics and time of death and/or disease-free interval. *Patients and methods*: We first analyzed the characteristics of a population of patients that had died from breast cancer ($n = 295$), then of a population including patients still alive 7 years after surgery ($n = 344$). We used correspondence analysis (CA) which is based on χ^2-metrics, does not assume linear relationships, and provides graphic overviews. *Results*: The CA mapped variables (clinical stage, histoprognostic grade, node status, receptor positivity) in a way that fits in well with available knowledge on their importance as prognostic factors. We observed, however, that death occurred during three main periods (1–3, 4–7, ≤ 8 years after surgery) defined by different mixes of variables as if the disease progressed by stage rather than continuously. The CA distinguished long-term survivors (> 7 years) from patients who died 8–10 years after surgery. Long-term survivors tended to be node-negative; those who died at 8–10 years tended to be the youngest patients (under 40). *Conclusions*: Because correspondence analysis combines the advantages of multidimensional and non-linear methods, it is a valuable exploratory tool for describing multiple correlations within a population before attempting to establish statistical significance of selected variables by more classic methods.

Keywords: Breast cancer; Correspondence analysis; Multivariate analysis; Prognostic factor; Steroid receptor; Survival

1. Introduction

Clinical trials relating to prognostic factors and survival in breast cancer patients tend to follow a fairly standard pattern. Increasingly, they are prospective randomised controlled trials (RCT) or cohort studies, whose results are usually analysed by accepted statistical methods such as Cox' proportional hazards regression model to assess the relative importance of the chosen covariates, the Kaplan–Meier method to estimate survival, and the log-rank test to compare survival curves. This type of approach is considered by physicians and biometricians alike to provide some of the best and strongest evidence.

Although this may often be the case, it is unfortunate when a single prevailing paradigm tends to overshadow other ways of obtaining evidence. There are important questions that need to be addressed such as: How can we learn about the natural history of diseases? How can we best extract useful information from retrospective data? RCTs tell us about selected populations, but how can we find out more about individuals and their profiles? Moreover, are the statistical methods currently in use always the most appropriate? The Kaplan–Meier estimator is a non-parametric estimate of survival distribution but it only considers a single covariate at a time. The Cox' proportional hazard method assumes that the logarithm of the hazard rate is a linear function of the covariates, but is this true as often as we are led to believe? The assumption of linearity can be examined [1] and fit can be improved, but why rely exclusively on

* Corresponding author. Tel./fax: +33-1-42537981.
E-mail address: tiiu.ojasoo@free.fr (T. Ojasoo).

0960-0760/00/$ - see front matter
PII: S0960-0760(00)00123-0

a linear model for the study of a complex dynamic system such as the development and progression of breast cancer? Since complex systems are rarely linear [2], why assume that the time-schedule governing disease development is continuous and linear? Might not the effects of covariates change over time?

Our contention is that before embarking upon lengthy and costly clinical trials, it might be useful to perform more preliminary exploratory research on retrospective data collected in clinical practice. Such data does contain evidence; it is part of the basis of our experience. The fundamental question is how to extract this evidence and how to judge its relevance.

In the present paper, we have looked at observational data collated over many years on breast cancer patients. We analysed these data using correspondence analysis (CA) [3–6], a factor analysis which has analogies with principal components analysis (PCA). Whereas PCA uses covariance, CA is based on χ^2-metrics and is thus ideally suited to the analysis of contingency tables. Advantages of CA are that it (i) provides simple graphic overviews of complex systems [7], (ii) does not arbitrarily choose an independent variable, and (iii) does not assume linearity [8].

CA is in increasingly widespread use in a variety of scientific fields essentially outside the realm of medicine. In the cancer literature, we are aware of a couple of early papers [9,10] and of a mention in a review [11] referring to results in abstract form only [12]. More recently, CA was used to select variables for building a predictive index of axillary nodal involvement in operable breast cancer [13] and in a study of patients with a high-risk of relapsing from prostate cancer [14]. We have provided a brief introduction to CA as a mapping aid to clinical judgment [15] and already used CA to analyse (i) structure-activity data on potential anti-cancer agents and other molecules [6,16,17] and (ii) changes in plasma hormones with treatment and age in patients with prostate cancer [18,19]. Here we emphasise its role as a research tool to describe patient populations.

2. Patients and methods

The medical files of patients who had undergone surgery for breast cancer (mastectomy or tumorectomy) at the Jean Paoli-Irène Calmettes Institute in Marseilles (Department of the late Professor J.M. Spitalier and of Dr D. Hans) between May 1976 and May 1988 were retrospectively reviewed in order to count (i) the number of women deceased from cancer within 1, 2, 3, etc., years of surgery and (ii) the number alive with a disease-free survival (DFS) of more than 7 years. All the women had been followed up on a regular basis either at the Institute or in a private oncology clinic for at least 3 years (unless deceased earlier) and a maximum of 15 years.

For each year of death and for a disease-free survival of more 7 years, the number of patients in each of several categories was recorded (Table 1). These categories were based on age (four categories), thermography results (hot or cold), tumor location (uni- or bilateral), clinical disease stage (I–IV), number of invaded lymph nodes (seven categories), histoprognostic SBR (Scarff-Bloom and Richardson) tumor grade (I–III), and, when available, estrogen (ER), progesterone (PR), and androgen (AR) receptor positivity. Receptors had been assayed in a single regional laboratory using a routine radioligand assay; AR assays were initiated in 1981 [20]. Most importantly, for each variable, a 'missing data' category was included to check whether these data were indeed missing at random (i.e. were not selected in any way) and thus did not bias the study.

2.1. Multivariate factor analysis

Correspondence analysis (CA) is a method of multidimensional data reduction which uses χ^2-metrics and which is therefore highly suited to the analysis of categorical variables. The most common application of CA is the analysis of frequency (contingency) tables such as Table 1. In the first seven columns of Table 1, the 295 women who died of breast cancer have been classified according to several variables (traits) and by year of death. The data of this part of the table were converted into percentages (total of each row = 100%), then into χ^2-distances between each trait and year of death. The resultant symmetrical semi-square probability matrix was analysed by CA (for full mathematical details, see Refs. [3–6]) using an in-house program written by one of us (J.C.D.). (CA is nowadays included in commercial software packages from SAS Institute; BMDP Statistical Software; SPSS …).

CA uses the original variables (traits) to calculate new variables (factorial axes) which are different mixes of the original variables (i.e. composite traits). Unlike the initial variables, the composite traits are orthogonal, i.e. each factorial axis describes an independent chunk of information. Furthermore, CA filters total information by breaking it down into a ranked series of $n-1$ factorial axes of decreasing inertia τ ($\varphi_1,\varphi_2,\varphi_3,\ldots,\varphi_{n-1}$ where n is the number of matrix columns).

CA is a graphic method. We can plot the original variables using their calculated coordinates to the factorial axes. Moreover, we can plot both row and column variables (traits and death years) on the same graph because the χ^2-distance matrix is symmetrical. (An interesting option not addressed in this paper is to depict patients and traits on the same plot [19]). In a factorial plot, the more clustered the variables, the stronger the

correlations between them (as long as the variables contribute substantially to the chosen factorial axes). The absolute contribution (AC) of a variable to a factorial axis is its degree of involvement in the constitution of the axis (ΣACs of all variables = 100% for any axis); the relative contribution (RC) of a variable tells us how its information content is distributed over the axes (ΣRCs of each variable to all axes = 1). A theoret-

Table 1
Retrospective data on 639 breast cancer patients (295 dead, 344 alive)[a]

	Number of deaths								DFS >7 years
	1 year	2 years	3 years	4 years	5 years	6–7 years	8–10 years	Total	Total
Age									
<40 years	4	5	10	5	3	6	6	39	36
40–55 years	11	12	23	12	8	12	4	82	142
55–70 years	10	15	37	14	7	14	5	102	123
>70 years	16	11	14	2	1	5	1	50	27
Missing	6	9	3	1	1	2	0	22	16
Thermography									
1.2 (cold)	9	9	10	7	6	10	3	54	80
4.5 (hot)	30	35	60	18	10	27	7	187	153
Missing	8	8	17	9	4	2	6	54	111
Laterality									
Unilateral	42	45	78	29	16	33	12	255	309
Bilateral	5	7	9	5	4	6	4	40	35
Stage									
I	14	17	25	17	14	18	9	114	218
II	11	20	33	12	4	14	7	101	99
III	11	13	25	5	1	4	0	59	26
IV	11	2	4	0	1	3	0	21	1
Tumor grade									
I	3	4	6	4	2	7	3	29	96
II	19	22	40	19	7	17	7	131	150
III	22	20	39	9	7	10	4	111	45
Missing	3	6	2	2	4	5	2	24	53
Lymph nodes									
0	5	6	5	8	6	5	0	35	93
1	1	3	3	2	1	1	1	12	38
2	2	2	9	3	2	2	3	23	20
3–4	3	3	13	2	2	7	3	33	5
5–7	2	8	11	4	1	1	0	27	8
8–10	3	4	7	2	0	0	0	16	1
>10	11	5	12	2	1	1	0	32	1
Missing	20	21	27	11	7	22	9	117	178
Steroid receptors									
ER^-	18	18	25	8	5	6	2	82	49
ER^+	21	24	49	18	11	18	1	142	105
ER missing	8	10	13	8	4	15	13	71	190
PR^-	28	19	46	12	9	9	1	124	72
PR^+	10	23	26	14	6	12	2	93	59
PR missing	9	10	15	8	5	18	13	78	213
AR^-	17	16	30	4	2	2	2	73	NA
AR^+	1	2	1	1	0	1	0	6	NA
AR missing	29	34	56	29	18	36	14	216	NA
Supplementary data									
ER^-PR^-	15	15	22	6	4	4	0	66	
ER^+PR^-	11	12	23	6	5	5	1	63	
ER^+PR^+	4	9	24	12	5	10	0	64	

[a] DFS, disease-free survival; NA, not available. The cut-off levels for positivity were ER >10 fmol/mg prot., PR ≥10 fmol/mg prot., AR ≥50 fmol/mg prot.

ical index (λ) indicates how well founded are the relationships among the variables for a given axis. A maximum value of 1 for λ means that one group of variables under study is totally distinct from the others; a value for λ that approaches zero is a sign of a random distribution.

In summary, CA produces plots of the inherent structure of a complex system described by a dataset by stratifying the information within this dataset according to its inertia (creation of factorial axes) and by highlighting correlations (linear or not) among all variables (creation of factorial plots).

3. Results

Our correspondence analyses concern 295 women who had died of cancer, and 344 who had survived for more than 7 years after surgery. At the time of analysis, 38 women were lost to follow-up and 17 had died of unknown causes. The remainder were alive but had not yet reached 7 years of survival. The median follow-up was 10 years.

3.1. Profiles of the patients who died

The break-down of the 295 deaths from breast cancer according to time of death (columns) and traits (rows) is given in Table 1. The results of the CA of the first seven columns of this table (excluding supplementary steroid receptor data) are shown in the $\varphi_1\varphi_2$ factorial plot of Fig. 1 which depicts 70% of the total inertia (φ_1: 52.7%; φ_2: 17.3%). This is the most high-ranking, coherent and discriminatory information; quirks and noise have been relegated to lower axes. The overall plot has been subdivided into five superimposable panels (A–E) for the sake of readability. The relative positions of the column variable, year of death, are shown in panel 1A and those of row variables (traits) in panels 1B–E. Mock variables (missing data categories in Table 1), although they were included in the calculation, are not shown. Missing data did not constitute a bias in the study (i.e. were not selected) but occurred randomly.

In panel 1A, the years of death follow a typical Guttman curve [21]. Their projections on the φ_1 axis are in perfect chronological order with a single inversion. Year 3 comes before year 2. These years are close because the studied variables can hardly distinguish them. The projections of death years 4, 5 and 6–7 onto the φ_1 axis are also clustered. It thus seems that we are dealing with three major time-phases: death at 1–3 years, 4–7 years, and 8–10 years after surgery. In other words, if we consider the predominant 70% of the inertia ($\varphi_1\varphi_2$), the disease seems to progress through stages defined by different sets of variables. If, on the other hand, we consider just the first 53% of the inertia (φ_1), we find that the plot of the φ_1 coordinates of the timepoints against calendar years is virtually linear ($r = 0.981$) (not shown). Apparently, a mix of variables accounting for the main φ_1 factorial axis evolves linearly.

Clustered variables within factorial plots are correlated as long as they do not lie too far above or below the plane of the plot. Thus a trait located close to a given year of death may be an indicator of death at this time. Patients over 70 tended to die of their cancer within 1 year of surgery whereas some of the youngest patients displayed the longest survival times (over 6 years) (panels A and B). Bilateral cancers tended to occur in the under 40s (panel B). More than four invaded lymph nodes was a sign of a bad prognosis (death within 3 years) but there was no clear-cut relationship between node number and death for four or less invaded nodes (panels A and C). Death 8–10 years after surgery was not correlated with node status but with young age (< 40 years old). The more poorly differentiated the tumor and, in particular, the more advanced the clinical stage of the disease, the sooner the patients died (panels A and D). The presence of any one of the three steroid hormone receptors (PR, ER, AR) postponed death slightly (panel E). PR postponed it most (furthest to the right). However, the long-term prognostic value of no receptor reached that of either clinical disease stage or tumor grade. This may be because the numbers of patients in whom receptors were determined in the longer surviving groups were small (Table 1) and/or because steroid receptors lose their relevance as molecular targets for therapy after ~ 4 years when some tumors had acquired hormone independence [22]. Thermography results (not shown) were not discriminatory in line with the conclusions of the consensus conference on mammography held in 1977 [23].

In summary, the $\varphi_1\varphi_2$ factorial plot provides a picture of the core information relating to the population of deceased patients. Later death was a feature of the youngest patients, those with low stage disease, low grade tumors, receptor-positive tumors, and/or those with four or less invaded nodes.

3.2. Correspondence analysis of steroid receptor data

The receptor data form a perfectly valid multidimensional system in their own right. A CA of these data for the deceased patients of Table 1 yielded Fig. 2 in which years of death are represented by hollow circles and receptors by black dots. Using this CA as a mathematical model, we introduced the supplementary data on combined ER.PR positivity given at the bottom of Table 1 (see stars). Despite the shortcomings of the data (number of missing values especially for AR), this

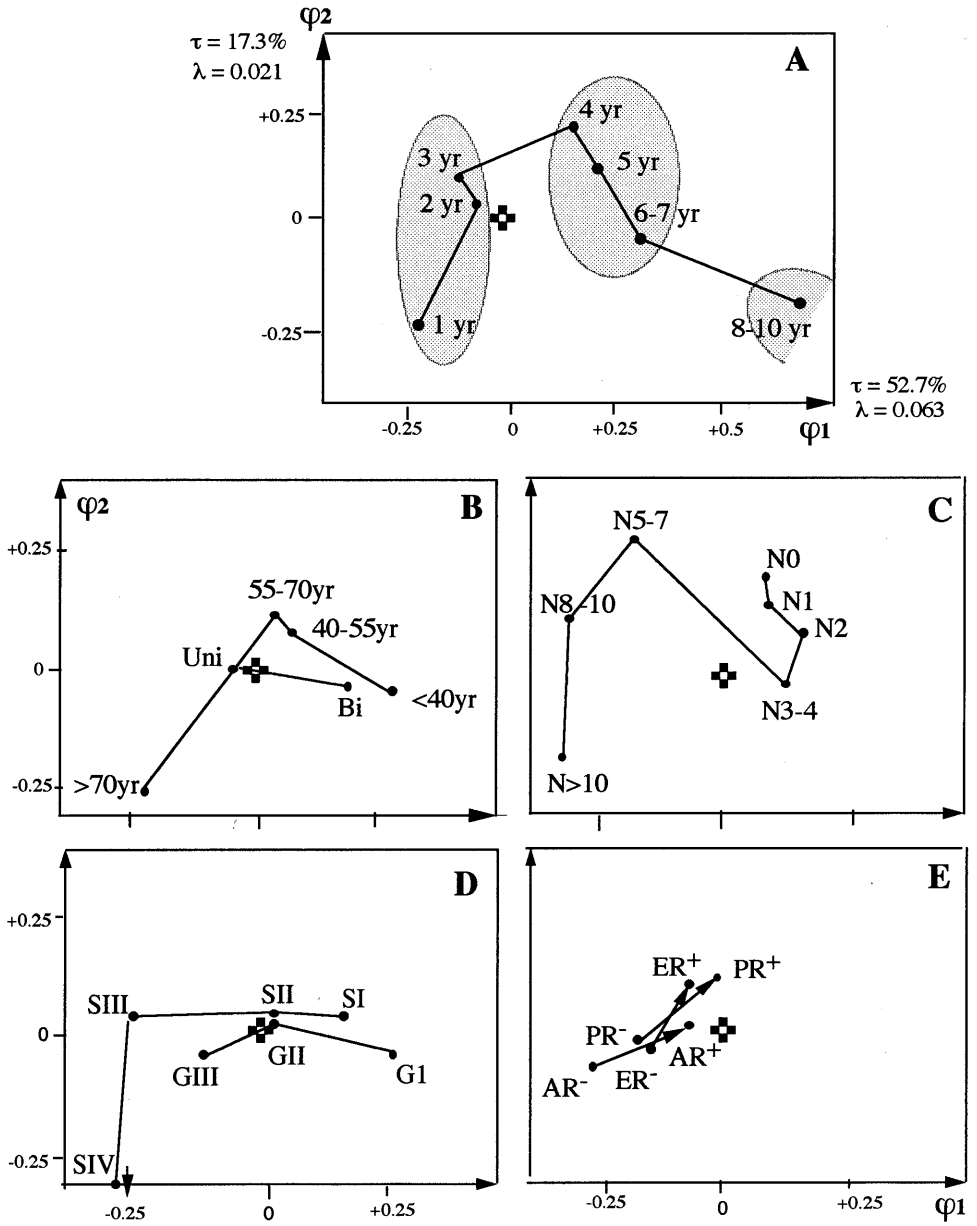

Fig. 1. CA plots of cancer death data (first seven columns in Table 1). (A) Plot showing death years (column variable). (B–E) Plots showing row variables: age, uni- or bilateral breast cancer, clinical stage (S), node status (N), histoprognostic grade (G), steroid receptor positivity (ER, PR, AR: estrogen, progesterone and androgen receptors; + positive, − negative). All plots are superimposable (✠, origin of the axes). The plots show the two main factorial axes that are derived from the data matrix and which together represent 70% of the most discriminant information in the matrix.

CA plot highlights several important points: (a) receptors are relevant for up to 5 years only with a difference between 1–3 years and 4–5 years, (b) receptor positivity is a sign of a slightly better prognosis, (c) PR is a better prognostic factor than ER, (d) AR is equivalent to PR but this result should be viewed with caution because of the limited number of AR determinations, (e) ER and PR are independent prognostic factors because, in Fig. 2, ER^+PR^+ (or ER^-PR^-) is located at the apex of a parallelogram that can be drawn using the three points: origin, ER^+ and PR^+ (in analogy to a parallelogram of vectors).

3.3. Patients still alive after 7 years

Table 1 also provides data on 344 patients with a disease-free survival (DFS) of more than 7 years and who were still alive at the time of analysis. A global CA on the 295 dead women and these 344 women — excluding however the rather fragmentary AR data — was performed in order to find out if and how inclusion of these women into the analysis affected the results.

Fig. 3 depicts 83% (φ_1: 72.0%, φ_2: 10.9%) of the inertia of the combined data. There is a surprisingly small shift in the positions of most variables when we

compare this figure to Fig. 1. The 5 and 6–7 death years have moved somewhat closer to each other. However, the introduction of the patients with >7 year-DFS has essentially reorganised node status when there are less than five invaded nodes and repositioned the under 40s. This is because the profiles of the >7 year-DFS patients and 8–10-year death patients are most different with respect to these variables. The >7 year-DFS patients have acted as an attractor in a force field. They have drawn toward them the traits that describe them best (grade I tumors, stage I disease, no or only one invaded lymph node (see shaded area)) and repelled others (chiefly three to four involved nodes). Clearly, nodal involvement is detrimental to long-term survival. Patients who died between 8 and 10 years tended to be the youngest patients (<40 years of age) (see shading). This correlation was highlighted by the φ_3 axis (not shown).

4. Discussion

Most of the correlations between either death from cancer or long-term DFS, on the one hand, and patient or tumor characteristics, on the other, noted in this study are hardly novel. This is because we analysed variables with relatively well-established prognostic value. The interesting point is that CA of retrospective clinical practice data supported current knowledge and, moreover, provided a visual overview. It confirmed the limitations of steroid receptor assays as prognostic factors, the relative value of each receptor, and the importance of lack of nodal involvement in long-term DFS. Patients with four or more involved nodes at initial diagnosis suffered from a more aggressive form of disease [24,25]; they died earlier. Unfortunately, no systematic data on tumor size was available.

In many statistical analyses, time is taken as a rather special variable, a yardstick that is set apart and against which changes in other variables are measured. However, CA treats time no differently from the other variables; it derives correlations among all variables including time, and seeks the timescale ('inner clock') that best relates to the data. It discovered, on the basis

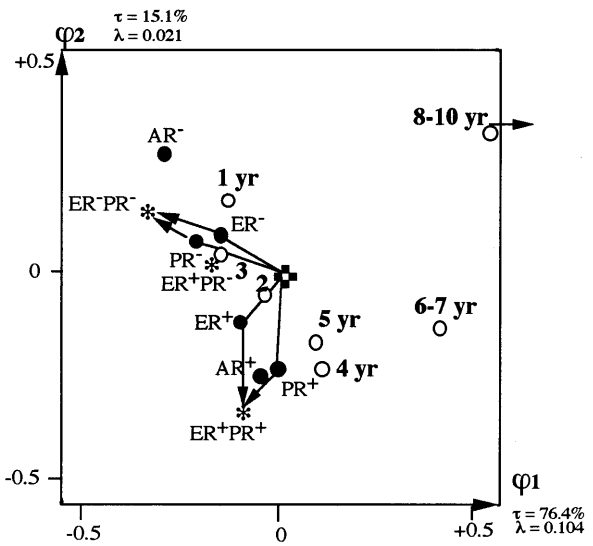

Fig. 2. CA plot of the steroid receptor data in the first seven columns of Table 1 (○, death years; ●, single receptor; *, receptor combinations; ✠, origin of the factorial axes). The plot shows the two main factorial axes that are derived from the receptor data matrix and which together represent 91.5% of the most discriminant information in the matrix. The receptor combinations (ER PR) were added as supplementary variables to the CA performed on single receptors and used as a mathematical model.

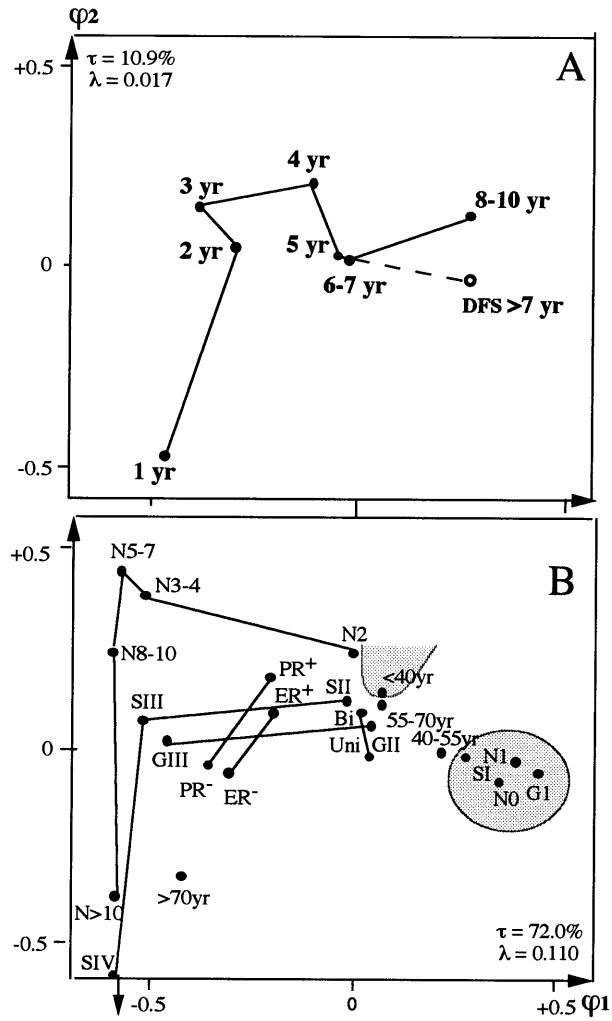

Fig. 3. Superimposable CA plots of a data matrix of deaths and DFS (first seven columns and last column of Table 1, excluding androgen receptor data). The two main factorial axes represent 83% of the most discriminant information in the data matrix (for abbreviations, see legend to Fig. 1).

of our data, that years have a chronological order but decided that there were three different epochs for dying each described by different mixes of variables (composite traits). There is a first stage from 1 to 3 years, a second from 4 to 7 years, and a third between 8 and 10 years as if there were different disease states. It is interesting to relate this observation to the biphasic trend in relapse (a peak at 3 years and one between 7 and 9 years) noted in at least two studies [26,27]. CA of a greater volume of data might afford an optimal partitioned categorical coding of time that might prove valuable in future hypothesis-testing analyses.

Our CA has also highlighted a difference between the profiles of women who died between 8 and 10 years after surgery and those who survived beyond 7 years. Most of the survivors had grade I tumors, no or only one invaded lymph node, and stage I disease. Those who died between 8 and 10 years were characterised mainly by their youth (<40 years) as if this, rather than any other feature, determined the nature of their disease. They were also the patients who tended to present bilateral disease. Age at first primary has been shown to be a determinant of the incidence of bilateral breast cancer [28]. This observation is in line with the possible hereditary nature of some cancers in younger patients.

When the results for an additional variable are available for all patients of a study, these can be introduced into existing CA biplots used as mathematical models [6]. The principle of adding extra variables is briefly illustrated in this study using supplementary data on steroid receptors. To be able to add data means that records have been well kept and are complete. Although our data have shortcomings, they are nevertheless reasonable for a hospital-cum-private practice in the 1970–1980s not taking part in a controlled study. They were breast cancer patients from one and the same catchment area followed by the same clinical team. Moreover, in our analyses, we took the precaution of checking that missing data were not selected in any way and did not create a bias.

The major criticism that can be directed against the use of CA in the study of retrospective survival data is that it does not account for censoring. Much justified emphasis is laid on possible biases in the analysis of censored data [29]. However, in our mind, the criticism is only partly valid for CA since CA filters information and extracts that which has the most coherence. It finds out how the major relationships within a complex dynamic system are organised. The quirks, background noise, and also the bias due to censoring would be relegated to lower-order factorial axes as long as the amount of censored data is not overwhelming. Combining differently censored data (deaths over 10 years and long-term survival at 7 years) did not upset our core analysis but highlighted major differences between population subsets.

The important point is that CA is not a statistical method for testing hypotheses, but a method of data analysis for seeking new hypotheses. Its role is not to find answers to predefined questions but, by highlighting correlations, to help experts define appropriate research questions. For our demonstration, we selected run-of-the-mill data and known variables and showed that the results were in line with published data. The method's full potential lies in less well explored fields. CA can deal with vast populations, a rather mixed bag of variables (quantitative and qualitative), and, like neural networks, with non-linear situations. It supplies graphic displays of correlations among unselected variables and combines non-linearity with multidimensionality. Such advantages should not be ignored.

References

[1] T.M. Therneau, P.M. Grambsch, T.R. Fleming, Martingale-based residuals for survival models, Biometrika 77 (1990) 147–160.
[2] F. Griffiths, B. Byrne, General practice and the new science emerging from the theories of 'chaos' and complexity, Br. J. Gen. Pract. 48 (1998) 1697–1699.
[3] M. Jambu, Exploratory and Multivariate Data Analysis, Academic Press, New York, 1991.
[4] M. Greenacre, Correspondence analysis in medical research, Stat. Methods Med. Res. 1 (1992) 97–117.
[5] M.J. Greenacre, Correspondence Analysis in Practice, Academic Press, New York, 1993.
[6] J.C. Doré, T. Ojasoo, Molecular taxonomy by correspondence factorial analysis (CFA), in: H. van de Waterbeemd (Ed.), Advanced Computer-Assisted Techniques in Drug Discovery. In: Methods and Principles in Medicinal Chemistry, vol. 3, VCH, Weinheim, 1994, pp. 190–227.
[7] C. Robertson, P. Boyle, Age-period-cohort models of chronic disease rates. 11: Graphical approaches, Stat. Med. 17 (1998) 1325–1339.
[8] L. Lebart, Méthodes factorielles, in: S. Thiria, Y. Le Chevallier (Eds.), Statistique et Méthodes Neuronales, Dunod, Paris, 1997, pp. 132–150.
[9] J.P. Nakache, B. Asselain, P. Pouillart, Contribution of correspondence analysis in survival studies, in: J. Janssen, J.F. Marcotorchino, J.M. Proth (Eds.), New Trends in Data Analysis and Applications, Elsevier (North-Holland), Amsterdam, 1983, pp. 261–272.
[10] R. Rousseau, B. Augereau, A. Daver, D. Leguay, Méthodologie de la régression et de la prédiction fondée sur la classification automatique, Cahiers Analyse Données XVI (1991) 479–488.
[11] G.M. Clark, S.G. Hilsenbeck, P.M. Ravdin, M. De Laurentiis, C.K. Osborne, Prognostic factors: rationale and methods of analysis and integration, Breast Cancer Res. Treat. 32 (1994) 105–112.
[12] S.G. Hilsenbeck, G.M. Clark, G. Chamness, C.K. Osborne, Untangling the multiple prognostic factor problem, Proc. Am. Ass. Cancer Res. 34 (1993) 193 (abstract no. 1152).
[13] M. De Laurentiis, C. Gallo, S. De Placido, F. Perrone, G. Pettinato, G. Petrella, C. Carlomagno, L. Panico, P. Delrio, A.R. Bianco, A predictive index of axillary nodal involvement in operable breast cancer, Br. J. Cancer 73 (1996) 1241–1247.
[14] B.G. Cipolla, J. Ziade, J.Y. Bansard, J.P. Moulinoux, F. Staerman, V. Quemener, B. Lobel, F. Guille, Pretherapeutic erythro-

cyte polyamine spermine levels discriminate high risk relapsing patients with M1 prostate carcinoma, Cancer 78 (1996) 1055–1065.

[15] D. Hans, R. Amalric, T. Ojasoo, J.C. Doré, Maps as an aid to clinical judgment, Cancer J. 9 (1996) 145–148 (Erratum: 9 (1996) 191).

[16] E. Bignon, M. Pons, A. Crastes de Paulet, J. Gilbert, J. Abecassis, J.F. Miquel, T. Ojasoo, J.P. Raynaud, Effect of triphenylacrylonitrile derivatives on estradiol receptor binding and on human breast cancer cell growth, J. Med. Chem. 32 (1989) 2092–2103.

[17] J. Gilbert, M. Fuentes, T. Ojasoo, J.C. Doré, M. Pons, Cytotoxicity and antiestrogenicity of a novel series of basic diphenylethylenes, J. Med. Chem. 40 (1997) 1104–1111.

[18] J. Fiet, J.C. Doré, A. Le Gô, T. Ojasoo, J.P. Raynaud, Multivariate analysis of plasma hormones in patients with metastatic prostate cancer receiving combined LH-RH analog and anti-androgen therapy, Prostate 23 (1993) 291–313.

[19] T. Ojasoo, J.C. Doré, J. Fiet, J.P. Raynaud, Visualization of data on age-related endocrine and metabolic changes in men, Endocrinol. Metab. 4 (1997) 329–337.

[20] P.M. Martin, H. Magdelenat, P. Kelly, Simultaneous determination of estrogen, progestin and androgen receptors on small amounts of breast tumor cytosol, in: F. Cavalli, et al. (Eds.), International Symposium on Medroxyprogesterone Acetate, Excerpta Medica, Amsterdam, 1982, pp. 544–588.

[21] L Guttman, The principal components of scale analysis, in: S.A. Stouffer, et al. (Eds.), Measurement and Prediction, Princeton University Press, Princeton, NJ, 1950.

[22] K.B. Horwitz, Mechanisms of hormone resistance in breast cancer, Breast Cancer Res. Treat. 26 (1993) 119–130.

[23] National Institutes of Health/National Cancer Institute, National Institutes of Health/National Cancer Institute consensus development meeting on breast cancer screening: issues and recommendations, J. Natl. Cancer Inst. 60 (1978) 1519–1521.

[24] C.A. Quiet, D.J. Ferguson, R.R. Weichselbaum, S. Hellman, Natural history of node-positive breast cancer: the curability of small cancers with a limited number of positive nodes, J. Clin. Oncol. 14 (1996) 3105–3111.

[25] I. Jatoi, S.G. Hilsenbeck, G.M. Clark, C.K. Osborne, Significance of axillary lymph node metastasis in primary breast cancer, J. Clin. Oncol. 17 (1999) 2334–2340.

[26] P. Valagussa, G. Bonadonna, U. Veronesi, Patterns of relapse and survival following radical mastectomy. Analysis of 716 consecutive patients, Cancer 41 (1978) 1170–1178.

[27] M. Baum, M.A. Chaplain, A.R. Anderson, M. Douek, J.S. Vaidya, Does breast cancer exist in a state of chaos?, Eur. J. Cancer 35 (1999) 886–891.

[28] H.O. Adami, R. Bergstrom, J. Hansen, Age at first primary as a determinant of the incidence of bilateral breast cancer. Cumulative and relative risks in a population-based case-control study, Cancer 55 (1985) 643–647.

[29] D.G. Altman, J.M. Bland, Time to event (survival), Br. Med. J. 317 (1998) 468–469.