ESTIMATE OF EXPECTED SURVIVAL AT DIAGNOSIS IN HODGKIN'S DISEASE: A MEANS OF WEIGHTING PROGNOSTIC FACTORS AND A TOOL FOR TREATMENT CHOICE AND CLINICAL RESEARCH
A report from the International Database on Hodgkin's Disease (IDHD)

Paolo G. Gobbi, Mario Comelli*, Giovanni E. Grignani, Carla Pieresca, Daniele Bertoloni, Edoardo Ascari

Clinica Medica II, Dipartimento di Medicina Interna, IRCCS S. Matteo and *Cattedra di Statistica Medica, Istituto di Scienze Sanitarie e Statistiche, University of Pavia, Italy

ABSTRACT

Purpose. **The aim was to identify a mathematical model that, when fitted with the survival time distribution of a Hodgkin's disease population, would provide a reliable estimate of expected survival at diagnosis for any new Hodgkin patient. This model would be based upon a multivariable selection of the best prognostic factors evaluable at diagnosis and its forecast could be of assistance in the choice of treatment.**

Methods. **The study sample consisted of the 5,023 patients whose basic clinical information was collected into the IDHD. These were people treated with standard protocols over the last two decades in 18 institutions. Several survival time distributions (exponential, Weibull, Gompertz, log-logistic and log-normal) were investigated to find the one that best fit the data and to relate its parameters to patient prognostic characteristics.**

Results. **The log-normal model provided the best fit for the data. The most statistically significant prognostic covariates were stage, age, histotype, B symptoms, serum albumin, sex and involved area distribution. Mediastinal, extranodal or bone marrow involvement, erythrocyte sedimentation rate, hemoglobin, serum alkaline phosphatase and lactate dehydrogenase did not add significant information. An equation containing these seven variables was derived to estimate median survival. Five distinct prognostic classes were identified by four cut-off values for this estimate.**

Conclusions. **Direct use of estimated median survival or allocating each patient into one of the five prognostic classes allows better tailoring of clinical strategies according to prognostic characteristics, more accurate patient stratification and evaluation of results in clinical trials and meta-analyses. Instructions are given for using this tool for both clinical and investigational purposes.**

Key words: Hodgkin's disease, modelling, prognostic factors, survival

Hodgkin's disease (HD) represents a historical challenge that still offers intriguing fields for investigation in spite of the fact that several goals in its management have been reached. Many general clinical problems have been the object of pioneer investigational efforts in HD over the last few decades, and substantial – although not definitive – progress has been recorded in some of them.[1]

Prognostic evaluation plays an important role by supplying a rational working linkage

between the wide complexity of disease presentations and available therapeutic resources. An improved prognostic evaluation able to identify the likely outcome of a given patient would be of considerable clinical importance. A wide number of studies[2] have pointed out a limited series of selected primary clinical features at presentation that contribute independently to patient survival. However, the exact influence on survival of each of these factors, and their precise contribution to correct treatment choices and outcome evaluation are still undefined.

A large sample from the Western world's HD population of the 70's and 80's collected in the IDHD offered us an opportunity to verify (a) how much of the overall prognostic variability can really be explained by the tumor- and patient-related characteristics commonly recorded at diagnosis; (b) the best choice of the clinical and staging parameters that prove to be most useful in clinics; (c) the possibility of adding one or more of the commonly used biological parameters, because of their easy measurability, to the list of prognostically important factors and (d) the possibility of constructing a mathematical model able to characterize the course of the disease on the basis of a few pretherapeutic clinical variables.

The main usefulness of such a model would lie in its ability to provide an improved method for estimating survival and a more accurate tool for making both therapeutic decisions and statistical comparisons between differently treated patient series.

Materials and methods

Collection criteria of the IDHD cases

In 1988 a large cooperative database was undertaken to assess – with respect to treatment response and survival – the exact relevance of parameters commonly considered in the management of HD patients. Although criteria, purposes and a first statistical evaluation of this database are better detailed elsewhere,[3] a concise general summary will be presented here.

The IDHD collected basic clinical information on 14,315 patients from 20 of the largest institutions or cooperative groups in the Western world with experience in the treatment of HD. Patients fulfilled the following inclusion criteria: (a) previously untreated; (b) age > 14 years at diagnosis; (c) initial treatment completed before the end of 1987.

The record of initial patient characteristics was restricted to those previously presented as prognostically important according to a consolidated amount of clinical experience and literature reports, i.e. sex, age, histological type (Htp), initial clinical presentation (CP), clinical stage (Stg), pathological stage, if performed, number of major lymph node areas involved (NLA), mediastinal (Med) and localized extranodal involvement (LEI), systemic symptoms (SS, categorized as the absence, A, or presence, B, of one or more of the following: night sweats, fever > 38° C and weight loss > 10% in the last 6 months[4]), erythrocyte sedimentation rate (ESR), hemoglobin concentration (Hb), serum lactate dehydrogenase (LDH), serum alkaline phosphatase (AP) and serum albumin (Alb). Staging information was taken according to the specific Ann Arbor Conference recommendations.[4,5]

The large majority of patients was treated according to randomized trials or well-known protocols, as reflected by the reports available in the literature and concisely reviewed elsewhere.[3]

Radiation therapy was classified into three simple groups: localized, regional and extended radiotherapy, respectively. Chemotherapy regimens were subgrouped into some broad categories: single agent, MOPP (mechlorethamine, vincristine, procarbazine and prednisone), or MOPP-like and adriamycin-containing regimens. Treatment response, disease-free or relapse-free survival, and overall survival were recorded according to Dixon et al.[6]

For disease course evaluation the following time data were recorded: a) date treatment began (survival evaluation starts from this point, not from the date of evaluation of clinical response, which was not recorded); b) date of relapse; c) date of last known vital condition (dead or alive); d) date of second cancer onset.

Causes of death were grouped under 5 headings: related to HD, related to treatment without evidence of disease, due to second malignancy, intercurrent, due to an unspecified cause. Clinical status was judged as either in complete remission or with active disease.

Population of the present study

Eighteen of the 20 IDHD centers agreed to the purposes and methods proposed for the present investigation, thus data were available for 14,074 of the total 14,315 patients.

We chose to exclude from the study those subjects treated before 1970 (in all 1,427 cases), since their overall survival proved to be lower than that of the patients treated both during the decade 1970-1979 and after 1979 (see also Henry-Amar et al.[3,7]). For the study population that remained (12,647 patients), the acceptable prognostic homogeneity, the basic rationality of the therapeutic plans,[3] the good common standards of efficiency in clinical management reached by the different centers and the very large recruitment definitely seemed to allow the drawing of general conclusions regarding prognostic matters. So, results can be considered reliably derived from a broad representative sample of the world's HD population that has been managed with pooled standard protocols from the last 20 years.

Laboratory data were carefully considered because of their demonstrated individual prognostic importance, easy measurement and comparability. Unfortunately, such values were often unreported in the files of many centers (see Henry-Amar et al.,[3] pages 179-190) and this greatly reduced the number of patients who presented a complete set of data and were thus suitable for this study. For instance, the number of patients who presented Alb, regardless of other biological variables, was 5,023.

The overall survival of patients presenting these biological data in their records proved to be identical with that of the ones without such information, thereby demonstrating that these cases do not represent a selected sample of the whole population in this respect (the Kaplan and Meier estimate[8] of the proportion of patients surviving at 200 months is, respectively, 0.60 for those lacking and 0.57 for those having Alb in their records).

Statistics

General techniques and modelling. The IDHD file was managed on a personal computer with the SPSS,[9] SAS[10] and BMDP[11] statistical packages.

Survival was the only prognostic outcome considered for the aims of this study.[6] The 75 patients who died at the age of 65 or over either from unspecified causes or conditions unrelated to HD or to its therapy were taken as censored. This criterion was adopted in order to avoid seriously overestimating HD mortality due to the generally short expected survival of the normal 65 and over population. Indeed, life expectancy for the general 65 and over age group is much shorter than the median survival of HD patients.[12] However, patients who died at the age of 65 or over from causes related to HD, to treatment or to second malignancy were not censored.

The whole population was subdivided into two casual and independent subsets of patients. In the first – the training sample of 6,502 patients – we studied and comparatively evaluated the goodness of fit of several parametric survival models and selected one of them. In the second – the test sample of 6,145 patients – we cross-validated the accuracy of the selected model.

The plot of the empirical hazard vs. time, the plots of the logarithm of the empirical hazard, the cumulative hazard and the logarithm of the cumulative hazard vs. time and the logarithm of time were obtained to check the fit of the exponential, Weibull, Gompertz, log-logistic and log-normal models.[13]

Inspective evaluation of the linearity of these curves was completed by an analysis of the log-likelihood of all the distribution models considered (except that of Gompertz),[14] performed by means of the SAS LIFEREG program[15] without the use of explanatory variables.

A further check was carried out by analyzing the generalized residuals of the model according to the method suggested by Blossfeld et al.[16] Briefly, this method consists of checking the

exponentiality in the distribution of the theoretical cumulative hazard computed, using the different models, at each death or censure time.

Transformation and selection of explanatory variables (covariates). The following 16 pretreatment parameters were considered for analysis in the training sample: sex, age (in years), Htp (4 types: lymphocyte predominance, LP; nodular sclerosis, NS; mixed cellularity, MC; lymphocyte depletion, LD), CP (3 levels: above, below or on both sides of the diaphragm), Stg (4 levels, I to IV), NLA (possible numbers: 0 to 10), Med (involved or uninvolved, bulk information not provided), LEI and BM (both with 2 possible levels), SS (A or B category), ESR (mm at 1st hr), Hb (mMol/L), LDH (mMol/L), AP (mMol/L), and Alb (mMol/L).

The distributions of ESR and Hb were similar when compared among centers,[3] thus they were handled as direct quantitative data. The distributions of LDH, AP and Alb were different among centers so it was necessary to use them as percentiles of the frequency distribution observed within each center.

Through analysis of the variations in the goodness of fit (log-likelihood ratio) of the selected survival model, the following covariate modifications were adopted:

– the variable *stage* (Stg) corresponds to the pathological stage, when available (5,796 patients); otherwise, it refers to the clinical stage;

– age was square transformed $[(age)^2]$ in accordance with its clinical weight, which is commonly accepted to be higher as the patient gets older;

– the value of Alb, as percentile of the frequency distribution within each center, was further converted into its natural logarithm $[\ln(Alb)]$. This is consistent with the decreasing clinical significance of high levels of this variable;

– Htp were regrouped into two classes: LD vs. all other types ($\neq$LD).

LDH was excluded in advance from the selection of covariates for two reasons: first, its consideration would have dramatically reduced the

number of observations; second, LDH lost all statistical importance when it was included in the model together with stage and age in a preliminary evaluation on the 1,746 patients presenting LDH information.

In the course of the selection procedures to determine the best variables, evidence showed that collapsing two of them – CP and NLA – into one would be appropriate. The newly-formed variable, called *involved area distribution* (IAD), presents two distinct categories: ≤ 3 involved areas above the diaphragm vs. all other conditions (i.e., ≥ 4 supradiaphragmatic areas or any subdiaphragmatic ones, or any number of areas on both sides of the diaphragm).

Selection of the best covariates was accomplished by means of both marginal and conditional testing of each possible explanatory variable. These tests evaluate the difference between the log-likelihood of the model when a given variable is or is not included in the presence or the absence of other covariates.[17] The significance level used was 0.05.

Possible interactions between explanatory variables were systematically investigated within the group of the best covariates selected.

Validation of the model and identification of homogeneous prognostic classes. The best theoretical model chosen using the training sample was cross-validated in the test sample. Its accuracy in fitting the data of the test sample was evaluated, and the statistical importance of each covariate was verified. Moreover, the values of the coefficients estimated in the test sample were compared with the estimates obtained in the training sample by means of a Z test, using the estimated standard errors.[18]

The results of these tests should demonstrate that the theoretical model reflects structural relationships between the survival of HD patients and the prognostic factors, and not just chance aspects of the sample used.

Theoretical survival distribution can be similar for a number of patients with different combinations of prognostic factors. So, using cluster analysis, we investigated a set of classes, well differentiated for theoretical survival, containing subjects with relatively homogeneous

expected prognosis. It is noteworthy that the homogeneity within each class is limited to the prognosis-conditioned treatment because patients with different clinical presentations – which implies different treatments – can belong to the same class. The prognostic classes identified were validated on both the training and test samples.

The various prognostic classes identified presented an opportunity for checking how the model could forecast empirical survival distribution. To this end, in each prognostic class of both the training and test samples the largest group of patients presenting identical prognostic factors was sought. It should be emphasized that the choice of the largest group in each class was deliberately independent of that which best fit the model or differentiated the classes.

The 95% confidence band of its theoretical survival distribution was computed[10] for each of the selected groups. Then, the survival curve of the corresponding subset of patients in the test sample (i.e. those with exactly the same explanatory prognostic variables) was calculated according to the Kaplan and Meier technique.[8] The model is validated by the inscription of the survival curve of the test sample patients within the confidence band of the expected survival calculated from the training sample. This criterion of validation is rather severe, since the calculated confidence bands do not account for the multiple testing artifact.[19]

After this validation, a new estimation of both the parameters of the model and the cut-off values for the prognostic classes identified was performed on the total available population (5,023 patients), in order to increase the general accuracy of the model for possible users.

Results

Visual evaluation of the plots of the empirical hazard, cumulative hazard and their logarithmic transformations suggested that the log-logistic and the log-normal models fit the data of the training sample better than the other models. Despite the fact the graphical analysis tended to exclude all models other than the log-normal and log-logistic, we further checked the

Weibull and exponential distributions with a log-likelihood analysis. The results indicated a definite advantage for the log-normal model in best fitting the data. The generalized residuals plots, shown for the different models in Figures 1-4, confirm this conclusion. Thus the log-normal model was definitively chosen as the most adequate, and it was utilized for the selection of the explanatory variables.

All covariates were statistically significant at univariate analysis. Only seven of them were statistically significant at a multivariate analysis, as shown in Table 1.

No interaction between variables demonstrated a statistically significant role in the model.

The cross-validation results obtained for the model with the 7 selected covariates in the test sample are shown in Table 1, which reports the coefficients with standard errors and corresponding P values for the 7 variables when the model was applied on both the training and the test samples.

In the test sample, 6 out of 7 parameters confirmed their prognostic importance by showing high statistical significance, while IAD retained a 0.09 p value. Moreover, the differences in the absolute values of the estimated parameters in both samples are often negligible and never statistically significant when checked with the Z test. These results mean the model describes reproducible structural features of HD mortality and not just casual features of the data.

The right hand section of Table 1 reports the parameters of the model estimated on the total available population. These coefficients allow a more accurate calculation of the distribution of survival probabilities for each patient. From this distribution the following equation for median expected survival can be derived:

$$T \text{ (months)} = \exp [3.75 + 1.25 \star \text{Stg I} + 0.77 \star \text{Stg II} + 0.46 \star \text{Stg III} - 0.00046 \star (\text{age})^2 + 0.85 \star \text{Htp} + 0.42 \star \text{SS} + 0.24 \star \ln(\text{Alb}) + 0.25 \star \text{sex} + 0.25 \star \text{IAD}]$$

Using this equation for a given patient, the values of the quantitative variables have to be directly introduced into the model, while the qualitative features have to be inputed as 1

Table 1. Final model parameters with standard errors and statistical significance.

Variable		Training sample (2542 pts)			Test sample (2481 pts)			Whole sample (5023 pts)		
		Parameter	SE	P	Parameter	SE	P	Parameter	SE	P
Stage	I	1.32629265	0.192593	0.0001	1.16192148	0.191687	0.0001	1.25480611	0.135966	0.0001
	II	0.7523949	0.147552	0.0001	0.77510785	0.150865	0.0001	0.76884839	0.105584	0.0001
	III	0.34746573	0.123778	0.0050	0.56242705	0.124096	0.0001	0.46168073	0.087618	0.0001
Age^2		-0.0004599	0.000033	0.0001	-0.0004581	0.000033	0.0001	-0.0004585	0.000023	0.0001
Htp		1.11418815	0.19988	0.0001	0.60472612	0.192716	0.0001	0.84583432	0.13608	0.0001
SS		0.45914521	0.104552	0.0001	0.38275433	0.104845	0.0001	0.42134335	0.074138	0.0001
ln(Alb)		0.24682946	0.050623	0.0001	0.23110039	0.051129	0.0001	0.23859284	0.036029	0.0001
Sex		0.2330325	0.095093	0.0143	0.2471226	0.094039	0.0086	0.24782	0.066877	0.0002
IAD		0.29263854	0.1223	0.0167	0.2122107	0.124898	0.0908	0.2471129	0.087452	0.0047
Intercept †		3.43333934	0.259449	0.0001	4.02208959	0.258752	0.0001	3.74650849	0.183254	0.0001
Scale ‡		1.73756961	0.048699		1.71414372	0.048623		1.72962772	0.03449	

† This parameter is the natural logarithm of the expected median survival (months) of pts with all unfavorable prognostic factors.
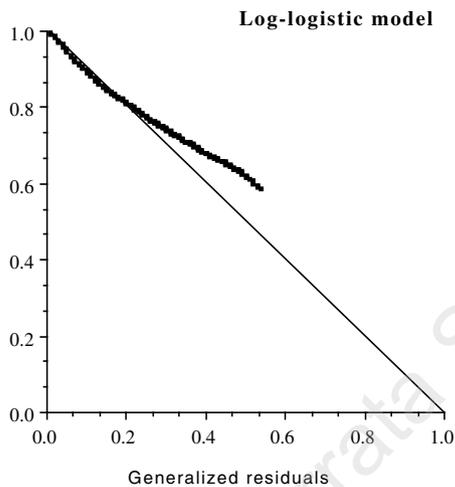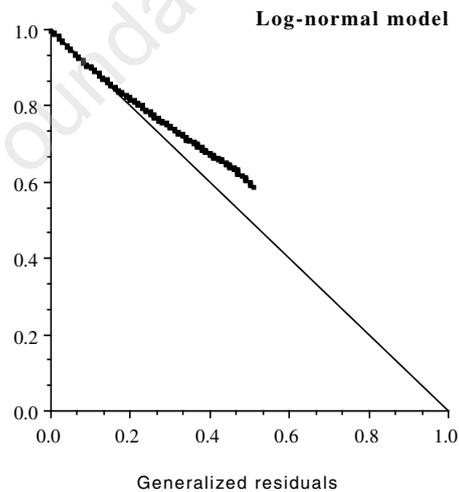‡ This parameter measures the spread of the expected survival distribution.
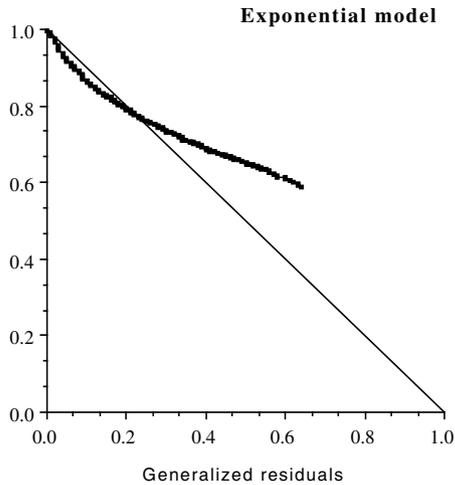
1
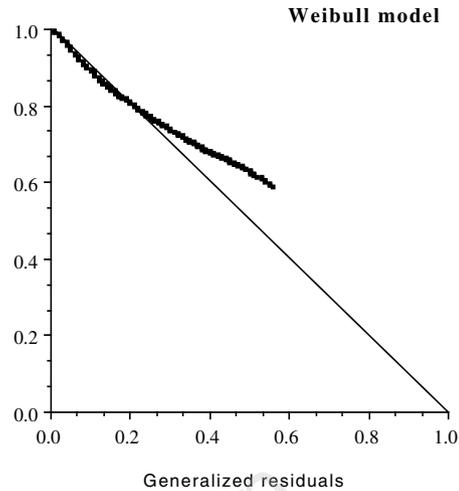
2

3

4

Figures 1-4. Generalized residuals in the four models evaluated.

when present or 0 when absent. For example, the median expected survival of a 55-year-old male patient with stage III lymphocyte depleted disease, with B symptoms, presenting 4 involved lymph node areas above the diaphragm and a serum albumin concentration equal to the 20th percentile is:

$$T \text{ (months)} = \exp(3.75 + 1.25 * 0 + 0.77 * 0 + 0.46 * 1 - 0.00046 * 55^2 + 0.85 * 0 + 0.42 * 0 + 0.24 * \ln 20 + 0.25 * 0 + 0.25 * 0) = 34$$

The unfavorable basal clinical categories (those entered as 0 in the equation) are: stage IV disease, lymphocyte depletion, the presence of B symptoms, male gender, > 3 involved supradiaphragmatic lymph node areas, any subdiaphragmatic ones or any on both sides of the diaphragm. One can expect that when a patient does not present a basal category for a qualitative clinical variable, his median survival will be multiplied by exp(coefficient of the variable). So, the expected median survival of a

Table 2. Expected quartiles of survival and observed percentages of deaths in the prognostic classes identified in the whole sample.

Prognostic classes	I	II	III	IV	V
Number of patients	825	1244	1361	1041	552
Deaths %	9.58	17.68	28.07	43.71	66.85
1st quartiles †	322	160	81	35	9
2nd quartiles †	1051	524	266	116	30
3rd quartiles †	3438	1713	870	379	99

† Quartiles measured in months.

patient with stage I disease will be exp(1.25) times higher than that expected for a patient with stage IV disease, all other clinical variables being equal.

The probability that a patient will survive at least **t** number of months is given by:

$$P = 1 - \Phi \{[\ln \mathbf{t} - (3.75 + 1.25 * \text{Stg1} + 0.77 * \text{Stg2} + 0.46 * \text{Stg3} - 0.00046 * (\text{age})^2 + 0.85 * \text{Htp} + 0.42 * \text{SS} + 0.24 * \ln(\text{Alb}) + 0.25 * \text{sex} + 0.25 * \text{IAD})] / 1.73\}$$

where Φ is the standard Gaussian probability integral. The probability of the patient in the preceding example surviving at least 120 or 60 months is given by:

$$P = 1 - \Phi \{[\ln 120 - (3.75 + 1.25 * 0 + 0.77 * 0 + 0.46 * 1 - 0.00046 * 55^2 + 0.85 * 0 + 0.42 * 0 + 0.23 * \ln(20) + 0.25 * 0 + 0.25 * 0)] / 1.73\} = 0.23$$

$$P = 1 - \Phi \{[\ln 60 - (3.75 + 1.25 * 0 + 0.77 * 0 + 0.46 * 1 - 0.00046 * 55^2 + 0.85 * 0 + 0.42 * 0 + 0.23 * \ln(20) + 0.25 * 0 + 0.25 * 0)] / 1.73\} = 0.63$$

Cluster analysis identified 5 prognostic classes characterized by high within-group and low inter-group homogeneity with respect to estimated survival. These are separated by the following 4 cut-off values for median expected survival: 723, 377, 178 and 64 months. The number of patients from the total available study population allocated into each prognostic class, the percent of deaths and the median sur-

vival estimate quartiles are reported for each class in Table 2.

The first class contains patients whose expected HD-related survival is so high that mortality very near that of the normal population can be hypothesized. Note that a 50% probability of death at 1051 months does not mean that 50% of the patients will be dead at 87.5 years from diagnosis (this would exceed the expected survival of even a healthy young population). It simply means that 50% of HD-specific deaths can be expected after that time, and if this interval is longer that the normal expected survival for subjects at a given age, then death from other causes will be more probable. In this respect, the second class includes patients with a 25% probability of dying within 160 months (about 13.5 years), which represents a poorer outlook than that of the normal population. On the other hand, patients in the fifth class have a 50% probability of dying within 2.5 years, and only 25% of them will survive for little more than 8 years.

When the unique covariate, that given by the ordinal number of the prognostic class to which every patient belongs, is substituted for the 7 explanatory variables in the model, the accuracy of the fit decreases quite negligibly in the training sample.
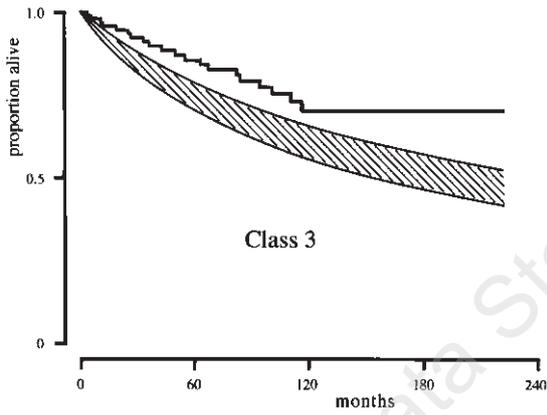
Figures 5-9 illustrate the comparison between the theoretical survival curves estimated in the training sample (95% confidence band) and survival in the test sample as estimated by the Kaplan and Meier curves. This comparison was performed in each prognostic class between the largest patient groups presenting the same
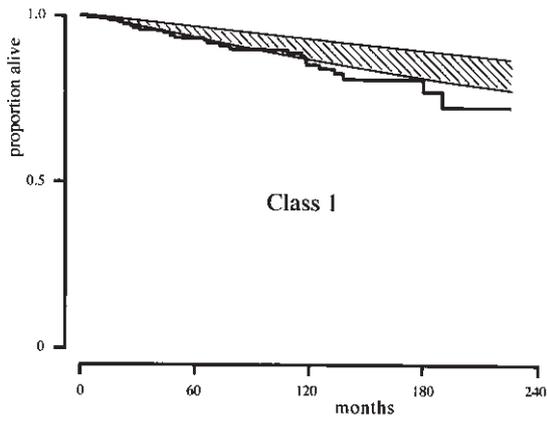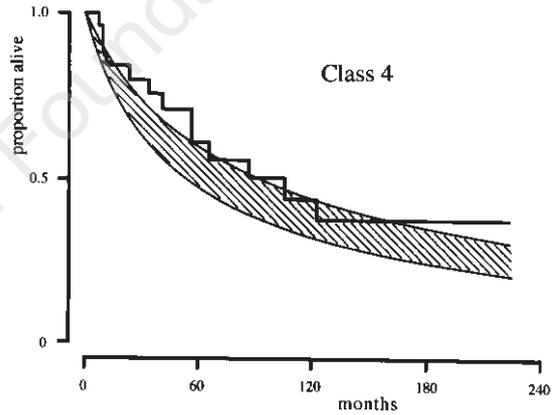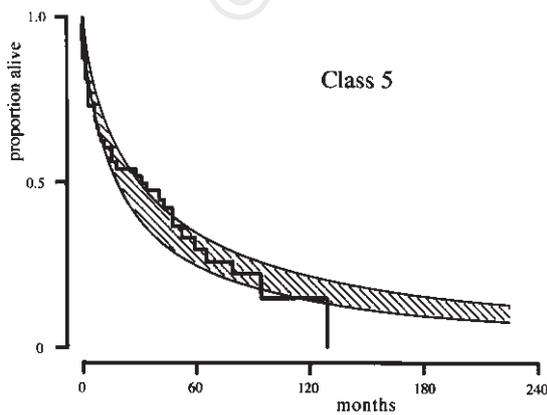
5
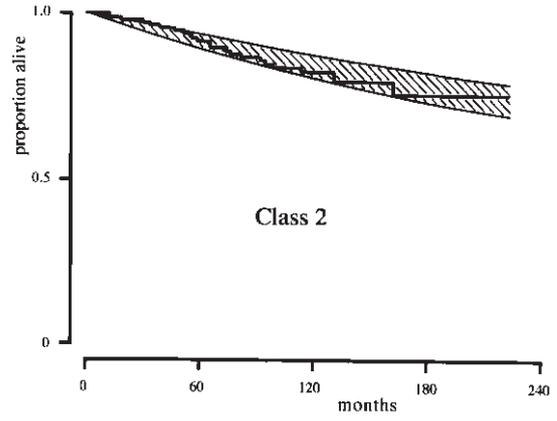
6

7

8

9

Figures 5-9. Comparison within the largest group of patients in each prognostic class between the theoretical survival estimated in the training sample and the survival observed in the test sample. The thin lines encompass the 95% confidence band for expected survival. The thick line is the observed Kaplan-Meier curve.

tion of broad clinical experience in the hope that they will prove to be a useful tool for clinical research of the future.

The results can be commented along the following lines.

Modelling

There are several reasons[20] why a parametric model of survival time distribution is a valuable object in HD. First, a good-fitting model can economically characterize the survival experience of a group of individuals in terms of a few parameters. This in itself represents a goal in HD, a disease marked by well-known, emblematic variability in clinical presentation and prognosis. Second, the interpretation of clinical events in subsequent studies is improved when a model fitting data according to the experience of a precedent study is available. In general, more powerful tests of the differences in survival distribution can be made if a model for survival time distribution is known. Third, when the parameters of a validated model correspond to clinical characteristics, the best insight is available on the true prognostic factors of the disease. Finally, a good parametric model provides a direct estimate of the survival of a given patient in terms of time units with known accuracy, which is a more flexible and directly useful output than the prognostic indices generated by nonparametric analyses. To give the same information in terms of survival, a nonparametric modelling procedure would require hundred of different diagrams of parameter estimates.

However, more nonparametric[21-28] than parametric[29] models have been proposed so far in HD, and none of them is given much consideration in clinical practice or research.[30-32]

The log-normal distribution of survival time data represents a common choice among research workers in the life sciences. It is characterized by a hazard function having a maximum after time zero, a shape that agrees with the general experience of a decreasing death risk after a higher peak in the first few months[20] in the course of HD. However, the evaluable variations of the hazard function in HD are not very ample when plotted against time and are evident only in large samples of patients. Some of us, in a preceding modelling approach on a ninefold smaller population of HD patients,[29] accepted the exponential distribution as best fitting the data, despite a mild deviation from linearity. However, it is well known that it is often difficult to distinguish a log-normal from an exponential distribution with relatively small samples. Moreover, when the hazard function is nearly constant and other distributions might also be acceptable, the exponential distribution is a reasonable choice because of its simplicity.[33]

In the present study the availability of a large sample made the selection of the model unquestionable. Lastly, the use of the log-normal distribution model for survival has a few theoretical justifications.[34]

Prognostic factors

Stage, age, histology, B symptoms, albumin, sex and involved area distribution were considered the relatively best prognostic factors evaluable before treatment. The aims of this study were extremely clinically oriented. It follows that if a survival model is to help in the choice of therapy, it should be applicable to all patients at diagnosis, without restriction of stages or age (at least in the adult range). For the same reason, it should not include factors evaluable only during treatment or after its completion (e.g. clinical response, drug dose intensity, etc.). Therefore the model included stage with all its distinct levels (which still primarily orients the selection of radiotherapy, chemotherapy or a variable combination of the two), as well as age with whole year distribution beyond 14 years.

The square transformation of the age datum proved to be very suitable both according to the results of Proctor et al.[28] and to the clinical significance already stressed by other authors.[35,36] The only report that refused to attribute individual prognostic importance to age in HD[37] described a modest population in which deaths due to other causes were censored, a procedure we consider very difficult and dangerous in prognostic studies.

Furthermore, Alb – the second quantitative factor selected – was handled in such a way as

to avoid the loss of information related to a categorization of these values. Apart from the clinical validity of the logarithmic transformation (see *Materials and Methods*), the very different distribution of Alb levels in each center, which required the use of percentiles within each distribution, might explain the contrasting results found with regard to the prognostic importance of this parameter in malignant lymphomas.[27,38-40] The relative importance of Alb proved to be higher than that observed by some of us[29] in a preceding parametric approach, while that of sex was substantially unmodified. The IAD variable was devised to integrate prognostic information from both CP and NLA, each of which separately exerted an interesting (not statistically significant, but clinically relevant) importance in the model. No laboratory data other than Alb (ESR, Hb, AP, LDH), which we relied on to substitute for the record of B symptoms with more quantifiable clinical parameters, retained statistical importance when included in the model with any one of the 7 clinical variables selected.

The reasons already expressed in favor of the general validity of the present results should settle any questions on possible discrepancies with the outcome of earlier multivariate research on HD. As a matter of fact, many multivariate studies have found different series of major survival determinants, as was partially reviewed by Gobbi et al.[29] and further expanded by other papers.[27,28,38,41-43] Stage, age and histology recur as primary factors in the large majority of these works; sex, symptoms and ESR are frequently found, while a number of other clinical characteristics sporadically enter the group of the most important factors (hemoglobin,[28] mediastinal involvement,[26,44] lymphocyte count,[27,28] HLA phenotype,[45] immunocompetence,[46] tumor burden[43]).

We believe that small differences in the choice of the clinical parameters to evaluate prognostically, among those commonly considered, or varying accuracy in the clinical evaluation or biochemical measurement of some of them can strongly influence the independent role attributed to each parameter or their resulting hierarchical order. However, we also think that such

differences have little effect on the absolute portion of the overall prognostic variability that can be explained by the entire complex of whatever common clinical factors are evaluated, because of their multiple partial interrelationships.

Some advantages can probably be expected by taking into account parameters that directly evaluate certain biological features of the tumor or its responsiveness to therapy (i.e. tumor cell kinetics or cell resistance to drugs), rather than serologic tests invariably correlated among themselves and with staging parameters.[3]

So, we are convinced that a large part of the prognostic significance brought by the the semi-quantization of tumor burden according to Specht et al.[41] is retained in our model by the 4 levels of Stg together with the 2 of IAD, and probably those of Alb.[47] Likewise, the possible prognostic weight of Med, if there is any, might have been replaced by B SS, female sex and younger age – all factors correlated with Med.[3,48] It is unlikely that bulky Med, a parameter not available in the IDHD and not yet univocally defined, might have assumed a more definite prognostic role, since its relationship with either irradiation techniques[49] or with non-administration of combined chemotherapy plus localized radiotherapy[30,49] has already been demonstrated (it could be considered a therapy-amendable prognostic factor). In a similar way, the strong intercorrelations between ESR and Alb, Hb and Alb, and among SS, Alb, ESR and Hb demonstrated in the IDHD[3] explain our selection of only two of these factors as offering the best contributions to the model.

The results from the parametric model approach to the whole evaluable IDHD population substantially agree with those obtained from Cox's model on separate early and advanced HD patients in the database by Meerwaldt et al.[51] and Löffler et al.,[52] respectively. However, the higher prognostic importance of ESR and Alb than B SS found by some of us in 586 cases[29] was not confirmed in the very large IDHD sample. The presence or absence of SS still proved to be a valuable prognostic factor that can be usefully integrated by some laboratory parameters, but can hardly be replaced by

a few of them.

On biological grounds the presence of SS probably reflects the cytokine-mediated interaction between host and tumor more directly than the so-called acute-phase or chronic disease biologic indicators that are routinely evaluated among blood laboratoty tests. Unfortunately, the clinical evaluation of SS is rather subjective and less accurate than the measurement of a laboratoty test. Moreover, prognostic and biological evidence has been collected on the opportuneness of recoding the truly unfavorable symptoms.[53,54]

A reclassification of SS, excluding night sweats and including severe pruritus, should probably enhance the individual prognostic role of a new, recoded B symptomatology category. One possible way of replacing SS with measurable and easily comparable parameters might be the determination of the production of some cytokines or their products. For instance, analysis of the largely stratified quantity of clinical experience collected in the IDHD demonstrated that Alb represents the most important, definite and sufficiently independent laboratory datum in HD, and is worth adding to the consideration of SS and the other major prognostic factors.

Possible uses of the survival model

Clinical uses. Hopefully, the use of a quantitative estimate of survival for each patient by considering some selected clinical findings presented at diagnosis will integrate the international staging information recommended at the Ann Arbor Conference in 1971.[4,5] This staging information is somewhat unsatisfactory from a prognostic point of view, and was not substantially improved in this regard by the recent modifications at the Cotswolds Meeting in 1989.[55] The log-normal model of HD survival is able to fulfill the expectations of a working, therapy-oriented prognostic system that can integrate new risk factors drawn from laboratory investigations with consolidated staging parameters, as has been predicted for the near future in HD research.[56]

Prognosis can now be estimated for each patient along a continuous distribution of predicted survival times related to a given set of initial clinical parameters. This can be considered a more accurate and flexible prognostic evaluation than that allowed by the series of independent stratifications for a few parameters (stage, histology, symptoms, etc.) currently made for the choice of therapy and evaluation of clinical results. Thus, the model can be considered a powerful clinical tool for pretherapeutic measurement of individual prognostic severity, for better tailoring the treatment plan to the patient's therapeutic needs, and for controlling the prognostic disparity of patients to be treated with the same therapy.

A number of possible working cut-off values for expected survival can be fixed for particular presentations, treatment policies, or definite therapeutic resources to treat prognostically homogeneous subsets of patients. Interesting fields of application are offered by the clear identification of those stage I or stage II disease patients with a very favorable estimate of survival on whom therapy reductions might be explored, or the definition of a certain poor survival estimate for homogeneous selection of patients undergoing eradicating megachemotherapy followed by bone marrow transplantation.

The five classes into which total prognostic variability was subdivided represent only one example of the possible uses of the information from prognostic factors, irrespective of the therapeutic possibilities related to particular clinical presentations. From this point of view, these classes cannot be used as direct indicators of specific treatments, since by themselves they regroup patients who show similar survival as the effect of dissimilar treatments for different clinical presentations. So, it would be wrong to consider all class 1 patients as good candidates for mild therapies, since among them subjects with advanced stage disease successfully treated with standard multiple drug chemotherapy associated with radiotherapy can also be found. Of course, administering such combined therapy is the only way to assure the patient of a class 1 prognosis. Likewise, not all class 5 individuals have to be addressed to bone marrow transplantation, due to the presence in this class of

many older patients who cannot tolerate this procedure. Nevertheless, a suitably age-restricted subset of patients (e.g. <45-50 years), identified through a ≥50% probability of death at 30 months (class 5 patients), might now be proposed for autologous bone marrow rescue after megachemotherapy. This strategy would be prognostically more accurate and better able to be compared than that followed up to now. On the other hand, it should be correct that all patients with the same presentation and destined to a specific treatment belong to the same class.

Thus, the clinical utilization of a directly predictive model can help in reaching the goal of individual therapy adopted according to more rational and objective pretherapeutic evaluation.

Investigational uses. Pretreatment knowledge of the expected survival for each patient represents the most powerful tool for reliable stratification of patients in clinical trials.

There are several possible ways to use the parametric model in the design of clinical research.

First, a definite range of estimated survival can be preliminarily chosen to stratify patients undergoing a clinical trial. This method should warrant the prospective selection of the most prognostically homogeneous population for the study (under standard therapy protocols).

Second, a retrospective analysis can be performed to evaluate the role differences in expected pretreatment survival play on the results obtained. For this purpose one can check the number of patients in the treatment arms who could be allocated into the 5 prognostic classes. An easy technique for adjusting observed survival figures for an unbalanced distribution of the five classes among the treated groups could be the one devised by Axtell et al.[24]

Another possible method would be to utilize the median expected survival generated by the model as a distinct prognostic covariate to be entered in a Cox's proportional hazard regression model. In this technique the covariate would synthetically replace the whole complex of the most important prognostic factors; it would make the analysis simpler and easier and, most importantly, it would allow reasonable accuracy in the evaluation of results, even with decidedly smaller samples. Since the advisable number of predictor variables in a Cox's model should not be more than 5 to 10% of the number of endpoints (deaths, relapses, or other events),[57] the use of only one covariate besides that testing the null hypothesis should be correct when starting from a number of 20 to 40 endpoints.

With these techniques, even published trials might be re-evaluated for better control of prognostic homogeneity in the randomized arms.

References

1. Urba WJ, Longo DL. Hodgkin's disease. N Engl J Med 1992; 326:678-87.
2. Hagemeister FB. Prognostic factors in decision-making in the clinical management of Hodgkin's disease. Hematol Oncol 1988; 6:257-69.
3. Henry-Amar M, in collaboration with Aeppli DM, Anderson J, et al. Workshop statistical report. In: Somers R, Henry-Amar M, Meerwaldt JK, Carde P eds. Treatment strategy in Hodgkin's disease. Montrouge, France: Editions John Libbey Eurotext, 1990, 169-425.
4. Carbone PP, Kaplan HS, Musshoff K, Smithers D, Tubiana M. Report of the committee on Hodgkin's disease staging classification. Cancer Res 1971; 31:1860-1.
5. Rappaport H, Berard CW, Butler JJ, Dorfman RF, Lukes RJ, Thomas LB. Report of the committee on histopathological criteria contributing to staging of Hodgkin's disease. Cancer Res 1971; 31:1864-5.
6. Dixon DO, McLaughlin P, Hagemeister FB, et al. Reporting outcomes in Hodgkin's disease and lymphoma. J Clin Oncol 1987; 5:1670-2.
7. Henry-Amar M, Somers R. Survival outcome after Hodgkin's disease: a report from the International database on Hodgkin's disease. Semin Oncol 1990; 17:758-68.
8. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Ass 1958; 53:457-81.
9. Norusis MJ/SPSS Inc. Manual SPSS/PC. Chicago, Illinois: SPSS Inc, 1988.
10. Hudnell MZ, McCullogh BC, Fitzgerald M. SAS Users' Guide 1986. Cary, North Carolina: SAS Institute Inc, 1986, 508-28.
11. Dixon WJ, Brown MB, Engleman L, Jenrich RI. BMDP Manual 1990. Berkeley, California: University of California Press, 1990.
12. Chin Long Chiang. Introduction to stochastic processes in biostatistics. New York, NY: John Wiley & Sons, 1968: 201-34.
13. Cox DR, Oakes D. Analysis of survival data. London: Chapman & Hall, 1984, 25-87.
14. Aitkin M, Anderson D, Francis B, Hinde J. Statistical modeling in GLIM. Oxford: Clarendon Press, 1989: 267-331.
15. Hudnell MZ, McCullogh BC, Fitzgerald M. SAS Statistics 1986. Cary, North Carolina: SAS Institute Inc, 1986.
16. Blossfeld H-P, Hamerle A, Mayer KU. Event history analysis. Hillsdale, NY: Lawrence Erlbaum Associated Publishers 1989, 82-4, 194-9 and 238-59.

17. Benedetti JK, Brown MB. Strategies for the selection of log-linear models. Biometrics 1978; 34:680-6.

18. Armitage P, Berry G. Statistical methods in medical research. 2nd ed. Oxford, GB: Blackwell Scientific Publications, 1987, 93-140.

19. Fleiss JL. The design of clinical experiments. New York, NY: John Wiley & Sons, 1986:41-70.

20. Gehan EA. Statistical methods for survival time studies. In: Staquet MJ, ed. Cancer therapy: prognostic factors and criteria of response. New York, NY: Raven Press, 1975:7-34.

21. Westling P. Studies of the prognosis in Hodgkin's disease. Acta Radiol 1965; 243 (Suppl 1):1-125.

22. Papillon J, Croizat P, Revol L, Chassard JL, Chassard E, Contamin A. Le pronostic de la maladie de Hodgkin. J Radiol Electrol 1966; 47:381-90.

23. Keller AR, Kaplan HS, Lukes RJ, Rappaport H. Correlation of histopathology with other prognostic indicators in Hodgkin's disease. Cancer 1968; 22:487-99.

24. Axtell LM, Myers MH, Thomas LH, Berard CW, Kagan AR, Newell GR. Prognostic indicators in Hodgkin's disease. Cancer 1972; 29:1481-8.

25. Haybittle JL, Hayhoe FGJ, Easterling MJ, et al. Review of British National Lymphoma Investigation studies of Hodgkin's disease and development of prognostic index. Lancet 1985; i:967-72.

26. Tubiana M, Henry-Amar M, Van der Werf-Messing B, et al. A multivariate analysis of prognostic factors in early stage Hodgkin's disease. Int J Rad Oncol Biol Phys 1985; 11:23-30.

27. Wagstaff J, Gregory WM, Swindell R, Crowther D, Lister TA. Prognostic factors for survival in stage IIIB and IV Hodgkin's disease: a multivariate analysis comparing two specialist treatment centres. Br J Cancer 1988; 58:487-92.

28. Proctor SJ, Taylor PT, Donna P, Boys R, Lennard A, Prescott RJ with members of the Scotland and Newcastle Lymphoma Group (SNLG) Therapy Working Party. A numerical prognostic index for clinical use in identification of poor-risk patients with Hodgkin's disease at diagnosis. Eur J Cancer Clin Oncol 1991; 27:624-9.

29. Gobbi PG, Cavalli C, Federico M, et al. Hodgkin's disease prognosis: a directly predictive equation. Lancet 1988; i:675-9.

30. Landberg T, Larsson LE. Hodgkin's disease. Retrospective clinico-pathologic study in 149 patients. Acta Radiol 1969; 8: 390-414.

31. Gobbi PG, Gobbi PG, Mazza P, Zinzani PL. Multivariate analysis of Hodgkin's prognosis. Fitness and use of a directly predictive equation. Haematologica 1989; 74:29-38.

32. Bettini R, Quadrelli CM, Quadrelli G, Biganzoli L, Venco A. Prognosis in Hodgkin's disease: verification of a new predictive equation (letter). Haematologica 1990; 75:482.

33. Gehan EA. Estimating survival functions from the life table. J Chronic Dis 1969; 21:629-44.

34. Johnson NL, Kotz S. Continuous Univariate Distributions-1. New York, NY: J. John Wiley & Sons, 1969:113-4.

35. Terblanche APS, Falkson G, Matzner L. The prognostic significance of age in patients with Hodgkin's disease. Eur J Cancer Clin Oncol 1988; 24:1805-9.

36. Walker A, Schoenfeld ER, Lowman JT, Mettlin CJ, Mac Millan J, Grufferman S. Survival of the older patient compared with the younger patient with Hodgkin's disease. Influence of histologic type staging and treatment. Cancer 1990; 65:1635-40.

37. Guinee VF, Giacco GG, Durand M, et al. The prognosis of Hodgkin's disease in older patients. J Clin Oncol 1991; 9: 947-53.

38. Bosi A, Messori A, Casini C, Bellesi G, Longo G, Rossi Ferrini P. Prognostic factors in Hodgkin's disease. Lancet 1988; i: 1117.

39. Coiffier B, Gisselbrecht C, Vose JM, Hervé Tilly Herbrecht R, Bosly A, Armitage JO for the Groupe d'Etudes des Lymphomes Agressifs. Prognostic factors in aggressive malignant lymphomas: description and validation of a prognostic index that could identify patients requiring a more intensive therapy. J Clin Oncol 1991; 9:211-9.

40. Wagstaff J, Steward WP, Jones M, et al. Factors affecting remission and survival in patients with advanced Hodgkin's disease treated with MVPP. Haematol Oncol 1986; 4:135-47.

41. Horwich A, Easton D, Nogueira-Costa R, Liew KH, Colman M, Peckham MJ. An analysis of prognostic factors in early stage Hodgkin's disease. Radiother Oncol 1986; 6:1-14.

42. Loeffler M, Pfreundschuh M, Hasenclever D, et al. Prognostic risk factors in advanced Hodgkin's disease. Report of the German Hodgkin Study Group. Blut 1988; 56: 273-81.

43. Specht L, Nordentoft AM, Cold S, Tøffner Clausen N, Nissen NI, for the Danish National Hodgkin Study Group. Tumor burden as the most important prognostic factor in early stage Hodgkin's disease. Relation to the other prognostic factors and implications for the choice of treatment. Cancer 1988; 61:1719-27.

44. Fuller LM, Gamble JF, Velazquez WS, et al. Evaluation of the significance of prognostic factors in stage III Hodgkin's disease treated with MOPP and radiotherapy. Cancer 1980; 45: 1352-64.

45. Osoba D, Falk JA, Sousan P, Ciampi A, Till JE. The prognostic value of HLA phenotypes in Hodgkin's disease. Cancer 1980; 46:1825-32.

46. Faguet GB, Davis H. Survival in Hodgkin's disease: the role of immunocompetence and other major risk factors. Blood 1982; 59:938-45.

47. Gobbi PG, Gendarini A, Crema A, et al. Serum albumin in Hodgkin's disease. Cancer 1985; 55:389-93.

48. Lee CKK, Bloomfield CD, Goldman AI, Lewitt SH. Prognostic significance of mediastinal involvement in Hodgkin's disease treated with curative radiotherapy. Cancer 1980; 46:2403-9.

49. Cosset JM, Henry-Amar M, Carde P, Clarke D, Le Bourgeois JP, Tubiana M. The prognostic significance of large mediastinal masses in the treatment of Hodgkin's disease. The experience of the Institut Gustave-Roussy. Hematol Oncol 1984; 2: 33-43.

50. Longo DL, Russo A, Duffey PL, et al. Treatment of advanced-stage massive mediastinal Hodgkin's disease: the case for combined modality treatment. J Clin Oncol 1991; 9:227-35.

51. Meerwaldt JH, van Glabbeke M, Vaughan-Hudson B. Prognostic factors for stage I and II Hodgkin's disease. In: Somers R, Henry-Amar M, Meerwaldt JK, Carde P, eds. Treatment strategy in Hodgkin's disease. Montrouge: John Libbey Eurotext, 1990:37-50.

52. Löffler M, Dixon DO, Swindell R. Prognostic factors for stage III and IV Hodgkin's disease. In: Somers R, Henry-Amar M, Meerwaldt JK, Carde P, eds. Treatment strategy in Hodgkin's disease. Montrouge: John Libbey Eurotext, 1990: 89-104.

53. Gobbi PG, Cavalli C, Gendarini A, et al. Reevaluation of prognostic significance of symptoms in Hodgkin's disease. Cancer 1985; 56:2874-80.

54. Gobbi PG, Pieresca C, Ricciardi L, et al. Night sweats in Hodgkin's disease. A manifestation of preceding minor febrile pulses. Cancer 1990; 65:2074-7.

55. Lister TA, Crowther D, Sutcliffe SB, et al. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds Meeting. J Clin Oncol 1989; 11:1630-6.

56. Ultmann JE. Classification of Hodgkin's disease: yesterday, today and tomorrow. Eur J Cancer Clin Oncol 1992; 28: 2074-9.

57. Harrel FE. The PHGLM procedure, in SAS supplemental users' guide. Cary, North Carolina: SAS Institute Inc., 1983: 267-94.

Appendix

International Database on Hodgkin's Disease (IDHD): list of contributors.

IDHD cooperating centers and cooperating groups

British National Lymphoma Investigation (BNLI), London, UK: K.A. MacLennan, B. Vaughan Hudson, G. Vaughan Hudson; **EORTC Lymphoma Cooperative Group:** M. Henry-Amar, J.H. Meerwaldt, R. Somers, J. Thomas; **Stanford University Medical Center, USA:** R.S. Cox, R.T. Hoppe, S.A. Rosenberg; **Princess Margaret Hospital, Toronto, Canada:** D.E. Bergsagel, M. Gospodarowicz, S. Sutcliffe; **Southwest Oncology Group (SWOG), USA:** C.A. Coltman, S.J. Dahlberg; **University of Texas M.D. Anderson Cancer Center, Houston, USA:** L.M. Fuller, F.B. Hagemeister; **Royal Marsden Hospital, London, UK:** S. Ashley, A. Horwich; **St. Bartholomew's Hospital, London, UK:** T.A. Lister; **Grupo Argentino de Tratamiento de la Leucemia Aguda (GATLA):** S. Pavlosky, M.T. Santarelli; **Università di Pavia, Italy:** P.G. Gobbi, M. Comelli; **Joint Center for Radiation Therapy, Boston, USA:** N.C. Coleman, P. Mauch; **Finsen Institute, Copenhagen, Denmark:** N.I. Nissen, L. Specht; **Fondation Bergonié, Bordeaux, France:** F. Bonichon, H. Eghbali; **German Hodgkin Study Group, Federal Republic of Germany:** V. Diehl, D. Hasenclever, M. Löffler, M. Pfreundschuh; **Groupe Pierre et Marie Curie, France:** H. Eghbali, A. Najman, R. Zittoun; **Christie Hospital & Holt Radium Institute, Manchester, UK:** D. Crowther, R. Swindell; **The Institute of Oncology, Ljubljana, Slovenia:** V. Pompe Kirn, M. Vork; **University of Nebraska, Omaha, USA:** J. Anderson.