Pergamon

*European Journal of Cancer* Vol. 32A, No. 3, pp. 480–490, 1996

Printed in Great Britain
0959–8049/96 $15.00+0.00

0959-8049(95)00635-4

## Original Paper

# The Follicular Non-Hodgkin's Lymphomas—II. Prognostic Factors: What do They Mean?

J.W. Denham,[1] E. Denham,[1] K.B. Dear[2] and G.V. Hudson[3]

[1]Radiation Onocology Department, Newcastle Mater Misericordiae Hospital, Hunter Region Mail Centre, NSW 2310, Australia; [2]Department of Statistics, The University of Newcastle, Callaghan, NSW 2308, Australia; and [3]British National Lymphoma Investigation Offices, Middlesex Hospital, Mortimer Street, London W1, U.K.

**The Ann Arbor staging classification has long been recognised to have shortcomings when used to stage the follicular lymphomas. To date, the identification of important prognostic variables has not succeeded in producing a superior staging classification that reflects the stages of dissemination of these processes in a way that can be used in the testing of new therapeutic strategies. A fresh look is taken at these factors. Data from 398 patients entered into the British National Lymphoma Investigation trials between 1974 and 1980, were analysed to evaluate the performance of the Ann Arbor staging classification. Multiple regression and proportional hazards techniques were used to determine what factors independently influence response to initial treatment, the durability of that response and ultimate survival, and to isolate factors that relate to disease progression from those that have other mechanisms of action. The Ann Arbor staging classification fared poorly, minimally separating relapse-free and cause-specific survival probabilities in patients with the largest staging groupings, III and IV. Significant prognostic heterogeneity was seen in both of these stage groupings, with 22% of patients with stage IV disease on the basis of marrow involvement having slightly better outcomes than patients with stage III disease. Significant differences in outcome were also observed between patients of different age and sex in each Ann Arbor stage grouping. Increasing number of lymph node regions involved, constitutional symptoms, the presence of splenomegaly and increasing age were observed to have powerfully independent adverse influence on probability of complete response to treatment and cause-specific survival. The evolution of the follicular lymphomas is reflected at the clinical level by an increase in the number of lymph node regions involved and splenomegaly. Simple classifications based on simple counts of lymph node regions involved and splenomegaly are more successful than the Ann Arbor staging classification in subdividing the series into patient subgroups that, regardless of gender or age, experience significantly different probabilities of responding completely to therapy and, as a consequence, relapse-free and cause-specific survival expectations. The definition of poor prognosis in subgroups may be of value in selecting patients for newer and more intensive therapeutic approaches.**

**Key words: non-Hodgkin's lymphoma, follicular, nodular, prognostic factors, staging**
*Eur J Cancer*, Vol. 32A, No. 3, pp. 480–490, 1996

## INTRODUCTION

UNLIKE HODGKIN'S disease, which appears to remain localised for considerable periods before spreading systematically via the lymphatics [1], the follicular lymphomas are frequently disseminated at presentation and often appear to spread randomly. Marrow involvement is a very common finding especially in patients with the predominantly small follicle cell variant, but in spite of this the patient may live for many trouble-free years without any treatment [2, 3].

The Ann Arbor staging classification, which was designed for use in Hodgkin's disease [4], is also widely used throughout the world for staging non-Hodgkin's (NHLs) lymphomas. However, as Rosenberg pointed out in 1977, the classification is not particularly satisfactory for the low-grade 'lymphocytic' NHLs, which include most cases of follicular lymphoma,

Correspondence to J.W. Denham.
Revised 2 Oct. 1995; accepted 13 Nov. 1995.

because the majority of patients have disseminated disease at presentation and because the prognosis of patients with stage III or IV disease can be extremely variable [5]. Rosenberg compared the situation with chronic lymphatic leukaemia in which every case could be considered to have Ann Arbor stage IV disease but where the prognosis could be anything from months to many years. Rosenberg concluded that the Ann Arbor classification should be replaced in the low-grade NHLs by a classification reminiscent of the one proposed by Rai and associates [6] for use in chronic lymphatic leukaemia.

In 1978, Galton and associates [7], Warnke and Levy [8] and Weissman and colleagues [9] suggested that some, or all, neoplastic follicle cells that compose the follicular lymphomas retain some of the migratory and 'homing' properties of normal follicle cells and speculated that the observed anatomical distribution of disease in B-cell domains of the lymphoid system and in marrow might be explained on this basis. Yet in spite of this now widely accepted [10] conceptual breakthrough, a new staging classification has not emerged. This is largely because the different stages of disease progression remain difficult to identify in a disease process that is frequently disseminated at presentation.

In the present paper, a fresh look is taken at factors that seem to have prognostic importance in the follicular lymphomas. In attempting to isolate factors that reflect disease progression, an effort has been made to understand how factors that are independent of disease stage exert their influence. It is argued that without such knowledge a rational staging classification that will be of value in treatment design as well as for prognostication will be unlikely to evolve.

## PATIENTS AND METHODS

*Patients*

This report is based on 398 patients with follicular non-Hodgkin's who, after giving their consent, were entered from 38 institutions in the United Kingdom into the British National Lymphoma Investigation (BNLI) trials between January 1974 and October 1980. The analysis that forms the basis of this report commenced in October 1992 providing a minimum follow-up period of 12 years and a maximum of almost 19 years.

The histopathological appearances of pathological material from all of these patients have been re-reviewed by the authors of 'British' classification of the NHLs [11] (Drs Michael Bennett, Kristin Henry and Geoffrey Farrar-Brown) and have subsequently been reviewed by Dr Bennett. The relationship between the British classification and the Working Formulation [12] is described in the first in this series of two reports (pages 470–479). Although the two classifications of the follicular lymphomas are comparable, it is likely that some patients with predominantly small cell histology in this report would be classified as having follicular mixed cell histology according to the Working Formulation, and some with mixed histology would be classified as follicular large cell according to the Formulation.

The patients reported herein were entered into three prospective randomised controlled clinical trials addressing management options in the low-grade NHLs (Table 1) organised by the BNLI during that period [13, 14] after the staging work-up protocol listed in Table 2. Details of these trials have been described in the first report in this series (pages 470–479).

*Overview of the analytical methodology used*

The first part of the analysis was to determine the performance of the Ann Arbor staging classification in this series of patients. Examined in this part of the analysis was the practical utility in terms of size of the staging categories, the prognostic discrimination offered by the categories in patients of both sex, at all ages and with all histological subtypes, and the presence of important prognostic heterogeneities within the staging categories.

The second part of the analysis was to determine the relative influence of and interaction between prognostic variables. A series of multiple regression and Cox proportional hazards regression analyses [15] were performed to determine which disease 'stage-related' variables (i.e. clinical manifestations of involvement) consistently reflect disease advancement, regardless of the age or sex of the patient or the histological subtype of the disease. To do this, it was necessary to establish which variables have an important influence both on response to treatment and ultimate survival that is independent of age, sex and histological subtype. First, therefore, multiple logistic regression procedures were performed to establish which factors independently influence response to initial therapy. Then proportional hazards models were derived in completely responding patients to determine which factors influence durability of remission. Finally, proportional hazards models based on the whole series of patients (regardless of response to treatment) were used to establish which patient subgroups enjoy better cause-specific survival expectations.

*Definitions adopted*

In analyses designed to determine whether associations exist between an increasing number of lymph nodes sites with features such as marrow involvement, splenomegaly, constitutional symptoms, etc., and outcome measures such as response to therapy, relapse-free and cause-specific survival, a simplification of the classification of lymph node sites involved was used in this study. This simplification involved the definition of six lymph node 'regions' (groups of adjacent sites). Three regions above the diaphragm and three below were defined to reduce the probability of an artificial over-emphasis in the importance of supradiaphragmatic sites in such analysis due to the fact that approximately 20% of patients in the series had not been subjected to investigations to assess the intra-abdominal nodes. The 'regions' included two 'peripheral' regions on either side of the diaphragm (above the diaphragm: both cervical/axillary 'regions', below the diaphragm: both inguino-iliac 'regions') and one 'central' region on either side of the diaphragm (above the diaphragm: the mediastinal 'region'; below the diaphragm: the para-aortic/mesenteric/porta hepatic 'region'; Figure 1).

The definitions chosen for splenomegaly and hepatomegaly were based upon the probability of involvement of these two organs being involved by their palpable enlargement. In follicular lymphoma, it is highly probable that enlargement of the spleen will indicate its involvement and this was therefore used as the criterion for scoring 'splenomegaly' present. Studies on spleens removed at staging laparotomy or post mortem [16–19] indicate that involvement occurs in probably 35% of spleens weighing less than 400 g, 26–39% of spleens up to 200 g, 44–66% of those weighing between 200 and 400 g, and in nearly 100% of spleens weighing 400 g or more. Data of this type suggests that a palpable spleen is highly likely to be involved, whereas an impalpable spleen is likely to be involved

*Table 1. British National Lymphoma Investigation (BNLI) trials in low-grade NHL* (1974–1980). *The BNLI grade 1 trials* (n = *number of patients with follicular lymphoma and considered in this report who were accrued into the various trial arms. A further* 13 *patients included in this report were treated with COP between the first and second disseminated disease trials*)

| BNLI trials in low-grade NHL (1974–1980) |
|---|
| Limited disease trial (Ann Arbor stages I and II)<br>Involved field radiotherapy to 35 Gy ($n = 55$)<br>*versus*<br>Involved field radiotherapy to 35 Gy then chlorambucil 0.2 mg/kg/day orally for 8 weeks then chlorambucil 0.1 mg/kg/day orally for 16 weeks ($n = 50$) |
| Disseminated disease trials<br>*First trial*<br>Chlorambucil 0.2 mg/kg/day orally until complete remission and then for 8 weeks following this, then chlorambucil 0.1 mg/kg/day orally until a total of 2 years had elapsed since starting treatment ($n = 50$)<br>*versus*<br>i.v. cycles of COP at intervals of 2–4 weeks until complete remission obtained then three further cycles following this to a minimum of six cycles administered ($n = 47$)<br>COP:<br>Cyclophosphamide 600 mg/m$^2$ (max 1 g) i.v. days 1 and 8<br>Vincristine 1.4 mg/m$^2$ (max 2 mg) i.v. days 1 and 8<br>Prednisone 50 mg/m$^2$ (max 100 mg) orally days 1 to 8<br>*Second trial*<br>Chlorambucil ($n = 90$) versus COP ($n = 93$) as above. Patients not achieving complete remission within 3 months of starting therapy were switched to the opposite trial arm.<br>A second randomisation to chlorambucil 0.1 mg/kg/day orally until 2 years had elapsed after commencing treatment or no further treatment occurred in completely responding patients. |

*Table 2. Staging protocol for patients accrued into the British National Lymphoma Investigation trials*

- History and examination
- ENT examination
- Nodal biopsy
- Weight
- Full blood count (including white cell differential and ESR)
- Liver function studies (including albumen and globulin levels)
- Chest X-ray, PA and lateral
- Mediastinal tomograms*
- Abdominal lymphangiogram
- Bone marrow
- Scintigraphy of liver and spleen*
- Ultrasound of liver, spleen and abdominal nodes*
- CT scanning abdomen and thorax*
- Liver biopsy*
- Laparotomy

*Optional, depending upon clinical indications and availability.

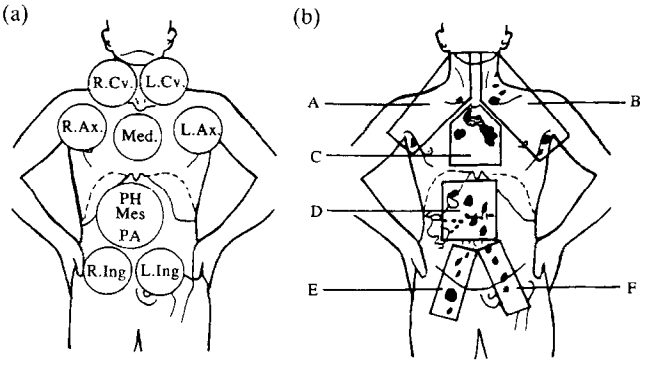


**Figure 1. (a) The arbitrary grouping of lymph node sites used in this study. (b) The six arbitrary node regions (groups of lymph node sites) used in this study (see text for further details).**

in only 30–40% of cases. This has been confirmed in the studies of Moran and associates [20] and Goffinet and associates [21]. Since the assessment of spleen size on isotope scan is not particularly reliable [22–25], clinical enlargement of the spleen to 2 cm below the costal margin was adopted as the criterion for splenomegaly in this study, with scintigraphic evidence of splenomegaly accepted only in the obese patient where palpation of the abdominal contents proved difficult. It was felt that the application of this definition of splenomegaly would make it extremely likely that the spleen was not only enlarged but also involved in patients so designated.

The decision regarding the definition of hepatomegaly was less straightforward. The liver edge is often palpable in normal individuals and is felt even more frequently in patients with emphysema or congestive cardiac failure. However the liver is frequently involved in follicular lymphoma [17, 21, 26–29] and significant involvement is likely to be manifest by hepatic enlargement [19]. The criterion of hepatomegaly used in this study was a liver edge palpable 4 cm below the costal margin. Hass and associates [29], Goffinet and associates [30], Chabner and asssociates [31] and Moran and associates [20]

found that the incidence of liver involvement correlates strongly and positively with degree of splenic involvement (assessed by its weight). Goffinet and colleagues [21] found the liver to be involved in more than two thirds of patients with palpable splenomegaly (spleen size 400–900 g or more: 12 out of 18 cases liver involved; Spleen size 900 g or more: 9 out of 9 cases liver involved too). Although the measurement of liver volume using emission computed tomography may prove to be an extremely accurate method [32], conventional scintigraphy (as used in the 1970s) is extremely disappointing in the estimation of liver size and involvement in follicular lymphoma. It was considered that hepatomegaly, so defined, in the presence of splenomegaly would indicate a probability of involvement in excess of 0.66.

*Statistical methodology*

The clinical data supplied by clinicians who participated in the BNLI trials were entered on to a PC-based temporal database package (MEDLOG) and were analysed using some of the statistical options available on this database (which include Kaplan–Meier survival curve options [33], the log-rank test [34] for differences between curves, parametric and non-parametric tests for comparing variable value distribution between subgroups, multiple logistic and Cox proportional hazards regression analyses) [15]. Survival was estimated from the date of starting treating to the date of death, or to the date last seen alive. Death was assumed to be a result of lymphoma unless the patient was in first remission up to the time of death and the cause of death was certified as being due to another (intercurrent) cause. Relapse-free survival was based on the time to estimated date of relapse (or to date that relapse was first recorded—if an estimate was unavailable) from the date of starting treatment in patients who responded completely to therapy. The resulting curves therefore commence at the percentage of patients who responded completely to initial therapy.

## RESULTS

*Demographic information*

In this series, males exceeded females by a ratio of 222:176. However, neither age distribution of patients within the series nor distribution of histological subtype differed with respect to gender. Almost exactly one third of male and female patients were younger than 50 years and one third were over 62 years. 270 (67.8%) had predominantly small cell histology, 108 (27.2%) had mixed cell and 20 (5.0%) predominantly large cell. Undoubtedly the incidence of marrow involvement in this series (20.9%) is under-represented. In the early years of accrual into these trials, it is quite possible that a failure to appreciate that involvement is commonly paratrabecular in distribution [19, 35] may provide an explanation for the low incidence of detection. Ililac crest trephine biopsies replaced sternal aspiration examinations once the higher yield of positive findings from doing so [1, 11, 13, 14, 35] became widely recognised in the late 1970s throughout the United Kingdom. Marrow involvement and splenomegaly were more frequently involved in male patients and those with predominantly small cell histology. However, after multiple regression analysis, which took into account number of lymph node regions involved only the association between the presence of splenomegaly, and gender and small cell histology retained independent significance. Lower haemoglobin values were seen in female patients and lower serum albumen levels were recorded in patients who were over 62 years. As number of lymph node regions involved increased, almost linear increases in incidence of marrow involvement, splenomegaly and constitutional symptoms were observed in patients of both sexes with all histological subtypes (Figure 2). Decreases in haemoglobin and albumin levels were observed to accompany these increases. Increasing Ann Arbor stage also correlated with an increasing incidence of splenomegaly and constitutional symptoms and decreasing levels of haemoglobin and albumen, but the relationships were not linear.

*The performance of the Ann Arbor staging system*

*Overall performance.* The Ann Arbor staging system was observed to have some prognostic validity in the series taken as a whole (Figure 3). Although it did not discriminate well between the two most numerically large categories (stages III and IV) in terms of relapse-free or cause-specific survival, it did separate these from the two more favourable categories. The presence of constitutional symptoms was observed to influence adversely prognosis in most stages although to a considerably varying degree.

When the series was divided according to patient gender, the prognostic discrimination provided by the classification broke down. In female patients, minimal differences between Ann Arbor stages II, III and IV could be detected in relapse-free or cause-specific survival at 5 or 10 years. This appeared to be because female patients with Ann Arbor stages III and IV disease had better relapse-free and cause-specific expectations than male patients with these stages (10-year figures with 95% confidence intervals: for relapse-free survival, males $14.4 \pm 4.8$ versus females $21.4 \pm 7.2\%$, $P = 0.046$; cause specific survival, males $33.2 \pm 7.6$ versus females $50.7 \pm 9\%$, $P = 0.0002$). Patients over 62 had worse relapse-free and cause-specific survival expectations than younger patients in every Ann Arbor stage category (10-year relapse-free figures: $16.7 \pm 7.3\%$ versus $27.5 \pm 5.6\%$, $P = 0.046$; cause-specific, $29.4 \pm 8.5$ versus $54 \pm 6.4\%$, $P < 0.0001$, Figure 4). The prognostic discrimination between the stage III and IV categories was particularly poor for patients over 62 years (a non-significant advantage in favour of stage IV disease was observed in both male and female patients).

*Utility of marrow involvement as a criterion for stage IV disease.* The lack of independent influence of marrow involvement apparent in the proportional hazards model was well illustrated by stratifying the series according to the number of lymph node regions involved. The prognosis of patients with marrow involvement (who by definition had Ann Arbor stage IV disease) varied considerably according to the number of lymph node regions involved both in terms of relapse-free and cause-specific survival (Figure 5). In fact, patients with marrow involvement who had only one or two lymph node regions involved (i.e. 22% of patients in the stage IV category) had slightly, but not significantly, better relapse-free and cause-specific survival expectations than patients with stage III disease (who by definition had not been found to have marrow involvement)—10-year relapse-free and cause-specific figures with 95% confidence intervals were $26.7 \pm 21.4$ versus $20 \pm 6.1\%$ ($P = 0.28$) and $48.5 \pm 24.1$ versus $45.1 \pm 8.2\%$ ($P = 0.45$), respectively.

*Utility of nodal distribution with respect to the diaphragm as a staging criterion.* The Ann Arbor staging classification dis-

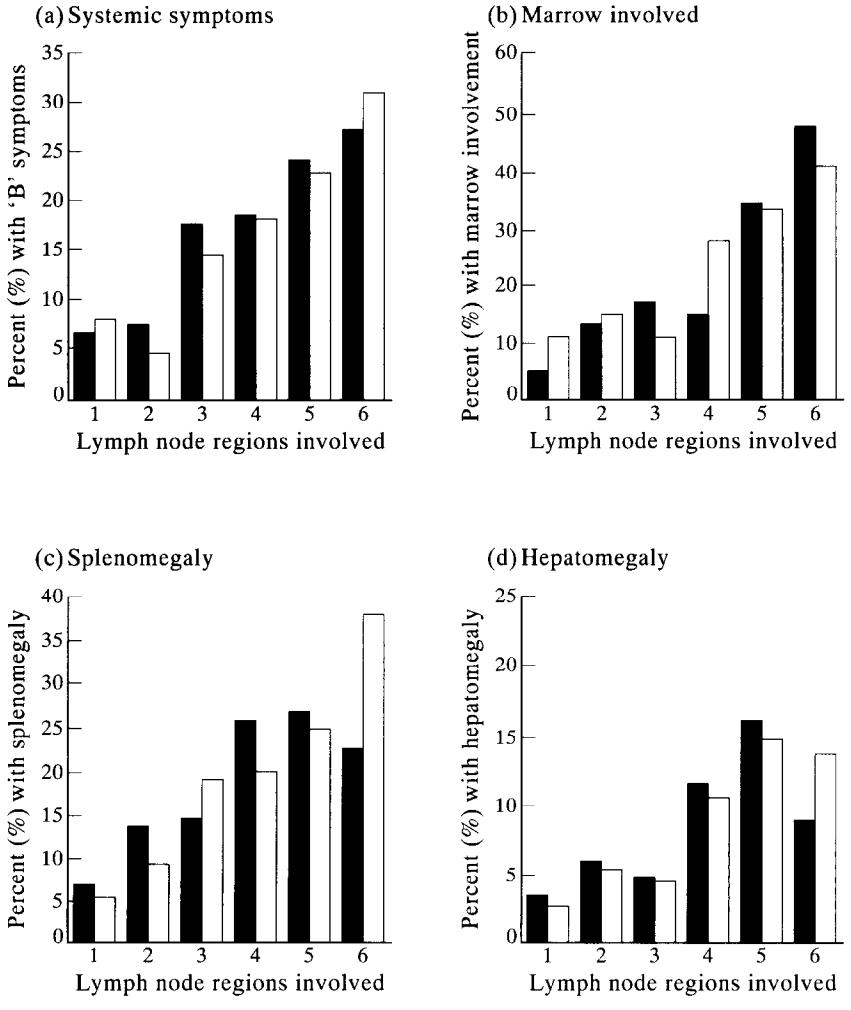


**Figure 2. The incidence of (a) systemic symptoms, (b) marrow involvement, (c) splenomegaly and (d) hepatomegaly according to the number of lymph node regions involved. The closed bars refer to the number of regions counted. The open bars take into account the assumption when counting regions involved that the para-aortic nodes were involved in cases where no assessment of the retroperitoneum had been carried out.**

criminates between patients with several lymph node groups involved on one side of the diaphragm, i.e. stage II (but without extra-lymphatic involvement), with several lymph node groups involved but distributed anatomically on both sides of the diaphragm (who also do not have extra-lymphatic involvement), i.e. stage III. However, patients that have two or three lymph node regions involved that are situated on both sides of the diaphragm were not found to have worse relapse-free or cause-specific survival expectations than patients with two or three regions involved on one side of the diaphragm only (at 10 years: relapse-free, 24 ± 10 versus 21.6 ± 13.3%, $P = 0.37$; cause-specific, 50 ± 9.7 versus 44.2 ± 15.6%, $P = 0.72$).

*Prognostic factors*

*Influence of patient-related and disease-subtype-related variables on prognosis.* Presented in Table 3 are the results of logistic regression and Cox proportional hazards regression analyses that took into account the influence of patient-related factors, such as age and sex as well as various stage-related variables and histological subtypes, on the expectation of complete response to therapy, relapse-free survival following complete remission and cause-specific survival in the whole series. Proportionality considerations aside, the events examined probably do not respond linearly to increasing age as the models assume. Within the limitations of the methodology, the models do demonstrate quite clearly, however, that increasing age does have an adverse influence on probability of complete response to therapy and cause-specific survival. They also suggest that male gender is an important adverse factor that needs to be taken into account when looking at cause-specific survival, because it seems to reduce the odds of obtaining remission as well as reducing its durability. Histological classification also emerged as an independent predictor of response to treatment and cause-specific survival. The model predicts that patients with predominantly small cell histology will respond completely to therapy more frequently than patients with mixed cell histology, and enjoy better survival expectations. Similarly, patients with mixed histology were predicted to benefit more from treatment than patients with large cell histology. These predictions were reflected by better survival figures for patients with predominantly small cell histology than for patients with mixed cell histology (at 10 years: 49.6 ± 6.2 versus 36.7 ± 9.8%, $P = 0.064$). Unfortu-

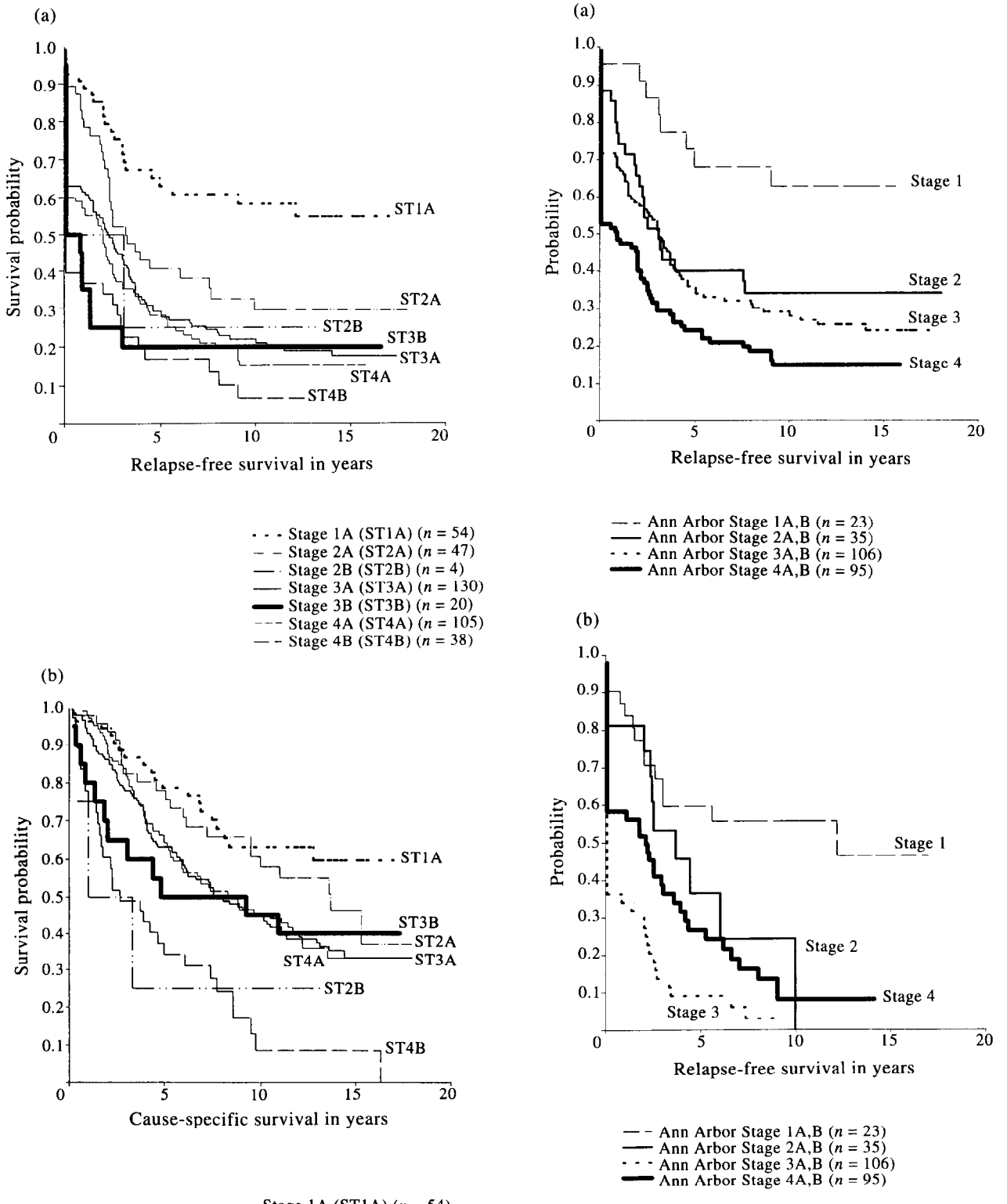


**Figure 3. (a) Relapse-free and (b) cause-specific survival according to Ann Arbor stage.**




**Figure 4. Relapse-free survival according to Ann Arbor stage in patients younger than 62 years (a) and in patients 62 years or more (b).**

nately, too few patients with large cell histology were at risk at 10 years to draw a meaningful comparison between this subtype and the others.

*Influence of disease-stage-related variables.* After the important influence of increasing age, male sex and histologi-

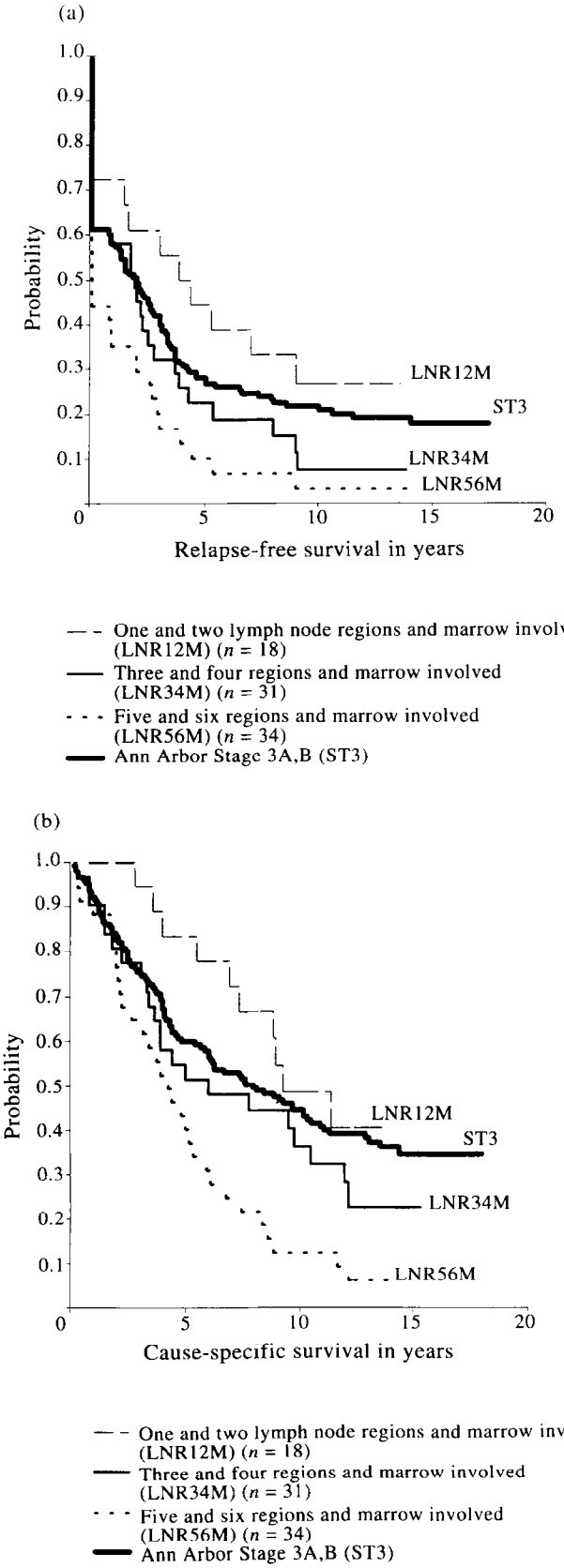


**Figure 5. (a) Relapse-free and (b) cause-specific survival in patients found to have marrow involvement according to the number of lymph node regions involved. For comparative purposes, curves for patients with Ann Arbor stage III disease (i.e. who were not found to have marrow involvement) are shown.**

cal subtype are taken into account, the presence of constitutional symptoms and splenomegaly emerge as having an important independent adverse influence on probability of complete response to therapy and cause-specific survival. The same is true of an increasing number of lymph node regions involved. In addition, however, this is the only variable shown to have a major independent influence probability of relapse once remission is obtained. The influence of Ann Arbor stage is less consistent. Although it appears to have an important independent influence on response to therapy, it does not appear to have an influence on cause-specific survival that is independent of the other variables considered. Marrow involvement did not emerge in any of the models as having an adverse influence independent of other variables.

*Reclassification of the series based on prognostic variables.* In the multiple regression and proportional hazards analyses, whose results have been described herein, increasing number of lymph regions involved, the presence of splenomegaly and the presence of constitutional symptoms emerged as candidates to provide a better prognostic model than the Ann Arbor classification. Figure 6 provides a reclassification of the series simply based on the number of lymph node regions involved. Although an obvious prognostic gradient may be discerned, clear separation of patients with two, three and four lymph node regions involved does not occur. In addition, the categories vary considerably in numerical size.

Increased prognostic discrimination can be achieved by a reduction in the number of categories to four by combining groups of patients with two and three regions involved and patients with four and five regions involved. Further improvement takes place if patients with splenomegaly are included in the category of patients with six regions involved (Figure 6). It will be noted that this reclassification produces categories of more even size than the Ann Arbor classification as well as producing statistically significant differences in relapse-free and cause-specific survival expectation between each category (10 year relapse-free and cause-specific figures: category I, 52.2 ± 11.8 and 66.4 ± 11.2%; category II, 27.1 ± 8.8 and 50.9 ± 9.8%; category III, 16.5 ± 6.9 and 41.7 ± 9.2%; category IV 7 ± 5.6 and 27.5 ± 9.7%; relapse-free long-rank comparisons: I versus II, $P = 0.005$; II versus III, $P = 0.023$; III versus IV, $P = 0.002$; cause-specific log-rank comparisons: I versus II, $P = 0.026$; II versus III, $P = 0.049$; III versus IV, $P = 0.014$). It also predicts an almost linear relationship between failure to obtain complete remission to therapy and increasing category. Not shown, but of relevance, are relapse-free and cause-specific Kaplan–Meier survival plots after the series has been stratified according to age and sex, which show that the categories provided by this classification separate consistently throughout the series but most strongly in male patients. Although constitutional symptoms would not appear to be common in category 1 and 2 patients, they were noted to exert important adverse influence on prognosis in categories 3 and 4.

## DISCUSSION

After more than 20 years of doubt it will not be a surprise that the Ann Arbor staging classification has not performed particularly well in a large series such as this one that has a follow-up exceeding 12 years. Its most important failure in this series has been in separating the two largest stage groupings III and IV. Perhaps the most important reason for this, as Rosenberg predicted in 1977, is the fact that the finding of

*Table 3. Results from a logistic regression analysis to determine the influence of various factors on probability of obtaining complete remission following therapy. The table also provides results from Cox proportional hazards analyses performed to determine the influence of these variables on relapse-free survival once remission is obtained and on cause-specific survival*

| Factors | Failure to respond completely to therapy ($n = 398$) Odds ratio (95% CI) $P$ value | All patients in all trial arms Relapse in complete responders ($n = 264$) Hazard ratio (95% CI) $P$ value | Decreasing cause-specific survival ($n = 398$) Hazard ratio (95% CI) $P$ value |
|---|---|---|---|
| Increasing age | 1.02 (1.00–1.04) $P = 0.02$ | 1.01 (1.00–1.02) $P = 0.13$ | 1.03 (1.02–1.05) $P < 0.0001$ |
| Female gender | 0.83 (0.51–1.34) $P = 0.44$ | 0.77 (0.56–1.06) $P = 0.11$ | 0.66 (0.50–0.86) $P = 0.003$ |
| Histological subtype | 1.50 (1.00–2.25) $P = 0.05$ | 1.06 (0.80–1.40) $P = 0.70$ | 1.34 (1.07–1.68) $P = 0.012$ |
| Systemic symptoms absent | 0.54 (0.29–1.00) $P = 0.054$ | 1.07 (0.63–1.80) $P = 0.81$ | 0.60 (0.43–0.85) $P = 0.004$ |
| Increasing Ann Arbor stage | 1.71 (1.19–2.45) $P = 0.003$ | 1.03 (0.82–1.30) $P = 0.77$ | 0.98 (0.80–1.20) $P = 0.87$ |
| Increasing lymph node regions | 1.21 (1.01–1.44) $P = 0.03$ | 1.25 (1.10–1.42) $P = 0.0004$ | 1.25 (1.12–1.39) $P < 0.0001$ |
| Marrow not involved | 1.57 (0.83–3.00) $P = 0.17$ | 1.25 (0.79–1.97) $P = 0.34$ | 0.77 (0.54–1.10) $P = 0.16$ |
| Splenomegaly absent | 0.29 (0.16–0.53) $P < 0.0001$ | 1.12 (0.67–11.87) $P = 0.68$ | 0.68 (0.48–0.96) $P = 0.03$ |

marrow involvement is not an ominous finding in all cases. Some patients with marrow involvement have been shown to have a median survival of just over 9 years—slightly longer than the median survival of all patients with stage III disease (who by definition were not found to have marrow involvement). As suggested previously [5, 36], the finding of marrow involvement would therefore not seem to be a good reason, by itself, to allocate patients into the highest staging category. In part, this is the reason for the failure of the classification to separate between stages III and IV. Many patients with stage III disease who have adverse features such as five or six lymph node regions involved and splenomegaly fare worse than the not insignificant number of patients with Ann Arbor stage IV disease who have less than five lymph node regions involved.

The analysis of prognostic factors operative in this series has not produced major differences with conclusions drawn by other investigators using data from other contemporary series. As in other analyses, increasing age (particularly over 60 years) [37–46], histological subclassification, the presence of constitutional symptoms [36, 39, 46–50] and increasing number of lymph node regions [45, 51] have been found to exert the most powerful independent adverse influences upon the probability of obtaining complete remission to therapy, and subsequent relapse-free and cause-specific survival. The adverse impact of the presence of splenomegaly on probability of obtaining complete remission in particular is interesting and has not been extensively documented previously [40]. The important question to ask is what these factors mean. The adverse influence of advancing age on the odds of obtaining complete clinical remission following therapy is hardly surprising. Perhaps decreasing marrow reserve and host resistance, amongst others, are the important operative factors. The adverse influence of male gender on odds of failure to respond completely and remission durability is more difficult to explain. It is a factor that is prognostically important for many other diseases, both benign and malignant.

It has not been possible to account for the presence of other intercurrent life shortening afflictions in this series and this may be a consideration. Results from this series suggest that after one lymph node site has become involved, it matters little whether subsequent sites of involvement (up to three regions) are confined to one side of the diaphragm or distributed on both sides. Classification according to spatial (anatomical) distribution of involvement, as provided by the Ann Arbor staging classification, breaks down at this point. The powerful and independent adverse influence of increasing number of lymph node regions involved in odds of obtaining initial remission, then remaining in remission and ultimately dying from follicular lymphoma, is of particular interest in this respect. It is now held that the evolution of the follicular lymphomas is to be characterised by an accumulation of neoplastic lymphocytes throughout the lymphoid system. Perhaps, therefore, the prognostic importance of an increasing number of lymph node regions involved reflects this phenomenon and is more an expression of increasing disease burden than of anatomical dispersion of the disease.

The lack of independent adverse significance of marrow involvement will probably not surprise the many investigators that are critical of its use as a criterion for placing patients in the highest staging category (Ann Arbor stage IV). Nevertheless, the incidence data in this series are interesting in that an almost linear correlation between increasing number of lymph node regions involved and increasing incidence of marrow involvement could be demonstrated. This finding adds weight to similar findings from other series which suggest that the discovery of marrow involvement is a stochastic phenomenon, i.e. that marrow involvement will be found with increasing frequency if a greater number of marrow samples from each patient are examined [1, 21, 28, 32, 33] and will be found as

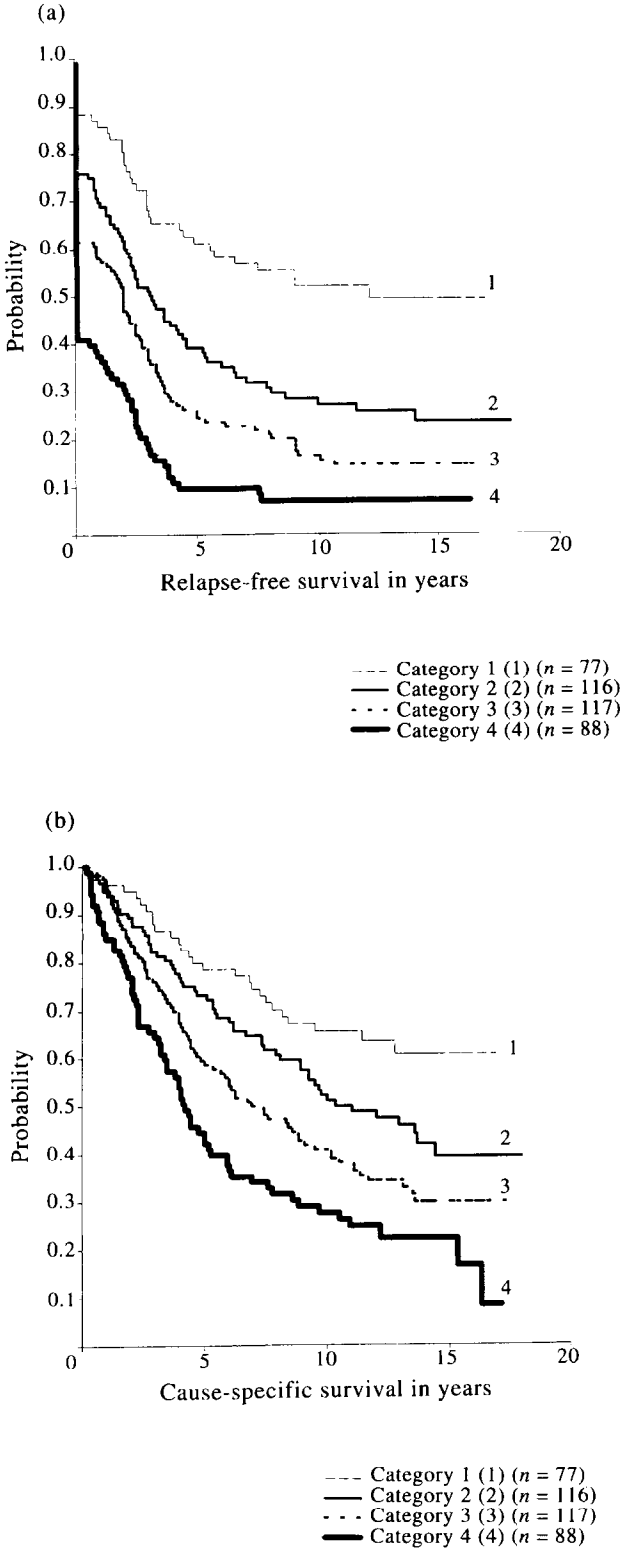


**Figure 6. (a) Relapse-free and (b) cause-specific survival according to lymph node involvement and splenomegaly. Category 1—one lymph node site without splenomegaly $n = 77$ (19%). Category 2—two or three lymph node regions without splenomegaly $n = 116$ (29%). Category 3—four or five lymph node regions without splenomegaly $n = 117$ (29%). Category 4—six lymph node regions and/or splenomegaly present $n = 88$ (22%). $n$ = number of patients in present series who would be allocated to each category.**

accumulation of neoplastic follicle cells [9] in the lymphoid system takes place. Further support for this proposal comes from the proportional hazard analysis which showed that the finding of marrow involvement did not have prognostic significance independent of the number of lymph node regions in any subgroup of patients classified according to age, sex and histological subtype. Of course, it could be legitimately argued that this null finding has come about because marrow involvement was under-diagnosed and because the resulting subgroups were too small to detect differences that were really there. However, this explanation overlooks the fact that the marrow involved subgroup was composed of patients with markedly different relapse-free and overall survival expectations and that these differences depended in an entirely coherent manner on the number of lymph node regions involved. Patients who had a larger number of regions involved had a worse prognosis regardless of whether or not their marrows were shown to be involved. Although the frequency of splenomegaly was observed to correlate almost linearly with increasing number of lymph node regions involved, just as marrow involvement was found to do, it is less easy to explain the apprarently independent adverse influence of splenomegaly on probability of complete remission and cause-specific survival observed in this series on the basis of a purely stochastic phenomenon. Patients with splenomegaly had expectations of complete remission to therapy that were inferior to those of patients with the same number of lymph node groups involved but without splenomegaly. Perhaps the presence of splenomegaly might be viewed as an additional but highly significant site of disease 'bulk' in some or all of the patients. Its disappearance following therapy may require particularly high levels of cell kill. It must be pointed out, however, that the proportional hazards analysis that assesssed the influence of factors responsible for relapse in completely responding patients provided no support for this proposal. An alternative possibility is that splenomegaly may be a feature of a follicular lymphomatous subtype that is more resistant to therapy and therefore has a worse outlook than the more common presentations. However, although splenomegaly was found to be more common in male patients and patients with predominantly small cell histology, no clear-cut evidence emerged from the series of the presence of a specific disease subtype in which splenomegaly is a manifestation.

It will be noted that the re-classification of this series using the number of lymph node regions involved and splenomegaly into categories that perform better than the Ann Arbor classification staging categories is a simple matter. Not only do four prognostic categories of substantial size result, but each of these categories differs to a highly significantly different extent with respect to response to treatment and subsequent outcome. The question is whether doing this has any clinical utility in the modern era where, for example, computerised tomography has produced major improvements in the diagnosis of intra-abdominal and intrathoracic disease. More accurate information concerning number of lymph node regions involved could quite obviously impact on the results of an analysis of prognostic factors such as this one. Also unknown is the independent impact of factors, such as pre-treatment serum LDH, performance status, sites of bulky disease and extra-nodal spread, which have been shown in the 'International non-Hodgkin's lymphoma prognostic factors project' to be of important predictive value for patients with

aggressive non-Hodgkin's lymphoma [52]. More recent analyses suggest that these factors also have importance for the follicular lymphomas [53–55]. Unfortunately, LDH and performance status were not routinely evaluated at the time the trials in this report were undertaken. The clinical utility of prognostic scheme based on a simple count of lymph node regions involved and splenomegaly, which are shown to be potent prognostic variables in this report and are easy to estimate without recourse to invasive investigation, therefore awaits evaluation in a reliably staged modern series in which serum LDH and performance status estimates, in particular, are available. If a simple count of lymph node regions involved and an assessment of splenomegaly were to retain their independent potency following such an evaluation, the question would then arise as to whether they could be used to derive a new staging system that would be useful in the assessment of new treatment methods.

The development of a staging system for low-grade non-Hodgkin's lymphoma with reasonable predictive value for outcome, particularly in the Ann Arbor Stage IV category, has become increasingly important in the era of new therapies such as high-dose chemotherapy and biological response modifiers [54, 55]. Currently, the majority of the data available to assess more intense treatment come from phase II studies. The inability of the Ann Arbor staging system to predict outcome has made it difficult to regard the comparisons of the results of these high-dose strategies with historical controls with any certainty. In addition, a randomised design for a study of high-dose therapy would be unlikely to compare like with like if Ann Arbor stage was a major criterion used for stratification. The results of this re-analysis and efforts such as the 'International non-Hodgkin's lymphoma prognostic factors project' will hopefully make interpretation of the results of new therapies for non-Hodgkin's lymphoma more reliable.


1. Rosenberg SA, Kaplan HS. Evidence for an orderly progression in the spread of Hodgkin's disease. *Cancer Res* 1966, **26**, 1225–1231.
2. Portlock CS, Rosenberg SA. No initial therapy for stage II and IV non-Hodgkin's lymphomas of favourable histologic types. *Ann Intern Med* 1979, **90**, 10–13.
3. Horning SJ, Rosenberg SA. The natural history of initially untreated low-grade non-Hodgkin's lymphomas. *N Engl J Med* 1984, **311**, 1471–1475.
4. Carbone PP, Kaplan HS, Mushoff K, Smithers DW, Tubiana M. Report of the committee on Hodgkin's disease staging classification. *Cancer Res* 1971, **31**, 1860–1861.
5. Rosenberg SA. Validity of the Ann Arbor staging classification for the non-Hodgkin's lymphomas. *Cancer Treat Rep* 1977, **61**, 1023–1047.
6. Rai KR, Sawitzky A, Cronkite EP, Chanana AD, Levy RN, Pasternak BS. Clinical staging of chronic lymphocytic leukaemia. *Blood* 1975, **46**, 219–234.
7. Galton DAG, Catovsky D, Wiltshaw E. Clinical spectrum of lymphoproliferative disease. *Cancer* 1978, **42**, 901–910.
8. Warnke R, Levy R. Immunopathology of the follicular lymphomas—a model of 'B' lymphocyte homing. *N Engl J Med* 1978, **298**, 481–486.
9. Weissman IL, Warnke R, Butcher EC, Rouse R, Levy R. The lymphoid system: its normal architecture and the potential for understanding the system through the study of lymphoproliferative diseases. *Human Pathol* 1978, **9**, 25–45.
10. Groves ES, Longo LD. Progression of lymphoproliferative disorders and haematological malignancies. *Cancer Growth Prog* 1989, 7, 122–147.
11. Bennett MH, Farrer-Brown G, Henry K. A classification of the non-Hodgkin's lymphomas. Presented at the Workshop on Classification of non-Hodgkin's Lymphomas, University of Chicago. *Clin Radiol* 1973, **20**, 339–343.
12. Rosenberg SA, Berard CW, Brown BW, *et al.* Study of classificaitions of non-Hodgkin's lymphomas: summary and description of a working formulation for clinical usage. *Cancer* 1982, **49**, 2112–2135.
13. Phillips DL. Radiotherapy in the treatment of localised non-Hodgkin's lymphoma (Report No. 16). *Clin Radiol* 1981, **32**, 543–546.
14. Hayhoe FGH. Chemotherapy in the management of stage III/IV grade I non-Hodgkin's lymphomas (Report No. 17). *Clin Radiol* 1981, **32**, 547–552.
15. Cox DR. Regression models and life tables. *J R Stat Soc* 1972, **34**, 187–220.
16. Veronesi V, Musemeci R, Pizzett F, Gennari L, Bonnadonna G. The value of staging laparotomy in non-Hodgkin's lymphomas (with emphasis on the histiocytic type). *Cancer* 1974, **33**, 446–454.
17. Rosenberg SA, Dorfman RF, Kaplan HS. The value of sequential bone marrow biopsy and laparotomy and splenectomy in 127 consecutive untreated patients, with non-Hodgkin's lymphoma. *Br J Cancer* 1975, **31**, 221–227.
18. Lotz MJ, Chabner B, De Vita VT, Johnson RE, Berard CW. Pathologic staging of 100 consecutive untreated patients with non-Hodgkin's lymphoma. *Cancer* 1976, **37**, 266–270.
19. Risdall R, Hoppe RT, Warnke R. Non-Hodgkin's lymphoma. A study of the evolution of the disease based on 92 autopsied cases. *Cancer* 1979, **44**, 529–542.
20. Moran EM, Ultmann JE, Ferguson DJ, Hoffer PB, Ranniger K, Rappaport H. Staging laparotomy in non-Hodgkin's lymphoma. *Br J Cancer* 1975, **31**, 228–236.
21. Goffinet DR, Warnke R, Dunnick NR, Castellino R, Glastein E, Nelsen TS. Clinical and surgical (laparotomy) evaluation of patients with non-Hodgkin's lymphomas. *Cancer Treat Rep* 1977, **61**, 981–992.
22. McCready VR, Gwyther MM, Stringer AM. Scintigraphy in the diagnosis and treatment of malignant lymphomas. In Anonymous. *Medical Radioisotope Scintigraphy*. Vienna, IAEA, 1969, 419–430.
23. Keeling DH, Newman GB, Baker JT, Millett Y. The measurement of splenic disease in vivo. *Br J Radiol* 1970, **44**, 403.
24. Parmentier C, Di Paola R, Bazin JP, *et al.* Le splenographie dans les hemtosarcoms. In *Medical Radioisotope Scintigraphy*. Vienna, IAEA, 1969, 101–127.
25. Seligson A. Some reported equations for relating spleen weight in adults to the length (or other dimension) on scan. In *CRC Handbook in Clinical Laboratory Science*. Florida, CRC, 1981.
26. Rappaport H, Winter WJ, Hicks EB. Follicular lymphoma—a re-evaluation of its position in the scheme of malignant lymphoma, based on a survey of 253 cases. *Cancer* 1956, **9**, 792–821.
27. Rosenberg SA, Diamond HD, Jaslowitz B, Craver LF. Lymphosarcoma: a review of 1269 cases. *Medicine* 1961, **40**, 31–84.
28. Chabner BA, Johnson RE, De Vita VT, *et al.* Sequential staging in non-Hodgkin's lymphoma. *Cancer Treat Rep* 1977, **61**, 993–997.
29. Hass AC, Brunk SF, Gulesserian HP, Givler RL. The value of exploratory laparotomy in malignant lymphoma. *Radiology* 1971, **101**, 157–165.
30. Goffinet DR, Castellino RA, Kim H, Dorfman RF, Fuks Z, Rosenberg SA. Staging laparotomy in unselected previously untreated patients with non-Hodgkin's lymphomas. *Cancer* 1973, **32**, 672–686.
31. Chabner BA, Johnson RE, Chretien PB, *et al.* Percutaneous liver biopsy, peritoneoscopy and paparotomy: an assessment of the relative merits in the lymphomata. *Br J Cancer* 1975, **31**, 242–247.
32. Kan MK, Hopkins GB. Measurement of liver volume by emission computed tomography. *J Nucl Med* 1979, **20**, 514–520.
33. Kaplan EL, Meier P. Non-parametric estimation from incomplete observation. *J Am Stat Assoc* 1958, **53**, 457–481.
34. Peto R, Pike MC, Armitage P, *et al.* Design and analysis of randomised clinical trials requiring prolonged observations of each patient. II. Analysis and examples. *Br J Cancer* 1977, **35**, 1–39.
35. Ersboll J, Schultz HB, Pedersen-Bjergaard J, Nissen NI. Follicular low-grade non-Hodgkin's lymphoma: long term outcome with or without tumour progression. *Eur J Haematol* 1989, **42**, 155–163.

36. Bennett JM, Cain KC, Glick JH, Johnson GJ, Ezdinli E, O'Connell MJ. The significance of bone marrow involvement in non-Hodgkin's lymphoma: The Eastern Cooperative Oncology Group experience. *J Clin Oncol* 1986, **4**, 1462–1469.
37. Spiro S, Galton DAG, Wiltshaw E, Lohmann RC. Follicular lymphoma: a survey of 75 cases with special reference to the syndrome resembling chronic lymphocytic leukaemia. *Br J Cancer* 1975, **31**, 60–72.
38. Qazi R, Aisenberg AC, Long JC. The natural history of nodular lymphoma. *Cancer* 1976, **37**, 1923–1927.
39. Ciampi A, Bush RS, Gospodarowicz M, Till J. An approach to classifying prognostic factors related to survival experience for non-Hodgkin's lymphoma patients. *Cancer* 1981, **47**, 621–627.
40. Gallagher CJ, Lister TA. Follicular non-Hodgkin's lymphoma. *Baillieres Clin Haematol* 1987, **1**, 141–155.
41. Gospodarowicz MK, Bush RS, Brown TC, Chua T. Prognostic factors in nodular lymphomas: a multivariate analysis based on the Princess Margaret Hospital experience. *Int J Radiat Oncol Biol Phys* 1984, **10**, 489–497.
42. Jones SE, Fuks Z, Bull M, *et al.* Non-Hodgkin's lymphomas: IV clinicopathologic correlation in 405 cases. *Cancer* 1973, **31**, 806–823.
43. Nissen NI, Pajak T, Glidwell O, Blom H, Flaherty M, Hayes D. Overview of four clinical studies of chemotherapy for stage III and IV non-Hodgkin's lymphomas by the Cancer and Leukaemia Group B. *Cancer Treat Rep* 1977, **61**, 1097–1107.
44. Rudders RA, Kaddis M, De Lellis RA, Casey H. Nodular non-Hodgkin's lymphoma. Factors influencing prognosis and indications for aggressive treatment. *Cancer* 1979, **43**, 1643–1651.
45. Hoppe RT, Rosenberg SA, Berard CW, Bloomfield CD, Bonnadonna G, Glatstein E. A working formulation of non-Hodgkin's lymphoma for clinical usage. Clinicopathologic correlations of low-grade and follicular lymphomas. *Proc Am Soc Clin Oncol* 1981, **C732**.
46. Straus DJ, Gaynor JJ, Leiberman PH, Filippa DA, Kozinec B, Clarkson BD. Non-Hodgkin's lymphomas: characteristics of long-term survivors following conservative treatment. *Am J Med* 1987, **82**, 247–256.
47. Anderson T, Chabner BA, Young RC, *et al.* Malignant lymphoma: 1. The histology and staging of 473 patients at the National Cancer Institute. 2. Prognostic factors and response to treatment of 473 patients at the National Cancer Institute. *Cancer* 1982, **50**, 2699–2721.
48. Bloomfield CD, Goldman A, Dick F, Brunning RD, Kennedy BJ. Multivariate analysis of prognosis factors in the non-Hodgkin's malignant lymphomas. *Cancer* 1974, **33**, 870–879.
49. Patchefsky AS, Brodovsky HS, Menduke H, *et al.* Non-Hodgkin's lymphoma: a clinicopathologic study of 293 cases. *Cancer* 1974, **34**, 1173–1186.
50. Ezdinli E, Anderson JR, Melvin F. Moderate versus aggressive chemotherapy of nodular lymphocytic poorly differentiated lymphoma. *Clin Oncol* 1985, **3**, 769–775.
51. Paryani S, Hoppe RT, Cox R, Colby T, Rosenberg SA, Kaplan HS. Analysis of non-Hodgkin's lymphomas with nodular and favourable histologies, stages I and II. *Cancer* 1983, **52**, 2300–2307.
52. Shipp MA, Harrington DP. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med* 1993, **329**, 987–994.
53. Bartlett NL, Rizeq M, Dorfman RF, Halpern J, Horning SJ. Follicular large-cell lymphoma: intermediate or low grade? *J Clin Oncol* 1994, **12**, 1349–1357.
54. Lopez-Guillermo A, Montserrat E, Bosch F, Terol MJ, Campo E, Rozman C. Applicability of the international index for aggressive lymphomas to patients with low-grade lymphoma. *J Clin Oncol* 1994, **12**, 1343–1348.
55. Romaguera JE. Can we identify patients with low-grade lymphoma for frontline ABMT? *Leuk Lymphoma* 1993, **10**, 9–15.

**Acknowledgements**—The authors are grateful to the many past and present members of the British National Lymphoma Investigation who have participated in the trials organised by the BNLI but particularly to Dr A.M. Jelliffe, past Chairman. The authors are also grateful to Bryan Vaughan Hudson and other past and present staff members of the BNLI office. Dr Peter O'Brien is thanked for his help in reviewing the manuscript, as are the reviewers of the European Journal of Cancer. The BNLI is supported financially by the Lymphoma Research Trust, the Cancer Research Campaign, the Lisa Lear Fund and The Isle of Man Anti-Cancer Association. Its offices are supported by the Middlesex Hospital National Health Trust, London.