MIT/LCS/TR-271

A COMPUTER SYSTEM

FOR

DECISION ANALYSIS IN HODGKINS DISEASE

Cynthia J. Rutherford

Byron Davies

Arnold I. Barnett

Jane F. Desforges

This blank page was inserted to preserve pagination.

A Computer System for Decision Analysis in Hodgkins Disease

Final Report

Cynthia J. Rutherford
Byron Davies
Arnold I. Barnett
Jane F. Desforges

Tufts-New England Medical Center Hospital, Boston

and Massachusetts Institute of Technology, Cambridge, Massachusetts

This research was supported in part by the Health Resources Administration, U.S. Public Health Service under grant number MB 0017 from the Bureau of Health Manpower, by U.S. Public Health Service Grant number CA 19122 from the National Cancer Institute, by National Institutes of Health Grant No. 1 P41 RR 01096-03 from the Division of Research Resources.

Massachusetts Institute of Technology
Laboratory for Computer Science
Cambridge, Massachusetts 02139

This empty page was substituted for a blank page in the original document.

CONTENTS

0. Introduction

This report draws together the diverse strands involved in developing a unique computer-based system to stage and manage Hodgkins Disease (HD). Those of us who worked on the final version of this project included two hematologists, a computer scientist, and a statistician. We have all contributed to this report, and our respective sections reflect not only our own special interest in and approach to the problem, but also the type of readers with whom we hope to communicate. We expect that many will not read this report "in toto", but that others working on similar problems may find particular sections helpful and interesting. In addition, because the program is not being maintained, we have included much data to insure that information from our extensive patient database is not lost.

The report begins with a general discussion of the principles of management of HD, covered in sufficient detail to explain the problem to those not familiar with this field. This section is aimed towards the non-physician or the physician who is not involved with HD management; it explains the rationale for a structured approach to the problem.

Section 2 describes the patient database, and how this is used to predict the likelihood of the various stages of HD. A general discussion of Bayes Rule then leads to a description of the particular way this rule is applied to revise the probabilities of pathologic stage prior to management decisions.

Section 3 deals with the technique of decision analysis. No previous knowledge of decision analysis is assumed, and detailed descriptions of all the steps involved are given, with the examples based on the HD system. The latter part of this section requires an understanding of Bayes Rule as described in section 2.

Sections 4 and 5 are concerned with the statistical validity of the program. After describing the assumptions made in developing the program, this section deals with the way both these assumptions and the general validity of the program were tested. Details of the methods used are supplied, and wider application of the

method is discussed.

Section 6 is included to show exactly how the Bayesian and decision analysis techniques are applied. It gives a detailed description of a typical patient session, including all the special features which can be used to tailor the decision plans to a particular patient and the experience of a particular hospital.

In contrast, sections 7 and 8 describe more general applications of our system. We use the program to approach some important management dilemmas in HD and we draw conclusions pertinent to large number of patients. Although these sections contain information mainly of interest to the practising hematologist, the discussion of decision-making thresholds has a more general application.

Rigorous application of Bayes Rule and decision analysis to the inexact science of clinical management is bound to result in some interesting problems. These are addressed in section 9, in which we discuss the limitations we found in this approach.

Section 10 is written as a guide to computer programmers interested in the implementation of such a program. The program is described in sufficient detail to enable a similar program to be set up. Certain original and sophisticated techniques developed for this program are described.

Finally, there are four appendices. The first is a comprehensive list of prior probabilities for the basic clinical findings, the conditional probabilties which are used to modify these prior probabilities, and posterior probabilities for every combination of clinical findings, all of which are derived from our 1200 patient database. We feel that these will be very valuable to the physician managing HD. The second appendix gives the general format of the statistical test used to evaluate probabilities of stage as described in section 5. The short Appendix 3 documents the information used in the program to derive the conditional probabilities of stage from the lymphangiogram, and shows how these probabilities are derived from the data. Finally Appendix 4 gives an annotated trace of the decision tree analysis program in operation.

1. Hodgkins Disease - Modern Approaches to Staging and Treatment

Although Hodgkins Disease (HD) is a malignant disease of the lymphoid system, a malignant lymphoma, its biologic course is different from the other lymphomas, and from neoplasms in general. In the majority of patients the disease arises in a single lymph node area, typically cervical, and spreads by lymphatic channels to involve contiguous lymph node areas in an orderly way. Local spread of disease to structures contiguous to affected lymph nodes is noted, particularly with involvement of mediastinal and hilar lymph nodes; occasionally the disease may begin in a localized extranodal site. Splenic involvement usually occurs as the disease progresses, and is thought to result from hematogenous spread, since the spleen has no afferent lymphatics. Hematogenous spread to other extranodal structures such as bone marrow, liver or lung does not usually occur until late in the disease.

The extent of the disease determines the type of treatment which is indicated. To achieve uniformity in describing this extent the Ann Arbor staging system was adopted in 1970, and is now well-established (Carbone et al, 1971) (Table 1.1). This stage of the disease, together with the presence (B) or absence (A) of the specific symptoms is the most important guide to prognosis, with the outlook becoming less favorable with advancing stage, and if symptoms are present.

There is now good evidence that proper treatment of HD in certain presentations results in long term remissions and probably cure. For localized disease radiotherapy may be curative in a high proportion of patients if all the tumor is encompassed with a tumoricidal dose of radiation - usually in the vicinity of 3500-4000 rad; for disease which has spread beyond the lymph nodes and spleen, combination chemotherapy has given a high proportion of patients such prolonged remissions that they may be considered cured. For other stages, where neither radiotherapy nor chemotherapy when given alone has proven extremely successful, combinations of these two modalities are being tried to effect a cure.

The diagnosis of HD often marks the beginning of a series of increasingly invasive diagnostic tests to

determine this stage in order to select treatment. This intensive staging may include an intravenous pyelogram, percutaneous bone marrow and liver biopsies, a gallium scan, a lymphangiogram, and computerized tomography; since all these tests have an appreciable false negative rate, and, with the exception of the biopsies, all have a false positive rate, (Table 1.2) the series often culminates in an exploratory laparotomy, with extensive lymph node sampling and splenectomy. If the extent of disease is established by laparotomy, or by histological confirmation of disease in extranodal structures by a biopsy, the patient is said to be "pathologically-staged". Staging without exploratory laparotomy, where biopsies of extranodal sites are negative for disease, is termed "clinical staging".

Our basic purpose in applying the techniques of decision analysis to HD has been to tailor staging investigations to the individual patient, and thus to use the minimum number of investigations to select treatment. This has involved studying a large series of patients in order to draw general conclusions about the management of HD which could then be accessible to a wide audience of physicians, with or without access to computer facilities for individual patient study.

For any patient the value in performing an investigation is related to three factors:

1. the likelihood that the patient has the abnormality for which (s)he is being tested.
2. the sensitivity or true-positive rate of the test.
3. the specificity or true-negative rate of the test.

All too often in HD the whole battery of available tests are used without analyzing their value in a particular patient, and without considering that they may provide overlapping information. There are a number of compelling reasons for performing the minimum number of tests required:

Table 1.1. The Ann Arbor Staging System for Hodgkin's Disease

STAGE	CHARACTERISTICS
Stage I	Involvement of a single lymph node region, or of a single extralymphatic organ or site (designated by E).
Stage II	Involvement limited to one side of the diaphragm either of two or more lymph node regions, or localized involvement of an extralymphatic (E) site and one or more lymph node regions.
Stage III	Involvement of lymph node regions on both sides of the diaphragm, which may include localized involvement of an extralymphatic organ or spleen.
Stage IV	Diffuse or disseminated involvement of one or more extralymphatic organ, or any liver involvement, with or without associated lymph node involvement.

Each case is further classified as:

A	if asymptomatic
B	if any of the following are present: - unexplained weight loss of more than 10% of body weight in the preceding 6 months. - unexplained fever, with temperatures above 38 C. - night sweats.

modified from Carbone et al 1971.

Table 1.2. False Positive and False Negative Rates of Tests

	FP(%)	FN(%)	REF
Gallium scan (abd.nodes)	5	49	McCaffrey(1976)
Lymphangiogram	13	16	Appendix III
Liver scan (abnormal uptake pattern)	23	38	Mildner(1973) Lipton(1972)
Spleen scan (i) filling defects	0	81	Silverman(1972) Mildner(1973)
(ii) size >15cm	7	60	
Computerized tomography (abd. nodes)	17	13	Lee(1977), Best(1977) Alcorn(1977) Breiman(1978)

1. with the exception of the gallium scan and computed tomography, all the tests have a mortality risk. This may be extremely remote in the case of a bone marrow biopsy, but there is a small mortality risk for such tests as the lymphangiogram and laparoscopy, and an appreciable risk for the staging laparotomy. Data collated from the medical literature for 2345 patients who underwent staging laparotomy is summarized in Table 1.3, and shows a mortality rate of 0.5% overall for the laparotomy and its immediate post-operative period.

2. all tests, again with the above exceptions, have a definite risk of morbid complications; here too the highest incidence of serious complications is seen after staging laparotomy, where 6.7% of patients experienced major complications as serious as pulmonary embolism, subphrenic abscess, intestinal obstruction etc., and a further 15% had less severe complications. This data is also summarized in Table 1.3.

3. All tests are expensive. A full work-up which uses all the tests could cost in the vicinity of $10,000. Individual costs for the different procedures are given in Table 3.3.

Information on the specificity and sensitivity of the tests used to stage HD is readily available in the medical literature. Information on the likelihood of positive findings for the various tests under different conditions, however, is not.

In the following section we describe how we derived this information from data obtained from laparotomy-staged patients. We also describe how, from the same data, we use Bayes' Rule to predict the probabilities of the different stages for a particular patient with Hodgkins disease, so that decisions about staging investigations and treatment may be made using the technique of decision analysis.

Table 1.3. Mortality and Morbidity from Staging Laparotomy for Hodgkins Disease

Reference	No. of Patients	No. of Deaths	No. of Major Complications
Aisenberg et al (1974)	100	0	2 (2%)
Andrassy et al (1977)	76	0	1 (1%)
Beretta et al (1976)	110	0	1 (1%)
Brogadir et al (1978)	90	0	16 (18%)
Bruntsch et al (1977)	275	0	28 (10%)
Cannon et al (1975)	400	1	15 (3.7%)
Coleman et al (1976)	31	1	7 (23%)
Ferguson et al (1973)	31	1	1 (1%)
Gamble et al (1975)	139	1	5 (4%)
Garcia et al (1971)	20	0	0 (0%)
Gazet (1973)	65	1	4 (6%)
Hermreck et al (1975)	50	0	13 (26%)
Jelliffe et al (1970)	22	0	1 (5%)
Lowenbraun et al (1970)	12	0	0 (0%)
Marston (1972)	60	2	4 (7%)
Meeker et al (1972)	30	2	4 (13%)
Mitchell et al (1972)	45	1	3 (7%)
Paglia et al (1973)	51	0	6 (12%)
Piro et al (1973)	114	1	6 (5%)
Poulsen et al (1977)	91	0	5 (5%)
Prosnitz et al (1972)	40	0	4 (10%)
Roberts et al (1976)	82	0	3 (4%)
Rozman et al (1973)	56	1	2 (4%)
Smith et al (1972)	70	0	2 (3%)
Sutcliffe et al (1976)	98	0	19 (1%)
Urlaub et al (1979)	107	0	5 (4%)
Zarembok et al (1972)	30	0	0 (0%)
Total	2345	11 (0.5%)	157 (6.7%)

2. Deriving Probabilities of Stage

2.1 The Database

A great deal of information about the patterns of spread of Hodgkins disease (HD) has been amassed in the decade since exploratory laparotomy has been regularly used to stage the disease; from this information has come our knowledge of the predictable pattern of spread of HD according to normal lymphatic distribution. The different histological subtypes have been found to have different patterns of spread, with the degree of dissemination increasing from lymphocyte predominant (LP) to nodular sclerosis (NS) to mixed cellularity (MC) to lymphocyte depleted (LD). We know also that the presence of symptoms is related to the spread of disease, with symptomatic (B) patients having more advanced disease than asymptomatic (A). Other factors known to affect the extent of the disease are the sex and the age of the patient. The younger age group, corresponding to the first incidence peak in the characteristic bimodal age:incidence curve of HD, shows an almost equal male:female ratio, a predominance of the NS histologic pattern, a higher frequency of mediastinal involvement and a more benign clinical course. In contrast, the older group, corresponding to the second peak of the age:incidence curve, shows a higher male:female ratio, and a number of inter-related clinical features including a greater proportion of the MC histological type, a much higher incidence of infradiaphragmatic involvement at presentation and a more aggressive clinical course (Gutensohn et al 1977). The pattern of involved supradiaphragmatic lymph nodes is related to the frequency of associated infradiaphragmatic disease; left cervical node involvement is associated with a higher incidence of disease below the diaphragm whereas with mediastinal involvement a lower incidence is seen.

We have used data from approximately 1200 patients who have been pathologically staged, to help predict the likelihood of the various stages of Hodgkins disease according to the basic clinical data known about the patient.

All information about the following findings which are available for a given patient have been collected and stored:

```
Histologic subtype

Presence (B) or absence (A) of symptoms

Sex

Age

Specific supradiaphragmatic lymph node groups involved

Alkaline phosphatase levels

Clinical splenomegaly

Absolute and percentage lymphocyte count

Liver/spleen scan

Percutaneous bone marrow biopsy

Percutaneous liver biopsy

Gallium scan

Computer-assisted tomography (CAT scan)

Lymphangiogram

Peritoneoscopy

Laparotomy
```

A typical string of patient data may be seen below:

```
((SOURCE TUFTS) (ID XX) (AGE 20) (SEX FEMALE) (HISTOLOGY NS) (A-OR-B A)
 (LOCATION-PRESENTING-NODES (LEFT-NECK RIGHT-NECK))
 (MEDIASTINUM NEGATIVE) (LIVER-SCAN NEGATIVE)
 (SPLEEN-SCAN NEGATIVE) (SPLENOMEGALY NORMAL)
 (ALK-PHOS (BODANSKY 3.0)) (GAL POSITIVE-ABD-NODES) (LAG POSITIVE)
 (BMBX NEGATIVE) (LBX UNKNOWN)
 (LAP ((LAP-SPLEEN NEGATIVE)
       (LAP-MARROW NEGATIVE)
       (LAP-LIVER NEGATIVE)
       (LAP-NODES POSITIVE))))
```

Patient data from the following institutions was obtained:

Tufts-New England Medical Center	91 patients
Joint Center for Radiation Therapy	89 patients
Massachusetts General Hospital	113 patients
Stanford Division of Radiation Therapy	504 patients
Harvard School of Public Health	310 patients

The remainder came from individual patient data in the medical literature by Bearman et al (1978), Hanks et al (1971), Jelliffe et al (1970), Lowenbraun et al (1970), Mitchell et al (1972), Prosnitz et al (1972), Zarembok et al (1972).

The patient with the uncommon infradiaphragmatic presentation of Hodgkins disease was poorly represented in the initial database. Furthermore, in this type of patient certain test results have a different meaning in terms of staging, e.g. such findings as involved abdominal lymph nodes or involved spleen define the patient as having stage II disease only. For these two reasons, we have excluded the patients with an infradiaphragmatic presentation from the database, and do not attempt to analyze this type of case.

Initially the proportion of patients in each pathological stage was calculated from the raw data. Stages I and II were analyzed together, as therapeutic decisions for these stages are almost always identical. Each finding was

then assessed to see whether it was helpful in predicting the pathological stage. Certain findings - alkaline phosphatase and clinical assessment of splenomegaly - were not found to be helpful in this respect, so they were not used in the final version of the program.

For those findings which were found to be helpful in distinguishing among the different stages, the number of patients in each pathological stage was ascertained and was expressed as a proportion of 1.0 e.g. for the finding Histological Subtype:

Histological subtype	Pathological stage	No of patients	Proportion of patients
LD	I+II	7	0.32
	III	9	0.41
	IV	6	0.27
LP	I+II	49	0.76
	III	14	0.22
	IV	2	0.03
MC	I+II	105	0.42
	III	112	0.44
	IV	35	0.14
NS	I+II	451	0.62
	III	235	0.32
	IV	44	0.06

One can use this data to calculate the probability of being in a particular pathological stage, given the histological subtype, e.g. the probability that a patient demonstrating the NS subtype will be in stage III is 0.32, or simply stated P(III|NS) = 0.32

Similarly the conditional probabilities of STAGE given SYMPTOMS can be obtained from the database:

	Pathological stage	No of patients	Probability
A	I+II	485	0.86
	III	225	0.3
	IV	31	0.04
B	I+II	151	0.41
	III	157	0.42
	IV	63	0.17

From this one may derive the probability of STAGE given SYMPTOMS e.g P(IV|A) = 0.04. Data can be similarly expressed for all the findings entered.

If the findings were all independent of one another, then one could find the probability of having a combination of findings in stage (I+II), stage III, or stage IV for any patient simply by multiplying together the conditional probabilities of each finding given the stage. For example, for an asymptomatic (A) male patient who has the MC histological subtype:

p(A MALE MC|III) would be p(A|III) X p(MALE|III) X p(MC|III)

Clinical information, however, suggested that some findings were likely to be interdependent. Chi-squared testing (see Chapter 4) did confirm this, and dependence was shown among the the patient's sex, the presence or absence of symptoms and the histological subtype, though it was not seen among the other findings used by the program. This dependence precluded the simple type of calculation shown above. Instead, prior probabilities of stage were calculated from the data for the triad of interdependent findings -- the patient's sex, the presence or absence of symptoms and the histological subtype -- making sixteen different combinations of prior findings. For any patient, Bayes' rule was then used to modify these prior probabilities by the conditional probabilities of FINDING|STAGE for the subsequent findings used in the program -- the patient's age, involvement/non-involvement of cervical and mediastinal lymph nodes, and the results of the various tests, including liver-spleen scan, bone marrow and liver biopsies, gallium scan, and lymphangiogram.

2.2 The Use of Bayes Rule To Revise Probabilities of Stage in Hodgkins Disease

Bayes Theorem is a mathematical method for modifying probabilities when new information is available. We have used this method to revise the probabilities of stage, as we take into account the basic clinical information known about the patient and the results of testing. See section 2.2.2.

2.2.1 General Application of Bayes Rule

Consider a test with two possible outcomes (T+ or T-) which is used to modify the probability of a patient being in a disease state, D. If P(D) is the probability of the disease state prior to the test, the revised probability P(D|T+) i.e. the probability of D given a positive outcome, and of P(D|T-), i.e. the probability of D given a negative outcome, can be expressed as follows:

$$P(D|T+) = \frac{P(D)\ P(T+|D)}{P(D)\ P(T+|D) + P(\sim D)\ P(T+|\sim D)}$$

$$P(D|T-) = \frac{P(D)\ P(T-|D)}{P(D)\ P(T-|D) + P(\sim D)\ P(T-|\sim D)}$$

Where P(T+|D) = Probability of a positive test, given the disease

P(~D) = probability of no disease

P(T+|~D) = probability of a positive test, given no disease, i.e. the false positive rate.

P(T-|~D) = probability of a negative test given no disease

P(T-|D) = probability of a negative test given disease, i.e the false negative rate.

Knowing the prior probability of a disease state, P(D), and the false positive and false negative rates of a test used to detect it, one can use Bayes' rule to revise P(D) according to the outcome of the test.

2.2.2 Use of Bayes Rule in the Hodgkins Disease Program

In our program we have used Bayes Rule to revise the probabilities of each of the stages of HD, taking into account both basic clinical information known about a patient and the results of any tests. Prior probabilities for each stage are those derived directly from the database for the three basic findings which have been found to be interdependent, namely histologic subtype, sex and the presence or absence of symptoms. Prior probabilities for each of the sixteen possible combinations of the findings are given in Appendix I.

To demonstrate how we apply Bayes Rule we will show the calculations involved in the case demonstrated in Chapter 6. This patient is a male with B symptoms - weight loss - and the nodular sclerosis histologic subtype. The prior probabilities of stage for this patient are:

Stages I+II	0.34
Stage III	0.50
Stage IV	0.16

We wish to revise these probabilities of stage knowing the patient is in the age group 12-39 years. For this we need to know the conditional probabilities of AGE|STAGE which have been calculated from the database, and which have been shown to be independent of other findings. These conditional probabilities are:

Younger than 12 years	I+II 0.03	III 0.05	IV 0.08
12 to 39 years	I+II 0.84	III 0.74	IV 0.46
Older than 39 years	I+II 0.14	III 0.23	IV 0.47

To update the prior probability of stage (I+II) for example

$$p(I+II|AGE12\text{-}39) = \frac{p(I+II).p(AGE12\text{-}39|(I+II)}{p(I+II).p(AGE12\text{-}39|(I+II) + p(III).p(AGE12\text{-}39|(III) + p(IV).p(AGE12\text{-}39|(IV)}$$

The term p(I+II).p(age|(I+II)) corresponds to the "true positive" if one considers age in the sense of a

"test", with being in the 12-39 age group a "positive" result and the other age groups a "negative" result. This makes the other terms in the eqation - p(III).p(AGE12-39|III) and p(IV).p(AGE12-39|IV) - the equivalent of false negative results. Substituting the probability results in the equation:

$$p(I+II)|AGE12\text{-}39 = \frac{0.34 \times 0.84}{(0.34 \times 0.84) + (0.5 \times 0.74) + ((0.16 \times 0.45)}$$

$$= 0.3925$$

Similar calculations are carried out to revise the prior probabilities of stage III and stage IV according to the conditional probabilities of age|stage.

These posterior probabilities now become the new "prior" probabilities which can then be revised by the next piece of information available, such as the specific sites of supra-diaphragmatic lymph node involvement.

Bayes Rule is applied to the tests used in staging HD according to the general formula given in section 2.2.1.

For example, suppose we have calculated that the probability of abdominal node involvement in a patient with supradiaphragmatic HD is 0.30. A gallium scan is found to be positive for abdominal nodes, and we wish to know how this positive result has modified our initial probability.

P(+NODES) = 0.30 therefore p(-NODES) = 0.70.

From the literature we find that when the gallium scan is used to detect involved abdominal nodes in HD:

False positive rate = p(+test|-NODES) = 0.10

False negative rate = p(-test|+NODES) = 0.50

where the symbols + and - are used for positive and negative, i.e. +NODES are lymph nodes involved with HD.

From these we deduce:

True positive rate = p(+test|+NODES) = 0.68

True negative rate = p(-test|-NODES) = 0.90

We can now calculate the revised probability of abdominal node involvement:

$$p(+NODES|+test) = \frac{p(+NODES)\ p(+test|+NODES)}{p(+NODES).p(+test|+NODES) + p(-NODES).p(+test|-NODES)}$$

$$= \frac{.3 \times .6}{(.3 \times .6) + (.7 \times .10)}$$

$$= 0.68 \text{ or } 68\%$$

The use of Bayes Rule here has revised the probability of abdominal node involvement upwards, from 0.30 to 0.68 in the light of the positive result for gallium scan.

In our application of Bayesian methods to the staging of Hodgkins disease, we use the tests to modify information about a stage of the disease, rather than the specific sites of involvement, as demonstrated in the example above. If we wish to find the revised probability of stage III disease in a patient, knowing that the gallium scan is positive, we can write the basic calculation as:

$$p(III|+GAL) = \frac{p(+GAL|III).p(III)}{p(+GAL|I+II).p(I+II) + p(+GAL|III).p(III) + p(+GAL|IV).p(IV)}$$

However, the literature does not contain data in the form of conditional probabilities of test results in certain stages of the disease. This calculation, therefore must be carried out in two stages:

$$p(+GAL|III) = \underbrace{p(+GAL|+NODES)\ p(+NODES|III)}_{\text{TRUE POSITIVE}} + \underbrace{p(+GAL|-NODES)\ p(-NODES|III)}_{\text{FALSE POSITIVE}}$$

The terms for p(+GAL|I+II (or IV)) are expressed in the same way. The probability of positive and

negative nodes in stage III is obtained from the database, i.e.

```
In stage IIIA:

p(+ABDOMINAL-NODE-INVOLVEMENT) =  0.62
p(-ABDOMINAL-NODE-INVOLVEMENT) =  0.38
```

These lead to the conditional probabilities:

```
p(+GAL|III) = 0.6 X 0.62    +    0.10 X 0.38
              TRUE POSITIVES     FALSE POSITIVES

            = 0.31 + 0.038

            = 0.35
```

That is the probability of finding a positive Gallium scan in a stage IIIA patient is 0.35.

Similar calculations are performed for all stage/symptom combinations for this test and all the other tests used, so that conditional probabilities of the test given stage and symptoms are available for the Bayesian calculations. For the gallium scan these are:

```
NEGATIVE GALLIUM SCAN

  I+II   A 0.90   B 0.90
  III    A 0.65   B 0.56
  IV     A 0.61   B 0.54

POSITIVE GALLIUM SCAN

  I+II   A 0.10   B 0.10
  III    A 0.35   B 0.44
  IV     A 0.39   B 0.46
```

2.3 Defining the Age Groups

Comparing the incidence of Hodgkins disease with the age at presentation shows a bimodal age/incidence curve with one peak in early adult life, and a second sustained increase in incidence from the fifth decade. The two age groups defined by this curve show important differences, with the age at presentation showing correlation with the extent of the disease. The younger group shows an approximately equal male:female ratio, a predominance of the nodular sclerosis (NS) histologic pattern and a more benign clinical course, with mediastinal involvement quite common; in contrast, the older group shows a higher male:female ratio, and a number of inter-related clinical features including a greater proportion of the MC histological type, a much higher incidence of infradiaphragmatic involvement at presentation and a more aggressive clinical course.

We have, therefore, used age as one variable for estimating the probabilities of stage in Hodgkin's disease patients. Conditional probabilities of AGE given STAGE are used in applying Bayes rule to calculate posterior probabilities of STAGE given AGE. Use of this formula requires the assumption that age is independent of the findings used to calculate the prior probabilities. This assumption was checked for the arbitrarily-selected age groups - 0-15 years, 16-30 years, 31-45 years, and older than 45 years - initially used in the program by means of Chi square analysis, and found tenable (Eisen 1977).

Subsequent analysis of the larger database showed that there were essentially three age groups distinguished by distribution of stage: 0 to 11 years, 12 to 39 years, and 40 years and over. Homogeneity was observed within each group, with some minor statistical fluctuation, and distinct differences were observed between groups (as determined by Chi square testing).

3. Decision Analysis

Decision Analysis is a technique for evaluating complex decisions by breaking them down into their simplest component parts, quantitatively evaluating these, then reassembling them in a logical structure.

Decision analysis involves four sequential steps:

a) constructing the decision tree which displays, in chronological order, all the options for a particular decision, and every possible outcome of each option, whether determined by chance, or by further decision-making.

b) evaluating the likelihood of those outcomes in the decision tree which are determined by chance.

c) scoring each possible option, according to a system of utilities.

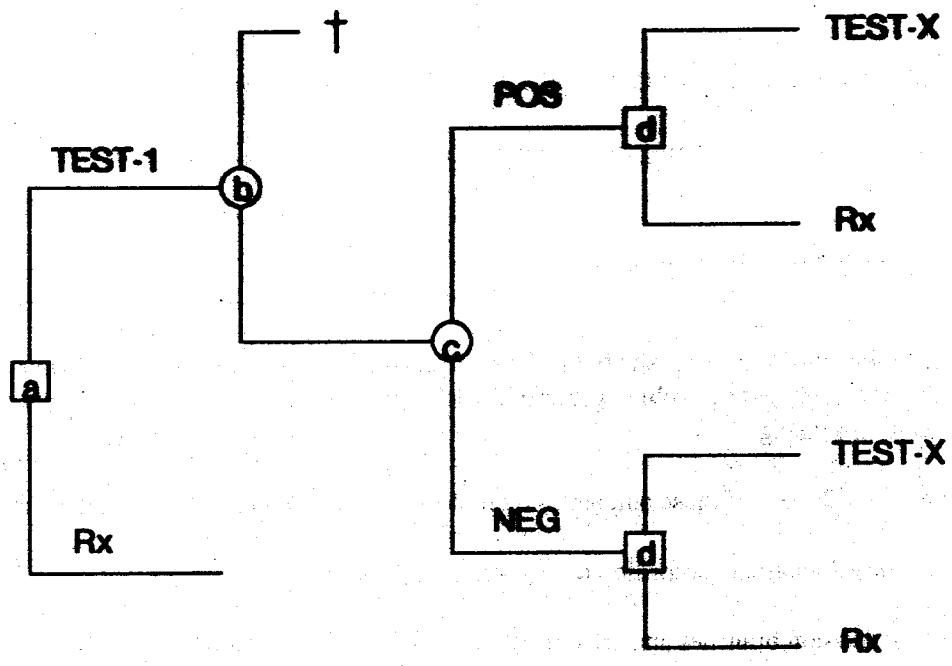
d) evaluating all decision branches and selecting those with the highest utility.

3.1 Constructing the Decision Tree

The decision tree is displayed on a branching framework, and includes all the options among which one must choose, with all possible outcomes of each option displayed. Subsequent outcomes include both those determined by chance, and those which involve later decisions.

At each branch point in the tree where a decision must be made, a small square is used to denote a *decision node.* At any point in the tree where the outcome of a decision is determined by chance, a small circle denotes a *chance node.*

Below is shown the simplest type of decision tree used in this work, the basic TEST versus TREAT decision, in which a physician has to decide whether to perform a test to define the patient's disease more accurately or whether to treat (Rx) the patient without further ado. By convention, the decision tree is drawn with the early decisions on the left, the later outcomes on the right.



At decision node a, the decision is between treatment (Rx) or performing TEST-1. At chance node b, the patient may die (+) from the test - this risk being the observed test mortality rate - or may survive the test. At chance node c, having survived the test, the patient may have either a positive or a negative result for the test. At each decision node d, one must now, in the light of information gained from the test, make another TEST versus TREAT decision, for some other test.

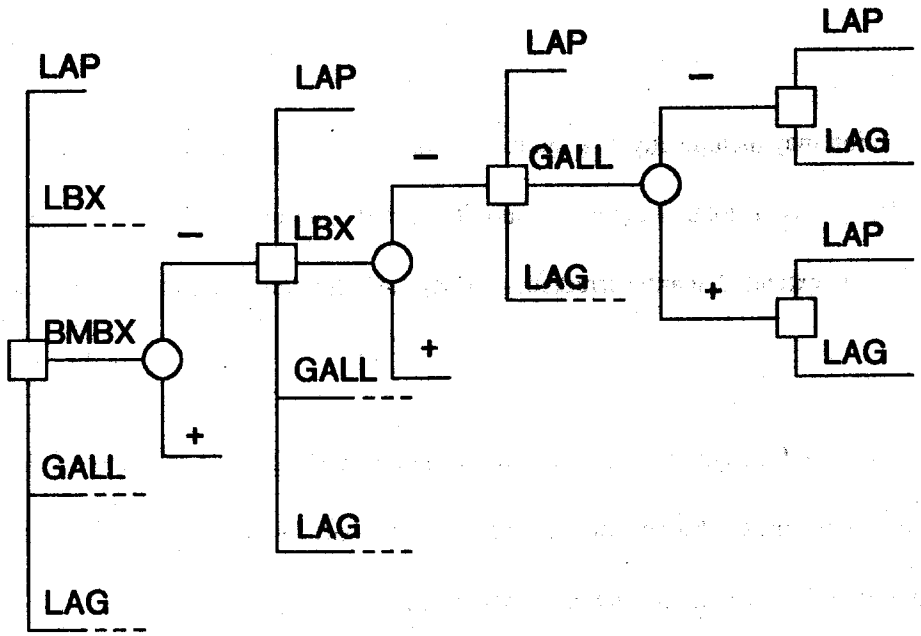
The actual decision tree involved in the selection of test procedures and treatments in Hodgkins disease is very complex, because:

(i) there are a number of different tests among which one must choose - namely bone marrow biopsy (BMBX), percutaneous liver biopsy (LBX), peritoneoscopy with guided liver biopsy (PTX), gallium scanning for abdominal node involvement (GALL), lymphangiogram (LAG), and staging

laparotomy (LAP) - and one may choose any number of these.

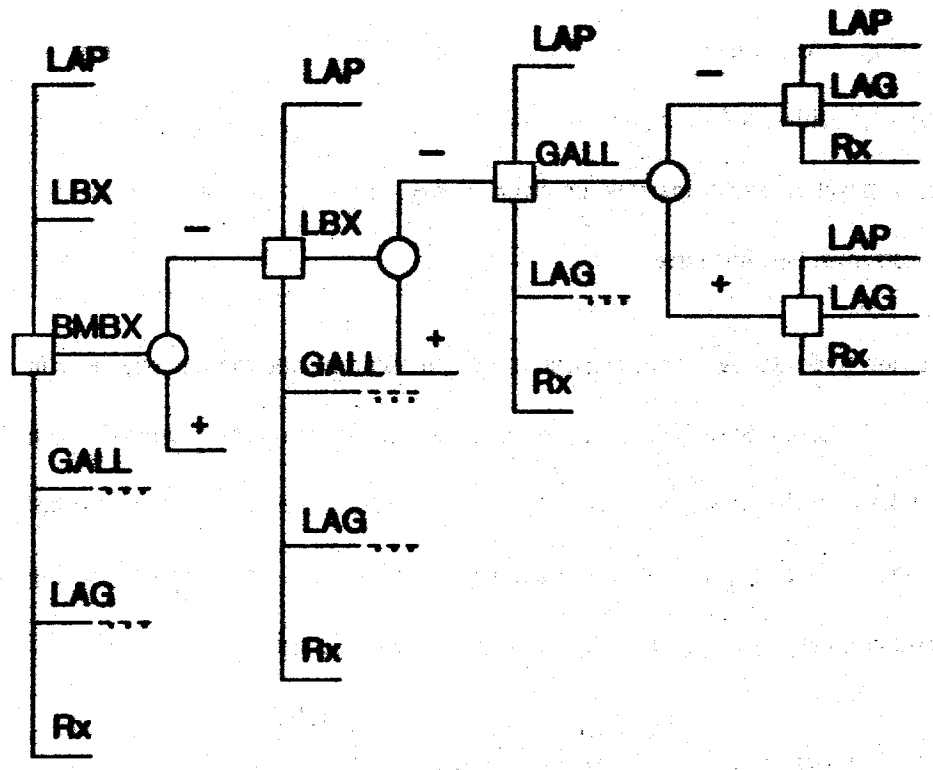
(ii) The order in which the tests are performed may be important. A test sequence stops with a positive bone marrow or liver biopsy which constitute definite evidence of disseminated (stage IV) disease, or with the staging LAP, which is considered for practical purposes as "perfect information", i.e. it is assigned neither a false positive nor a false negative rate and its results are always final. Following a negative bone marrow or liver biopsy, or after gallium scan or lymphangiogram further tests must be considered.

Below we show part of the decision tree relating to the selection of tests, to demonstrate the recursive way in which any tests not yet performed are considered. This decision tree displays only one of the branches which occur at each decision node, following the initial decision to select bone marrow biopsy. Where further decision "branching" follows, the test name is followed by dots, e.g. "GALL. . ." . A similar branching structure exists for each of the other initial tests, except for laparotomy, where results are final.



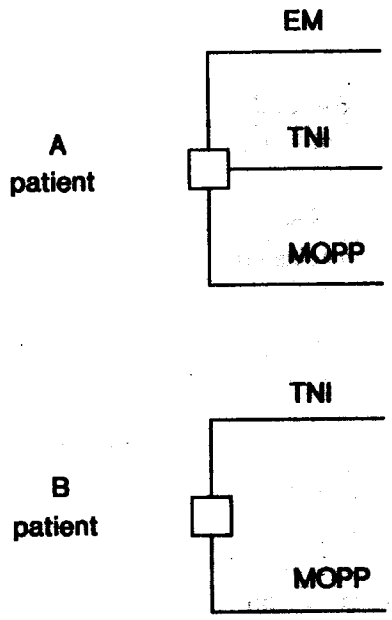
The actual tree evaluated is even more complex since at every decision node, the decision involves not only

further TESTING, but must also include an evaluation of the TREATMENT option. The actual framework of the tree is shown below:

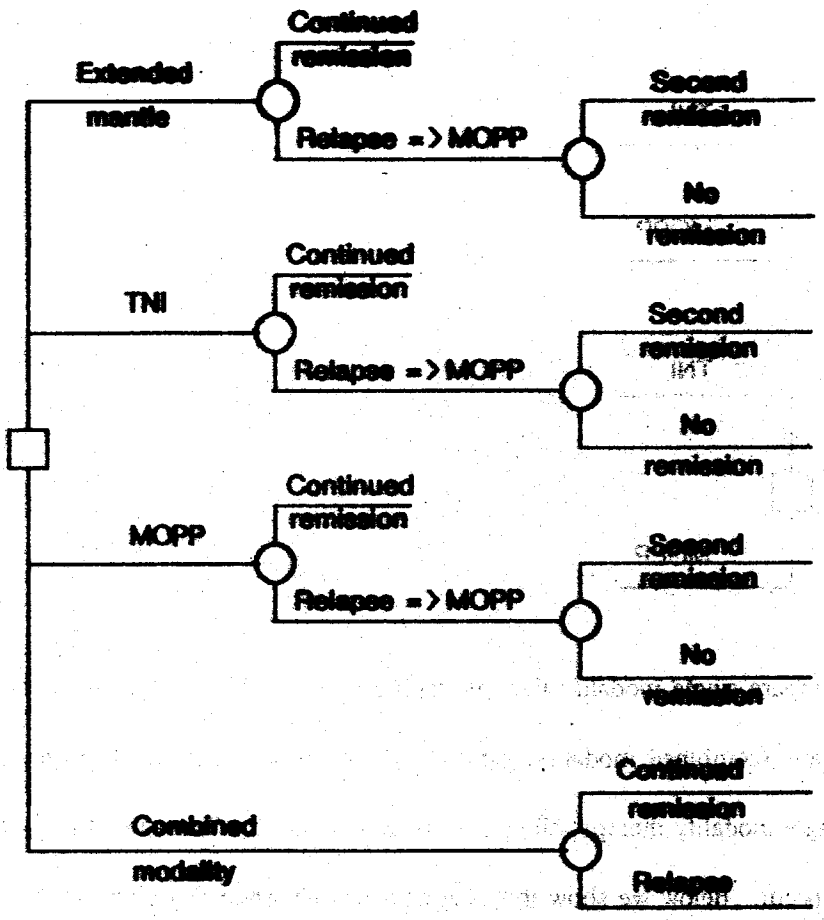


The TREATMENT option may include any of a number of different treatments; these include radiotherapy to all lymph-node areas, known as total nodal irradiation (TNI), radiotherapy which spares the pelvic lymph nodes, which is known as extended mantle irradiation (EM), and the combination chemotherapy regime MOPP.

The TREATMENT branches for asymptomatic (A) and symptomatic (B) disease are shown below. For A disease there are three treatments to choose amongst, for B patients only two in our basic decision tree. The tree may easily be modified, however, to include other treatments.



It is also possible to compare single modality therapy, with regimens which include both irradiation and combination chemotherapy (combined modality therapy). If this option is selected, however, for proper comparison with the single modality therapy, allowance must be made for "salvage" therapy (usually with MOPP) should relapse occur. Below we show the treatment branch when combined modality therapy is included in the treatment being evaluated for patients with symptomatic (B) disease.



3.2 Evaluating the Likelihood of the Outcomes

In this program we start with those probabilities of the pathologic stage derived from the repeated use of Bayes Rule, according to the basic findings known about the patient. For example, for a 20 year old symptomatic male patient, who has mixed cellularity (MC) Hodgkins disease involving the right cervical nodes and the mediastinum, and who has a normal liver/spleen scan, the prior probabilities are modified as shown in Table 3.1 before decision analysis is used to decide further management:

For this patient, these probabilities of stage derived from the basic clinical findings now become the "prior probabilities" for the decision analysis. As each test is evaluated, these probabilities are modified by:

Table 3.1. Results of Bayes Rule, Incorporating Basic Clinical Findings

Finding	Stage I+II	Stage III	Stage IV
M MC B	0.20	0.44	0.36
Age >11<40 years	0.26	0.50	0.24
L cervical nodes -ve	0.30	0.46	0.24
R cervical nodes +ve	0.27	0.42	0.31
Mediastinal nodes +ve	0.30	0.40	0.30
Spleen scan normal	0.43	0.34	0.23
Liver scan normal	0.48	0.37	0.15

---

1. the known mortality rate for the test (see decision node b in the first basic TEST versus TREAT decision tree in section 3.1).

2. the known false negative and, where applicable, false positive rates for the tests, applied according to Bayes Rule.

Suppose in the example, we are considering the branch of the tree which begins with bone marrow biopsy; note that this test has no false positive rate, and from the reported false negative rate, we have derived the following conditional probabilities:

```
Probability of NEGATIVE bone marrow biopsy (-BMBX) with
    I+IIB = 1.0, IIIB = 1.0, IVB = 0.56
Probability of POSITIVE bone marrow biopsy (+BMBX) with
    I+IIB = 0.0, IIIB = 0.0,  IVB = 0.44
```

The basic probabilities are modified by Bayes rule, using these conditional probabilities, e.g.:

$$p(IV|-BMBX)$$

$$= \frac{p(IV)\ p(-BMBX|IV)}{p(I+II)\ p(-BMBX|I+II) + p(III)\ p(-BMBX|III) + p(IV)\ p(-BMBX|IV)}$$

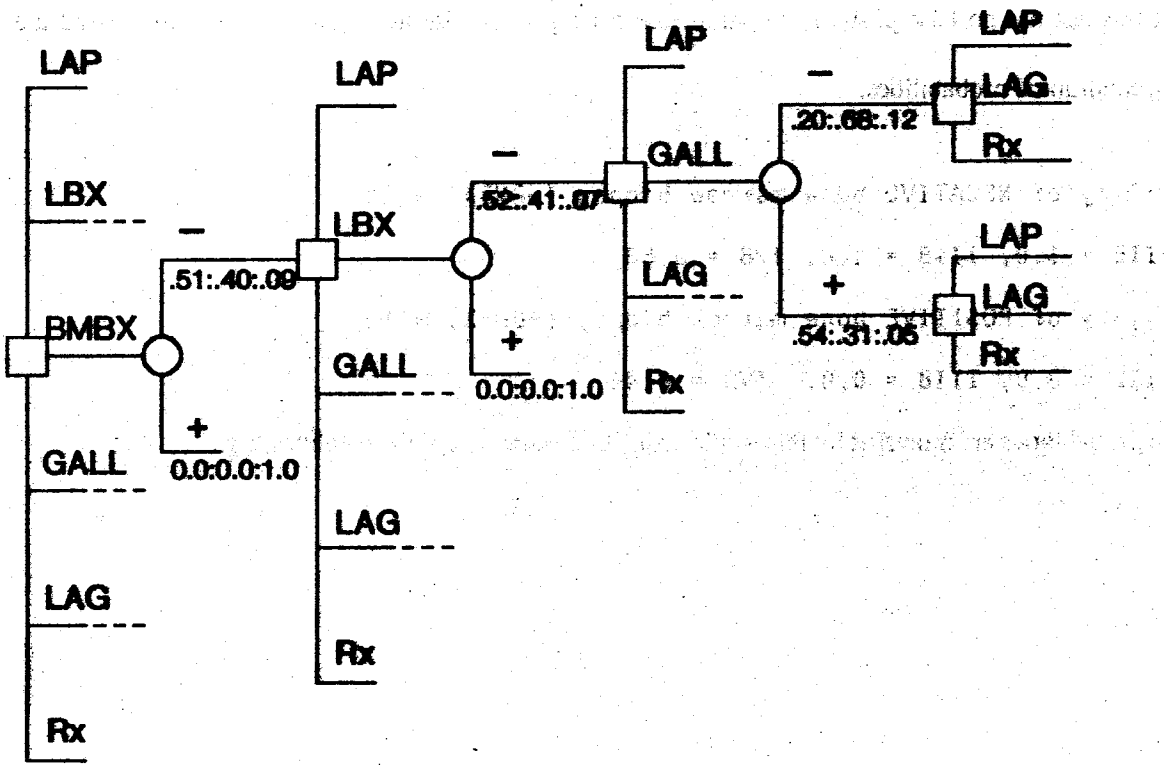
$$= \frac{0.15 \times 0.56}{0.48 \times 1.0 + 0.37 \times 1.0 + 0.15 \times 0.56}$$

$$= \frac{0.084}{0.48 + 0.37 + 0.084} = \frac{0.084}{0.934} = 0.09$$

Similar calculations are carried out for stages I+II and III giving revised probabilities if the test is negative of

I+IIB 0.51 IIIB 0.40 IVB 0.09

If positive, the test constitutes definite evidence of stage IV disease. These revised probabilities are entered into the next stage of the decision tree.



3.3 Assigning Utilities

The utility is the value placed on a particular outcome. Utilities may be expressed on an arbitrary scale, or a measurable quantity may be used. We have used several different types of utilities to assess our decisions:

i) Disease-Free Survival Five Years After Treatment (5 yr DFS). This figure is used because results of treatment are very frequently reported in the medical literature in this form. In addition, in HD, very few patients relapse after this period, and the 5 yr DFS can be used as an index of cure.

ii) Morbidity Units - in addition to its mortal risk, each diagnostic test and each treatment has an associated risk of morbidity. This includes both the risk of severe complications, such as pulmonary embolism after staging laparotomy, and the duration of pain and incapacity associated with the test or treatment under normal circumstances.

For each test and treatment we asked a number of physicians, directly involved in the investigation and treatment of patients with HD, for their estimate of the "pain and discomfort", and the "duration of incapacity", for the "average" patient undergoing a given test or treatment. An arbitrary standard was made which defined the "pain and discomfort" of a bone marrow biopsy or of a single day's incapacity each equivalent to 10 units. The literature was then searched for data on the complication rates of tests and treatments. Each complication was assessed according to the same criteria as the basic test; this figure, multiplied by the incidence of the complication is included in the morbidity values shown in Table 3.2.

iii) Dollar Cost of Tests and Treatments. These are shown in Table 3.3.

Each of these types of utilities is evaluated separately. At present, we have not found a satisfactory way to incorporate these disparate elements in the same scale of values, and no attempt has been made to "trade-off" one type of utility with another.

Table 3.2. Morbidity Units for the Staging Procedures and Treatments

Gallium scan	7 units
Bone marrow biopsy	10 units
Liver biopsy	25 units
Lymphangiography	40 units
Peritoneoscopy	80 units
Laparotomy	875 units
Extended mantle radiotherapy	1000 units
Total nodal irradiation	1500 units
MOPP chemotherapy	3000 units

Table 3.3. Dollar Cost of Tests Used in Staging Hodgkins Disease

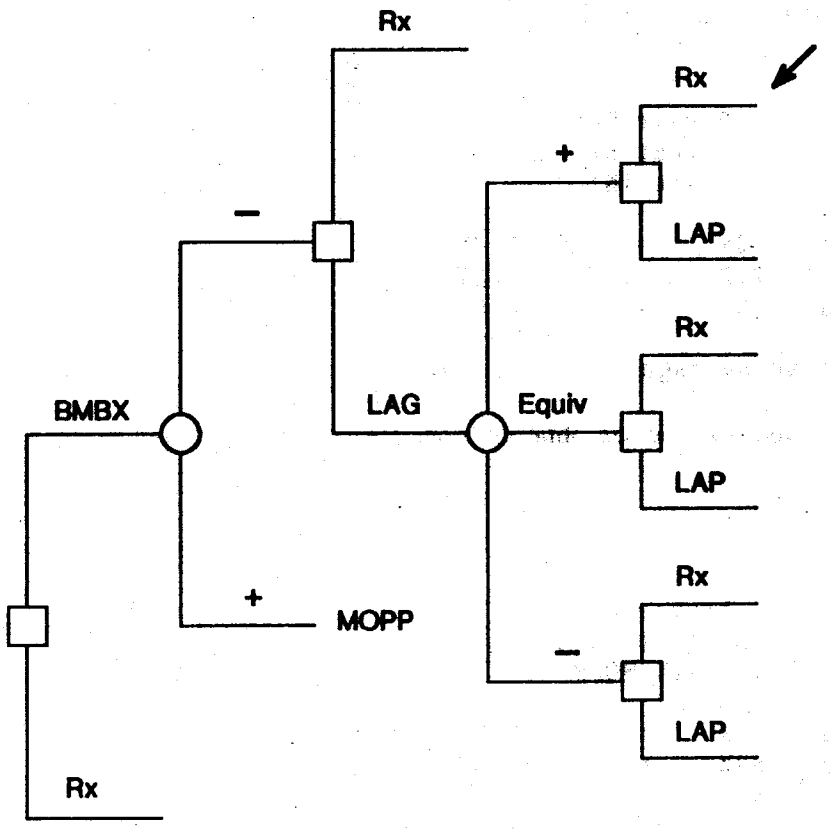
	Test Cost	Professional Fee	Total
Liver-Spleen Scan	213	60	273
CT Scan			
- with contrast	168	100	268
- without contrast	118	70	188
Gallium Scan	259	60	319
Liver Biopsy	72	155	900*
Bone Marrow Biopsy	94	60	154
Lymphangiogram	421	275	696
Laparotomy			6000-8000**

* includes 2 days in hospital, type and cross-match, etc.

** this includes 7-10 days in hospital, fees for surgeon, physician, and anesthesiologist, use of operating and recovery room, laboratory tests, X-rays, and pathology fees for operative specimens.

3.4 Evaluating Decision Branches

This step is known as "averaging out and folding back", or in more technical terms "the process of backwards induction in the theory of dynamic programming". Starting with the most peripheral decision nodes, all outcomes of a particular "sub-decision" are evaluated, and the option with the highest utility ascertained; this value is then assigned to the decision node. For example, in the decision tree shown below, one would initially evaluate each of three branches after lymphangiogram.



To evaluate these, one must know the probabilities of pathologic stage at that particular point in the tree, when the prior probabilities have been modified by a negative BMBX result and by the result of lymphangiogram. We will use the symptomatic male patient from the example in Section 3.2. We will evaluate the branch indicated by the arrow: the probabilities of pathologic stage for this patient at this

decision node, with negative bone marrow and positive lymphangiogram are:

Stage I+II:	0.22
Stage III:	0.64
Stage IV:	0.15

When the patient's stage is not known for certain, assessing the utility of the option to proceed directly with treatment, involves calculating the average disease-free survival for each treatment, and choosing the treatment with the highest value. The utility for a particular treatment is calculated as a weighted average, using the probabilities of stage and the five year disease free survival values for each stage for that treatment, e.g. assessing the value of total nodal irradiation for the patient:

Stage	Probability of that stage		5-year DFS for that stage with TNI		
I+IIB	0.22	times	70%	=	15.4%
IIIB	0.64	times	25%	=	16.0%
IVB	0.15	times	0%	=	0.0%
		Weighted average			= 31.4% DFS

This is then compared with the weighted average obtained from the corresponding values for MOPP chemotherapy:

Stage	Probability of that stage		5-year DFS for that stage with MOPP		
I+IIB	0.22	times	60%	=	13.2%
IIIB	0.64	times	50%	=	32.0%
IVB	0.15	times	35%	=	5.25%
		Weighted average		=	50.45% DFS

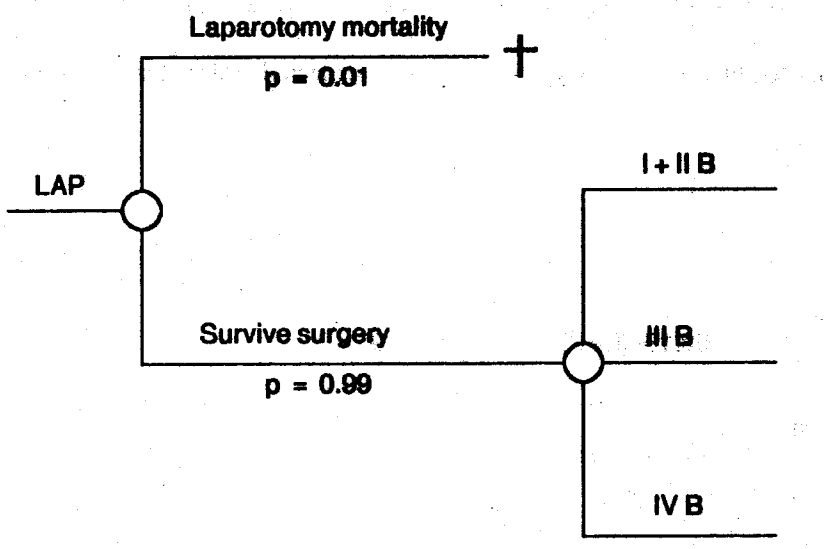
In this example the weighted average for MOPP chemotherapy is greater than that for total nodal irradiation, and would thus be preferred to TNI as the treatment option. The same techniques are used to evaluate combined modality therapy as one of the treatment options, but the calculations are more complicated because, as already described, one must allow for salvage of patients who initially receive single modality

therapy.

To assess the value of laparotomy, the probability of each stage, is multiplied by the *best* Treatment for that stage, e.g. to assess the utility of laparotomy before any other tests have been performed, again a weighted average is used as follows:

Stage	Probability of that stage		5-year DFS for that stage with BEST Rx	
I+IIB	0.22	times	70% TNI	= 15.4%
IIIB	0.64	times	50% MOPP	= 32.0%
IVB	0.15	times	35% MOPP	= 5.25%
		Weighted average		= 52.65%

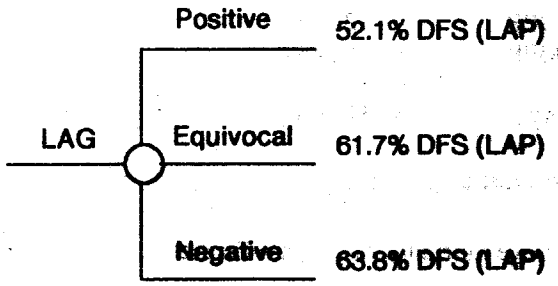
To obtain the true utility for laparotomy, one must also take into account its mortality risk - this is usually assessed as 1%, which includes 0.5% peri-operative mortality risk, and a further 0.5% late mortality risk from post-splenectomy sepsis. The laparotomy branch of the decision tree actually looks like:



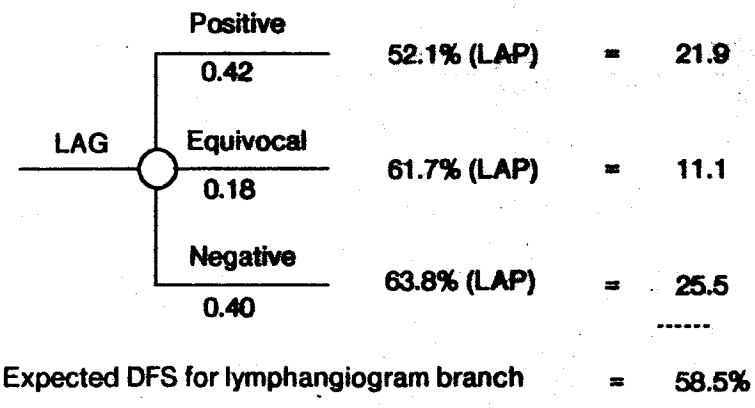
The value obtained in the calculations above for the average utility of LAP must therefore be multiplied by 0.99 to give the actual value which should be compared with the best treatment option (52.65% X 0.99 = 52.12%). When this modified value is compared with that derived for immediate treatment prior to any

testing, it is apparent that the laparotomy option (52.1%) still gives a higher chance on the average for disease-free-survival than the preferred treatment option MOPP (50.45%), so its value is assigned as the value for that decision node.

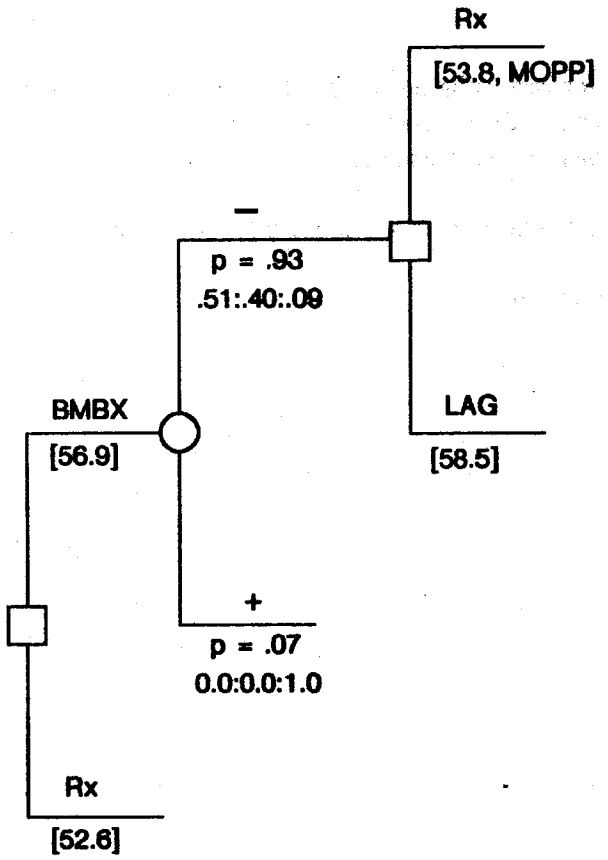
The other decision nodes which follow lymphangiogram are evaluated similarly, the only difference being the probabilities of stage which occur with equivocal and negative lymphangiograms.



To determine the value of the whole lymphangiogram branch - or the value of any branches after a chance node - one must first calculate the likelihood of each outcome after the chance node, and then multiply the expected utility of each branch by its likelihood. This gives the overall expected utility for the whole lymphangiogram branch.



Calculations for the whole tree are carried out similarly:



When all the calculations are performed, the utility of the BMBX option is 56.9% DFS, which exceeds that of immediate therapy (52.6% DFS).

The best course of action in this situation is therefore:

> To perform bone marrow biopsy; if positive give MOPP, if negative further testing is needed. Regardless of the results of lymphangiogram, laparotomy has a higher utility than the treatment options. This indicates that lymphangiogram is not necessary for this patient, but that laparotomy is needed to select appropriate treatment.

In summary, any decision tree is evaluated in the following way. Starting with the "leafiest" part of the tree, i.e. that furthest away from the fundamental decision, utilities are calculated in two ways:

1. at a *decision node* utilities are calculated for each branch and the *highest* one is selected as the value of the node.

2. at a *chance* node utilities are calculated for each subsequent branch; the utility of a particular branch is then multiplied by its probability and *all* the products obtained are added together to give a weighted average which is the expected utility for that chance node.

In a decision tree as complicated as that for HD, where there are more than 2000 decision branches to evaluate, it is necessary to use a computer for the calculations.

4. Conditional Independence Assumptions

For each patient with Hodgkins disease, we start out with information about histology, symptoms, sex and age. As further clinical information is considered, and diagnostic tests are performed more clues about the spread of disease are obtained. In the decision analysis program we repeatedly ask the question: given the information available so far, what are the probabilities of each stage in this patient?

Suppose we have a large number of past patients whose stages are now known, and who are identical to the current patient in all relevant respects. Probabilities about the new patient's stage could be estimated directly from the actual distribution of stage among the earlier patients. Thus, for example, if 30% of the former patients were in stage III, 30% is a sensible estimate for the probability of stage III disease for the patient at hand.

As more and more findings are added to the database, however, the cohort of past patients highly similar to the present one becomes smaller. Probabilistic predictions obtained under the method just described would be subject to such large sampling error as to be almost useless. If our calculations of staging probabilities are to be at all informative, certain independence assumptions must be fairly accurate. These assumptions, crudely speaking, assert that the way a particular finding is used in modifying a patient's prognosis does not depend on any other information about the patient.

In this section, we discuss our initial assumptions of independence, the tests we performed to validate them, and a modification of the program that arose when one of the independence assumptions was found inconsistent with our data.

4.1 The Independence Hypothesis

Suppose that, for a given patient, we have a set Y of prior findings and a new finding C. Bayes' Theorem asserts that the probability the patient is in stage i (i = 1,2,3 or 4) given this information follows:

$$P(i|Y \text{ and } C) = \frac{P(Y|i) \times P(C|i \text{ and } Y)}{P(Y \text{ and } C)} \times P(i) \quad (1)$$

where P(i) = probability of i, P(i|Y) = conditional probability of i given Y.

The independence assumption we make is that P(C|i and Y) = P(C|i), which means that, among those in stage i there is no correlation between those with finding C and those with finding Y. Under this independence assumption, Equation (1) becomes:

$$P(i|Y \text{ and } C) = \frac{P(Y|i) \times P(C|i)}{P(Y \text{ and } C)} \times P(i) \quad (1A)$$

The value using equation (1A) instead of (1) is that is allows us to use larger data sets, less subject to sampling error, in the calculation of needed probabilities.

Of course, if the independence assumption is false, calculations based on it cannot be trusted. Thus it was necessary to perform statistical tests of the null hypothesis H_0 that P(C|i and Y) = P(C|i) where C and Y were varied. The tests we employed were tied to the familiar Chi-squared test.

4.2 Tests of Independence

For all patients with Hodgkins disease we start with information about histology, symptoms, sex and age. Our initial estimates of staging probabilities were based on histology and symptoms alone. Thus, for example, the data-based estimate of the probability of being in stage III given mixed cellularity histology and the presence of symptoms (e.g. night sweats) was 0.30.

The data make clear that the very fact of being male makes one more likely to be in a later stage. Probabilities of stage given sex as well as symptoms and histology were estimated by the program under (1A), under the assumption that P(male given i) = P(male given i and Y), where Y is any one of the eight possible combinations of findings about histology and symptoms. The number 8 arises because there are four possible histologic subtypes and key symptoms are either present (B) or absent (A).

The following example will show how we used the Chi-squared test to examine the null hypothesis of independence among the basic findings sex, symptoms and the histological subtype (Eisen, 1977). In our database, 51% of the patients in stage I are male. 164 of the patients are asymptomatic (A), of whom 132 have the Nodular sclerosis (NS), 17 have Mixed cellularity (MC), and 15 have the Lymphocyte predominant (LP) histological subtypes. Under the independence hypothesis, the number of males in stage I who are also NS and A would be expected to be 132 X 0.51 = 67. The expected number of males who are A and MC, and A and LP, would be 9 and 8 respectively. Consider the table below:

Stage I Asymptomatic Patients

Histology	Expected number of Males (E_j) Under Independence Hypothesis	Actual number of Males (A_j)
NS (132 patients)	67	60
MC (17 ")	9	11
LP (15 ")	8	13

The viability of the null hypothesis depends on whether the observed differences between the expected (E_j) and the actual (A_j) numbers of males can reasonably be dismissed as chance fluctuations. We calculate the "discrepancy index" B_j under the rule $B_j = (E_j - A_j) / E_j$ for each of the categories above. We make similar calculations for those males who did have symptoms, and also for the females in stage I. The sum of the B_j's for all categories is compared to some cutoff number Z, with the null hypothesis rejected if the sum exceeds Z and accepted if the sum is at or below Z. Z is dependent on the number of different categories, and the extent to which certain categories are "redundant" of others; both these considerations are very familiar to users of Chi-squared tests.

It turns out that, for sex vs. histology and symptoms, the independence assumption was not acceptable under the statistical test. Thus to include the finding "sex" in the prediction, we had to go back to Equation (1) rather than use the simpler form (1a). We then proceeded in similar fashion to see whether age could be treated as independent of sex, symptoms, and histology in the estimation of staging probabilities. Here, the independence assumption passed its test easily.

We also investigated whether the errors in various diagnostic tests could be viewed as independent events. (E.g. does a false positive result in a Gallium scan increase the probability of a similar error in a lymphangiogram?) The data did not compel us to reject the independence assumptions, but often we had too few data points to allow any definitive statements.

In summary, we did not forget that our Bayesian procedure for revising staging probabilities was based on a strong independence assumption. Some tests of the assumption's validity were in order and were performed. Except in one case (sex vs. histology and symptoms), the tests indicated that the assumption was close enough to the truth that using it did not introduce serious errors in the calculations. In the case just mentioned, where the independence assumptions failed[1] , we recalculated probabilities. A less direct method of testing our assumptions was the full-scale evaluation of the program as described in Chapter 5. This evaluation showed that our predictions of stage, even when many findings were used by applying Bayes' Rule repeatedly, were reasonably accurate. This would not have occurred had our assumptions of independence proved false.

We feel that our procedure, including the tests of independence, came as close as possible to avoiding the twin evils of biased predictions and predictions with very high variance, either of which would have doomed our efforts.

---

1. One might wonder whether, when many tests of independence are performed, chance alone would lead one of them to yield a negative result. In this situation, the rejection was sufficiently emphatic that we did not think this was the case.

5. Testing the Bayesian Model for Predicting Pathological Stage

In Chapter 4, we showed that the independence assumptions that underlie the Bayesian calculations of staging properties were consistent with the data used to calibrate our computer program. But, there is another issue that must be addressed before the program can be considered for general use. Can a model based largely on the experiences of patients from large metropolitan teaching hospitals reliably be used in making prognoses at other hospitals?

There are several reasons one might fear this question should be answered in the negative. More than half the patients in the database came from a well-known radiotherapy referral center, which is perhaps more likely than other centers to attract patients with localized disease. This circumstance raises the possibility that probabilistic predictions based on data dominated by these patients might be overly optimistic for patients elsewhere. Furthermore, the data base consists solely of laparotomy-staged patients. This could lead to underestimation of the prevalence of stage IV disease, for some patients in this stage could have had their extra-nodal involvement confirmed without laparotomy. Moreover, the predictions depend on the histologic subtypes reported for the patients. Pathologists tend to differ somewhat in their assessments of histologic subtype (Jones et al, 1977); to the extent that pathologists differ from their counterparts elsewhere, the "transferability" of results is diminished.

Finally, each of the prior and conditional probabilities used in the program has an uncertainty due to sampling error, an uncertainty which may be increased by the repeated use of Bayes Rule. In principle, we could estimate the uncertainty in our final staging probabilities as a function of the uncertainties in the numbers that determine it. However, the calculations would be very difficult in practice, and ultimately of quite limited value if we could demonstrate that the program makes accurate predictions of stage.

The considerations suggest the importance of testing the general predictive power of the decision analysis program. In this section, we describe 2 separate empirical tests, one conducted in 1977 and the other in 1980.

Both tests, as we will see, suggest that the program's predictions are accurate to within an irreducible margin-of-error caused by usual chance fluctuations.

5.1 The 1977 Test

The test population we used consisted of 156 Hodgkins disease patients, 82 of whom were treated at Center A and 74 of whom were treated at the Tufts-New England Medical Center Hospital. We thought it appropriate to consider the two groups of patients separately, so that interesting, but opposite patterns, if they arose, would not cancel one another out. For each patient, information was obtained about histology, symptomatology, age, sex, presenting lymph nodes, spleen and liver scan results, and alkaline phosphatase levels[1]. These data were fed into our program, which used them to specify a probability distribution for the patient's stage under the Ann Arbor criteria. To assess the accuracy of the program, we compared these probabilistic predictions with the patient's actual stage as determined by laparotomy.

An obvious question arises: how does one tell if a probabilistic prediction is accurate? Suppose, for example, the program predicts that patient A has a 60% chance of being in stage I or II and a 40% chance of being in stage III or IV. Even if the patient is found through laparotomy to be in stage III, the prediction is not really wrong, for it had indicated considerable uncertainty about how far the disease had spread. But, while it is difficult to talk meaningfully about the accuracy of a given patient's prediction, it is somewhat easier to do so, given a large set of similar patients.

Suppose, for instance, that there are 50 patients for whom the program makes the 60%/40% prediction just mentioned. If these predictions are all accurate, one would expect that, except for random fluctuations, the number of patients who actually emerge in stages I and II would be 50 X .6 = 30. Thus, if the number observed in the two early stages is significantly greater than 30, this means the program was too pessimistic for

1. Not all information was available for every patient.

these patients. If, on the other hand, the number is far below 30, the program was too optimistic.

Our statistical test of the program's accuracy for the two group of patients is based on the principles cited above. In performing the test, we combined stages I+II because, for clinical purposes, the two stages are virtually the same. We also combined stages III and IV because, given the small test populations, we expected so few stage IV patients that we will be unable to make meaningful statements about the stage IV predictions. For each patient in the test, we recorded that program's estimate of "r", the probability the patient is in stage I+II. The probability that the patient was in stage III or IV is simply [1-r]. We then ordered the patients according to their r-values, from highest to lowest.

According to the criteria discussed in Appendix II, we broke the patients into groups within which the r values are close together. Finally, we compared the expected number of stage I+II patients in each group with the actual number as learned from laparotomy. We present some of our results in Tables 5.1-5.4:

Table 5.1. Evaluating the TUFTS patients, using Histology, Symptoms and Sex

Group	Number in Group	Average probability in group	No. in Stages I+II		W-Statistic
			Expected	Actual	
1	22	0.71	15.64	15	-0.26
2	26	0.57	14.90	17	0.84
3	26	0.3	7.72	8	0.12
	74		38.17	40	
			Z = 0.79		

In the tables the W-statistic for any group is a measure of whether the discrepancy between expected and actual outcomes can reasonably be attributed to chance alone. The statistic has only a 5% chance of exceeding 2 in absolute value if all predictions in the corresponding group are correct. Under the null hypothesis that all

Table 5.2. Evaluating the Center A patients, using Histology, Symptoms and Sex

Group	Number in Group	Average probability in group	No. in Stages I+II Expected	Actual	W-Statistic
1	16	0.7	11.24	12	0.41
2	28	0.6	16.84	19	0.84
3	21	0.44	9.28	8	-0.53
4	22	0.31	6.73	8	0.60

Z = 1.64

Table 5.3. Evaluating the TUFTS Patients, using Histology, Symptoms, Sex, Age and Alkaline Phosphatase Levels

Group	Number in Group	Average probability in group	No. in Stages I+II Expected	Actual	W-Statistic
1	23	0.74	16.91	15	-0.81
2	14	0.61	8.52	10	0.66
3	14	0.5	7.06	8	0.25
4	21	0.28	5.81	6	-0.41

Z = 1.89

Table 5.4. Evaluating the Center A Patients using Histology, Symptoms, Sex, Age and Left-Neck Nodes

Group	Number in Group	Average probability in group	No. in Stages I+II Expected	Actual	W-Statistic
1	17	0.76	12.74	14	0.7
2	15	0.67	10.1	8	-1.16
3	18	0.57	10.28	13	1.30
4	15	0.47	7.08	6	-0.56
5	22	0.32	7.03	6	-0.48

Z = 3.60

---

predictions in all groups are correct, the statistic Z obtained by summing the squares of the W-values should follow a Chi-squared probability distribution, with as many degrees of freedom as there are groups. Thus at the usual 5% level, one would reject the null hypothesis only if Z exceeds 7.81 when there are three groups, 9.49 when there are 4 and 11.07 when there are 5. Tables 5.1 and 5.2 indicate that the projected distributions on stage, based on histology, symptoms and sex were accurate at both the New England Medical Center

Hospital and Center A. Tables 5.3 and 5.4 suggest that the initial probabilities, modified in Bayesian fashion as additional findings become available, are also accurate. Revised predictions for the subset of patients who also had liver and spleen scans were as consistent with the data as those summarized in the tables. These tables imply that our concern that the predictions largely derived from a radiotherapy referral center would not be relevant elsewhere has not been realized.

At this point the reader might be wondering why, if the initial predictions in Tables 5.1 and 5.2 were accurate, we bothered to revise them. The reason is that we need predictions that are not only accurate, but as helpful as possible. To a clinician deciding between treatments it is far more informative to know that the probability of a particular contingency is 0.99 rather than 0.50. Since including new findings in the predictions tend to polarize them towards 1 (certainty) and zero (impossibility), the revised predictions are generally somewhat less equivocal than the early ones.

Once the Tufts and Center A patients were shown to be compatible with the original database, it seemed appropriate to include their data in an expanded data base. The program that resulted from merging the data was based on 900 rather than 700 patients. Of course the very compatibility of the new patients with the original ones means that the merger brought only slight changes to the program.

5.2 The 1980 Test

In 1980, we decided to perform another test of the validity of the program outside the hospitals involved in its calibration. The reasons for this new test were that (i) the 1977 data samples were too small to allow the consideration of stage IV predictions in their own right, and (ii) the problems inherent in the variable assessments of histology may have been artificially reduced in the 1977 study because the test hospitals were similar in terms of histopathology.

The 1980 test population consisted of all new pathologically-staged cases of Hodgkins disease arising in 100

cities and towns in Eastern Massachusetts over a defined period of time. Confidentiality of data was maintained, with both patients and hospitals indentified by study numbers only. The original database was searched to insure that none of this test population was included in the existing database. Clinical information about these patients came from 70 hospitals in the study area; in all, data on histology, sex, symptoms, age, and sites of lymphatic involvement was available for 301 pathologically-staged patients, 206 of whom had lymphangiogram tests.

The method for testing accuracy was similar to that used in 1977, except that we wanted to learn about the reliability of all three of the predictions of stage rather than just the two groupings used previously. The test procedure, which is described in detail in Appendix II, involved choosing at random one of the three predictions made for each patient. We called the stage to which the chosen prediction for a given patient referred, the "stage of interest". Thus, given 301 patients, we were really assessing the accuracy of 100 predictions about stage I+II, 100 about III and 100 about IV.

Tables 5.5 and 5.6 summarizing our test results appear below. The W-statistics can be interpreted the same way as those that appeared in tables 5.1-5.4. The hypothesis that all predictions are accurate is statistically unacceptable at the 5% level if Z is greater than 16.92 for 9 groups, and 18.31 for 10.

Tables 5.5 and 5.6 make clear that the program's predictions are accurate to within the level of chance fluctuations. Thus, once again, we can have confidence in the original probabilities assigned to various stages, and in the somewhat modified estimates that arise when additional information (generally including lymphangiogram results) is taken into account.

In summary, our tests in 1977 and 1980 offered strong evidence that the various potential problems in our Bayesian program are of little practical importance. The results increase our confidence that the program can be useful in a wide variety of hospitals.

Table 5.5. 1980 Test of predictions based on symptoms, sex and histologic subtype

Number of staged patients with this information available: 301
Actual number overall in "stage of interest": 99
Expected number overall in "stage of interest": 100.9

Group	Number in Group	Expected in "stage of interest"	Actual in "stage of interest"	W-Statistic
1	37	24.94	28	1.21
2	20	12.84	11	-0.70
3	11	5.92	5	-0.06
4	11	5.47	3	-1.42
5	14	6.73	9	1.08
6	12	5.56	4	-1.08
7	13	5.91	5	-0.76
8	19	7.51	6	-0.50
9	34	11.07	9	-0.78
10	17	5.18	5	-0.27
11	113	9.78	14	1.60

Z = 10.39

Table 5.6. 1980 Test of predictions using clinical findings and lymphangiogram results

Number of staged patients with this information available: 206
Actual number overall in "stage of interest": 63
Expected number overall in "stage of interest": 66.46

Group	Number in Group	Expected in "stage of interest"	Actual in "stage of interest"	W-Statistic
1	28	22.80	24	0.73
2	17	11.87	9	-1.44
3	12	6.86	7	0.18
4	11	5.04	4	-0.83
5	14	5.05	5	0.06
6	18	5.02	3	-1.02
7	106	9.81	11	0.63

Z = 4.66

---

6. Example of a Typical Patient Session

We will demonstrate the program with a typical patient session. The patient is a 25 year old male who has dropped 20lb in weight from his usual 170lb. He has no other generalized symptoms, and has sought medical attention because of a slowly-increasing mass on the left side of his neck. Biopsy of this left-neck mass revealed Hodgkins disease of the nodular sclerosis histologic subtype. Investigations performed at this time include a chest X-ray and a liver-spleen scan, both of which were normal. We will use this basic information to predict the likelihood of each of the stages of the disease, and use this knowledge to plan further investigations tailored to treatment.

First the program asks the three question which give information relating to the three interdependent findings from which the prior probabilities are estimated. In the dialogue between computer and user, the computer's questions are in `this typeface`, the user's responses in *italic typeface*.

```
Is the patient A (asymptomatic) or B (symptomatic)?      B
What is the histologic subtype of the patient?           Nodular sclerosis
What is the sex of the patient?                          Male
```

The program then refers to its database to give the appropriate prior probabilities, i.e.

```
Stages I+II      0.34
Stage III        0.50
Stage IV         0.16
```

Having established these, further questions are asked to characterize the patient. With each question the probabilities are modified by Bayes Rule, using the conditional probabilities of finding given stage.

```
In what age group is the patient?                        12 to 39 years
```

This is the correct answer for this patient, but by pressing "?" all possible answers the program will accept are displayed, i.e.

```
Unknown
Younger than 12 years
12 to 39 years
Older than 39 years
*Display data about fact...
*What if....
```

If the response "*WHAT IF..." is selected, the program displays the posterior probabilities for each possible answer.

```
Younger than 12 years:          I+II 0.21   III 0.52   IV 0.27
12 to 39 years:                 I+II 0.39   III 0.51   IV 0.10
Older than 39 years:            I+II 0.20   III 0.48   IV 0.32
```

The "*DISPLAY DATA ABOUT FACT" option displays the data known about the relationship between the patient's age and the stage of the disease, and the conditional probabilities of age given stage, by which the prior probabilities are modified.

```
    Younger than 12 years       I+II 0.03   III 0.05   IV 0.08
    12 to 39 years              I+II 0.84   III 0.74   IV 0.45
    Older than 39 years         I+II 0.14   III 0.23   IV 0.47
```

After one enters the appropriate age group for this patient -- 12 TO 39 YEARS -- the prior probabilities are modified by Bayes Rule, using these conditional probabilities, in this case altering them in favor of early stage disease.

```
Stages I+II       0.39
Stage III         0.51
Stage IV          0.10
```

The user is then asked consecutively about left neck/right neck and mediastinal nodes, which make modest changes in the probabilities. Again the same explanatory features are available.

```
Were there any nodes detected on the left
supraclavicular or cervical areas?                    Yes

Stages I+II       0.37
Stage III         0.53
Stage IV          0.10
```

```
Were there any nodes detected on the right
supraclavicular or cervical areas?                    No

Stages I+II       0.30
Stage III         0.54
Stage IV          0.06

Is there mediastinal involvement?                     No

Stages I+II       0.39
Stage III         0.59
Stage IV          0.07
```

Next, information is asked separately about the spleen scan and the liver scan. Once again the same explanatory features are available. Although these liver-spleen scan results were obtained simultaneously, because of their different interpretation, questions about their results are asked separately.

```
What is the appearance of the spleen on scan?         Normal
```

Acceptable answers include:

```
Normal scan
Enlarged (> 15 cm) on scan
Filling defects seen on scan
Both enlargement and filling defects seen on scan
*Display data about fact...
*What if...
```

With Normal scan the probabilities are modified to:

```
Stages I+II       0.48
Stage III         0.48
Stage IV          0.05
```

```
Was the uptake pattern of liver scan normal or abnormal?

Possible answers:
Normal
Abnormal
Unknown
*What if...
*Display data about fact...
```

Normal

```
Stages I+II       0.49
Stage III         0.49
Stage IV          0.03
```

The false positive and negative rates for the liver/spleen scan are given in Table 1.2. They have been combined with information from the database regarding the likelihood of an involved liver or spleen in the various stages to give conditional probabilities for the the probability of a particular scan result, given the pathologic stage.

Having used this basic information to characterize the patient, and obtained the above "posterior" probabilities of stage, the program now enquires which, if any, of the more invasive tests have been performed. For any that have, probabilities are further modified according to Bayes Rule, using false positive and false negative rates that have been derived from the literature, again modified where necessary to link results of the tests directly to pathologic stage as in the example shown for Gallium scan in Section 2.2.2.

The tests asked about at this stage include bone marrow biopsy (BMBX), percutaneous liver biopsy (LBX) or peritoneoscopy with guided liver biopsy (PTX), gallium scan of abdominal nodes (GALI.) lymphangiogram (LAG), and staging laparotomy (LAP).

The user is then asked whether any of these test procedures are contraindicated. For each procedure the user is prompted with a list of the recognized contraindications of the test, e.g. for lymphangiogram:

```
Does the patient have any of the following contraindications for
lymphangiogram?

          Previous radiotherapy to lungs
          Vascular shunts
          Iodide hypersensitivity
          Respiratory impairment
          Active thrombophlebitis
```

Any procedure which is contraindicated at this stage will not be used in constructing management plans.

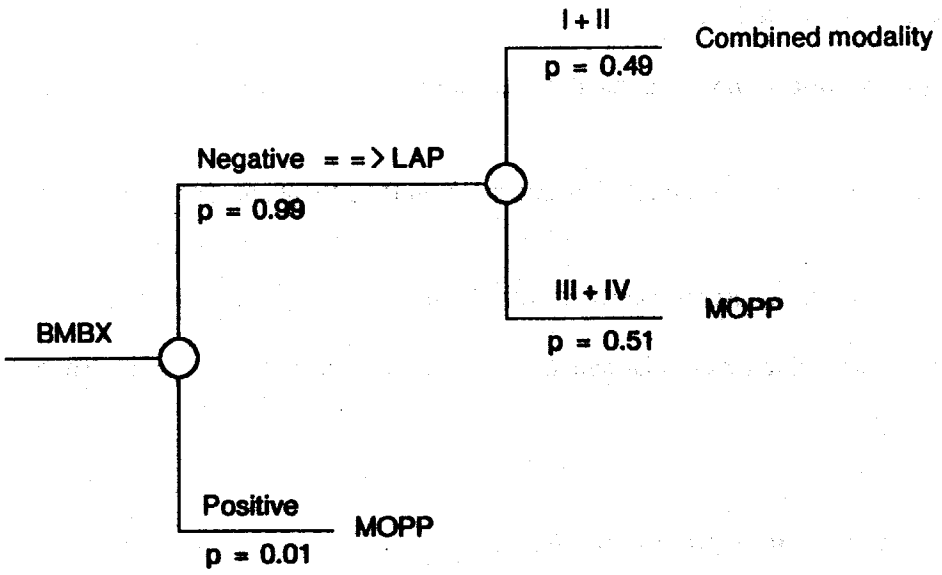
Decision analysis is now used to approach the TEST versus TREAT decision, i.e. whether it is necessary at this stage to perform some or all of the more invasive tests, or whether treatment can be given without more ado. The option of using combined modality therapy is also offered. The user is reminded that, when comparing 5 year disease-free survival (DFS) values after this therapy and after single modality therapy, allowance should be made for a second remission with chemotherapy if relapse occurs after single modality therapy. If the user elects to use combined modality therapy it is added to those considered in the TREAT branches of the decision tree; for symptomatic patients, such as the one under consideration, these treatments would then include combined modality therapy (CM), total nodal irradiation (TNI), and MOPP chemotherapy.

As all the possible branches of the decision tree are assessed, their utilities are displayed in tabular form, as below, with the first step of the plan in the left hand column.

First step in plan	Average 5-year DFS	Avg morbidity of testing	Avg morbidity of treatment	Dollar cost of plan
Total nodal	0.6955	0.0	1500	3500
MOPP	0.6742	0.0	3000	1000
CM	0.6788	0.0	4500	4200
Laparotomy	0.7306	865	3729	10565
Gallium scan	0.7307	878	3729	10550
Lymphangiogram	0.7298	906	3729	10565
Bone marrow	0.7307	900	3729	10580
Liver biopsy	0.7305	881	3729	10929

First, note that the plans are listed according to their first step. If this is a treatment, then there are, of course, no further steps. If the first step is a test, the utilities shown are calculated for the *entire* Plan, and include any

subsequent tests and treatment. For example, for the plan beginning with bone marrow biopsy, the plan with the highest utility, the whole plan looks like this:



This plan shows that when bone marrow biopsy is performed, there is a 0.01 chance of this test being positive; this chance is low because the probability of stage IV in this particular patient was only 0.02. Since a positive bone marrow biopsy is conclusive proof of stage IV disease, MOPP can then be given without any further testing. Ninety-nine percent of the time the bone marrow biopsy will prove negative. At this point the whole TEST versus TREAT decision is approached again. The best step next, with the probabilities revised by the negative bone marrow biopsy, is to perform a laparotomy, with the chances of I+IIB disease being .49, and III & IV being 0.51. Once laparotomy has been performed, stages are known for certain, and no further decision is necessary.

At this point the user is given the option of displaying a number of other items including:

`All plans` - these are displayed as decision trees similar to that above

`One plan` - any individual plan can be displayed by designating its first step

`Summary of plans` - very similar to the previous tabular display except that plans are displayed ranked according to their DFS utility.

`Verbose version of one plan` - a particular plan, specified by its first step, is described in words.

`Patient summary` - all the clinical and test data known about the patient are displayed.

Following this DISPLAY option, the user is presented with the further options which make the program so flexible that its use can be adapted to a particular patient, or to the experience of any hospital. These options include:

`Change patient information` - *further specified as:*

- `Clinical information`
- `Test information`

`Modify information about tests or treatments` - *further specified as:*

- `Prior probabilities`
- `False positive/negative rates for questions or tests`
- `Laparotomy as a perfect test` - *can assign a FN rate*
- `Mortalities for tests` - *the rates used are displayed together with an option to change them*
- `Morbidities for tests` - *see Table 3.2.*
- `Morbidities for treatments` - *see Table 3.2.*
- `Survival data for treatments` - *these can be changed according to institutional experience.*
- `Consider a different treatment`
- `Delete a treatment from consideration`

`Reanalyze current patient` - *if any of the MODIFY options are selected one must use this option to recalculate plans.*

`Supply your own plan` - *here the user is given the opportunity to supply his/her own plan for management, and compare it with those calculated by the computer.*

We will try some of these options for our patient. Imagine that there has been some doubt about his exact histologic subtype, and the pathologist notes it has some of the features of the mixed cellularity subtype. We want to see what difference this will make to the probabilities of stage, and hence to the management plans.

```
What do you want to do?                    Change patient information
What do you want to change?                Clinical information
```

The clinical information known about the patient will then be displayed, and the item to be changed designated. In this case, changing the histologic subtype from nodular sclerosis to mixed cellularity changes the probabilities from I+II 0.49 : III 0.49 : IV 0.03 to I+II 0.37 : III 0.55 : IV 0.08, i.e. it increases the probability of more advanced disease.

Further, supposing you are not very happy about the patient's lung function -- he is prone to asthmatic attacks, which you feel may affect the mortality risk for both the lymphangiogram and staging laparotomy.

```
What do you want to do?                    Modify information about tests or
                                           treatments

What do you want to modify?                Mortality rates for tests

Which test?                                Lymphangiogram

The current mortality for lymphangiogram is 0.001.
What would you like to change it to? 0.005

Which test?                                                    Laparotomy

The current mortality for laparotomy is 0.01.
What would you like to change it to? 0.02
```

Also you wish to consider as an alternative treatment for this patient, the "sandwich" form of combined modality therapy, in which 3 courses of MOPP are given, then a course of radiotherapy, then a second 3 courses of radiotherapy. Since this option replaces the more conventional combined modality therapy, it is first advisable to delete combined modality therapy from consideration, before indicating you wish to consider a different treatment. You will be asked at this point for the name of the new treatment and for appropriate disease free survival and morbidity data.

```
What would you like to do?                 Consider a different treatment.

Type the name of this treatment            "Sandwich"

Five year disease free survival for "Sandwich" given I+IIB:    0.75
```

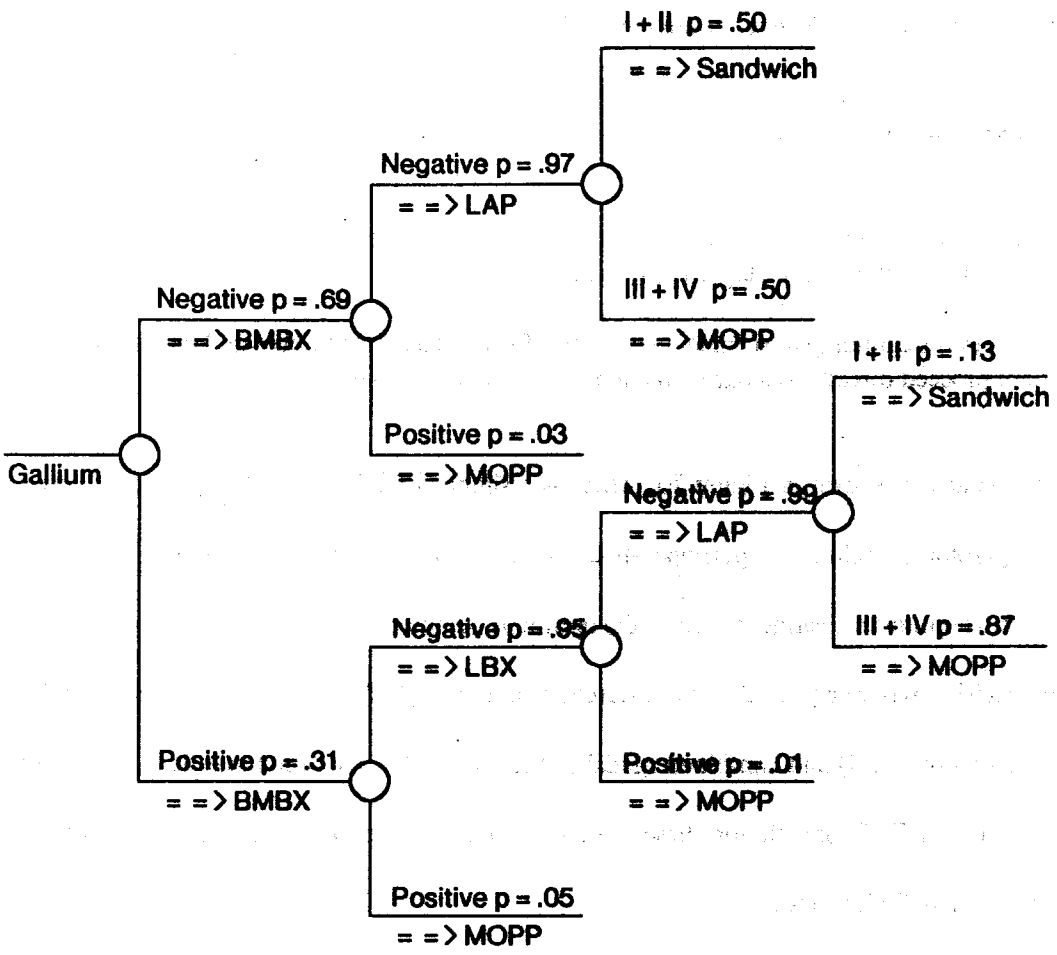
```
Five year disease free survival for "Sandwich" given IIIB:    0.4
Five year disease free survival for "Sandwich" given IVB:     0.35

What is the expected morbidity associated with Sandwich       3500

What would you like to do?          Reanalyze current patient
```

Now, the program again asks whether any invasive tests have been done already, or are contraindicated. It also asks whether combined modality therapy is indicated. Although standard combined modality therapy was deleted from the program, it was replaced by the "sandwich" combined therapy so we need to answer YES to this question so that salvage will be considered after the single modality therapies.

The computer then calculates a new series of plans, which incorporate the changes just made, i.e. the different prior probabilities, the increased mortality rate for the lymphangiogram and laparotomy, and the new "sandwich" treatment.

First step in plan	Average 5-year DFS	Avg morbidity of testing	Avg morbidity of treatment	Dollar cost of plan
Total nodal	0.6386	0.0	1500	3500
MOPP	0.6610	0.0	3000	1000
"Sandwich"	0.6819	0.0	3500	
Laparotomy	0.6938	865	3185	8628
Gallium scan	0.6942	859	3185	8814
Lymphangiogram	0.6909	614	3248	6048
Marrow biopsy	0.6942	850	3185	8473
Liver biopsy	0.6939	871	3185	8919

The plan with the highest disease-free-survival is the one beginning with the gallium scan:



Altering these factors has led to different diagnostic plans. Once again, further testing before selecting treatment gives better average disease-free survival values. However, there is now less emphasis on the tests, laparotomy and lymphangiogram, which have been estimated to have increased mortality rates.

7. Analysis: Immediate Laparotomy in Young Asymptomatic Patients

We wished to use decision analysis to answer two related questions:

1. Are there certain presentations of Hodgkins' disease (HD) which have such a low probability of infradiaphragmatic disease that staging laparotomy is not justified?

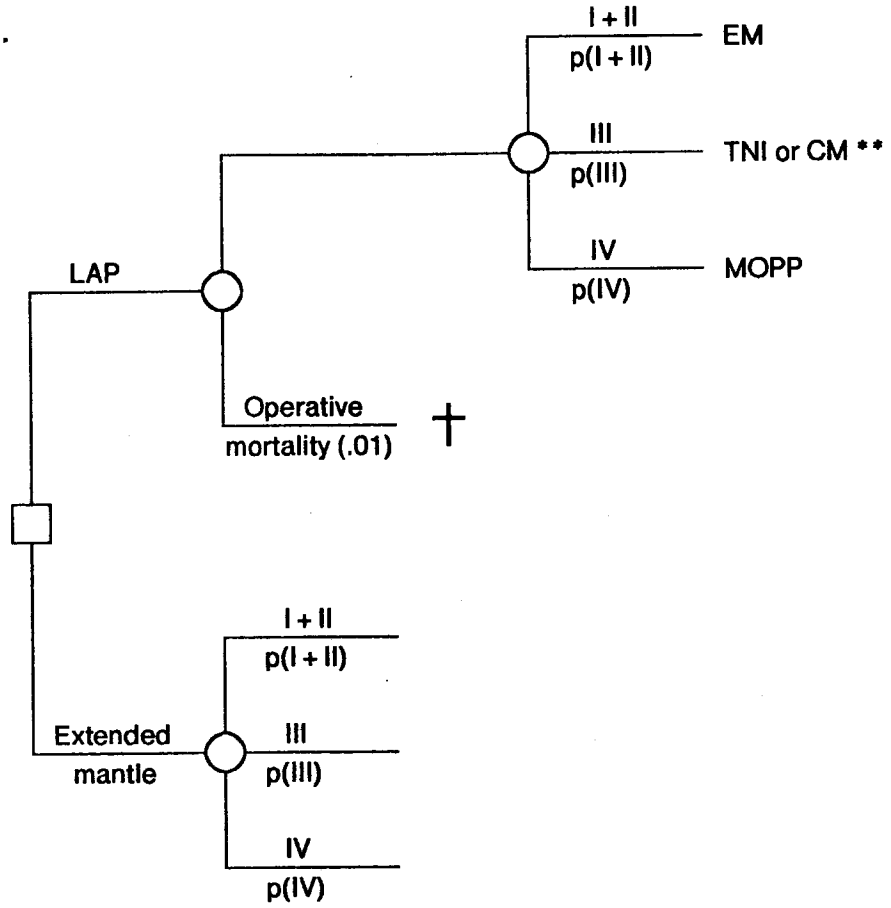
2. If negative test results cannot give sufficient certainty of early stage HD to obviate laparotomy, should the patient proceed directly to laparotomy and not be subjected to other staging tests?

The decision being evaluated is shown in Figure 7.1, which is a simple TEST versus TREAT decision tree. The test branch for laparotomy, includes its peri-operative and late infective mortality of 1%, and the option to treat each stage with its optimum treatment. Since the optimum treatment for stage IIIA is being debated, we will evaluate the decision separately for the two different treatments which are now being used for IIIA, namely total nodal irradiation (TNI) and combined modality therapy (CM). Extended mantle radiotherapy (EM) is evaluated in the TREAT branch for these young asymptomatic patients, as this would be the treatment appropriate to their clinical stage.

If the utility of immediate treatment is higher than that after laparotomy followed by stage-specific treatment, the tests are indicated since, if they prove to be negative, they may obviate laparotomy. If, on the other hand, laparotomy has the higher utility even when all the tests are negative, then one might proceed directly to laparotomy, with the knowledge that testing can never provide enough certainty of early stage disease to allow immediate treatment.

This particular analysis concerns asymptomatic patients in the age range 12-39 years, the age group with the highest chance of early stage disease. The probabilities of stage used to calculate the utilities are derived from the findings listed below - one criterion is taken from each category:

Fig. 7.1.



** When combined modality therapy is being evaluated as the optimal treatment for stage IIIA, MOPP salvage of relapse after the single modality therapy will be included for fair comparison.

Histologic subtype - only nodular sclerosis (NS) and mixed cellularity (MC) are considered.

Male or female

Involvement of right neck nodes (RN+/RN-)

Involvement of left neck nodes (LN+/LN-)

Involvement of mediastinal nodes (MED+/MED-)

There are 32 possible combinations of these findings; each set of probabilities is further modified in accordance with results of all tests prior to laparotomy being negative. For simplicity we will consider only

eight of these combinations i.e. for each sex/histology combination we will show the calculated utility for the presentations with:

the *most* "FAVORABLE" clinical findings - LN-, RN-, MED+

the *most* "UNFAVORABLE" findings - LN+, RN+, MED-

Table 7.1 shows the average five year disease-free-survival (5 yr DFS) calculated for each branch of the decision tree in Figure 7.1, when the treatment for stage IIIA after laparotomy is TNI.

Table 7.2 shows the average 5 yr DFS for each branch of the decision tree when combined modality is used

---

Table 7.1.

5 year Disease Free Survival with Laparotomy followed
by Single Modality Therapy or with Immediate Extended Mantle Radiotherapy

PATIENT	LAP WITH TNI FOR STAGE III	EXTENDED MANTLE RADIOTHERAPY
NS FEMALE "FAVORABLE" (0.93,0.07,0.0)	79.9%	78.9%
NS FEMALE "UNFAVORABLE" (0.85,0.15,0.00)	78.6%	75.7%
NS MALE "FAVORABLE" (0.91,0.08,0.00)	79.6%	78.3%
NS MALE "UNFAVORABLE" (0.82, 0.17, 0.01)	77.9%	74.0%
MC FEMALE "FAVORABLE" (0.88,0.12,0.01)	78.9%	76.6%
MC FEMALE "UNFAVORABLE" (0.75,0.22,0.03)	76.6%	70.6%
MC MALE "FAVORABLE" (0.82,0.18, 0.01)	77.8%	74.0%
MC MALE "UNFAVORABLE" (0.66,0.32,0.02)	74.8%	66.5%

for those patients found to be in stage IIIA at laparotomy. For comparison with immediate extended mantle radiotherapy, allowance is made for MOPP salvage of patients who relapse after any single modality therapy.

Table 7.2.

5 year DFS with Laparotomy followed by Single/Combined(IIIA) Modality Therapy or with Immediate Extended Mantle Radiotherapy

PATIENT FINDINGS (PROBABILITIES)	LAP WITH CM FOR STAGE III	EXTENDED MANTLE RADIOTHERAPY
NS FEMALE "FAVORABLE" (0.93,0.07,0.0)	87.7%	86.7%
NS FEMALE "UNFAVORABLE" (0.85,0.15,0.00)	87.1%	84.1%
NS MALE "FAVORABLE" (0.91,0.08,0.00)	87.5%	86.1%
NS MALE "UNFAVORABLE" (0.82, 0.17, 0.01)	86.8%	82.8%
MC FEMALE "FAVORABLE" (0.88,0.12,0.01)	87.2%	84.8%
MC FEMALE "UNFAVORABLE" (0.75,0.22,0.03)	86.1%	79.8%
MC MALE "FAVORABLE" (0.82,0.18, 0.01)	86.8%	82.8%
MC MALE "UNFAVORABLE" (0.66,0.32,0.02)	85.5%	76.8%

For all the combinations of clinical findings considered, and with both single and combined modality treatment options, laparotomy is calculated to have the highest average utility. However, for some of the more favorable clinical presentations there is only a small difference in outcome between laparotomy and immediate treatment. Only when the probability of stage III disease exceeds 0.20, e.g. the "UNFAVORABLE" MC presentations, is the difference in average five year DFS between the TEST and

TREAT options above 5%. Other factors than disease-free survival may influence the decision, for example the opportunity to perform oophoropexy in female patients.

If one cannot obviate laparotomy even when all other staging tests are negative, we feel it is preferable to proceed directly to laparotomy once the diagnosis of Hodgkins disease is made in this group of asymptomatic patients in the age group between 12 and 39 years. We particularly favor this course of action when certain unfavorable clinical findings give an appreciable chance of stage III disease.

8. Analysis: Laparotomy in Symptomatic Patients

In using the Bayesian diagnosis program we found that certain cohorts of symptomatic patients with a positive lymphangiogram have a very low probability of early stage disease. This study was carried out to determine whether, amongst these cohorts, there were any for whom the risks of staging laparotomy (LAP) outweighed the benefits it conferred through knowing the stage accurately. This work was first performed and published using the 900 patient database (Rutherford et al, 1980). The present study uses the prior probabilities calculated from the 1200 patient database.

For this study the probabilities of stage were calculated for cohorts of symptomatic patients, with a cohort being defined by one criterion from each of the following categories:

(1) Nodular sclerosis (NS) or mixed cellularity (MC) histologic subtypes

(2) Age 12-39 years, or older than 39 years

(3) Male or female

(4) Involvement/non-involvement left cervical lymph nodes.

(5) Involvement/non-involvement right cervical lymph nodes.

(6) Involvement/non-involvement mediastinal lymph nodes.

All patients were assumed to have a positive lymphangiogram and a negative percutaneous bone marrow biopsy.

There were thus 56 different combinations of findings, for which probabilities of stage were calculated, again using a simplified three stage version of the Ann Arbor criteria (Carbone et al, 1971), in which stages IB and IIB are combined. There were too few patients with the lymphocyte predominant and lymphocyte depleted subtypes for meaningful data in these categories.

8.1 Calculation of Decision Making Thresholds

A decision-making threshold was calculated at which the result of treatment after laparotomy, taking into account the mortality of the operation, was equivalent to immediate combination chemotherapy treatment with nitrogen mustard, vincristine (Oncovin), procarbazine and prednisone (MOPP) (DeVita et al, 1980). Treatments planned after laparotomy were total nodal irradiation (TNI) for stages IB AND IIB, and MOPP chemotherapy for IIIB and IVB. Estimates of the probability of five year disease free survival (DFS) for a given stage with these treatments are shown in Table 8.1. These results are obtained from data from the literature (Aisenberg et al, 1976, DeVita et al 1980, Goodman et al 1977, Rosenberg et al, 1975).

Table 8.1. Percentage 5 Year Disease-Free Survival for Treatments

	TNI	MOPP
Stage IB and IIB	70	60*
Stage IIIB	25	50
Stage IVB	0*	35

* = estimate

The probability of 5 year disease-free survival after laparotomy is calculated by adding the products of the probability of each stage, derived from the computer and the results of the best treatment for that stage; this is then corrected for the mortality of the operation (peri-operative and post-infective), i.e.

$$[DFS_{TNI,I+II}\ p(I+II) + DFS_{MOPP,III}\ p(III) + DFS_{MOPP,IV}\ p(IV)]\ (1 - mortality)$$

If we substitute the appropriate DFS values from Table 8.1 this expression becomes:

$$[0.70\ p(I+II) + 0.50\ p(III) + 0.35\ p(IV)]\ (1 - mortality)$$

The probability of 5 year disease-free survival if immediate MOPP therapy is given is:

$$DFS_{MOPP,I+II}\ p(I+II) + DFS_{MOPP,III}\ p(III) + DFS_{MOPP,IV}\ p(IV)$$

Again, substituting the values from Table 8.1 this becomes:

$$0.60\ p(I+II) + 0.50\ p(III) + 0.35\ p(IV)$$

P(IV) can be expressed as [1 - p(I+II) - p(III)]. The threshold between laparotomy and immediate MOPP therapy - found when the post-laparotomy disease-free survival is equal to that after immediate MOPP - can then be written as a linear equation. If, for example, one assumes a 1% total mortality from laparotomy, substitution of this in the laparotomy expression gives a linear equation for the threshold in terms of the probability of stage (I+II) and III:

$$9.65\ p(I+II) - 0.15\ p(III) = 0.35$$

This threshold is displayed in a graph with the probability of stage (I+II) and stage III on abscissa amd ordinate respectively and a line drawn between 1.0 on each axis, to outline the area relevant to the study (Figure 8.1).

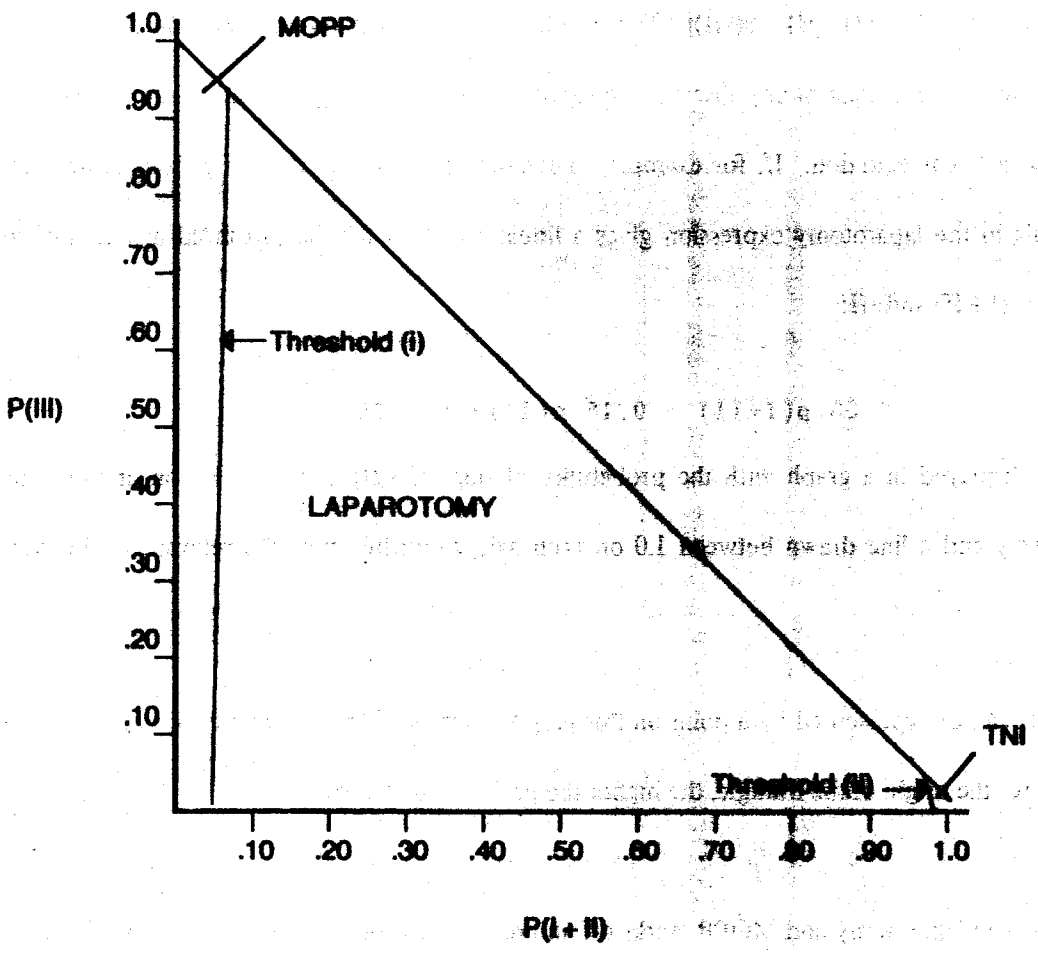
Any patient cohort can be represented by a point on this graph in terms of the probability of stage I+II and stage III. The nearer the origin of the triangle, the higher the probability of stage IV.

The threshold between laparotomy and MOPP marks off an area where there is a low probability of stage IB and IIB. To the right of this threshold line is the area where laparotomy is the best course of action, despite its mortality.

The equation for the threshold between laparotomy and MOPP is affected by the estimated mortality for laparotomy. As the estimated mortality of laparotomy increases, this threshold shifts to the right, diminishing the area where laparotomy is the preferred course of action, and increasing the area where immediate MOPP should be given (Figure 8.2).

The equation for the threshold if the laparotomy mortality is 2% is:

Fig. 8.1.
Decision thresholds between LAP and MOPP (Threshold i) and LAP and TNI (Threshold ii) when LAP mortality is estimated as 1%.



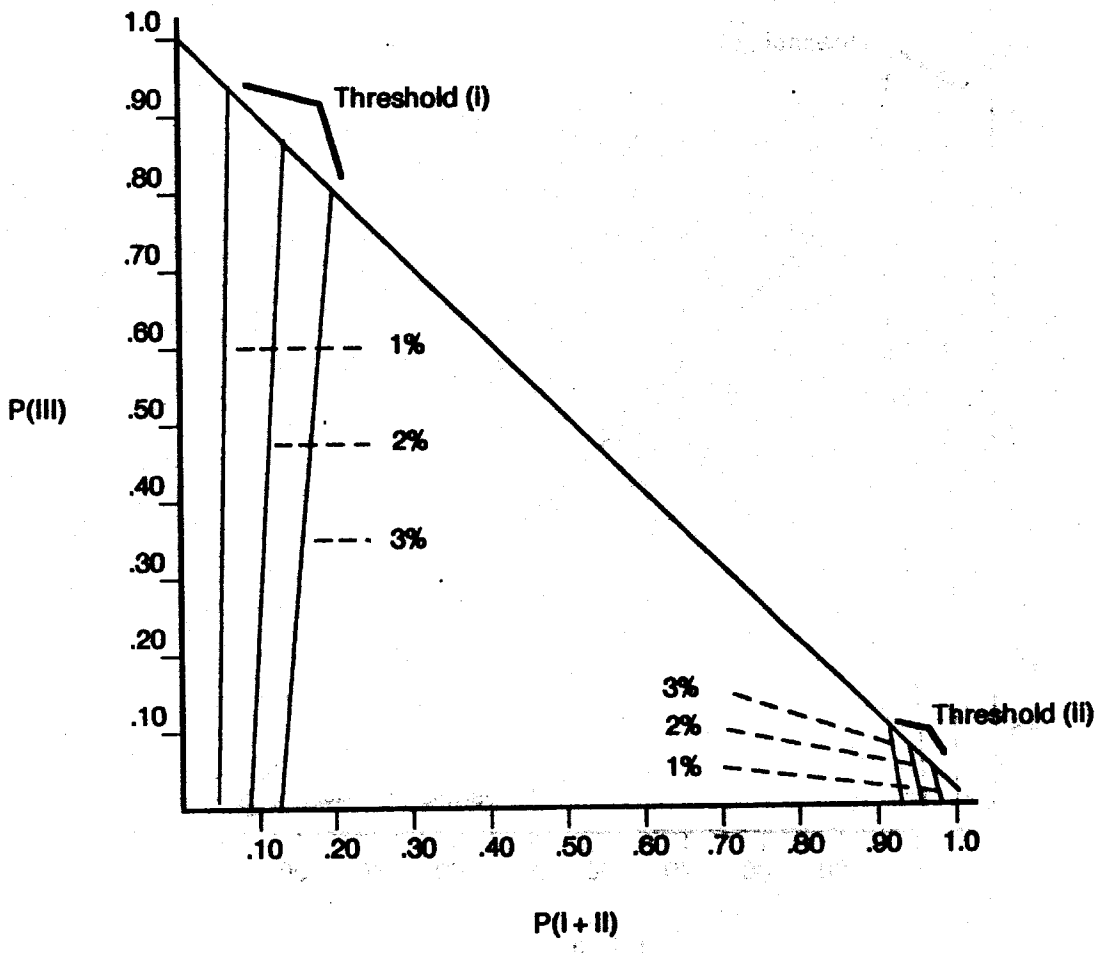
$$9.3\ p(I+II) - 0.3\ p(III) = 0.7$$

and if laparotomy mortality is 3%, the threshold equation becomes:

$$8.95\ p(I+II) - 0.45\ p(III) = 1.05$$

If there is uncertainty in the results of therapy, the threshold equation reflects this, and the threshold line becomes blurred. For example, if our estimates for DFS are off by 0.03 in opposite directions, and the difference between the results of the two treatments is at a maximum, the threshold line between MOPP and

Fig. 8.2. The effect of increasing LAP mortality estimates on decision thresholds



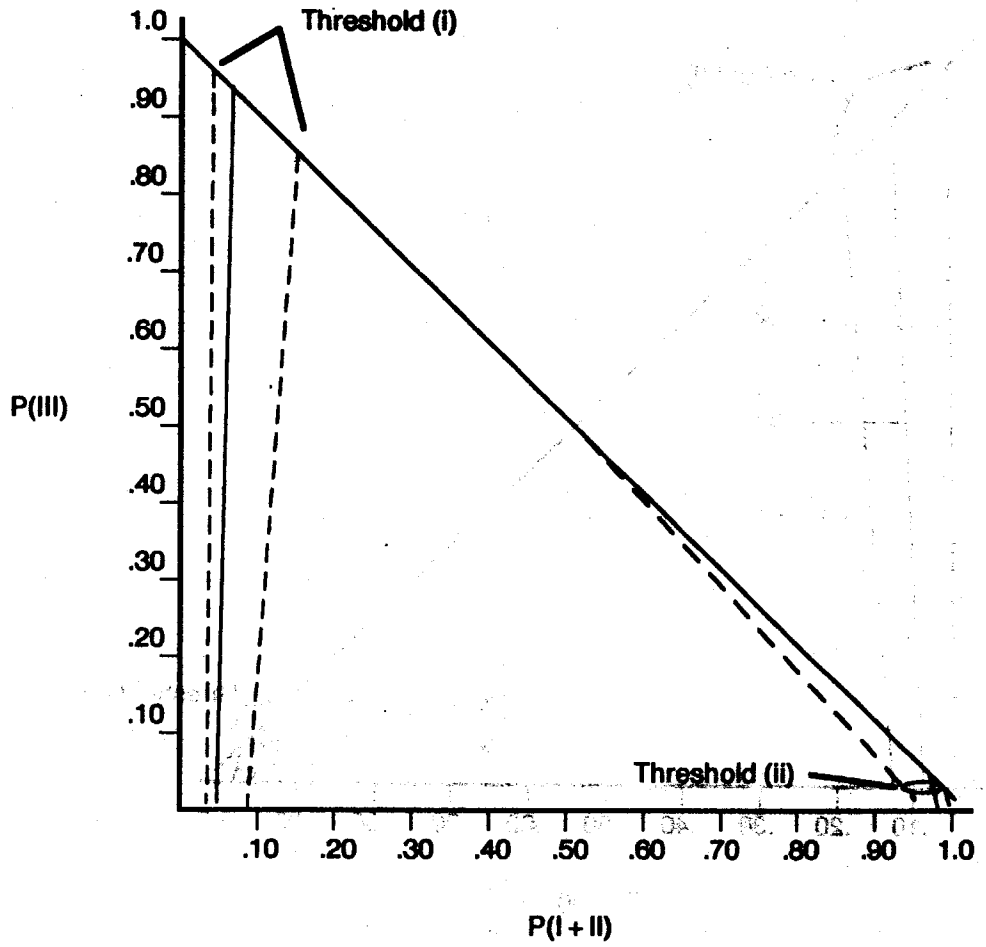
Values for varying LAP mortality are indicated.

---

laparotomy is shifted to the left. Conversely, when there is the least difference between results for each treatment, the threshold line is shifted to the right, making the value of laparotomy less. This "blurring" of the line is accentuated by increasing mortality of laparotomy (Figure 8.3).

If laparotomy has a false negative rate the decision threshold is shifted in such a way as to diminish its value. False negative rates of laparotomy of 1% and 5% are considered. In this context "false negative" refers to true stage III patients who are incorrectly called stage (I+II), and to stage IV patients mis-staged, with one quarter of the latter being called stage (I+II), and the remainder being called stage III.

Fig. 8.3. The effect on decision thresholds of 3% uncertainty in treatment results



The dotted lines on either side of the solid lines bound the areas included when 3% uncertainty in treatment results is taken into account. (LAP mortality is estimated as 1%.)

Analogous threshold lines can be drawn between laparotomy and immediate TNI; these mark off a triangular area where there is a very high probability of stages IB and IIB. We found, however, that there were no cohorts of symptomatic patients - not even those with the most favorable combinations of findings and negative lymhangiogram - whose probabilities fell within the area to the right of the laparotomy/TNI threshold, where immediate TNI would be the better option. This was so even when laparotomy mortality was increased to 3%, when uncertainty in therapy results was considered and when false negative rates were

considered for laparotomy.

8.2 Results

A point representing each of the fifty-six sets of probabilities of stage derived by the computer for symptomatic lymphangiogram-positive "patients" was plotted on a graph like those shown in figures 8.1-8.3 to determine which decision was appropriate for each cohort. These results are presented in tabular form. Table 8.2 shows results when the only variable affecting the threshold was the mortality of the staging laparotomy. With laparotomy mortality estimated at 1%, six of the 56 cohorts had such a high probability of late stage disease that the point representing their probabilities was plotted in the area where MOPP therapy could be given immediately. If the mortality of LAP was 2%, and the threshold line moved to the right (Figure 8.2), the "immediate MOPP area" included all male MC patients with positive lymphangiogram, together with several other cohorts which represented the less favorable findings for the other categories. If the laparotomy mortality was estimated to be 3%, all male patients with positive lymphangiogram are plotted in the "immediate MOPP area", together with more than half the cohorts of female patients considered.

Tables 8.2-8.5 show that the numbe of cohorts of patients for whom immediate MOPP was justified increased considerably, when some of the variables influencing the decision, namely the false negative rate for laparotomy and the uncertainty of the treatment results, were considered. The uncertainty in treatment results influenced the decision most, but considering false negative and false positive rates for laparotomy also increased the number of cohorts for whom immediate MOPP was indicated.

Table 8.2.
SYMPTOMATIC ADULT* PATIENTS WITH POSITIVE LYMPHANGIOGRAM FOR IMMEDIATE MOPP

LAP. MORT	HIST. TYPE	SEX	PATTERN OF NODAL INVOLVEMENT N-N- M+	N+N- M+	N+N+ M+	N+N- M-	N+N+ M-
1%	MC	M	-	>39	>39	>39	>39
2%	NS	M	>39	>39	>39	ALL	ALL
	NS	F	-	-	-	>39	>39
	MC	M	ALL	ALL	ALL	ALL	ALL
	MC	F	-	-	>39	>39	>39
3%	NS	M	ALL	ALL	ALL	ALL	ALL
	NS	F	-	>39	>39	>39	ALL
	MC	M	ALL	ALL	ALL	ALL	ALL
	MC	F	>39	>39	>39	ALL	ALL

*Adult - older than 11 years

>39 - refers to patients older than 39 years

M = Mediastinum N = Cervical lymph nodes

+ = involved with Hodgkins disease

- = not involved with Hodgkins disease

i.e. N+N+ = Bilateral cervical node involvement
N-N- = Cervical nodes not involved
N-N+ = Unilateral cervical node involvement - right- or left-sided.

8.3 Discussion

From basic clinical information and the results of the lymphangiogram we have been able to predict the likelihood of the various pathologic stages in symptomatic HD, and to compare the value of immediate chemotherapy with that of laparotomy in a group of patients who have a low probability of early stage disease. With the simple decision-making techniques used here it is possible to analyze the effect of three important variables on the decision to perform laparotomy - its mortality for a given patient, the validity of the results of laparotomy and the reliability of treatment results on which management decisions depend. Even with the conservative figures we have used for these variables, this approach could save an appreciable number of

Table 8.3.

SYMPTOMATIC ADULT* PATIENTS WITH POSITIVE LYMPHANGIOGRAM FOR IMMEDIATE MOPP IF THERE IS 3% UNCERTAINTY IN TREATMENT RESULTS

LAP. MORT	HIST. TYPE	SEX	PATTERN OF NODAL INVOLVEMENT N-N- M+	N+N- M+	N+N+ M+	N+N- M-	N+N+ M-
1%	NS	M	-	>39	>39	ALL	ALL
	NS	F	-	-	-	>39	>39
	MC	M	ALL	ALL	ALL	ALL	ALL
	MC	F	-	-	>39	>39	>39
2%				ALL			
3%				ALL			

For key to table, see Table 8.2

Table 8.4.

SYMPTOMATIC ADULT* PATIENTS WITH POSITIVE LYMPHANGIOGRAM FOR WHOM IMMEDIATE MOPP IS THE BETTER OPTION WHEN LAPAROTOMY IS CONSIDERED TO HAVE A 1% FALSE NEGATIVE RATE

LAP. MORT	HIST. TYPE	SEX	PATTERN OF NODAL INVOLVEMENT N-N- M+	N+N- M+	N+N+ M+	N+N- M-	N+N+ M-
1%	NS	M	-	-	-	-	>39
	MC	M	-	>39	>39	>39	>39
2%	NS	M	>39	>39	ALL	ALL	ALL
	NS	F	-	-	>39	>39	>39
	MC	M	ALL	ALL	ALL	ALL	ALL
	MC	F	-	>39	>39	>39	>39
3%	NS	M	ALL	ALL	ALL	ALL	ALL
	NS	F	>39	>39	>39	ALL	ALL
	MC	M	ALL	ALL	ALL	ALL	ALL
	MC	F	>39	>39	ALL	ALL	ALL

For key to table, see Table 8.2

patients the rigors, risks and expense of a LAP, and of the other tests frequently performed in routine staging.

In certain patients the risk of laparotomy mortality may well be higher than 1%. Immediate postoperative mortality for 2345 patients from 27 different series was 0.5% (Table 1.3); none of these series reported

Table 8.5.

SYMPTOMATIC ADULT* PATIENTS WITH POSITIVE LYMPHANGIOGRAM
FOR WHOM IMMEDIATE MOPP IS THE BETTER OPTION
WHEN LAPAROTOMY IS CONSIDERED TO HAVE A 5% FALSE NEGATIVE RATE

LAP. MORT	HIST. TYPE	SEX	PATTERN OF NODAL INVOLVEMENT N-N- M+	N+N- M+	N+N+ M+	N+N- M-	N+N+ M-
1%	NS	M	-	>39	>39	>39	ALL
	MC	M	>39	>39	ALL	ALL	ALL
	MC	F	-	-	-	-	>39
2%	NS	M	ALL	ALL	ALL	ALL	ALL
	NS	F	-	>39	>39	>39	ALL
	MC	M	ALL	ALL	ALL	ALL	ALL
	MC	F	>39	>39	ALL	ALL	ALL
3%	NS	M	ALL	ALL	ALL	ALL	ALL
	NS	F	>39	>39	ALL	ALL	ALL
	MC	M	ALL	ALL	ALL	ALL	ALL
	MC	F	>39	ALL	ALL	ALL	ALL

For key to table see legend for Table 8.2

---

age-specific rates, but there may be wide variation, with much lower rates in younger patients and higher rates in older patients, or those with other complicating factors. In addition to post-operative mortality, there may also be a delayed mortality due to the fulminant sepsis known to occur in a small proportion of splenectomized patients; this usually occurs within 2 years of splenectomy (Krivit, 1977). When Desser and Ultmann (1973) collated results from 1190 patients splenectomized as part of staging laparotomy, they found that late death from overwhelming sepsis, unrelated to treatment-induced leucopenia or other precipitating cause, occurred in 0.5%. Other series report higher rates of post-splenectomy sepsis; for example Singer (1973), reviewing 2975 splenectomized patients, found the overall late death rate from infection was 2.52%; even patients who had splenectomy incidental to another operation had an appreciable mortality from this complication (0.86%). Furthermore, susceptibility to infection after splenectomy may be exacerbated by treament-induced marrow suppression, (Schimpff & O'Connell, 1977), and by the defects in cellular immunity frequently noted in patients with HD (Desforges et al, 1979).

In this study, the mortality of laparotomy is considered to be its only disadvantage, and one must remember its other liabilities. In addition to mortality, 6.7% of the 2345 patients in Table 1.3. had serious complications of the staging laparotomy; these included such major post-operative problems as subphrenic abscess, pulmonary embolism or intestinal obstruction requiring reoperation.

Uncertainty about results of treatment should influence therapeutic decisions. For example there is no firm evidence that any treatment has a particular advantage for stages IB and HB HD; recent trials (Goodman et al 1977, Rosenberg et al 1978) have not demonstrated a statistically significant difference between TNI and combined modality treatment. All reported series of pathologically-staged IIB patients contain less than 25 patients in any particular treatment group (Aisenberg & Qazi, 1976, Fazekas et al 1975, Fuller & Madoc-Jones, 1977, Goodman et al 1977, Rosenberg et al 1978, Stoffel & Cox, 1977). In these small groups results for TNI vary between 79% and 48% 5 year disease free survival. For a series of 25 patients, 95% confidence limits for a disease free survival of 79% range between 90% and 60%, and when disease free survival is 48% they range between 72% and 27% (Greenwood & Hartley, 1962). The uncertainty rate we have used (3%) in these calculations is, therefore, a very conservative estimate for uncertainty in TNI results. Even this modest uncertainty, however, made a considerable difference to the numbers of patient cohorts for whom immediate treatment was appropriate. The MOPP chemotherapy regime for treating disseminated disease has large series (Aisenberg & Qazi, 1976, DeVita et al, 1980) reported with long-term follow-up and much narrower confidence limits; there is, however, little information on its use in localized disease. Some data on the use of MOPP for early stages has come from Uganda (Olweny et al, 1978), but this patient group is scarcely comparable, since their patients were predominantly young boys, who had minimal clinical staging; moreover, separate data for A and B patients is not given. Our figures for MOPP for stage IB and IIB are estimates, therefore, and subject to uncertainty, as are the estimates for TNI for disseminated disease.

A further complicating factor in the decision is the accuracy of laparotomy. Although physicians use the results of this operation as the final answer, there is likely to be a false negative rate, due to sampling error.

This sampling error is difficult to quantitate; Cotman et al (1977) showed that 18 of 88 Hodgkins disease patients still had lymphographically-abnormal nodes after LAP. A 1% false negative rate seems realistic for a surgeon experienced in this type of surgery; a 5% false negative rate may be appropriate for a surgeon not experienced in staging laparotomy.

In spite of these uncertainties in risks and outcome, decisions regarding staging and subsequent management must be made. Tables 8.2-8.5 give guidelines for patients who may proceed directly to treatment with MOPP, with provision to include the possible variables. The same principles of threshold analysis can be applied even if different treatments are used. Probabilities of pathological stage for a given patient can be obtained from Appendix 1; Bayes Rule can be used to incorporate the results of any test, using false positive and false negative rates appropriate to the institution in which they have been performed. New decision thresholds can be derived by substituting the new treatment results in the threshold equations.

9. Problems in Applying Decision Analysis to Hodgkins Disease

There are some limitations in the Hodgkins disease decision analysis program which may restrict its usefulness.

Utilities have posed the major problem. When we wish to evaluate our plans in terms of five year disease-free survival (5 yr DFS) we have had difficulty in finding large series of patients for whom disease-free survival data is available. All too often the series are small, with relatively few patients followed through to the five year stage, so that the data has considerable uncertainty. For some stages and treatments there is no disease-free survival information available, for example for chemotherapy in early stage disease or for radiotherapy in advanced disease. These values, therefore, have had to be estimated.

A further problem with the utilities of plans is that there is no reasonable way to "trade-off" or compare morbidity data with disease-free survival data. In the program these are, therefore, simply quantitated and expressed separately. Even comparing the different types of mortality is a problem. In our plans the "cost" of a post-operative death from laparotomy is balanced against survival disease-free at five years, where the latter is used to mean cure. It seems certain, however, that most patients would not regard immediate death and relapse at 5 years as having equal "cost".

The program is also not useful in those cases with unusual presentations. Since it is dependent on data to calculate both the prior and the conditional probabilities of stage, unusual presentations are poorly represented and we have little confidence in probabilities calculated from these data because of sampling error. The management of such presentations as bilateral axillary node involvement alone, or of inguinal or other isolated infra-diaphragmatic involvement is not clear cut. For this reason these are just the type of patient for whom the physician would like some guidance about the likely spread of disease, yet because of the relative rarity of such cases our program does not have sufficient data to give reliable probabilities of stage. The same problem arises with patients with the less common histological subtypes - lymphocyte

depleted and lymphocyte predominant.

There is also a problem with histologic interpretation. Our prior probabilities of stage are quite strongly influenced by the histological subtype of the HD. There is, however, considerable variation amongst pathologists in their reporting of the HD histology within the Rye histologic classification (Jones et al, 1977). Since our testing of the database showed that the program made accurate predictions of stage, we have concluded that this uncertainty in assigning histologic subtypes is not of major importance to our program.

10. The Computer Implementation

The goal of this chapter is to reveal enough of the workings of the Hodgkins disease program to allow an interested programmer to reproduce it. The following topics will be addressed:

1. representations of knowledge about Hodgkins disease, including stages, clinical findings, tests, and treatments, and probabilities
2. data structures for maintaining a description of a patient
3. data structures for representing decision trees
4. calculating probabilities of stage
5. growing decision trees and making decisions
6. subsidiary programs for displaying and modifying program data
7. interactive features to deal with the twin problems of low band-width data transmission and use of typewriter input by non-expert typists
8. additional features: threshold analysis, salvage analysis

10.1 Introduction

The Hodgkins disease computer system consists of two independent subsystems: a computer program for performing decision analysis in the management of patients with Hodgkins, and a database system for managing information about 2000 individual Hodgkins disease patients. Both subsystems are large, interactive Lisp programs devoted to ease and rapidity of use. The operation of these programs from the user's point of view has been discussed in other sections. This section deals with the details of the implementation of the decision analysis subsystem.

10.2 Maclisp as an Implementation Language

Several features of LISP, and of the Maclisp dialect in particular, facilitated the development of the Hodgkins disease program. LISP provides data structures appropriate to representing complex structures such as decision trees and diagnostic plans, namely *s-expressions*. Maclisp, together with the text editor Emacs, provides a remarkable interactive programming environment, with excellent debugging facilities. The LISP language, augmented by a macro package for iteration, provides control structures appropriate to creating and manipulating complex, repetitive, and recursive structures. The flexible input and output capabilities of Maclisp allow the easy development of dialog between computer and user.

10.3 Data Structures

The Hodgkins disease decision analysis program operates on three types of information: knowledge about Hodgkins disease, descriptions of the patient under consideration, and decision trees.

10.3.1 Knowledge about Hodgkins Disease

The program has fixed data structures representing its Hodgkins disease knowledge. The program's knowledge of Hodgkins disease consists of:

1. identifiers for important aspects of the disease, including symptoms, tests, and treatments (e.g. an important aspect of a particular patient's Hodgkins disease is its "symptomatology" -- whether the patient is asymptomatic (A) or symptomatic (B));
2. qualitative relationships between these aspects of the disease (e.g. the program contains information denoting the fact that symptomatology is an important clinical finding in Hodgkins disease); and
3. quantitative (probabilistic) relationships between aspects of the disease (e.g. the probability of positive left neck nodes in a symptomatic stage III patient is 0.75).

This knowledge is stored partly as the values of free (or global) variables in the Lisp environment, and partly as property lists. Free variables are used mainly to give names to useful lists of items, such as the list of

clinical findings or the list of tests. Examples of such free variables, and their default values, are shown below. The names of the variables are on the left, their values, in conventional LISP list notation, are on the right.

Variable	Value
`*prior-findings*`	`'(histologic-subtype symptomatology sex)`
`*conditional-clinical-findings*`	`'(age left-neck right-neck mediastinum spleen-size liver-scan)`
`*clinical-findings*`	`(append *prior-findings* *conditional-clinical-findings*)`
`*tests*`	`'(gal lag bmbx lbx)`
`*all-findings*`	`(append *clinical-findings* *tests*)`
`*treatments*`	`'(total-nodal mopp extended-mantle combined-modality)`
`*stages*`	`'(i+ii iii iv)`

These lists may be modified under user command. For example, if the user wants to have the program consider a treatment different from the four included in `*treatments*`, he invokes a request which runs a program which adds the name of the new treatment to the list `*treatments*`.

Property lists are used to maintain detailed information about each of the findings, tests, and so on. For a clinical finding, this information includes the list of *values* of the finding (e.g. `SEX` may be one of `(MALE FEMALE)`), plus directions to the program about how to ask about the sex of the patient and how to display it to the user. Such information is stored in a disk file, from which it is loaded when a new version of the Hodgkins program is constructed. For example, the information about SEX appears in a text file as the s-expression below:

```
(*load sex
      (kind clinical-specific)
      (results (male female))
      (inquire (|What is the sex of the patient?|)))
```

This expression is evaluated, to give SEX the property list shown below. (The "INQUIRE" and "KEYLST" properties provide the program with information for asking the user questions about the sex of the given patient.)

```
(RESULTS          (MALE FEMALE)
 KIND             CLINICAL-SPECIFIC
 INQUIRE          (|What is the sex of the patient?|)
 KEYLST           ((MALE NIL KEYSTRING |Male| RETURN 'MALE)
                   (FEMALE NIL KEYSTRING |Female| RETURN 'FEMALE)
                   (UNKNOWN NIL KEYSTRING |Unknown| RETURN 'UNKNOWN)
                   (|*WHAT IF ... ?| NIL KEYSTRING |*What if ... ?|
                         RETURN 'WHAT-IF)
                   (|*FINDINGS SO FAR| NIL KEYSTRING |*Findings so far|
                                 RETURN 'FINDINGS-SO-FAR)
                   (|*DISPLAY DATA| NIL KEYSTRING |*Display data|
                        RETURN DISPLAY-DATA)))
```

For a test, the detailed information consists of the *results* of the test, the *cost* of the test in terms of mortality, morbidity, and money, and the *conditional probabilities* relating test results and stage. For a treatment, it includes numerical values for survival values, morbidity, and dollar cost.

10.4 Patient Description

Throughout a decision analysis session, the program maintains a description of the patient, which is represented as the property list of the atom PATIENT-MODEL. It has entries for the known values of clinical findings (e.g. SEX=MALE), for the known results of tests previously carried out, and for the current estimates of stage probabilities. Once the decision analysis has been carried out, the set of recommended diagnostic plans is added to the patient description. For example, the patient described in the first paragraph of the last chapter would be represented internally as the property list below:

```
(PROBABILITIES              (0.486 0.485 0.029)
 A-OR-B                     B
 HISTOLOGIC-SUBTYPE         NS
 SEX                        MALE
 AGE                        11<AGE<40
 LEFT-NECK                  POSITIVE-LEFT-NECK
 RIGHT-NECK                 NEGATIVE-RIGHT-NECK
 MEDIASTINUM                NEGATIVE-MEDIASTINUM
 SPLEEN-SIZE                NORMAL-SIZE-NO-DEFECTS
 LIVER-SCAN                 NORMAL-LIVER-SCAN)
```

10.5 Probability Calculations

The program uses Bayes Rule to calculate stage probabilities from the patient description. For reasons discussed elsewhere in this article, the three findings SEX, HISTOLOGIC-SUBTYPE, and SYMPTOMATOLOGY were selected as the basis on which to estimate the *prior* probabilities. Once the program has asked for the values of each of these findings, it looks up the initial probabilities in a table indexed by the findings.

The program then asks for the other clinical findings and, using Bayes Rule, updates the probability estimates after each answer. Assumptions of conditional independence were made where intuitively reasonable and not contradicted by statistical data. Dependence was incorporated in two ways: by clumping dependent findings and by explicitly conditioning certain conditional probabilities on the values of other findings. Clumping was used in creating the group of findings for initial probabilities. Explicit conditioning was used to deal with the dependence of several findings on the symptomatology of the patient. Further, several of the findings were thought to be tests of more fundamental findings such as liver or spleen involvement. In these cases, a two stage version of Bayes Rule was used, as described in Section 2.1.

Our implementation of Bayes Rule in Maclisp calculates the *list* of posterior probabilities from the conditionals and the list of priors, following the mathematical statement of Bayes Rule, shown below for the probability of stage I+II given positive right neck nodes (+RN):

$$P(I+II|+RN) = \frac{P(+RN|I+II)*P(I+II)}{P(+RN|I+II)*P(I+II) + P(+RN|III)*P(III) + P(+RN|IV)*P(I+IV)}$$

This calculation is carried out by the function **Bayes**, whose definition in Lisp appears below:

```
(defun Bayes (finding prior)
       (unitize                    ; normalize the list
        (for prob in prior as stage in '(I+II III IV)
             collect
             (times (Conditional-Probability-of finding stage) prob)))))
                                    ;of the products of priors and conditionals
```

For example, if the prior probabilities are '(.3 .3 .4) corresponding to stages '(I+II III IV) in a B

patient with a certain set of clinical findings, then the list of posterior probabilities given that the patient has a positive right neck node is `(bayes '+RN '(.3 .3 .4))`.

The calculation of this quantity proceeds as follows:

1. `finding` is bound to `'+RN`
2. `prior` is bound to `'(.3 .3 .4)`
3. For each `prob` in `prior` corresponding to the stages, the following product is calculated:

   `P(+RN|stage) * prob`

   The list of these products is constructed, and then normalized to have sum 1.0 by the function `unitize`. This is exactly equivalent to dividing by the usual denominator in the expression for Bayes Rule, since that denominator is just the sum of the unnormalized products.

10.6 Decision Trees and Diagnostic Plans

Once all of the known information about the patient has been used to calculate the stage probabilities, the program constructs diagnostic plans for the patient from the available tests and treatments. A diagnostic plan is represented by a recursive list structure according to the following BNF-like syntax:

```
Plan := (Utility-Summary Utility-Data Branches)
Utility-Summary := (Utility . Action)
Utility is the numeric expected utility of the plan
Utility-Data is a breakdown of utility by attributes[1]
Action := Treatment | Test
Treatment := TNI | MOPP | EM
```

1. Though space in the structure has been left for this item, the current version of the program does not use multiple utility attributes except when summarizing an entire plan at top-level. The space is filled with the atom 'utility-data-place-holder.

```
Test := BMBX | LAG | LBX | GAL | LAP

Branches := (Branch) | (Branch . Branches)

Branch := (Test-Result Probability-of-Result Posteriors Plan)
```

`Test-Result` is a result of specified test. E.g. if test is `BMBX`, then `Test-Result` may be either `+BMBX` or `-BMBX`.

`Probability-of-Result` is a numerical value representing the probability that the result will occur, given the prior probabilities at that point in the tree.

`Posteriors` is the list of probabilities after the result of the test has been taken into account

`Plan` (second appearance) is the plan which is optimal if the given test result occurs

For example, the plan shown below is represented internally as the list structure following it.

```
Plan  1: The following diagnostic plan has an estimated DFS of 0.5873 for an
A patient with probabilities: I+II 0.0337 III 0.3716 IV 0.5946

                    / 0.04 I+II EM
     / 0.89 - LAP   | 0.42 III  TNI
     |              \ 0.54 IV   MOPP
BMBX |
     |
     \ 0.11 + MOPP
```

Internal representation:

```
((0.5872 . BMBX)
 utility-data-place-holder[2]
 ((-BMBX 0.8886
         (0.0379 0.4182 0.5438)
         ((0.5920 . LAP)
          utility-data-place-holder
          ((PATH-STAGE-I+II 0.0380
                            (1.0 0.0 0.0)
                            ((0.82 . EXTENDED-MANTLE)))
           (PATH-STAGE-III 0.4183
                           (0.0 1.0 0.0)
                           ((0.64 . TOTAL-NODAL)))
           (PATH-STAGE-IV 0.5438
                          (0.0 0.0 1.0)
                          ((0.55 . MOPP))))))
  (+BMBX 0.1115 (0.0 0.0 1.0) ((0.55 . MOPP)))))
```

The plans are created by exhaustively tracing a large decision tree of possibilities using a computer

2. See previous footnote.

implementation of "averaging out and folding back". The initiator of tree analysis is the function `Expand-Top-Level-Choice-Node`, which sprouts subtrees beginning with each of the possible treatments (listed in `*treatments*`), the special test `LAP`, and other tests not already done or otherwise contraindicated (`tests-left`), and evaluates the plan optionally supplied by the user.

```
(defun Expand-Top-Level-Choice-Node (tests-left prior lap-done?)
       (append
        ;Treatment plans
        (for treatment in *treatments*
             collect
             (ncons (cons (Treatment-E-U treatment prior) treatment)))
        ;Lap plan
        (ncons (Lap-Plan prior))
        ;Test plans
        (for tests on tests-left as test = (car tests)
             collect (Expand-Chance-Node test prior (cdr tests)))
        ;User plan
        (and (get 'controller 'user-plan-form)
             (ncons (Instantiate-Plan plan-form prior)))))
```

`Expand-Chance-Node` is called by the top-level node expander to grow the subtree beginning with a particular test. It constructs branches for each of the possible results of the tests, expands each branch, and calculates the expected utility for the subtree. It returns to the calling program the expected utility and the structure of the resulting subtree.

```
(defun Expand-Chance-Node (test prior tests-left)
       (for res in (test-results test)
            bind (cond-prob 0.0) posterior branch (utility 0.0)
                 (new-tests (cdr tests-left))
            collect
            (setq cond-prob (Conditional-Probability-of res prior)
                  posterior (Bayes res prior)
                  branch (if (eq test 'lap)
                             (Choose-Best-Treatment posterior)
                             (Expand-Choice-Node posterior new-tests))
                  utility (plus utility (times cond-prob (caar branch))))
            (list res cond-prob posterior branch)
            returning (list (cons (Account-For-Mortality test utility) test)
                            'utility-data-place-holder
                            result)))
```

`Expand-Choice-Node` constructs a subtree for each remaining test, beginning with that test. It evaluates each subtree, and returns to the calling program the subtree with the highest expected utility.

```
(defun Expand-Choice-Node (prior tests-left)
       (if (null tests-left)
           (Treatment-vs-Lap prior)
           (for tests on tests-left as test = (car tests)
                bind (best-treatment (Treatment-vs-Lap prior))
                choose-best (Expand-Chance-Node test prior tests-left)
                according-to '(lambda (x y) (lessp (caar x) (caar y)))
                returning (if (lessp (caar result) (caar best-treatment))
                              best-treatment
                              result))))

(defun Treatment-vs-Lap (prior)
       (let ((lap-plan (Lap-Plan prior))
             (best-treatment-plan (Choose-Best-Treatment prior)))
            (if (greaterp (:utility lap-plan) (:utility best-treatment-plan))
                lap-plan
                best-treatment-plan)))

(defun Choose-Best-Treatment (prior)
       (for treatment in *treatments*
            choose-best (cons (Treatment-E-U treatment prior) treatment)
            according-to '(lambda (p1 p2) (lessp (car p1) (car p2)))
            returning (ncons result)))
```

For an example of a program trace showing these programs being run to create a diagnostic plan, see Appendix 4.

10.7 Interactive Features

We paid particular attention to the interactive features of the program for two reasons. First, our "customer" is typically a doctor with little expertise in typing, so we want to minimize required input to the program. Secondly, access to the timeshared computer which runs the Hodgkins program is via slow-speed (30 or 120 char/sec) dial-up lines, hence it is important for output to be efficient as well as informative.

10.7.1 Input

To minimize typing without using graphical devices, we use *command completion* and *tabular data entry*. The former capability is used when asking the user a menu-selection question, a question with a fixed set of possible answers. The user typically has to type only one or two characters to uniquely determine a response

-- the program supplies the remaining characters. A simplified version of INTERLISP's ASKUSER package was used to implement this feature.

Tabular data entry is invoked when the user is required to supply or modify a table of numbers. For example, if the user wants to change the survival values for Hodgkins disease treatments, the computer displays the values as shown below:

	I+IIA	I+IIB	IIIA	IIIB	IVA	IVB
TOTAL-NODAL	.81	.7	.64	.25	.1	.05
MOPP	.7	.6	.62	.5	.55	.35
EXTENDED-MANTLE	.82	.62	.4	.15	.0	.0
CM	.87	.87	.85	.55	.55	.35

This feature allows the user to "edit" a table of numbers using a "real-time" editor. The user moves a pointer or cursor up or down or across the table by means of single character commands, and then changes the number at the pointer by typing in the new value.

For example, suppose the user wants to change the 5-year DFS figure for MOPP in IV A disease from 0.55 to 0.57. When the table is presented to the user, the cursor is in the upper left hand corner. Then the user types "D" to move the cursor to the second row and "R" 4 times to move the cursor to the fifth column. Then he simply types the new number over the old one. The new entry becomes permanent when the user types any character that is not a digit, "+", "-", or ".".

10.7.2 Output: Display of Diagnostic Plans

The primary output of the program is the diagnostic plan for the patient. In order to achieve the multiple goals of ease of comprehension, efficiency of display, and completeness of information, we employ two distinct representations of diagnostic plans: branching tree and outline format.

The branching tree format has been presented in figures throughout the paper, e.g. in section 11.6. The plan in that section beginning with BMBX is displayed in outline format as follows:

```
Perform BMBX
 1.1 -BMBX with P = 0.89. Posteriors = 0.04 0.42 0.54 . Then:
  Perform LAP
   1.1.1 PATH-STAGE-I+II with P = 0.04. Posteriors = 1.00 0.00 0.00.
         Then: Perform EXTENDED-MANTLE
   1.1.2 PATH-STAGE-III with P = 0.42. Posteriors = 0.00 1.00 0.00.
         Then: Perform TOTAL-NODAL
   1.1.3 PATH-STAGE-IV with P = 0.54. Posteriors = 0.00 0.00 1.00.
         Then: Perform MOPP
 1.2 +BMBX with P = 0.11. Posteriors = 0.00 0.00 1.00.
         Then: Perform MOPP
```

The advantage of this notation is that it can provide more detailed information than the branching tree format.

In its Lisp implementation, the outline display method is somewhat simpler than the branching tree format. The former utilizes a simple depth-first algorithm to traverse the tree in combination with a simple text formatting program which ensures that lines do not exceed the width of the screen. The latter, because it must satisfy more constraints, uses a more complicated algorithm to ensure that a given tree will fit on a screen both horizontally and vertically.

10.8 Additions to the Basic Hodgkins System

The basic modules of the Hodgkins system acquire the patient description, calculate probabilities of stage, and construct, evaluate and display decision trees. A number of features were added to the basic Hodgkins disease to perform other tasks associated with the decision analysis.

10.8.1 Tree Acquisition

Occasionally, the user will want to evaluate a diagnostic plan of his own choosing, to see how it stacks up against the plans selected by the program. We provide a facility to allow the user to enter the structure of his plan into the computer, so that the plan may be evaluated and displayed along with the program's selections.

The program requires the user to enter the plan depth first, using display formats similar to those used for

output of program-generated plans. The program begins by asking the user for the first test in the plan. Given the test, and its known possible results, the program can display the first level of a diagnostic plan, showing the test and branches corresponding each of the results. For each of the result branches in turn, the program asks for the action -- test or treatment -- to take next. If the action is a treatment, the branch is terminated. Otherwise, the action is a test, so the program recursively asks about each of its results. Such questioning continues until all leaves of the tree are filled in.

To integrate this facility into the remainder of the decision analysis system, there are program to calculate the utility of a prespecified diagnostic plan. Each time plans are constructed and compared, the utility of the user-supplied plan is calculated and ranked with the others.

10.8.2 Sensitivity and Threshold Analysis

The program contains facilities for performing *sensitivity* and *threshold analysis*. Sensitivity analysis is the study of the effects of small perturbations of the data on the results of decision analysis. For example, the expected utility of doing laparotomy depends on the cost, in terms of mortality, of subjecting the patient to the operation. Sensitivity analysis can determine what diagnostic plans would be recommended for mortality estimates ranging below and above the current estimate.

Sensitivity analysis is implemented as an additional top-level program on the Hodgkins disease system. This top-level program varies selected parameters and calculates the optimal plan for each value. Parametric variation may be of two types: numeric and non-numeric. In numeric variation, numeric parameters such as conditional probabilities are varied over a range. For example, the conditional probability of right-neck involvement might be caused to range from 0.4 to 0.6 to see what effect this has on the test/treatment decision. In non-numeric variation, the value of a finding is caused to vary over all possible values to see how the finding affects a decision.

11. Acknowledgments

We gratefully acknowledge the cooperation of Dr. Alan Aisenberg (Massachusetts General Hospital) Dr. Samuel Hellman (Harvard Joint Center for Radiation Therapy), Dr. Henry Kaplan (Stanford Division of Radiation Therapy), and Dr. Nancy Gutensohn (Harvard School of Public Health) who have allowed their patient data to be used in the computer database.

We also wish to thank Ms. Eileen O'Brien and Ms. Rena Angelo for their meticulous work in their capacity as research assistants. The early work on this program owes much to Dr. A.Z. Bluming, Dr. P. Tsichlis, and particularly to Dr. C.S. Safran whose continuing interest in the program has benefitted us all.

We are very grateful to Professor Peter Szolovitz and Dr. Ramesh Patil of the Laboratory for Computer Science for their advice and for their practical assistance in the preparation of this Technical Report.

We thank the Editor of the British Journal of Haematology for allowing us to reprint Figures 8.1-8.3. These figures initially appeared in a paper in this journal (Rutherford et al, 1980).

We acknowledge financial support from the National Cancer Institute Grant CA 19122, and the Division of Research Resources Grant 1 P41 RR 01096-02, both from the National Institutes of Health.

Appendix I - Comprehensive List of Probabilities for Hodgkins Disease Patients

A large amount of information important to decision-making in Hodgkins disease has been compiled as the basis of the decision analysis computer program. In order to make this information available to those who may not have access to our program, we present the following tables of probabilities of stage for:

1. PRIOR PROBABILITIES based on the three interdependent findings of sex, histological subtype and the presence (or absence) of symptoms.

2. CONDITIONAL PROBABILITIES for finding|stage which are used to modify these prior probabilities. Conditional probabilities are given for both the clinical findings, and the test results.

3. POSTERIOR PROBABILITIES for all combinations of the clinical findings.

N.B. Although probabilities of pathologic stage are given for all histologic subtypes and for all age groups, the numbers of patients in the "less than 12 years" age group, and with the "lymphocyte predominant" and "lymphocyte depleted" histologic subtypes are limited, so probabilities involving these parameters have considerable uncertainty. It must also be realized that these probabilities apply only to supra-diaphragmatic presentations of Hodgkins disease.

The following abbreviations have been used:

A = Asymptomatic

B = Symptomatic

NS = Nodular sclerosis

MC = Mixed cellularity

LD = Lymphocyte depleted

LP = Lymphocyte predominant

M = Male

F = Female

+ = Positive

- = Negative

LN = Left cervical lymph nodes

RN = Right cervical lymph nodes

MED = Mediastinum

1. PRIOR PROBABILITIES - for interdependent findings HISTOLOGY/SYMPTOMS/SEX

			STAGE I+II	STAGE III	STAGE IV
NS	A	FEMALE	0.72	0.27	0.01
NS	A	MALE	0.66	0.29	0.05
NS	B	FEMALE	0.56	0.37	0.08
NS	B	MALE	0.34	0.50	0.16
MC	A	FEMALE	0.55	0.36	0.09
MC	A	MALE	0.44	0.47	0.09
MC	B	FEMALE	0.50	0.56	0.04
MC	B	MALE	0.20	0.44	0.35
LP	A	FEMALE	0.80	0.18	0.02
LP	A	MALE	0.73	0.21	0.06
LP	B	FEMALE	0.48	0.41	0.11
LP	B	MALE	0.35	0.39	0.26
LD	A	FEMALE	0.41	0.47	0.13
LD	A	MALE	0.28	0.43	0.29
LD	B	FEMALE	0.35	0.48	0.17
LD	B	MALE	0.23	0.42	0.35

2. CONDITIONAL PROBABILITIES FOR FINDING|STAGE

AGE			
Younger than 12 years	I+II 0.03	III 0.06	IV 0.08
12 to 39 years	I+II 0.84	III 0.74	IV 0.45
Older than 39 years	I+II 0.14	III 0.23	IV 0.47
LEFT NECK NODAL INVOLVEMENT			
LN+	I+II 0.68	III 0.75	IV 0.73
LN-	I+II 0.32	III 0.25	IV 0.27
RIGHT NECK NODAL INVOLVEMENT			
RN+	I+II 0.47	III 0.49	IV 0.70
RN-	I+II 0.53	III 0.51	IV 0.30
MEDIASTINAL NODAL INVOLVEMENT			
MED+	I+II 0.62	III 0.53	IV 0.53
MED-	I+II 0.38	III 0.47	IV 0.47
SPLEEN SCAN			
Normal	I+II 0.93	III A 0.64 III B 0.66	IV 0.49
Enlarged >15 cm	I+II 0.07	III A 0.28 III B 0.29	IV 0.32
Filling defects	I+II 0.00	III A 0.09 III B 0.10	IV 0.11
Enlarged with filling defects	I+II 0.00	III A 0.06 III B 0.07	IV 0.08
LIVER SCAN			
-ve	I+II 0.23	III 0.23	IV 0.55
+ve	I+II 0.77	III 0.77	IV 0.45
PERCUTANEOUS LIVER BIOPSY			
-ve	I+II 1.00	III 1.00	IV 0.80
+ve	I+II 0.00	III 0.00	IV 0.20

LIVER BIOPSY VIA PERITONEOSCOPY

-ve	I+II 1.00	III 1.00	IV 0.39
+ve	I+II 0.00	III 0.00	IV 0.61

BONE MARROW BIOPSY

-ve	I+II 1.00	III 1.00	IV A 0.81
			IV B 0.56
+ve	I+II 0.00	III 0.00	IV A 0.19
			IV B 0.44

GALLIUM SCAN (Abdominal nodes)

-ve	I+II 0.90	III A 0.65	IV A 0.61
		III B 0.56	IV B 0.54
+ve	I+II 0.10	III A 0.35	IV A 0.39
		III B 0.44	IV B 0.46

LYMPHANGIOGRAM

+ve	I+II 0.18	III A 0.55	IV A 0.60
		III B 0.57	IV B 0.70
equivocal	I+II 0.23	III A 0.15	IV A 0.14
		III B 0.12	IV B 0.12
-ve	I+II 0.59	III A 0.20	IV A 0.26
		III B 0.21	IV B 0.18

Probability	(I+II)	III	IV
ASYMPTOMATIC (A) PATIENTS			
NS M AGE<12			
LN+RN+MED+	0.5	0.37	0.13
LN+RN+MED-	0.41	0.43	0.16
LN+RN-MED+	0.56	0.38	0.06
LN+RN-MED-	0.47	0.46	0.07
LN-RN+MED+	0.58	0.3	0.12
LN-RN+MED-	0.49	0.37	0.15
LN-RN-MED+	0.64	0.31	0.05
NS M 11<AGE<40			
LN+RN+MED+	0.69	0.27	0.04
LN+RN+MED-	0.61	0.34	0.05
LN+RN-MED+	0.73	0.26	0.01
LN+RN-MED-	0.65	0.33	0.02
LN-RN+MED+	0.76	0.21	0.03
LN-RN+MED-	0.69	0.27	0.04
LN-RN-MED+	0.79	0.2	0.01
NS M 39<AGE			
LN+RN+MED+	0.49	0.35	0.16
LN+RN+MED-	0.4	0.41	0.19
LN+RN-MED+	0.56	0.37	0.07
LN+RN-MED-	0.47	0.45	0.08
LN-RN+MED+	0.57	0.29	0.15
LN-RN+MED-	0.47	0.35	0.18
LN-RN-MED+	0.64	0.3	0.06
NS F AGE<12			
LN+RN+MED+	0.6	0.36	0.04
LN+RN+MED-	0.51	0.45	0.05
LN+RN-MED+	0.63	0.35	0.02
LN+RN-MED-	0.54	0.44	0.02
LN-RN+MED+	0.67	0.29	0.03
LN-RN+MED-	0.59	0.37	0.04
LN-RN-MED+	0.71	0.28	0.01
NS F 11<AGE<40			
LN+RN+MED+	0.75	0.24	0.01
LN+RN+MED-	0.67	0.31	0.01
LN+RN-MED+	0.77	0.23	0.0
LN+RN-MED-	0.7	0.3	0.0
LN-RN+MED+	0.81	0.18	0.01
LN-RN+MED-	0.74	0.25	0.01
LN-RN-MED+	0.82	0.17	0.0

NS F 39<AGE			
LN+RN+MED+	0.59	0.36	0.05
LN+RN+MED-	0.5	0.44	0.06
LN+RN-MED+	0.63	0.35	0.02
LN+RN-MED-	0.54	0.43	0.02
LN-RN+MED+	0.67	0.29	0.04
LN-RN+MED-	0.59	0.36	0.05
LN-RN-MED+	0.71	0.28	0.02
MC M AGE<12			
LN+RN+MED+	0.29	0.5	0.21
LN+RN+MED-	0.22	0.55	0.23
LN+RN-MED+	0.35	0.56	0.1
LN+RN-MED-	0.27	0.62	0.11
LN-RN+MED+	0.36	0.44	0.2
LN-RN+MED-	0.28	0.49	0.23
LN-RN-MED+	0.42	0.48	0.09
MC M 11<AGE<40			
LN+RN+MED+	0.48	0.45	0.07
LN+RN+MED-	0.39	0.52	0.08
LN+RN-MED+	0.52	0.45	0.03
LN+RN-MED-	0.43	0.53	0.03
LN-RN+MED+	0.57	0.37	0.07
LN-RN+MED-	0.47	0.45	0.08
LN-RN-MED+	0.61	0.37	0.03
MC M 39<AGE			
LN+RN+MED+	0.27	0.47	0.25
LN+RN+MED-	0.21	0.52	0.28
LN+RN-MED+	0.34	0.54	0.12
LN+RN-MED-	0.26	0.6	0.13
LN-RN+MED+	0.34	0.41	0.25
LN-RN+MED-	0.26	0.46	0.27
LN-RN-MED+	0.42	0.47	0.11
MC F AGE<12			
LN+RN+MED+	0.37	0.4	0.23
LN+RN+MED-	0.29	0.45	0.26
LN+RN-MED+	0.45	0.44	0.11
LN+RN-MED-	0.36	0.52	0.12
LN-RN+MED+	0.44	0.34	0.22
LN-RN+MED-	0.36	0.39	0.25
LN-RN-MED+	0.63	0.37	0.1
MC F 11<AGE<40			
LN+RN+MED+	0.59	0.33	0.07
LN+RN+MED-	0.5	0.41	0.09
LN+RN-MED+	0.64	0.33	0.03
LN+RN-MED-	0.55	0.41	0.04
LN-RN+MED+	0.67	0.27	0.07
LN-RN+MED-	0.58	0.34	0.08
LN-RN-MED+	0.71	0.26	0.03

MC F 39<AGE			
LN+RN+MED+	0.35	0.37	0.28
LN+RN+MED-	0.27	0.42	0.31
LN+RN-MED+	0.44	0.43	0.13
LN+RN-MED-	0.35	0.5	0.15
LN-RN+MED+	0.42	0.32	0.26
LN-RN+MED-	0.34	0.36	0.3
LN-RN-MED+	0.52	0.36	0.12
LP M AGE<12			
LN+RN+MED+	0.57	0.27	0.16
LN+RN+MED-	0.48	0.33	0.19
LN+RN-MED+	0.65	0.28	0.07
LN+RN-MED-	0.56	0.36	0.09
LN-RN+MED+	0.64	0.22	0.14
LN-RN+MED-	0.55	0.27	0.18
LN-RN-MED+	0.72	0.22	0.06
LP M 11<AGE<40			
LN+RN+MED+	0.76	0.19	0.04
LN+RN+MED-	0.69	0.25	0.06
LN+RN-MED+	0.8	0.19	0.02
LN+RN-MED-	0.73	0.25	0.02
LN-RN+MED+	0.82	0.15	0.04
LN-RN+MED-	0.76	0.2	0.05
LN-RN-MED+	0.85	0.14	0.01
LP M 39<AGE			
LN+RN+MED+	0.55	0.26	0.19
LN+RN+MED-	0.46	0.31	0.23
LN+RN-MED+	0.64	0.28	0.09
LN+RN-MED-	0.55	0.34	0.11
LN-RN+MED+	0.62	0.21	0.17
LN-RN+MED-	0.53	0.26	0.21
LN-RN-MED+	0.71	0.22	0.08
LP F AGE<12			
LN+RN+MED+	0.69	0.25	0.06
LN+RN+MED-	0.61	0.32	0.08
LN+RN-MED+	0.73	0.24	0.02
LN+RN-MED-	0.65	0.31	0.03
LN-RN+MED+	0.76	0.19	0.05
LN-RN+MED-	0.68	0.25	0.07
LN-RN-MED+	0.79	0.19	0.02
LP F 11<AGE<40			
LN+RN+MED+	0.83	0.16	0.01
LN+RN+MED-	0.77	0.21	0.02
LN+RN-MED+	0.85	0.15	0.01
LN+RN-MED-	0.79	0.2	0.01
LN-RN+MED+	0.87	0.12	0.01
LN-RN+MED-	0.82	0.16	0.02
LN-RN-MED+	0.89	0.11	0.0

LP F 39<AGE			
LN+RN+MED+	0.68	0.24	0.08
LN+RN+MED-	0.6	0.31	0.1
LN+RN-MED+	0.73	0.24	0.03
LN+RN-MED-	0.65	0.31	0.04
LN-RN+MED+	0.75	0.19	0.06
LN-RN+MED-	0.67	0.24	0.08
LN-RN-MED+	0.79	0.18	0.03
LD M AGE<12			
LN+RN+MED+	0.14	0.35	0.51
LN+RN+MED-	0.1	0.36	0.54
LN+RN-MED+	0.21	0.49	0.3
LN+RN-MED-	0.16	0.52	0.32
LN-RN+MED+	0.18	0.31	0.51
LN-RN+MED-	0.13	0.33	0.54
LN-RN-MED+	0.27	0.44	0.29
LD M 11<AGE<40			
LN+RN+MED+	0.33	0.43	0.24
LN+RN+MED-	0.25	0.48	0.27
LN+RN-MED+	0.4	0.49	0.11
LN+RN-MED-	0.32	0.56	0.13
LN-RN+MED+	0.4	0.37	0.23
LN-RN+MED-	0.31	0.42	0.26
LN-RN-MED+	0.48	0.41	0.11
LD M 39<AGE			
LN+RN+MED+	0.12	0.3	0.57
LN+RN+MED-	0.09	0.32	0.6
LN+RN-MED+	0.2	0.45	0.35
LN+RN-MED-	0.15	0.48	0.37
LN-RN+MED+	0.16	0.27	0.57
LN-RN+MED-	0.11	0.29	0.6
LN-RN-MED+	0.25	0.4	0.35
LD F AGE<12			
LN+RN+MED+	0.25	0.47	0.28
LN+RN+MED-	0.19	0.51	0.31
LN+RN-MED+	0.32	0.55	0.14
LN+RN-MED-	0.24	0.61	0.15
LN-RN+MED+	0.31	0.41	0.28
LN-RN+MED-	0.24	0.46	0.31
LN-RN-MED+	0.39	0.48	0.13
LD F 11<AGE<40			
LN+RN+MED+	0.45	0.45	0.1
LN+RN+MED-	0.36	0.52	0.12
LN+RN-MED+	0.5	0.46	0.04
LN+RN-MED-	0.41	0.54	0.05
LN-RN+MED+	0.53	0.37	0.1
LN-RN+MED-	0.44	0.45	0.11
LN-RN-MED+	0.58	0.38	0.04

LD F 39<AGE

LN+RN+MED+	0.23	0.43	0.34
LN+RN+MED-	0.17	0.46	0.36
LN+RN-MED+	0.31	0.52	0.17
LN+RN-MED-	0.24	0.58	0.19
LN-RN+MED+	0.29	0.38	0.33
LN-RN+MED-	0.22	0.42	0.36
LN-RN-MED+	0.38	0.46	0.16

SYMPTOMATIC (B) PATIENTS

NS M AGE<12

LN+RN+MED+	0.19	0.47	0.34
LN+RN+MED-	0.14	0.5	0.36
LN+RN-MED+	0.26	0.57	0.17
LN+RN-MED-	0.19	0.62	0.18
LN-RN+MED+	0.25	0.42	0.33
LN-RN+MED-	0.18	0.46	0.36
LN-RN-MED+	0.32	0.51	0.17

NS M 11<AGE<40

LN+RN+MED+	0.38	0.49	0.13
LN+RN+MED-	0.3	0.55	0.15
LN+RN-MED+	0.43	0.51	0.06
LN+RN-MED-	0.34	0.59	0.07
LN-RN+MED+	0.46	0.42	0.13
LN-RN+MED-	0.37	0.48	0.15
LN-RN-MED+	0.52	0.43	0.05

NS M 39<AGE

LN+RN+MED+	0.18	0.43	0.39
LN+RN+MED-	0.13	0.45	0.42
LN+RN-MED+	0.25	0.55	0.21
LN+RN-MED-	0.19	0.59	0.22
LN-RN+MED+	0.23	0.38	0.39
LN-RN+MED-	0.17	0.41	0.42
LN-RN-MED+	0.31	0.49	0.2

NS F AGE<12

LN+RN+MED+	0.39	0.42	0.19
LN+RN+MED-	0.3	0.48	0.22
LN+RN-MED+	0.46	0.46	0.09
LN+RN-MED-	0.37	0.53	0.1
LN-RN+MED+	0.46	0.36	0.18
LN-RN+MED-	0.37	0.42	0.21
LN-RN-MED+	0.54	0.38	0.08

NS F 11<AGE<40			
LN+RN+MED+	0.6	0.34	0.06
LN+RN+MED-	0.51	0.42	0.07
LN+RN-MED+	0.64	0.34	0.02
LN+RN-MED-	0.55	0.42	0.03
LN-RN+MED+	0.67	0.27	0.05
LN-RN+MED-	0.59	0.35	0.07
LN-RN-MED+	0.71	0.27	0.02
NS F 39<AGE			
LN+RN+MED+	0.37	0.4	0.23
LN+RN+MED-	0.29	0.45	0.26
LN+RN-MED+	0.45	0.44	0.11
LN+RN-MED-	0.36	0.52	0.12
LN-RN+MED+	0.45	0.34	0.22
LN-RN+MED-	0.36	0.39	0.25
LN-RN-MED+	0.53	0.37	0.1
MC M AGE<12			
LN+RN+MED+	0.09	0.33	0.58
LN+RN+MED-	0.07	0.34	0.6
LN+RN-MED+	0.15	0.49	0.36
LN+RN-MED-	0.11	0.52	0.37
LN-RN+MED+	0.12	0.3	0.58
LN-RN+MED-	0.08	0.31	0.61
LN-RN-MED+	0.19	0.45	0.36
MC M 11<AGE<40			
LN+RN+MED+	0.24	0.45	0.3
LN+RN+MED-	0.18	0.49	0.33
LN+RN-MED+	0.31	0.54	0.15
LN+RN-MED-	0.24	0.6	0.16
LN-RN+MED+	0.3	0.4	0.3
LN-RN+MED-	0.23	0.44	0.33
LN-RN-MED+	0.38	0.47	0.14
MC M 39<AGE			
LN+RN+MED+	0.08	0.28	0.64
LN+RN+MED-	0.06	0.29	0.65
LN+RN-MED+	0.14	0.45	0.41
LN+RN-MED-	0.1	0.47	0.43
LN-RN+MED+	0.1	0.26	0.64
LN-RN+MED-	0.07	0.27	0.66
LN-RN-MED+	0.18	0.41	0.42
MC F AGE<12			
LN+RN+MED+	0.36	0.54	0.1
LN+RN+MED-	0.28	0.61	0.11
LN+RN-MED+	0.4	0.56	0.04
LN+RN-MED-	0.32	0.64	0.05
LN-RN+MED+	0.44	0.47	0.1
LN-RN+MED-	0.35	0.54	0.11
LN-RN-MED+	0.48	0.48	0.04

MC F 11<AGE<40			
LN+RN+MED+	0.54	0.43	0.03
LN+RN+MED-	0.46	0.52	0.04
LN+RN-MED+	0.57	0.42	0.01
LN+RN-MED-	0.48	0.51	0.01
LN-RN+MED+	0.62	0.35	0.03
LN-RN+MED-	0.53	0.43	0.03
LN-RN-MED+	0.65	0.34	0.01
MC F 39<AGE			
LN+RN+MED+	0.35	0.52	0.12
LN+RN+MED-	0.27	0.59	0.14
LN+RN-MED+	0.4	0.55	0.05
LN+RN-MED-	0.31	0.63	0.06
LN-RN+MED+	0.43	0.45	0.12
LN-RN+MED-	0.34	0.52	0.14
LN-RN-MED+	0.48	0.47	0.05
LP M AGE<12			
LN+RN+MED+	0.18	0.33	0.49
LN+RN+MED-	0.13	0.35	0.52
LN+RN-MED+	0.27	0.46	0.28
LN+RN-MED-	0.2	0.5	0.3
LN-RN+MED+	0.22	0.29	0.48
LN-RN+MED-	0.17	0.32	0.52
LN-RN-MED+	0.33	0.4	0.27
LP M 11<AGE<40			
LN+RN+MED+	0.39	0.39	0.22
LN+RN+MED-	0.31	0.44	0.25
LN+RN-MED+	0.47	0.43	0.1
LN+RN-MED-	0.38	0.5	0.12
LN-RN+MED+	0.47	0.33	0.2
LN-RN+MED-	0.38	0.38	0.24
LN-RN-MED+	0.55	0.36	0.09
LP M 39<AGE			
LN+RN+MED+	0.16	0.29	0.55
LN+RN+MED-	0.12	0.31	0.58
LN+RN-MED+	0.25	0.42	0.33
LN+RN-MED-	0.19	0.46	0.36
LN-RN+MED+	0.2	0.26	0.54
LN-RN+MED-	0.15	0.28	0.58
LN-RN-MED+	0.31	0.37	0.32
LP F AGE<12			
LN+RN+MED+	0.31	0.43	0.26
LN+RN+MED-	0.23	0.48	0.29
LN+RN-MED+	0.38	0.49	0.12
LN+RN-MED-	0.3	0.56	0.14
LN-RN+MED+	0.37	0.37	0.25
LN-RN+MED-	0.29	0.42	0.29
LN-RN-MED+	0.46	0.42	0.12

LP F 11<AGE<40			
LN+RN+MED+	0.52	0.39	0.09
LN+RN+MED-	0.43	0.46	0.11
LN+RN-MED+	0.57	0.39	0.04
LN+RN-MED-	0.48	0.48	0.05
LN-RN+MED+	0.6	0.32	0.08
LN-RN+MED-	0.51	0.39	0.1
LN-RN-MED+	0.66	0.32	0.03
LP F 39<AGE			
LN+RN+MED+	0.29	0.4	0.31
LN+RN+MED-	0.22	0.44	0.34
LN+RN-MED+	0.37	0.48	0.15
LN+RN-MED-	0.29	0.54	0.17
LN-RN+MED+	0.35	0.35	0.3
LN-RN+MED-	0.27	0.39	0.34
LN-RN-MED+	0.45	0.41	0.14
LD M AGE<12			
LN+RN+MED+	0.11	0.31	0.58
LN+RN+MED-	0.08	0.32	0.6
LN+RN-MED+	0.17	0.47	0.36
LN+RN-MED-	0.12	0.5	0.38
LN-RN+MED+	0.13	0.28	0.58
LN-RN+MED-	0.1	0.3	0.61
LN-RN-MED+	0.22	0.42	0.36
LD M 11<AGE<40			
LN+RN+MED+	0.27	0.43	0.3
LN+RN+MED-	0.2	0.47	0.33
LN+RN-MED+	0.35	0.51	0.15
LN+RN-MED-	0.27	0.57	0.16
LN-RN+MED+	0.34	0.37	0.29
LN-RN+MED-	0.28	0.42	0.32
LN-RN-MED+	0.42	0.44	0.14
LD M 39<AGE			
LN+RN+MED+	0.09	0.27	0.64
LN+RN+MED-	0.07	0.28	0.66
LN+RN-MED+	0.16	0.43	0.41
LN+RN-MED-	0.11	0.45	0.44
LN-RN+MED+	0.12	0.24	0.64
LN-RN+MED-	0.08	0.25	0.66
LN-RN-MED+	0.2	0.38	0.41
LD F AGE<12			
LN+RN+MED+	0.2	0.45	0.35
LN+RN+MED-	0.15	0.48	0.37
LN+RN-MED+	0.27	0.56	0.17
LN+RN-MED-	0.2	0.61	0.19
LN-RN+MED+	0.25	0.4	0.34
LN-RN+MED-	0.19	0.44	0.37
LN-RN-MED+	0.34	0.49	0.17

LD F 11<AGE<40			
LN+RN+MED+	0.4	0.47	0.14
LN+RN+MED-	0.31	0.53	0.15
LN+RN-MED+	0.45	0.49	0.06
LN+RN-MED-	0.36	0.57	0.07
LN-RN+MED+	0.47	0.4	0.13
LN-RN+MED-	0.38	0.47	0.15
LN-RN-MED+	0.53	0.41	0.05
LD F 39<AGE			
LN+RN+MED+	0.19	0.41	0.4
LN+RN+MED-	0.14	0.44	0.43
LN+RN-MED+	0.26	0.53	0.21
LN+RN-MED-	0.19	0.58	0.23
LN-RN+MED+	0.23	0.37	0.4
LN-RN+MED-	0.17	0.4	0.43
LN-RN-MED+	0.32	0.47	0.21

Appendix II - A General Statistical Test For Evaluating Probabilities

Note: a more extensive discussion of this test, its application and power may be found in Barnett et al, 1981.

Suppose for a group of N patients numbered from 1 to N according to some neutral procedure (e.g. alphabetically), each is known to be in exactly one of M possible states of health. We assume N>M. A Bayesian model is used to estimate $P_i(j)$, the probability that patient i is in health state j (j = 1,....M). Given that the true state of health for each patient is ultimately learned, how might this information be used to assess the validity of the original probabilistic predictions?

The following procedure allows us to test the null hypothesis H_0 that the predictions are accurate.

1. Consider the model's estimates for $P_i(j)$. For patient 1 record $P_1(1)$, and for patient 2, $P_2(2)$, ...for patient M, $P_M(M)$. For patient M+1, we go back to the first stage and record $P_{M+1}(1)$, for patient M+2, $P_{M+2}(2)$, for patient 2M, $P_{2M}(M)$, for patient 2M+1, $P_{2M+1}(1)$, etc. Under this rule, we choose one of the predictions made for each patient, and over the entire set, choose roughly an equal number of predictions for each of the M states of health.

2.Take the N predictions just chosen and list them in decreasing order. Let us define the largest as r_1, the second largest as r_2, the smallest as r_N, etc. Note that r_1 is not necessarily associated with patient number 1.

3. Now the patients are to be divided into groups within which the chosen predictions are close together, even though, for different people, the predictions may refer to different states of health. For example, if the state of health is a stage of Hodgkins disease, one patient may have a 0.4 chance of being in stage I+II, while another has a 0.4 chance of being in state III. If these predictions are chosen for study in step 1, both patients will be grouped together.

Begin forming a first group with the patient with prediction r_1, the patient with prediction r_2, etc. Stop, upon

reaching K, the smallest number for which $\sum_{i=1}^{K} r_i$ and $\sum_{i=1}^{K} (1-r_i)$ are both at least five. If $r_{K+1} = r_K$, add the patient with prediction r_{K+1} to the group. Do the same with r_{K+2}, etc. This will ensure that people with identical predictions fall in the same groups.

4. Starting with the next lowest prediction, construct a second group in an analogous way. Proceed similarly until as many groups as possible are formed. Probably a few patients with the very lowest predictions (e.g. r_N) will constitute a "remainder" not large enough to form a full group; they should be added to the last group formed.

5. For each patient, note whether (s)he was finally found to be in the state of health whose estimated probability has been listed. For example, was patient 1 actually in stage I+II?

6. Now for each of the various groups formed in steps 3 and 4, calculate the quantity W, defined by:

$$W = \frac{Q - X}{A}$$

where $Q = \sum r_i$ for the group

$A = \sum r_i (1 - r_i)$

X = the number of group members who were in the state of health for which a prediction was made.

For example, suppose we wish to calculate W for a group of 12 patients with Hodgkins disease for whom one has the following information:

Patient	Stage for which prediction was recorded	Recorded probability estimate	Actual stage
1	I+II	0.25	III
3	IV	0.25	IV*
4	I+II	0.25	IV
7	I+II	0.25	I+II*
11	III	0.25	I+II
16	I+II	0.25	I+II*

21	IV	0.25	III
23	III	0.25	IV
28	I+II	0.25	III
33	IV	0.25	I+II
41	III	0.25	III*
43	I+II	0.25	I+II*

Note that, since we are considering three states of health for Hodgkins disease (stage I and II, stage III, and stage IV), the use of rule 1 leads to the recorded stages shown for the patients listed.

Here $X = 5$, because only the cases with a * meet the stated criterion. In this group all the r 's are 0.25.

then $Q = 12 \times 0.25 = 3$

and $A = (12 \times 0.25).0.75 = 2.25 = 1.5$

Thus $W = -2/1.5 = -1.33$ for the group.

Under the null hypothesis of correct predictions, the calculated W should have approximately a unit normal probability distribution.

7. Calculate Z from the rule $Z = \sum_{i=1}^{B} W_j$, where W_j is the W-value for the jth group, and B the number of groups formed. If the predictions are accurate then Z would be Chi-distributed with B degrees of freedom. One can use this fact to determine the significance level (p-value) of the value of Z calculated from the data.

Appendix III - Information on Lymphangiogram Used in the Hodgkins Program

III.1 Where Results are Reported as Positive/Equivocal/Negative

SOURCE	NODES HISTOLOGICALLY POSITIVE			NODES HISTOLOGICALLY NEGATIVE		
	LAG +ve	LAG -ve	LAG ?	LAG +ve	LAG -ve	LAG ?
Aisenberg('71)	8	2	0	5	5	5
Enright('70)	14	7	8	6	17	14
Glees('78)	18	2	0	8	20	2
Hanks('72)	6	0	0	1	3	0
Hass('71)	15	0	0	1	27	0
Jelliffe('70)	6	1	2	0	10	2
Kaplan(73)	61	5	5	19	137	43
Lowenbraun('70)	4	0	0	3	2	0
Mitchell('72)	10	2	0	11	16	2
Paglia ('73)	10	4	0	2	17	4
Prosnitz('72)	13	4	2	0	10	7
Sutcliffe('76)	22	4	2	2	54	6
	187	31	19	67	321	85

False negative = 31/237 = 13.1%
False positive = 67/463 = 12.3%

III.2 Where Results are Reported as Positive/Negative Only

SOURCE	NODES HISTOLOGICALLY POSITIVE		NODES HISTOLOGICALLY NEGATIVE	
	LAG +ve	LAG -ve	LAG +ve	LAG -ve
Banfi('74)	29	0	3	70
Castellino('74)	41	0	14	111
Cotman('77)	12	7	10	69
Garcia('71)	2	4	2	7
Hermreck(1976)	7	5	14	8
Hellman('74)	25	2	12	38
Kademian('77)	26	3	4	98
Martire('74)	13	0	9	38
Urlaub('79)	19	9	13	60
	174	40	81	489

False negative = 40/214 = 18.7%
False positive = 81/570 = 14.2%

Overall adding results from both series:

False negative = 71/451 = 15.7%
False positive = 138/1033 = 13.4%

Appendix IV - Annotated Trace of Decision Tree Analysis Program in Operation

In this example, `Expand-Top-Level-Choice-Node` is called with:

```
tests-left        =    '(LAG)
prior             =    '(.81 .17 .02)
lap-done?         =    NIL
symptomatology    =    'A
```

That is, we are considering an A patient with probability 0.81 of being in stage I+II, .17 of being in stage III, and .02 of being in stage IV. Tests other than lymphangiogram (LAG) and laparotomy (LAP) have either been done or are contradicted for other reasons.

The program trace feature of Lisp has been turned on to show the pattern of function calls which carries out the decision tree analysis. The functions Expand-Top-Level-Choice-Node, Expand-Chance-Node, Expand-Choice-Node, Treatment-vs-LAP, Choose-Best-Treatment, and Treatment-E-U have been selected for tracing. To save space and complexity, some calls to Choose-Best-Treatment and Treatment-E-U at lower levels have been deleted.

```
(1 ENTER Expand-Top-Level-Choice-Node ((LAG) (0.81 0.17 0.02)))
```

First the program calculates, for each treatment, the expected utility of carrying out that treatment immediately, without further testing.

```
  (1 ENTER Treatment-E-U (TNI (0.81 0.17 0.02)))
  (1 EXIT Treatment-E-U 0.77)
==> The utility of immediate total nodal irradiation is 0.77

  (1 ENTER Treatment-E-U (MOPP (0.81 0.17 0.02)))
  (1 EXIT Treatment-E-U 0.68)
==> The utility of immediate MOPP chemotherapy is 0.68

  (1 ENTER Treatment-E-U (EM (0.81 0.17 0.02)))
  (1 EXIT Treatment-E-U 0.73)
==> The utility of immediate extended mantle radiotherapy is 0.73
```

Next the program calculates the utility of performing LAP followed by optimal therapy defined by the results of LAP.

```
  (1 ENTER Lap-Plan ((0.81 0.17 0.02)))
    (1 ENTER Expand-Chance-Node (LAP (0.81 0.17 0.02) NIL))
      (1 ENTER Choose-Best-Treatment ((1.0 0.0 0.0)))
      (1 EXIT Choose-Best-Treatment ((0.82 . EM)))
```

```
      (1 ENTER Choose-Best-Treatment ((0.0 1.0 0.0)))
      (1 EXIT Choose-Best-Treatment ((0.64 . TNI)))
      (1 ENTER Choose-Best-Treatment ((0.0 0.0 1.0)))
      (1 EXIT Choose-Best-Treatment ((0.55 . MOPP)))
    (1 EXIT Expand-Chance-Node
       ((0.78 . LAP) utility-data-place-holder
        ((I+II 0.81 (1.0 0.0 0.0) ((0.82 . EM)))
         (III 0.17 (0.0 1.0 0.0) ((0.64 . TNI)))
         (IV 0.02 (0.0 0.0 1.0) ((0.55 . MOPP))))))
  (1 EXIT Lap-Plan
     ((0.78 . LAP) utility-data-place-holder
      ((I+II 0.81 (1.0 0.0 0.0) ((0.82 . EM)))
       (III 0.17 (0.0 1.0 0.0) ((0.64 . TNI)))
       (IV 0.02 (0.0 0.0 1.0) ((0.55 . MOPP))))))
```

==> The utility of LAP is 0.78

Finally, the program calculates the utility of performing lymphangiogram.

```
(1 ENTER Expand-Chance-Node (LAG (0.81 0.17 0.02) NIL))
```

For the result of LAG being +LAG (lymphangiogram positive for abdominal nodes), the program calculates the optimal subsequent plan and its utility.

```
    (1 ENTER Expand-Choice-Node ((0.67 0.38 0.05) NIL))
      (1 ENTER Treatment-vs-LAP ((0.67 0.38 0.05)))
        (1 ENTER Lap-Plan ((0.67 0.38 0.05)))
          (2 ENTER Expand-Chance-Node (LAP (0.67 0.38 0.05) NIL))
          (2 EXIT Expand-Chance-Node
             ((0.73 . LAP) utility-data-place-holder
              ((I+II 0.67 (1.0 0.0 0.0) ((0.82 . EM)))
               (III 0.38 (0.0 1.0 0.0) ((0.64 . TNI)))
               (IV 0.05 (0.0 0.0 1.0) ((0.55 . MOPP))))))
        (1 EXIT Lap-Plan
           ((0.73 . LAP) utility-data-place-holder
            ((I+II 0.67 (1.0 0.0 0.0) ((0.82 . EM)))
             (III 0.38 (0.0 1.0 0.0) ((0.64 . TNI)))
             (IV 0.05 (0.0 0.0 1.0) ((0.55 . MOPP))))))
        (1 ENTER Choose-Best-Treatment ((0.67 0.38 0.05)))
        (1 EXIT Choose-Best-Treatment ((0.71 . TNI)))
      (1 EXIT Treatment-vs-LAP
         ((0.73 . LAP) utility-data-place-holder
          ((I+II 0.67 (1.0 0.0 0.0) ((0.82 . EM)))
           (III 0.38 (0.0 1.0 0.0) ((0.64 . TNI)))
           (IV 0.05 (0.0 0.0 1.0) ((0.55 . MOPP))))))
    (1 EXIT Expand-Choice-Node
       ((0.73 . LAP) utility-data-place-holder
        ((I+II 0.67 (1.0 0.0 0.0) ((0.82 . EM)))
         (III 0.38 (0.0 1.0 0.0) ((0.64 . TNI)))
         (IV 0.05 (0.0 0.0 1.0) ((0.55 . MOPP))))))
```

For the result of LAG being ?LAG (equivocal lymphangiogram), the program calculates the optimal subsequent plan and its utility.

```
    (1 ENTER Expand-Choice-Node ((0.87 0.12 0.01) NIL))
      (1 ENTER Treatment-vs-LAP ((0.87 0.12 0.01)))
```

```
    (1 ENTER Lap-Plan ((0.87 0.12 0.01)))
      (2 ENTER Expand-Chance-Node (LAP (0.87 0.12 0.01) NIL))
      (2 EXIT Expand-Chance-Node
         ((0.79 . LAP) utility-data-place-holder
          ((I+II 0.87 (1.0 0.0 0.0) ((0.82 . EM)))
           (III 0.12 (0.0 1.0 0.0) ((0.64 . TNI)))
           (IV 0.01 (0.0 0.0 1.0) ((0.55 . MOPP))))))
    (1 EXIT Lap-Plan ((0.79 . LAP) utility-data-place-holder
       ((I+II 0.87 (1.0 0.0 0.0) ((0.82 . EM)))
        (III 0.12 (0.0 1.0 0.0) ((0.64 . TNI)))
        (IV 0.01 (0.0 0.0 1.0) ((0.55 . MOPP))))))
    (1 ENTER Choose-Best-Treatment ((0.87 0.12 0.01)))
    (1 EXIT Choose-Best-Treatment ((0.78 . TNI)))
  (1 EXIT Treatment-vs-LAP
     ((0.79 . LAP) utility-data-place-holder
      ((I+II 0.87 (1.0 0.0 0.0) ((0.82 . EM)))
       (III 0.12 (0.0 1.0 0.0) ((0.64 . TNI)))
       (IV 0.01 (0.0 0.0 1.0) ((0.55 . MOPP))))))
(1 EXIT Expand-Choice-Node ((0.79 . LAP) utility-data-place-holder
   ((I+II 0.87 (1.0 0.0 0.0) ((0.82 . EM)))
    (III 0.12 (0.0 1.0 0.0) ((0.64 . TNI)))
    (IV 0.01 (0.0 0.0 1.0) ((0.55 . MOPP))))))
```

For the result of LAG being -LAG (negative lymphangiogram), the program calculates the optimal subsequent plan and its utility.

```
(1 ENTER Expand-Choice-Node ((0.89 0.10 0.01) NIL))
  (1 ENTER Treatment-vs-LAP ((0.89 0.10 0.01)))
    (1 ENTER Lap-Plan ((0.89 0.10 0.01)))
      (2 ENTER Expand-Chance-Node (LAP (0.89 0.10 0.01) NIL))
      (2 EXIT Expand-Chance-Node
         ((0.79 . LAP) utility-data-place-holder
          ((I+II 0.89 (1.0 0.0 0.0) ((0.82 . EM)))
           (III 0.10 (0.0 1.0 0.0) ((0.64 . TNI)))
           (IV 0.01 (0.0 0.0 1.0) ((0.55 . MOPP))))))
    (1 EXIT Lap-Plan
       ((0.79 . LAP) utility-data-place-holder
        ((I+II 0.89 (1.0 0.0 0.0) ((0.82 . EM)))
         (III 0.10 (0.0 1.0 0.0) ((0.64 . TNI)))
         (IV 0.01 (0.0 0.0 1.0) ((0.55 . MOPP))))))
    (1 ENTER Choose-Best-Treatment ((0.89 0.10 0.01)))
    (1 EXIT Choose-Best-Treatment ((0.79 . TNI)))
  (1 EXIT Treatment-vs-LAP
     ((0.79 . LAP) utility-data-place-holder
      ((I+II 0.89 (1.0 0.0 0.0) ((0.82 . EM)))
       (III 0.10  (0.0 1.0 0.0) ((0.64 . TNI)))
       (IV 0.01 (0.0 0.0 1.0) ((0.55 . MOPP))))))
(1 EXIT Expand-Choice-Node
   ((0.79 . LAP) utility-data-place-holder
    ((I+II 0.89 (1.0 0.0 0.0) ((0.82 . EM)))
     (III 0.10 (0.0 1.0 0.0) ((0.64 . TNI)))
     (IV 0.01 (0.0 0.0 1.0) ((0.55 . MOPP))))))
```

The lymphangiogram plan is complete. Regardless of the result of LAG, as shown in the plan fragment below, the best action to take subsequently is laparotomy. Hence, LAG can be omitted,

as the plan beginning with with LAP will be better in all respects.

```
  (1 EXIT Expand-Chance-Node
     ((0.77 . LAG) utility-data-place-holder
      ((+LAG 0.25 (0.57 0.38 0.06)
             ((0.73 . LAP) utility-data-place-holder
              ((I+II 0.57 (1.0 0.0 0.0) ((0.82 . EM)))
               (III 0.38 (0.0 1.0 0.0) ((0.64 . TNI)))
               (IV 0.05 (0.0 0.0 1.0) ((0.55 . MOPP))))))
       (?LAG 0.22 (0.87 0.12 0.01)
             ((0.79 . LAP) utility-data-place-holder
              ((I+II 0.87 (1.0 0.0 0.0) ((0.82 . EM)))
               (III 0.12 (0.0 1.0 0.0) ((0.64 . TNI)))
               (IV 0.01 (0.0 0.0 1.0) ((0.55 . MOPP))))))
       (-LAG 0.53 (0.89 0.10 0.01)
             ((0.79 . LAP) utility-data-place-holder
              ((I+II 0.89 (1.0 0.0 0.0) ((0.82 . EM)))
               (III 0.10 (0.0 1.0 0.0) ((0.64 . TNI)))
               (IV 0.01 (0.0 0.0 1.0) ((0.55 . MOPP)))))))))
(1 EXIT Expand-Top-Level-Choice-Node 32)
```

References

1. Aisenberg AC, Lingood RM, Lewis RA: The Changing Face of Hodgkins Disease. Am J Med 67: 921-928 1979.

2. Aisenberg AC, Qazi R: Abdominal Involvement at the Onset of Hodgkins Disease. Am J Med 57:870-874

3. Aisenberg AC, Qazi Raman: Improved Survival in Hodgkins Disease. Cancer 37:2423-2429 1976

4. Alcorn FS, Mategramo VC, Petasnick JP, et al: Contributions of Computed Tomography in the Staging and Management of Malignant Lymphoma. Radiology 125:717-723, 1977.

5. Andrassy RJ, Haff Roderick C: Laparotomy for Staging of Hodgkins and non-Hodgkins Lymphoma. Surgery, Gynecology & Obstetrics 144:208-210, 1977.

6. Barnett AI, Rutherford CJ, Desforges JF, Gutensohn N, Davies B: Evaluating the Validity of a Bayesian Program for Predicting Stage in Hodgkin's Disease. Meth Inf Med (in press).

7. Bearman RM, Pangalis GA, Rappoport H: Hodgkins Disease, Lymphocyte Depleted Type. Cancer 41: 293-302, 1978.

8. Beretta G, Spinelli P: Sequential Laparoscopy and Laparotomy Combined With Bone Marrow Biopsy in Staging Hodgkins Disease. Cancer Treat Rep 60:1231-1237, 1976.

9. Breiman RS, Castellino RA, Haell GS, et al: C.T. - Pathologic Correlations in Hodgkins Disease and Non-Hodgkins Lymphoma. Radiology 126:159-166, 1978.

10. British National Lymphoma Investigation: Initial Treatment of Stage IIIA Hodgkins Disease. Lancet 2:991-995, 1976.

11. Cannon WB, Kaplan HS: Staging Laparotomy With Splenectomy in Hodgkins Disease. Surg Ann 7:103-114, 1975.

12. Carbone PP, Kaplan HS, Musshoff K, et al: Report of the Committee on Hodgkins Disease Staging Classification. Cancer Res 31: 1860-1861, 1971.

13. Coleman CN, Wiliams CJ, Flint A, Glatstein EJ, Rosenberg SA, Kaplan HS: Hematologic Malignancy in patients treated for Hodgkins disease. N Eng J Med 297:1249-1252 1977.

14. Coleman Morton, Lightdale Charles, Vinciguerra Vincent P, Degnan Thomas J, Goldstein Michael, Horwitz S Theodore, Winawer Sidney J, Silver Richard T: Peritoneoscopy in Hodgkin's Disease. JAMA 236:2634-7 1976.

15. Cotman HE, Bloomfield CD, Amplatz K et al:Lymphography as a guide during laparotomy in Hodgkin's and non-Hodgkin's lymphoma. Acta Radiol (Ther) (Stockh) 16:295-304 1977.

16. Desforges JF, Rutherford CJ, Piro A: Hodgkins Disease. N Eng J Med 301:1212-1222, 1979.

17. Desser RK, Ultmann JE: Risk of Severe Infection in Patients with Hodgkins Disease or Lymphoma After Diagnostic Splenectomy. Ann Int Med 77:143.

18. DeVita VT, Simon RM, Hubbard SM, Young RC, Berard CW, Moxley JH, Frei E, Carbone PP, Canellos GP: Curability of advanced Hodgkins Disease with Chemotherapy. Ann Int Med 92:587-595, 1980.

19. Eisen E: MIT Sloan School Master's Thesis.

20. Ferguson, Allen, Curren, Moran, Rappaport and Ultman: Surgical Experience with Staging Laparotomy in 125 Patients with Lymphoma. Arch Int Med 131:356 (1973).

21. Fisher RI, DeVita VT, Hubbard SP, Simon R, Young RC: Prolonged Disease-Free Survival in Hodgkins Disease with MOPP Reinduction After First Relapse. Ann Int Med 90:761-763 1979.

22. Fuller LM, Madoc-Jones H New Assessment of the Prognostic Significance of Histopathology in Hodgkins Disease for Laparotomy- Negative Stage I and Stage II Patients. Cancer 39:2174-2182, 1977.

23. Gamble JF, Fuller LM, Martin RG, Sullivan MP, Jing B-S, Butler JJ, Shullenberger CC: Influence of Staging Celiotomy in Localized Presentations of Hodgkins Disease. Cancer 35:817-825, 1975.

24. Garcia EV, Santiago PJ: Usefulness of Laparotomy and Splenectomy in Hodgkins Disease. Bol Assoc Med PR 63:104-112 1971.

25. Gazet JC: Laparotomy and Splenectomy. from "Hodgkin's Disease" ed. Prof. David Smithers Churchill Livingstone, p.190, 1973.

26. Goodman Robert, Mauch Peter, Piro Anthony, Rosenthal David, Goldstein Michael, Tullis James, Hellman Samuel: Stages IIB and IIIB Hodgkins Disease. Cancer 40:84-89 1977.

27. Greenwood JA, Hartley HO: Guide to Tables in Medical Statistics, Section 3.32. Princeton University Press, 1962.

28. Hancock BW, Bruce L, Dunsmore IR, Ward AM, Richmond J: Follow-up Studies on the Immune Status of Patients with Hodgkins Disease After Splenectomy and Treatment in Relapse and Remission. Brit J Cancer 36:347-354, 1977.

29. Hanks GE, Newsome JF: The Value of Laparotomy in Staging Lymphomas. Southern Med J 64:585-588, 1971.

30. Hellman S, Mauch P, Goodman RL, Rosenthal D, Moloney WC: The Place of Radiation Therapy in the Treatment of Hodgkins Disease. Cancer 42:971-978, 1978.

31. Hermreck AS, Kofender VS, Bell C: The Staging of Hodgkins Disease. Am J Surg 130:639-642, 1975.

32. Hoogstraten B, Glidewell O, Holland JF, Blom J, Stutzman L, Nissen NI, Perlberg HJ, Kramer S: Long Term Followup of Combination Chemotherapy Radiotherapy of Stage III Hodgkins Disease. Cancer 34: 1234-1244 1979.

33. Jelliffe AM, Millett YL: Laparotomy and Splenectomy as Routine Investigations in the Staging of Hodgkin's Disease Before Treatment. Clin Radiol 21:439-445 1970.

34. Jones SE, Butler JJ, Byrne GE Jr., et al: Histopathologic Review of Lymphoma Cases from the Southwest Oncology Group. Cancer 39: 1071-1076, 1977.

35. Kademian MT, Wirtanen GW: Accuracy of Bipedal Lymphography in Hodgkins Disease. Am J Roentgenol 129:1041-1042, 1977.

36. Krivit William: Overwhelming Post-Splenectomy Infection. Am J Hematology 2:193-201 1977.

37. Lee JK, Stanley RJ, Sagel SS, et al: Accuracy of Computed tomography in detecting intraabdominal and pelvic adenopathy in lymphoma. Am J Rad 131:311-315, 1978.

38. Lipton MJ, Denardo GL, Silverman S, et al: Evaluation of the Liver and Spleen in Hodgkins Disease. The Value of Hepatic Scintigraphy. Am J Med 52:356-361, 1972.

39. Lowenbraun S, Ramsey H: Diagnostic Laparotomy and Splenectomy for Staging Hodgkins Disease. Ann Int Med 72:655-663, 1970.

40. Marston, Adrian: Laparotomy and Splenectomy in the Diagnosis and Management of Hodgkins Disease. Proc Roy Soc Med 65:1111-1112, 1972.

41. McCaffrey JA, Rudders RA, Kahn PC, et al: Clinical Usefulness of Gallium Scanning in the Malignant Lymphomas. Am J Med 60:523-530, 1976.

42. Mecker WR, Richardson JD: Critical Evaluation of Laparotomy and Splenectomy in Hodgkins Disease. Arch Surg 105:222-229, 1972.

43. Mildner MS, Larson SM, Bagley CM, et al: Liver-Spleen Scan in Hodgkins Disease Cancer 31:826-834, 1973.

44. Mitchell RI, Peters MV: Laparotomy for Hodgkins Disease:Some Surgical Observations. Surgery 71:694-703, 1972.

45. Olweny CLM, Katongole-Mbidde, Kiire C, Lwanga SK, MacGrath I, Ziegler JL: Childhood Hodgkins Disease in Uganda - a ten year experience. Cancer 42:787-792, 1978.

46. Paglia MA, Lacher MJ, Hertz REL, Geller W, Watson RC, Lewis JL, Nisce LZ, Lieberman PH: Surgical Apects and Results of Laparotomy and Splenectomy in Hodgkins Disease. Amer J Roentgenol Radium Ther Nucl Med 117: 12-18, 1973.

47. Piro Anthony J, Hellman Samuel: International Symposium on Hodgkins Disease. Invited Discussion:Laparotomy Alters Treatment in Hodgkins Disease. Natl Cancer Inst Monograph 36:307-311 1973.

48. Portlock CS, Rosenberg SA, Glatstein E, Kaplan HS: Impact of Salvage Treatment on Initial Relapses in Patients With Hodgkins Disease, Stages I-III. Blood 51:825-833, 1978.

49. Poulsen Henrik, Bengmark Stig: Staging Laparotomy with Splenectomy in Hodgkins Disease. Acta Chir Surg 143:347-352 1977.

50. Prosnitz LR, Nuland SB: Role of Laparotomy and Splenectomy in the Management of Hodgkins Disease. Cancer 24:44-50, 1972.

51. Prosnitz LR, Montalvo RL, Fischer DB, Silberstein AB, Berger DS: Treatment of Stage IIIA Hodgkins Disease: Is Radiotherapy Alone Adequate? Int J Radiation Oncology Biol Phys 4:781-787, 1978.

52. Roberts SJ, Roeser HP: Hodgkins Disease: An Evaluation of Staging Laparotomy in 82 Patients. Australas Radiol 20:314-320 1976.

53. Rosenberg Saul A, Kaplan Henry S: The Management of Stages I, II and III Hodgkins Disease with Combined Radiotherapy and Chemotherapy. Cancer 35:55-63 1975.

54. Rozman C, Triginer J: The Value of Laparotomy and Splenectomy in the Staging of 56 Patients with Hodgkins Disease. Acta Haematologica 50:321-38 1973.

55. Rutherford CJ, Desforges JF, Davies B, Barnett AI: The Decision to Perform Staging Laparotomy in Symptomatic Hodgkins Disease. Brit J Haem 44:347-358, 1980.

56. Rutherford CJ, Desforges JF, Barnett AI, Safran CS, Davies B: The Decision Between Single and Combined Modality Therapy in Hodgkin's Disease. Am J Med (in press).

57. Safran C, Tsichlis PN, Bluming AZ, et al: Diagnostic Planning Using Computer Assisted Decision-Making for Patients With Hodgkins Disease. Cancer 39:2426-2434, 1977.

58. Santoro A, Zucali R, Bonfante V et al: Comparative therapeutic effects and morbidity of two combined treatment modalities in PS IIB-IIIA&B Hodgkins Disease. Proc Am Soc Clin Oncol 20:359, 1979.

59. Schimpff SC, O'Connell MJ: Infection in 92 Splenectomized Patients with Hodgkins Disease. Am J Med 59:695-701, 1977.

60. Silverman S, DeNardo GL, Glatstein E, et al: Evaluation of the Liver and Spleen in Hodgkins Disease. The Value of Splenic Scintigraphy. Am J Med 52:362-366, 1972.

61. Singer DB: Postsplenectomy Sepsis. In Rosenberg HS, Bolande RP (eds) "Perspectives in Pediatric Pathology" Chicago: Year Book Medical Publishers, 1973, p285-311.

62. Smith MD, Klebanoff G: Exploratory Laparotomy for Staging in Hodgkins Disease. Am J Surg 124:811-814, 1972.

63. Stein RS, Golomb HM, Diggs CH, Mauch P, Hellman S, Wiernik PH, Ultmann JE, Rosenthal DS: Anatomic Substages of Stage IIIA Hodgkins Disease, a Collaborative Study. Ann Int Med 92:159-165, 1980.

64. Stoffel TJ, Cox JD: Hodgkins Disease Stage I and II. Cancer 40:90-97, 1977.

65. Sutcliffe BJ, Wrigley PFM: Intensive Investigation in Management of Hodgkins Disease. BMJ 2:1343-1347 1976.

66. Urlaub BJ, Mack E: Evaluation and Complications of 107 Staing Laparotomies for Hodgkins Disease. Ann Surg 190:45-47, 1979.

67. Wiernik PH, Gustafson J, Schimpff, Diggs C: Combined Modality Treatment of Hodgkins Disease Confined To Lymph Nodes. Am J Medicine 67: 183-193, 1979.

68. Weitzman S, Aisenberg AC: Fulminant Sepsis After the Successful Treatment of Hodgkins Disease. Am J Med 62:47-49, 1977.

69. Zarembok I, Ramsey H: Laparotomy and Splenectomy in the Staging of Untreated Patients with Hodgkins Disease. Radiology 102:673-678, 1972.